Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences of the

Ruprecht - Karls - University

Heidelberg

Presented by

Cristina Conde Lopez, M.Sc.

born in: Granada, Spain

Oral examination: 30th of June 2025

Characterizing Sex Chromosome Dosage Differences in Head and Neck Squamous Cell Carcinoma Microenvironment

Referees: Prof. Dr. Sergio P. Acebrón
Prof. Dr. Jochen Hess

Abstract

Head and neck squamous cell carcinoma (HNSCC) is a biologically heterogeneous disease with marked differences in incidence and prognosis between sexes. While clinical sex differences have been increasingly recognized in HNSCC, the molecular consequences of sex chromosome dosage alterations, such as Loss of Y chromosome (LoY) and Extreme Downregulation of Y-linked genes (EDY), remain poorly understood. In this thesis, I investigate how such alterations contribute to tumor biology and shape the tumor microenvironment (TME), with particular focus on fibroblast and immune cell dynamics. By integrating bulk and single-cell transcriptomic datasets, I systematically explore the prevalence and implications of sex chromosome dosage variations across HNSCC patients.

Using TCGA bulk RNA-seq data, I identify widespread LoY and EDY events in male patients, with a notable enrichment in HPV-negative tumors. These alterations are tightly linked, with LoY emerging as a key driver of EDY. Importantly, patients stratified by sex chromosome dosage, XX, XY, and XØ (EDY/LoY males), display distinct transcriptomic and cellular profiles. To further dissect these differences, I constructed a harmonized single-cell HNSCC atlas, enabling the analysis of gene expression, chromosomal instability, and cell-cell interactions at single-cell resolution.

Tumor cells emerged as the primary site of Y chromosome downregulation, and their classification enabled refined stratification of the cohort. I show that the TME composition, particularly the abundance and activity of fibroblast subtypes (matrix and inflammatory cancer-associated fibroblasts (mCAFs and iCAFs, respectively)), varies significantly by sex chromosome dosage. Cell-cell communication analysis revealed that fibroblasts act as major signaling mediators, especially in XY tumors, where they exhibit enhanced outgoing interactions to tumor cells. Furthermore, XY iCAFs uniquely co-express COX2 and AR, suggesting a possible interaction between inflammatory and hormonal pathways that may support tumor progression. These findings were further

validated by deconvolution of bulk transcriptomic data, confirming higher fibroblast abundance in XY tumors.

Collectively, this work reveals that sex chromosome dosage is a biologically meaningful variable in HNSCC, driving differences in tumor–stroma communication and shaping the microenvironment. These insights highlight the importance of moving beyond binary sex classifications to uncover nuanced mechanisms of sex-biased tumor biology, with potential implications for therapeutic stratification and biomarker development.

Zusammenfassung

Das Plattenepithelkarzinom des Kopfes und Halses (HNSCC) ist eine biologisch heterogene Erkrankung mit deutlichen Unterschieden in der geschlechtsspezifischen Inzidenz und Prognose. Während das klinische Geschlecht weitgehend erforscht ist, sind die molekularen Folgen von Veränderungen der Geschlechtschromosomendosierung, wie z. B. der Verlust des Y-Chromosoms (LoY) und die extreme Herunterregulierung von Y-gebundenen Genen (EDY), nach wie vor kaum verstanden. In dieser Arbeit untersuche ich, wie solche Veränderungen zur Tumorbiologie beitragen und die Tumormikroumgebung (TME) formen, mit besonderem Augenmerk auf die Dynamik von Fibroblasten und Immunzellen. Durch die Integration von Massen- und Einzelzell-Transkriptomdatensätzen untersuche ich systematisch die Prävalenz und die Auswirkungen von Variationen der Geschlechtschromosomendosierung bei HNSCC-Patienten.

Unter Verwendung von TCGA-RNA-seq-Daten identifiziere ich weit verbreitete LoY- und EDY-Ereignisse bei männlichen Patienten, mit einer bemerkenswerten Anreicherung in HPV-negativen Tumoren. Diese Veränderungen sind eng miteinander verknüpft, wobei sich LoY als wesentlicher Treiber von EDY herausstellt. Wichtig ist, dass Patienten, die nach Geschlechtschromosomendosierung - XX, XY und XØ (EDY/LoY-Männer) - geschichtet sind, unterschiedliche transkriptomische und zelluläre Profile aufweisen. Um diese Unterschiede weiter zu entschlüsseln, habe ich einen harmonisierten Einzelzell-HNSCC-Atlas erstellt, der die Analyse der Genexpression, der chromosomalen Instabilität und der Zell-Zell-Interaktionen bei Einzelzellauflösung ermöglicht.

Tumorzellen erwiesen sich als der primäre Ort der Downregulation des Y-Chromosoms, und ihre Klassifizierung ermöglichte eine verfeinerte Stratifizierung der Kohorte. Ich zeige, dass die TME-Zusammensetzung, insbesondere die Häufigkeit und Aktivität von Fibroblasten-Subtypen (Matrix- und entzündliche krebsassoziierte Fibroblasten (CAFs)), je nach Geschlechtschromosomendosierung erheblich variiert. Die Analyse der Zell-

Zell-Kommunikation ergab, dass Fibroblasten als wichtige Signalvermittler fungieren, insbesondere in XY-Tumoren, wo sie verstärkte ausgehende Interaktionen mit Tumorzellen aufweisen. Darüber hinaus exprimieren XY-iCAFs in einzigartiger Weise COX2 und AR, was auf eine mögliche Interaktion zwischen entzündlichen und hormonellen Signalwegen hindeutet, die das Fortschreiten des Tumors begünstigen kann. Diese Ergebnisse wurden durch die Dekonvolution von transkriptomischen Massendaten bestätigt, die eine höhere Fibroblastenhäufigkeit in XY-Tumoren bestätigte.

Insgesamt zeigt diese Arbeit, dass die Geschlechtschromosomendosierung eine biologisch bedeutsame Variable bei HNSCC ist, die Unterschiede in der Tumor-Stroma-Kommunikation bewirkt und die Mikroumgebung prägt. Diese Erkenntnisse unterstreichen, wie wichtig es ist, über binäre Geschlechtsklassifizierungen hinauszugehen, um nuancierte Mechanismen der geschlechtsspezifischen Tumorbiologie aufzudecken, was sich möglicherweise auf die therapeutische Stratifizierung und die Entwicklung von Biomarkern auswirkt.

Table of Contents

1. Introduction

1.1 HNSCC Overview

Cancer is a heterogeneous and multifactorial disease characterized by uncontrolled cell growth and spread to other parts of the body. It remains one of the leading global health concerns, with millions of new cases diagnosed each year. Head and neck squamous cell carcinoma (HNSCC) is a group of cancers that develop from the mucosal epithelium in the oral cavity, pharynx, and larynx[1]. It is the sixth most common cancer worldwide, accounting for approximately 4.5% of all cancer diagnoses and deaths globally[2,3]. HNSCC is a prominent disease, comprising over 90% of head and neck cancers. While there has been significant progress in understanding the molecular and cellular mechanisms behind cancer, survival rates for HNSCC remain suboptimal[4]. In advance-stage or high-risk cases the five-year overall survival (OS) rate is below 50%, and recurrence rates remain high, with over 50% of patients experiencing relapse within two years of treatment[3]. This points to the need for more effective therapeutic strategies and improved patient stratification based on molecular and cellular insights.

HNSCC can be classified into two main categories based on the presence or absence of human papillomavirus (HPV) infection: HPV-negative and HPV-positive (Fig.1). These two forms of HNSCC exhibit distinct biological and clinical behaviors, with important implications for treatment and prognosis[5–7].

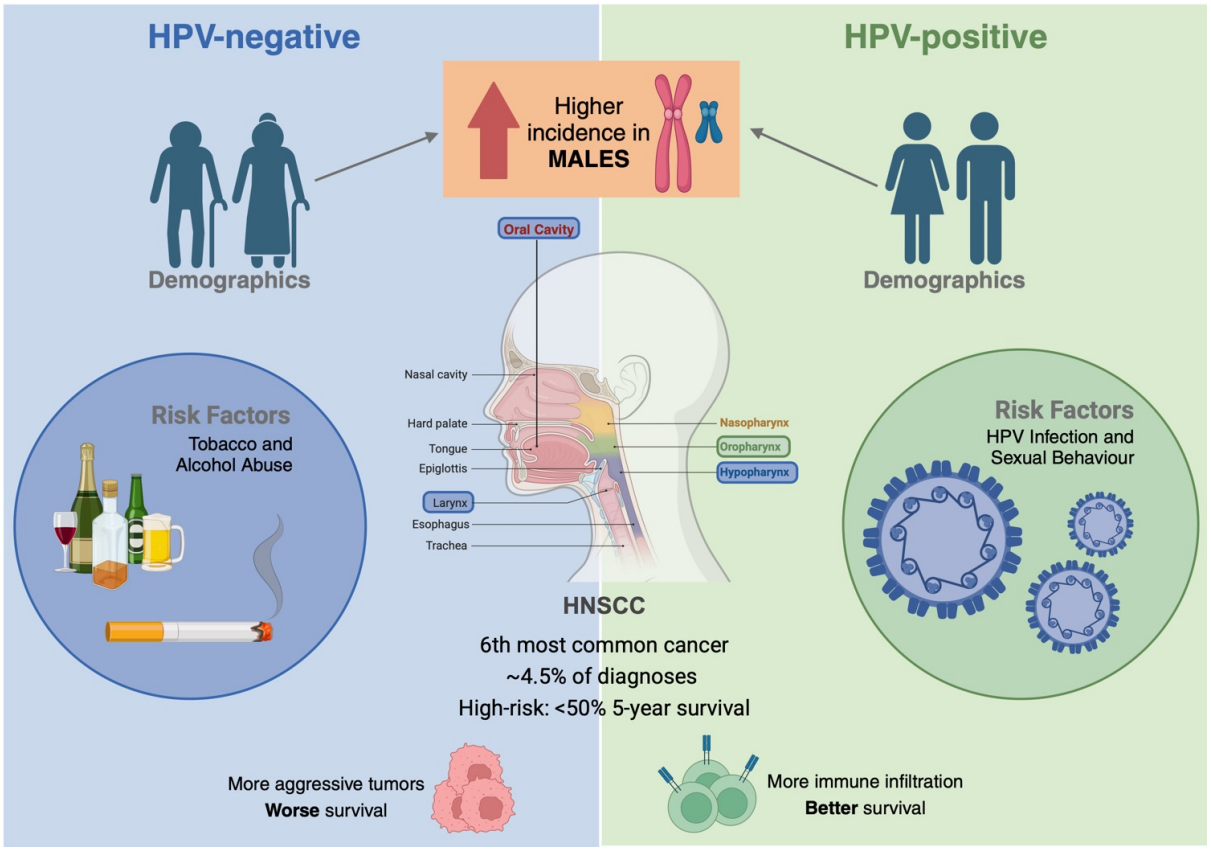


Figure 1. Overview of HPV-negative and HPV-positive HNSCC. Differences in demographics, risk factors, anatomical location, immune infiltration, and clinical outcomes.

HPV-negative HNSCC is traditionally associated with risk factors such as tobacco use and alcohol consumption[8,9]. This subtype is more common in older patients and is often diagnosed in the oral cavity, larynx, and hypopharynx[3,10]. HPV-negative tumors tend to be more aggressive, these tumors often harbor genetic mutations and alterations in pathways associated with carcinogenesis, such as the tumor suppressor p53 and the retinoblastoma (RB) pathway[11–13].

On the other hand, HPV-positive HNSCC is primarily associated with infection by high-risk HPV strains particularly HPV-16, other strains have been detected in smaller populations such as HPV-18, 31, 33 and 52[1,14]. HPV-positive tumors are most commonly found in the oropharynx and are more prevalent in younger, non-smoking individuals[3,15]. HPV-positive tumors tend to have a better prognosis than their HPV-negative counterparts, with improved response to treatment and a more favorable survival rate[16]. This difference is thought to be due to the unique biology of HPV-driven carcinogenesis,

where the viral oncoproteins silence tumor suppressors like p53 and RB, reducing host genome mutation in these genes leading to a simpler genomic mutational load compared to non-viral cancers[14,17]. HPV-positive HNSCC also creates a distinct immune tumor microenvironment (TME), characterized by increased T-cell infiltration, enhanced activation of immune effector cells, greater diversity of T-cell receptors, and higher immune cytolytic activity[18–20]. These features make HPV-positive tumors more responsive to immunotherapies, such as those targeting PD-1/PD-L1, and contribute to a stronger immune response, with higher immune cell infiltration and better immune activation. While HPV-negative tumors are generally associated with greater biological aggressiveness and poorer prognosis, findings regarding lymph node metastasis are more heterogeneous. Some studies report a higher incidence of lymph node involvement in HPV-positive tumors, whereas others have observed increased lymphatic spread in HPV-negative cases[21–23]. These discrepancies likely reflect differences in cohort composition, tumor subsite, and diagnostic criteria across studies.

The growing recognition of the different biological and clinical outcomes between HPV-positive and HPV-negative HNSCC has led to the consideration of HPV status as an important factor in patient stratification and personalized treatment planning. HPV-positive tumors often respond better to radiation therapy and certain immunotherapies and have higher cure rates compared to HPV-negative tumors, leading to ongoing efforts to tailor therapeutic approaches based on HPV status[24,25]. Furthermore, the increasing prevalence of HPV-positive HNSCC in non-smoking populations highlights the need for better screening, early detection, and targeted therapies for this subgroup of patients[26].

Moreover, HNSCC exhibits notable tumor heterogeneity at both the intra- and inter-tumor levels, which poses significant challenges for diagnosis, treatment, and prognosis[27,28]. Tumor heterogeneity refers to the variation in genetic and phenotypic characteristics of tumor cells within a single patient, as well as between different patients[29]. This complexity arises from the dynamic interactions within the TME, including stromal and immune cells, which can influence tumor behavior and response to therapies[30]. This heterogeneity can lead to therapeutic resistance and relapse, as subpopulations of cells within the same tumor may have distinct molecular profiles,

altering their sensitivity to treatment[31,32]. Understanding the heterogeneity of HNSCC is critical for developing more precise and personalized treatment strategies, as targeting only a homogeneous group of cells may not be sufficient for long-term tumor control. The challenges posed by tumor heterogeneity directly inform the need for more sophisticated, individualized therapeutic strategies.

Given these complexities, the management of HNSCC typically involves a combination of surgery, radiation therapy, and chemotherapy. Surgical resection is often the first-line treatment for localized tumors, followed by radiation therapy for patients with a higher risk of recurrence[1,33,34]. Chemotherapy is commonly used in advanced or metastatic disease, often in combination with radiation (chemoradiotherapy) to improve outcomes[1,35]. In recent years, targeted therapies and immunotherapies have also gained traction, particularly in HPV-positive HNSCC, where immune checkpoint inhibitors such as pembrolizumab and nivolumab have shown promise[1,36]. However, the treatment protocols for HNSCC can vary depending on the tumor subtype, stage, and HPV status, highlighting the need for more individualized approaches.

One of the goals of current research is to improve patient outcomes through treatment de-escalation, especially for HPV-positive patients who generally have a better prognosis[37,38]. De-escalation strategies aim to reduce the intensity of treatment in selected patients, thereby minimizing side effects while maintaining effective tumor control. This approach is based on the understanding that certain tumors can be less aggressive and therefore respond well to less aggressive therapies, providing an opportunity to reduce the long-term health burdens associated with over-treatment.

The complexity and heterogeneity of HNSCC highlight the need for personalized treatment strategies. While significant progress has been made in understanding the disease's molecular and clinical behaviors, challenges remain, particularly in addressing the diverse TMEs and the biological differences between HPV-positive and HPV-negative tumors. Additionally, the incidence of HNSCC shows notable sex differences, with a significantly higher incidence in males than females. While the disparity in incidence has traditionally been attributed to risk factors such as tobacco

and alcohol use, emerging evidence suggests that underlying biological mechanisms also contribute to these differences. Understanding these mechanisms is critical for developing more personalized and effective treatment strategies for both sexes. The ultimate goal of ongoing research is to improve patient outcomes, not only through better-targeted treatments but also by implementing de-escalation strategies that minimize unnecessary treatment burdens and enhance the quality of life for patients. Through personalized approaches and improved patient stratification, it is possible to advance the management of HNSCC and ultimately increase survival rates while reducing relapse.

1.2 Sex Dimorphism in HNSCC

Sex dimorphism in cancer refers to the observed differences in cancer prevalence, progression, and therapeutic responses between males and females. These disparities are driven by a complex interplay of genetic, hormonal, and immune factors, all of which contribute to the development and progression of the disease in different ways[39–41] (Figure 2). In HNSCC, sex differences are particularly pronounced, with males being significantly more prone to developing the disease compared to females[42]. This higher incidence in males is often coupled with more aggressive tumor progression and poorer prognosis. Although tobacco and alcohol use have long been recognized as key risk factors, emerging evidence suggests that underlying biological mechanisms, particularly those related to sex chromosomes and hormones, may also play a crucial role in the observed sex-based disparities in cancer risk and outcomes[43,44].

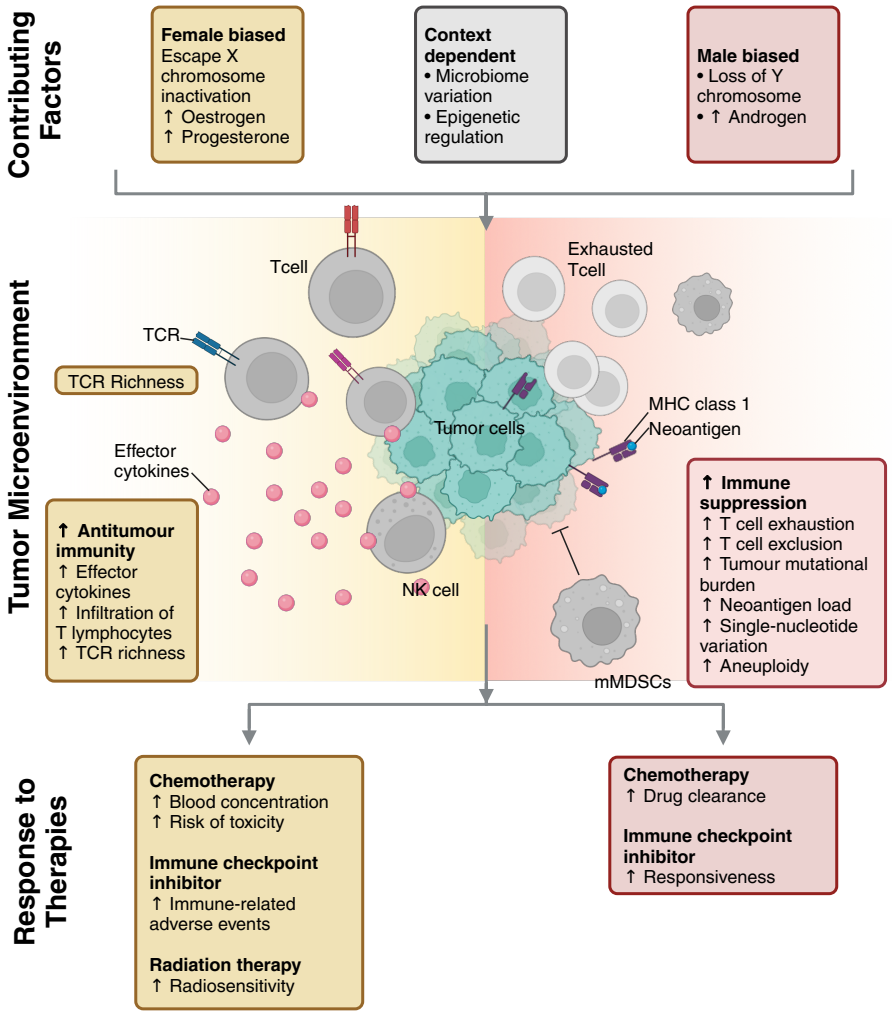


Figure 2. Overview of sex-related differences in antitumor immunity, TME, and response to therapies. This schematic illustrates key biological and immunological differences associated with sex. Yellow-labeled features represent female-biased processes, red-labeled features are male-biased, and grey-labeled ones are context-dependent, varying by tumor type, disease status, or oncogenic mechanism. The figure is organized around three levels of influence: (1) contributing factors such as sex chromosomes, hormones, and regulatory mechanisms; (2) differences in the immune landscape of the TME; and (3) sex-specific variability in treatment response. *Figure adapted from Xiao et al., Nat Rev Cancer (2024), reproduced with permission from Springer Nature (License N°: 6011310409719).*

Recent studies across various cancer types have highlighted the importance of sex-biased immune responses in contributing to these differences. For instance, in glioblastoma (GBM), male mice exhibit accelerated tumor growth along with decreased CD8+ T cell frequency and increased exhaustion in the TME, which is crucial for tumor progression and response to immunotherapy[45]. Moreover, a higher frequency of

progenitor exhausted T cells in males was associated with improved responsiveness to anti–PD-1 treatment. These findings indicate that sex-biased T-cell behaviors, such as increased exhaustion in males, play a key role in driving sex differences in cancer progression and therapeutic response. In renal cell carcinoma (RCC), studies have shown that male tumors tend to have more CD8+ T cells infiltrating the TME, but these cells are often in an exhausted and dysfunctional state, largely driven by androgen signaling. This finding suggests that sex hormones, particularly androgens, may contribute to immune suppression in male RCC tumors[46]. The role of the androgen–androgen receptor axis in sex differences further emphasizes the importance of considering sex when designing therapies that involve immune modulation[47]. A similar trend has been observed in esophageal cancer (EC), where immunotherapy has been associated with better outcomes in males compared to females. Studies show that males derive significantly more benefit from immunotherapy, with exhausted CD8+ T cells being highly infiltrated in male patients who responded to treatment[48]. In contrast, female patients showed little to no benefit from immunotherapy, which was linked to differences in immune cell composition and gene expression between the sexes.

This growing body of evidence underscores the need for sex-specific therapeutic strategies. A study utilizing data from The Cancer Genome Atlas (TCGA) classified HNSCC as part of the “strong sex-effect” group, which includes cancers such as thyroid carcinoma and lung squamous cell carcinoma[49]. This group is characterized by extensive sex-biased molecular signatures across multiple data types, including somatic mutations, gene expression, DNA methylation, and miRNA expression. These findings highlight the significant sex-based molecular differences in HNSCC, where male and female patients exhibit distinct tumor behaviors at the genetic and molecular levels. For example, differences in the expression of sex-biased genes and DNA methylation patterns may contribute to the more aggressive tumor behavior seen in male patients. Understanding how these sex-related molecular differences, whether in immune cell infiltration, hormonal regulation, or genetic pathways, affect tumor behavior and treatment response could lead to more effective, personalized therapies.

Males with HNSCC typically experience a more aggressive disease course, which is reflected in higher rates of metastasis and a greater likelihood of treatment resistance[50,51]. This has led researchers to examine the molecular mechanisms by which sex chromosomes, sex hormones, and the TME interact to influence cancer progression. In particular, the Y chromosome and the effects of male sex hormones such as testosterone are believed to contribute to the increased susceptibility of males to certain cancers, including HNSCC[50]. The TME itself also differs between males and females, with immune cells, stromal cells, and fibroblasts potentially exhibiting sex-specific behaviors that influence tumor growth and metastasis[43,52].

A critical aspect of sex dimorphism in cancer is the differential response to treatment. In HNSCC, the response to treatment may vary between males and females, and this difference is not always straightforward. While some studies have found no significant differences in OS or treatment outcomes between males and females in certain subgroups, emerging evidence suggests that sex-specific factors, including immune system activation and hormonal influences, play a role in treatment efficacy[53]. For example, males generally show a more favorable response to immune checkpoint inhibitors (ICIs), with studies indicating that male patients tend to benefit more from ICIs compared to females[54,55]. However, these responses can also vary depending on the type of treatment and cancer subtype. While immune responses differ between sexes, with females often showing more activated CD8+ T lymphocytes, the overall treatment response in males may be more robust, particularly with ICIs, suggesting that sex-based approaches to treatment could enhance therapeutic outcomes[54,56].

The discovery of these sex-related molecular mechanisms underscores the need for sex-specific therapeutic strategies in HNSCC treatment. Developing targeted therapies that take into account the biological differences between sexes can lead to more effective treatment regimens. For instance, exploring the role of sex hormones and immune checkpoint modulation could open new avenues for enhancing the efficacy of treatments, particularly in male patients who may benefit from therapies designed to address sex-specific immune responses or hormone-driven mechanisms.

In conclusion, the differences between males and females in terms of cancer prevalence, progression, and response to therapy are becoming increasingly recognized as important factors in developing more personalized cancer treatments. Understanding the underlying mechanisms of sex dimorphism in cancer is essential for improving patient outcomes, particularly in diseases like HNSCC where sex-related factors play a pivotal role. Personalized and sex-specific treatment approaches could hold the key to overcoming the current limitations in HNSCC management and improving long-term survival rates.

1.2.1 Loss or Extreme Down Regulation of Y Chromosome

The loss of the Y chromosome (LoY) or extreme downregulation of Y-linked genes (EDY) has emerged as an important phenomenon in cancer research, particularly in male-dominated cancers such as HNSCC[57,58]. These genetic or transcriptomic alterations are associated with tumor progression, immune evasion, and resistance to therapy[59–62]. In HNSCC, LoY has been shown to correlate with increased tumor aggressiveness, altered immune responses, and poor prognosis in male HNSCC patients[63]. It is also associated with distinct biological differences between male and female patients. Recent research, including cohort analyses, has shown that LoY occurs in about one-quarter of male HNSCC tumors, with higher prevalence in HPV-negative tumors, underscoring potential differences in tumor etiology[50]. Additionally, loss of Y chromosome expression impacts the regulation of several Y-linked genes, some of which are critical for male viability and could play a role in cancer biology[50].

LoY was initially identified as an age-related phenomenon in blood cells, where it was found to be one of the most frequent somatic mutations in aging males[60,64]. This observation led to its association with increased mortality risk and age-related diseases, including neurodegenerative disorders such as Alzheimer's disease[65]. However, subsequent studies revealed that LoY is not restricted to hematopoietic cells but is also present in a range of diseases, including cancer[57,59,66]. In tumors, LoY has been observed across diverse cell types, raising critical questions about its functional impact in different cellular compartments. While the depletion of Y chromosome genes in immune

cells has been linked to impaired immune surveillance and dysfunction, its presence in tumor cells has been associated with increased tumor aggressiveness and immune evasion[67–69]. The extent to which LoY contributes to tumor progression depends on the specific cellular context, making it essential to investigate in which cell populations this alteration exerts the most significant effects.

Recent studies across different cancer types have shown that LoY/EDY is not merely a marker of genetic instability but plays an active role in tumor biology. This alteration, observed in various cancers such as bladder cancer, has been linked to impaired immune responses, including reduced T-cell activation, which may contribute to tumor progression and resistance to therapy[69]. In male tumors, LoY/EDY could promote immune dysfunction, allowing tumors to evade immune surveillance. This suggests that males with LoY may have a diminished response to ICIs, which rely on immune activation to target cancer cells[70]. Additionally, LoY has been associated with the dysregulation of immune checkpoint pathways, making tumors more resistant to treatments like PD-1/PD-L1 inhibitors[71]. These findings highlight the potential of LoY as a biomarker for predicting treatment response and guide therapeutic strategies in cancer immunotherapy.

Further research into LoY, EDY and the molecular mechanisms behind Y chromosome function alteration is critical for understanding how these genetic or transcriptomic alterations affect tumor biology and immune responses. This research has the potential to uncover novel therapeutic targets for male patients and to improve patient stratification based on genetic and immune profiles. In particular, targeting the pathways disrupted by LoY/EDY could lead to more effective therapies for male patients with HNSCC.

1.3 Omics Technologies in HNSCC Research

Omics technologies, including genomics, transcriptomics, and epigenomics, have revolutionized the study of cancer by providing comprehensive insights into the molecular foundation of tumor biology. These technologies enable large-scale analyses

of genetic mutations, gene expression patterns, and epigenetic modifications, facilitating the identification of biomarkers and therapeutic targets[72–74]. In HNSCC, omics approaches have been important in uncovering the molecular alterations that contribute to disease progression, metastasis, and therapy resistance.

Genomic sequencing technologies, such as whole genome sequencing and targeted sequencing, have allowed for the identification of somatic mutations, copy number alterations, and structural variations in HNSCC tumors[12,75]. These genetic alterations can drive tumorigenesis and may contribute to resistance to targeted therapies[76,77]. Similarly, transcriptomics has revealed dysregulated gene expression patterns in HNSCC, often associated with poor prognosis and aggressive disease features[78,79]. Epigenomic approaches, including DNA methylation and histone modification profiling, have also provided valuable insights into the mechanisms regulating gene expression in HNSCC and its impact on tumor behavior and the immune microenvironment[80,81].

The integration of genomics, transcriptomics, and epigenomics has provided a deeper understanding of the molecular heterogeneity in HNSCC. This has led to the identification of distinct tumor subtypes with varying clinical outcomes[82–84]. However, traditional bulk RNA sequencing methods fail to account for the cellular heterogeneity within tumors, which can obscure critical molecular events. To address this challenge, single-cell RNA sequencing (scRNAseq) has emerged as a powerful tool for providing high-resolution insights into the cellular composition of tumors and the TME. scRNAseq allows for the characterization of individual cells, enabling researchers to identify previously unrecognized subpopulations of tumor, immune, and stromal cells, which contribute to cancer progression and therapeutic response[85].

1.3.1 Introduction to Single-Cell Transcriptomics

The introduction of single-cell transcriptomics has revolutionized the study of complex biological systems, particularly in cancer research. Unlike traditional bulk sequencing methods, which analyze mixed populations of cells, single-cell technologies allow for the dissection of individual cells, offering unparalleled insights into cellular

heterogeneity and the molecular processes driving cancer[86,87]. scRNAseq, chromatin accessibility sequencing, and spatial transcriptomics have become integral tools in studying tumor biology, immune cell interactions, and other cellular processes at a granular level[88,89].

The transition from bulk sequencing to single-cell technologies has significantly impacted the study of tumors. By enabling the analysis of individual tumor cells, scRNAseq has revealed distinct cell populations within the TME, including immune, fibroblast, endothelial, and cancer-associated stromal cells[87]. This technology has enabled a more comprehensive understanding of the cellular interactions that shape tumor progression and metastasis, as well as their contribution to treatment resistance.

Despite the promising potential of single-cell technologies, several challenges remain. These include inconsistencies in data annotation, experimental methodologies, and sample sizes across studies, which can complicate the interpretation of results. To address these challenges, community-driven initiatives such as Single-Cell Best Practices and scRNA-tools have been developed to promote standardized methodologies and improve data integration and analysis[90,91]. Additionally, with over a thousand computational tools dedicated to single-cell transcriptomics, including popular frameworks like Seurat and Scanpy, researchers now have robust methods to process, analyze, and visualize large-scale single-cell datasets[92,93].

In HNSCC, the application of scRNAseq has provided valuable insights into the immune landscape and stromal components of the TME[6,94]. By analyzing the TME at a single-cell resolution, researchers can identify novel immune and stromal cell types that interact with tumor cells, influencing their behavior and response to therapy. These insights are crucial for understanding the molecular interactions that drive cancer progression, immune evasion, and resistance to therapy. Additionally, integrating scRNAseq with other omics data types, such as genomics and epigenomics, offers a more comprehensive view of tumor biology, aiding in the identification of novel therapeutic targets and the development of precision medicine strategies for HNSCC[89,95]. The integration of spatial proteomics, which maps gene or protein expression patterns within

tissues, further enhances this understanding by providing a more detailed picture of how the TME affects tumor biology and treatment outcomes[96].

In this chapter, I will provide a comprehensive summary of the typical steps involved in analyzing single-cell transcriptomics, focusing on the methods I have applied throughout my research. This includes an overview of computational tools and best practices to enhance the rigor of data analysis. The insights from single-cell transcriptomics are crucial for advancing our understanding of HNSCC and improving patient outcomes through personalized therapeutic strategies.

Steps in Single-Cell Genomics Analysis

The process of analyzing single-cell genomics data can be broken down into several distinct steps (Figure 3). Each step is crucial for obtaining accurate, reliable, and meaningful results. The key steps in single-cell transcriptomics analysis, as applied in my research, are as follows:

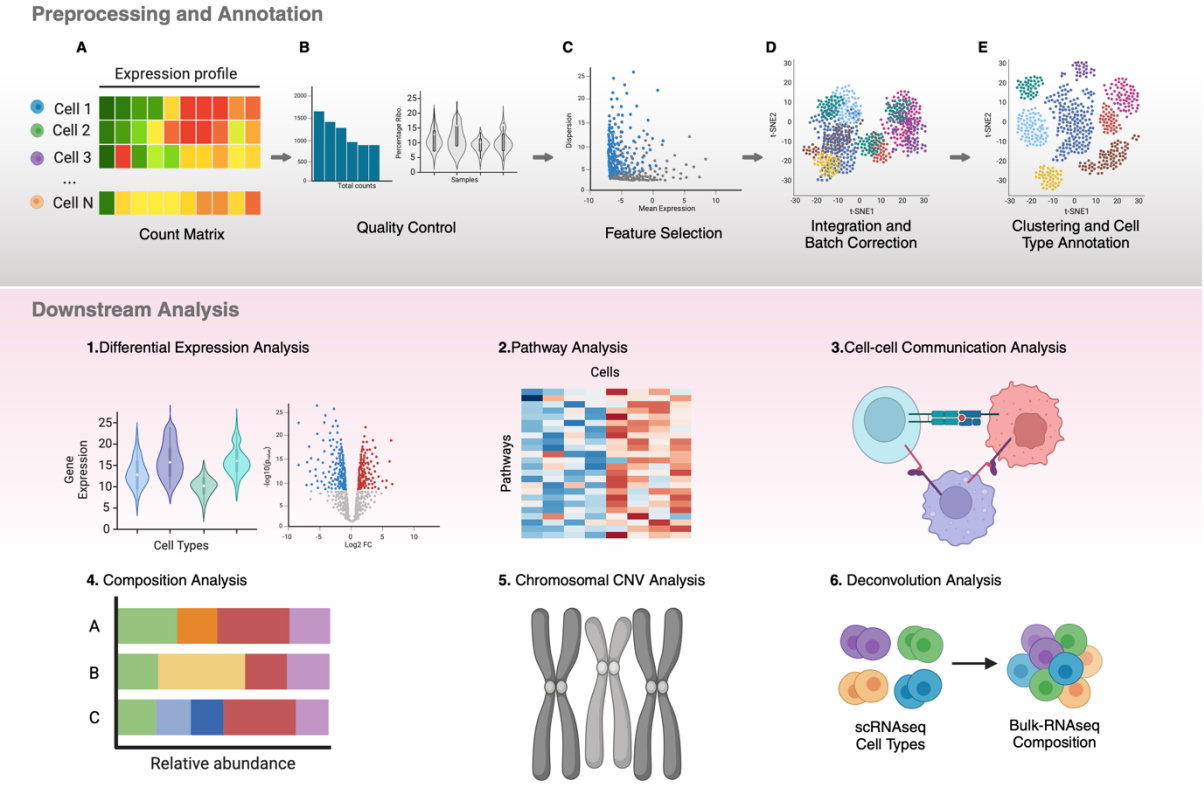


Figure 3. Overview of the single-cell RNA-seq analysis workflow. The figure summarizes the main steps of single-cell transcriptomic analysis, as applied in this study. The upper panel illustrates preprocessing and annotation, including generation of the count matrix *(A)*, quality control and filtering of low-quality cells *(B)*, feature selection of highly variable genes *(C)*, data integration and batch correction across datasets *(D)*, and clustering with subsequent cell type annotation *(E)*. The lower panel outlines key downstream analyses: (1) differential expression analysis between cell types or conditions, (2) pathway enrichment analysis, (3) cell–cell communication inference through ligand–receptor interaction modeling, (4) composition analysis across samples or groups, (5) inference of chromosomal copy number variations (CNVs), and (6) deconvolution of bulk RNA-seq data using scRNA-seq-derived reference profiles.

1. Data Pre-processing

Data preprocessing is the first and fundamental step in single-cell analysis. Raw data obtained from sequencing platforms typically require extensive cleaning before proceeding with further analysis. This includes filtering out low-quality cells, removing cells with low gene counts, and correcting for technical biases such as library size and batch effects. The goal of preprocessing is to ensure that the data is of high quality and

that any noise or artifacts from the sequencing process are minimized. During this step, cells are typically filtered based on the number of genes detected and the total number of unique molecular identifiers (UMIs), which helps to exclude dead or damaged cells[91].

2. Normalization

Normalization is a critical step that compensates for technical variations between cells, such as differences in sequencing depth. In scRNAseq, where sequencing depth can vary widely across individual cells, normalization methods are used to make gene expression levels comparable across cells. Common normalization methods include total-count normalization, log-transformation, and scaling[97]. A widely used approach in single-cell analysis is SCTransform, a normalization method implemented in Seurat, which accounts for gene-specific technical noise and reduces variability, allowing for downstream analyses like clustering to be more accurate[92].

3. Feature Selection

Feature selection identifies the most informative genes to use in downstream analyses. This step is crucial to reduce the dimensionality of the dataset and to avoid overfitting. Genes that are highly variable across cells are typically selected for further analysis because they are more likely to provide valuable information about cell types or states. Methods for feature selection often involve calculating the variance of each gene's expression across cells and selecting the top N genes that exhibit the most variation. This selection process ensures that the most biologically relevant signals are preserved for further analysis[92,98].

4. Dimensionality Reduction

After feature selection, dimensionality reduction is performed to reduce the complexity of the data while preserving important biological variation. This step allows researchers to visualize high-dimensional data in lower dimensions and to identify patterns such as clustering of similar cells. The most common methods for dimensionality reduction in

single-cell analysis include Principal Component Analysis (PCA) and t-SNE (t-Distributed Stochastic Neighbor Embedding), as well as UMAP (Uniform Manifold Approximation and Projection)[99–101]. These techniques allow us to explore the relationships between cells and visualize them in two or three dimensions, which is essential for interpreting the data.

5. Data Integration and Batch Correction

Data integration is a crucial step when working with multiple datasets, particularly when they originate from different experimental conditions or batches. The goal of this step is to remove batch effects while preserving biologically relevant differences, ensuring that datasets can be meaningfully compared. Common approaches include Harmony and Seurat's integration workflow[92,102]. Some methods allow for a semi-supervised integration that incorporates prior cell-type information to refine alignment, such as STACAS[103]. Batch correction methods ensure that differences observed between samples are not due to technical artifacts while maintaining biologically meaningful variation across datasets.

6. Clustering and Cell Type Annotation

Once the data has been preprocessed, normalized, and reduced in dimensionality, the next step is clustering cells into distinct groups based on their gene expression profiles. Clustering enables the identification of cellular subpopulations and provides insight into the heterogeneity of the tissue. Various clustering algorithms, such as Louvain or K-means, can be used to group cells with similar expression profiles.

After clustering, cell type annotation is performed to assign biological meaning to each cluster, typically by identifying known marker genes associated with specific cell types. This process can be done manually using literature or automated using databases like CellMarker or SingleR[104,105]. Some methods allow for more refined and systematic annotation by incorporating predefined marker sets and decision-tree-based classification, such as scGate and Ikarus[106,107]. These approaches help improve

annotation accuracy by filtering out unwanted populations and ensuring consistency across datasets.

Downstream Analysis

Once the cells have been clustered and annotated, several downstream analyses can be performed to explore biological questions in more depth. These analyses allow for the identification of key drivers of disease, immune interactions, and potential therapeutic targets.

1. Differential Expression Analysis

Differential expression (DE) analysis is used to identify genes that are expressed differently between clusters or conditions. This step is important for understanding how cell populations differ in terms of gene activity. By comparing expression levels across different clusters, researchers can identify markers that distinguish between cell types or states, which is crucial for understanding disease mechanisms. DE analysis typically involves statistical methods like DESeq2, edgeR, or limma to detect genes that show statistically significant changes in expression[108–110].

2. Gene Set Enrichment Analysis (GSEA)

Gene Set Enrichment Analysis is used to identify whether specific biological pathways or processes are significantly enriched in a given gene list, such as differentially expressed genes[111,112]. This method helps to understand the broader biological context of the data by linking genes to predefined gene sets. Tools like ClusterProfiler and Enrichr can be used for pathway enrichment analysis, offering insights into the functional roles of genes and their involvement in specific biological processes or diseases[113–115].

3. Cell-to-Cell Communication

Single-cell transcriptomics also allows for the study of intercellular communication by analyzing ligand-receptor interactions between different cell types. This analysis reveals how cells interact with each other and how signaling pathways might influence tumor progression or immune responses[116]. Tools like CellChat, CellPhoneDB and NicheNet can predict cell-to-cell communication by identifying potential ligand-receptor pairs that are active in specific cell populations, providing insights into cellular interactions and tissue homeostasis[117–119].

4. Copy Number Variant (CNV) Analysis

Copy number variants (CNVs) are important in cancer research as they indicate genetic alterations that may contribute to tumorigenesis. CNV analysis in single-cell transcriptomics involves detecting variations in the number of copies of specific genomic regions by comparing expression profiles between reference and target cell clusters[120]. Tools like XClone, sciCNV, copyKAT, and inferCNV can be used to infer CNVs[120–122]. These tools help classify cells as either healthy or tumor cells based on the presence or absence of specific CNV events. CNV analysis is especially useful in identifying tumorigenic alterations and understanding tumor heterogeneity.

5. Cell Deconvolution Analysis

Cell deconvolution is a technique used to estimate the proportions of different cell types in bulk tissue samples, which is crucial when studying heterogeneous samples like tumors. Deconvolution methods rely on a reference gene expression profile from annotated cell types, such as pseudobulked single-cell transcriptomics data, to infer cell-type proportions from bulk RNA sequencing data[123–125]. This approach is especially useful for repurposing existing large-scale datasets, such as The Cancer Genome Atlas (TCGA), for analysis of single-cell datasets, helping to validate findings from single-cell studies.

Single-cell technologies have made it possible to study HNSCC at a much higher resolution, revealing cellular and molecular features that bulk methods often miss. By

capturing gene expression and cell-cell interactions, these approaches help us better understand tumor heterogeneity, the immune and stromal landscape, and how these factors vary across patients. Together, these insights underscore the need to refine our understanding of HNSCC by accounting for the complex interplay between viral infection, inter- and intra-tumor diversity, sex-based biological differences, and the cellular architecture of the TME. The integration of bulk and single-cell omics technologies offers a powerful framework to uncover novel molecular mechanisms, identify prognostic biomarkers, and stratify patients more effectively. As the field moves toward precision oncology, deciphering how sex and TME factors shape disease progression is critical for developing targeted therapies and improving clinical outcomes in HNSCC.

2. Aims of the thesis

The central hypothesis of this thesis is that variations in sex chromosome gene expression significantly influence immune cell composition and stromal interactions within the TME, thereby impacting tumor progression and treatment response in HNSCC. To address this hypothesis, this study moves beyond conventional male vs. female comparisons and instead focuses on the role of sex chromosome dosage as a key biological factor shaping tumor behavior.

To test this hypothesis, the study integrates bulk and single-cell omics data to enable a detailed, cell-type-specific characterization of the TME. By stratifying patients based on HPV status and sex chromosome dosage, the aim is to uncover key immune and stromal cell populations that may serve as potential therapeutic targets. To achieve this, the study is structured around the following key aims:

1. Investigating Sex Chromosome Dosage Alterations in HNSCC
    - Characterize LoY and EDY in bulk RNA-seq data, determining their prevalence in HNSCC and across cancer types.
    - Assess whether LoY and EDY are tumor-driven events by comparing their occurrence in tumor versus normal tissue.
    - Evaluate their clinical significance, particularly in relation to HPV status and survival outcomes.
    - Establish sex chromosome dosage groups based on sex chromosome gene expression and stratify patients accordingly.
    - Characterize the composition of the immune and stromal compartments in bulk RNA-seq data according to sex chromosome dosage groups.
2. Building a Unified Single-Cell Atlas for HNSCC
    - Integrate multiple single-cell RNA-seq datasets into a standardized HNSCC atlas, enabling a cohort-wide analysis of tumor and microenvironmental differences.

- Ensure comparability between groups by harmonizing data across datasets, allowing for robust cross-study comparisons.
- Validate tumor cell classification and immune/stromal annotation using a combination of automated and manual curation approaches, ensuring accurate cell-type identification.

3. Enhancing the Resolution of Sex Chromosome Dosage Differences Using Single-Cell Transcriptomics
    - Translate the concept of sex chromosome dosage from bulk RNA-seq to single-cell data, enabling a more granular assessment of how these alterations manifest at the individual cell level.
    - Investigate whether Y chromosome gene downregulation occurs homogeneously or heterogeneously within tumors, identifying potential cellular subpopulations that retain or lose Y chromosome expression.
    - Leverage single-cell resolution to refine the classification of immune and stromal compartments, ensuring a more detailed characterization of differences between sex chromosome dosage groups.
    - Examine tumor-intrinsic and microenvironmental variability across groups, providing a higher-resolution framework for understanding tumor-immune-stroma interactions in HNSCC.
4. Identifying Targetable Cellular Interactions in the TME
    - Determine which immune and stromal cell types exhibit the most significant differences across groups, prioritizing those with known roles in tumor-immune interactions and therapy resistance.
    - Leverage single-cell transcriptomic data to assess cell-cell communication across sex chromosome dosage groups, identifying ligand-receptor interactions that may influence tumor progression.

By addressing these aims, this thesis provides new insights into the role of sex chromosome dosage in HNSCC, with implications for tumor biology, immune modulation, and patient stratification, ultimately contributing to a more personalized approach to cancer research and treatment.

3. Material and Methods

3.1 Bulk Methods

3.1.1 Bulk Transcriptomic Data Collection

I used RNA sequencing (RNA-seq) data from The Cancer Genome Atlas (TCGA) HNSCC (n = 500; 367 males, 133 females) cohort to investigate sex-based differences in gene expression[126]. Among these, 412 patients were HPV-negative (289 males, 124 females) and 88 were HPV-positive (58 males, 6 females). Data was downloaded using the `TCGAbiolinks` R package, specifically utilizing `GDCquery` to query RNA-seq data and `GDCdownload` to retrieve the files. Clinical data was also downloaded through `TCGAbiolinks`, allowing me to integrate patient characteristics with expression data[127].

3.1.2 Differential Expression Analysis

To analyze sex-based gene expression differences, I used the selected RNA-seq data from the TCGA HNSCC cohort, which included 367 male and 133 female patients. Differential expression analysis (DEA) was performed using three methods, DESeq2, edgeR, and Limma-Voom, each offering a unique approach to identifying expression differences, adding robustness to our findings[108–110]. Only genes identified as differentially expressed by all three methods were kept, reducing the chance of method-specific biases and providing a reliable set of sex-linked expression differences.

The raw count data underwent preprocessing to remove genes with low expression across samples. To normalize the data, I used edgeR's `calcNormFactors()` function, which accounts for library size variations. In DESeq2, I built a DESeqDataSet object with `DESeqDataSetFromMatrix()`, modeling sex as the main design variable, and used `DESeq()` to fit the model. Genes were ranked by adjusted p-values, with a threshold of FDR < 0.05 to identify significant changes. Similarly, edgeR analysis used a model matrix

to differentiate male and female samples, with dispersion estimates obtained via `estimateGLMCommonDisp()` and `estimateGLMTrendedDisp()`. For Limma-Voom, log-transformed counts-per-million (logCPM) were generated using the `voom()` function, followed by linear modeling with `lmFit()` to examine sex differences.

After each analysis, I extracted overlapping genes that met the significance criteria across all methods, selecting a final set of robustly differentially expressed genes between male and female samples for further examination of sex-linked expression patterns in HNSCC.

3.1.3 Copy Number Index (CNI) Calculation for Chromosome Analysis

To assess chromosomal aberrations, particularly in the Y chromosome, I used segmentation data from the TCGA cohort to compute the Copy Number Index (CNI) for each chromosome in every patient[126]. Following the methodology described by Hollows et al., I calculated the CNI by taking the weighted average of copy number segments, where each segment was weighted according to its genomic length[50]. Specifically, for each chromosome, the sum of segment copy numbers (multiplied by their lengths) was divided by the total length of the chromosome. For autosomal chromosomes, a CNI of approximately 2 was expected, indicating two copies. For sex chromosomes, expected CNIs were 2 for chromosome X in females and 1 for chromosomes X and Y in males.

To classify male patients as experiencing LoY, I established a threshold based on the CNI distribution for the Y chromosome. This threshold was set at the secondary peak of the CNI distribution, which occurred below 0.5, indicating that these patients were presenting less than half of the expected Y chromosome copy number. Male patients with a Y chromosome CNI below this threshold were classified as LoY, suggesting a partial or complete loss of Y chromosome copies in these individuals. This threshold allowed for a systematical identification of LoY events, enabling a more accurate comparison of chromosomal deviations across patients, particularly in cases where Y chromosome expression was lower than expected. This calculation and classification

provided a structured framework for detecting deviations from expected CNI values, allowing the identification of potential loss or amplification events specific to the Y chromosome in male samples.

3.1.4 Classification of Y Chromosome Gene Expression and Extreme Downregulation (EDY)

To assess Y chromosome-linked gene expression levels, I analyzed the expression of genes specific to the Y chromosome, including SRY, RPS4Y1, ZFY, TBL1Y, USP9Y, DDX3Y, UTY, TMSB4Y, NLGN4Y, KDM5D, and EIF1AY. These were the only Y chromosome genes expressed in the tumor samples analyzed. I applied Gene Set Variation Analysis (GSVA) using the `GSVA` package in R and the `gsva()` function with the following parameters: `method = "gsva", annotation = "org.Hs.eg.db", min.sz = 5, max.sz = 500, parallel.sz = 4,` and `kcdf = "Gaussian"`[128]. This approach generated pathway scores for Y-linked genes across both male and female patients.

To identify males with low Y chromosome expression, I set a threshold using the maximum expression levels observed in female patients, after assessing and removing any outliers. Male patients with scores below this threshold were classified as EDY, allowing to investigate potential regulatory differences in Y-linked gene expression across the cohort.

3.1.5 Survival Analysis

To examine the impact of Y chromosome gene expression on patient survival, I performed a survival analysis using overall survival (OS) as the primary endpoint. OS was defined as the time from diagnosis to death or last follow-up, limited to a five-year period. The analysis focused on male patients from the TCGA HNSCC cohort, who were categorized by levels of Y chromosome gene expression.

I used the `survival` and `survminer` R packages to generate Kaplan-Meier survival curves and conduct statistical comparisons between Y chromosome expression groups[129,130]. A survival object was created using the `Surv()` function from the survival package, which incorporated overall survival (OS) time (in months) and event status. Kaplan-Meier survival curves were then fitted with the `survfit()` function, stratifying by Y chromosome expression levels. Statistical significance between survival curves was assessed using the log-rank test, also available within the `survival` package. The `ggsurvplot()` function from the `survminer` package was used to visualize survival curves, displaying p-values to indicate the significance of observed differences across expression groups.

3.1.6 Validation Analysis: Other cohorts

CPTAC, CCLE and HIPO-HNC

To validate findings across multiple cohorts, I extended our analysis to the Clinical Proteomic Tumor Analysis Consortium (CPTAC), Cancer Cell Line Encyclopedia (CCLE), and HIPO-HNC datasets[131–133]. The CPTAC dataset provides detailed RNA-seq data for HNSCC cases annotated with sex, HPV status, and other clinical features, enabling us to examine Y chromosome gene expression across HPV-positive and HPV-negative samples within an annotated cohort.

In parallel, the CCLE dataset for HNSCC cell lines, which includes RNA-seq data annotated with sex and HPV status, served as an *in vitro* reference, allowing to compare Y chromosome expression patterns across cell lines in alignment with our TCGA findings. For both the CPTAC and CCLE datasets, I applied GSVA to assess Y-linked gene expression.

Furthermore, I incorporated data from the HIPO-HNC cohort, which consists of 79 HNSCC all HPV-negative samples with gene expression data available on the Gene Expression Omnibus (GSE117973). This cohort includes clinicopathological annotations

relevant to the study and has previously been analyzed for key pathway activations (e.g., EGFR and PI3K), providing a well-characterized sample set for validating gene expression and pathway activities related to the Y chromosome. The three datasets were processed similarly to TCGA data, applying GSVA for Y-linked gene expression and segmentation data for LoY status which enabled us the comparison with the TCGA dataset.

Retrieval and Analysis of Multi-Cohort TCGA Data

To broaden the analysis of the relationship between EDY and LoY, I included additional TCGA RNA-seq and genomic datasets from cancer types containing over 50 male patients. Selecting cohorts with more than 50 male samples ensured a wide spectrum of Y chromosome expression, increasing the statistical power of our findings across diverse cancer types. I used the `TCGAbiolinks` R package to retrieve RNA-seq expression data, segmentation data and relevant clinical annotations, including sex, tumor type, and other patient characteristics, across multiple TCGA cohorts[127].

The downloaded datasets were processed uniformly to calculate GSVA scores for Y chromosome-linked gene expression, employing the same EDY detection method as previously described. The LoY classification was conducted using the CNI approach, applying the threshold previously established to define chromosomal loss. These methodologies allowed consistent identification of EDY and LoY across all included cancer types, supporting a standardized comparison of chromosomal deviations and gene expression across the selected cohorts.

3.1.7 Classification of Patients by Sex Chromosome Dosage and Y Chromosome Gene Expression

To better our understanding of sex chromosome dosage beyond traditional male and female classifications, I developed a classification system based on Y chromosome gene expression levels. Y chromosome-linked gene expression levels were quantified using GSVA, with scores compared across male and female patients[128]. A threshold

based on expression levels in female patients was established, allowing us to identify male patients with Extreme Downregulation of Y chromosome-linked genes. Then male patients were ranked according to their GSVA score, top and low patients were selected. Patients were then categorized as follows: XX (female patients with no Y chromosome expression), XY (male patients with high Y chromosome expression), and XØ (male patients exhibiting low Y chromosome expression, mostly overlapping with EDY).

3.1.8 Classification of Cell Types in Bulk RNA-Seq Data Using xCell

To gain deeper insights into the composition of the TME, in bulk RNA-seq data, I applied xCell R package, a computational method that infers cell type abundances from transcriptomic data[134]. xCell utilizes gene signature enrichment analysis to estimate the relative proportions of various immune and stromal cell populations in a given sample. I applied xCell to bulk RNA-seq expression data from HNSCC patients in the TCGA cohort to deconvolute cell type proportions across tumors. This approach enabled the identification of major TME components, including tumor-infiltrating immune cells (e.g., macrophages, B cells, T cells, NK cells) and stromal populations (e.g., fibroblasts, endothelial cells). Cell type proportions were further analyzed in relation to previously defined patient classifications, enabling the identification of specific TME components enriched in XY, XØ, and XX tumors.

3.1.9 Statistical Analysis

All statistical analyses were performed using R (version 4.2.0). Depending on the type and distribution of the data, appropriate statistical tests were applied, including both parametric and non-parametric methods. Group comparisons were typically carried out using tests such as the Wilcoxon rank-sum test, Kruskal–Wallis test, or t-tests, and correlations were assessed using Spearman or Pearson correlation coefficients as appropriate. P-values were adjusted for multiple testing where necessary, and results were considered statistically significant at $p < 0.05$.

3.2 Single Cell Methods

3.2.1 Single-Cell Data Collection and Integration

To build a high-resolution single-cell atlas of HNSCC, I performed a comprehensive integration of multiple publicly available single-cell RNA sequencing (scRNA-seq) datasets. These datasets were carefully selected based on the availability of necessary clinical annotations, including sex and HPV status, to be able to explore the key questions researched in this manuscript. I systematically searched for datasets on the Gene Expression Omnibus (GEO) repository, ultimately identifying four suitable datasets, GSE234933, GSE182227, GSE164690, and GSE181919, based on the presence of such annotations like patient sex, HPV status, and sample origin (normal tissue, primary tumor, or metastasis)[32,94,135,136]. In total, the integrated dataset includes 79 patients with HPV and sex annotations, comprising 46 HPV-negative (32 males, 14 females) and 33 HPV-positive (31 males, 2 females) cases.

Some datasets were directly available as processed R objects, facilitating immediate loading, while others were provided in a three-part format (matrix, features, and barcodes), requiring the use of the `Read10X()` function in `Seurat` to construct the gene expression matrix accurately[92]. This allowed for a consistent starting point for integrating datasets from varying formats. I reviewed the clinical and experimental annotations to standardize metadata across all datasets, ensuring consistency in key features. For datasets provided in raw format, I incorporated metadata manually to align with our inclusion criteria, maintaining uniformity across cohorts. Datasets lacking sufficient annotations were excluded, reinforcing the quality of our integrated analysis.

Data Loading and Metadata Harmonization

The initial data processing involved importing each dataset into `Seurat` objects using `readRDS()`, followed by a series of metadata alignment steps to standardize column names and patient identifiers across studies. For instance, attributes such as *"Cell_Type"*, *"Source"*, and *"Patient"* were systematically adjusted within each dataset's

metadata, providing consistency across datasets and enabling reliable integration. I also refined metadata annotations, such as HPV status, using functions like `revalue()` from the `plyr` package, ensuring consistent representation across studies[137].

Quality Control and Pre-Processing

Quality control (QC) was applied to remove low-quality cells and prevent noise artifacts from damaged or apoptotic cells. First, I performed data normalization using `NormalizeData()` in `Seurat`, a step that scales each cell's expression data, minimizing technical variability in the UMI counts across cells. Next, I identified highly variable genes with `FindVariableFeatures()`, a critical step for detecting meaningful biological differences while disregarding noise. Cells with excessively high mitochondrial gene expression or low UMI counts were filtered out using Seurat's `PercentageFeatureSet()` and `subset()` functions, eliminating likely dead or dying cells.

Following this, I scaled the data with `ScaleData()`, which centers and scales gene expression values, making them suitable for downstream dimensionality reduction. Principal component analysis (PCA) was performed with `RunPCA()`, a step that identifies the main sources of variation in each dataset, enhancing the subsequent integration process by focusing on biologically relevant components.

Batch Correction and Data Integration

To ensure batch effects did not obscure true biological differences, I used multiple integration methods. Seurat's `FindIntegrationAnchors()` and `IntegrateData()` functions were employed to anchor datasets together by identifying and aligning shared biological signals across studies. `Harmony` was then applied to harmonize the datasets, effectively mitigating major sources of technical variability.

For cell-type alignment, I applied `STACAS`, a semi-supervised method that uses pre-established cell-type labels to guide the integration process[103]. This approach enabled finer adjustments to cell-type boundaries, ensuring that each cell cluster reflected consistent biological definitions across datasets. `STACAS` was specifically useful here, because it allowed the incorporation of known cell-type information to align cell clusters in a biologically meaningful way. I ran `Run.STACAS()` with `anchor.features` set to 1000 and `dims` to 1:30, allowing alignment across the primary features used for clustering. This method assured consistency in cell-type definitions, minimizing dataset-specific biases in the integrated dataset.

Following integration, I clustered the combined dataset using Seurat's `FindNeighbors()` and `FindClusters()` functions to reveal distinct cellular groups within the merged data. To visualize these clusters in a lower-dimensional space, I applied Uniform Manifold Approximation and Projection (UMAP) through `RunUMAP()`, which preserved the relationships between cells and facilitated clear and interpretable visualization of cellular clusters across the combined dataset. This approach allowed to retain both the fine cell-type distinctions and the broader biological patterns essential for our analyses.

Cell-Type Annotation and Marker-Based Refinement

To standardize cell-type annotation across the integrated datasets, I applied the `scGate` tool, which enabled consistent homogeneous relabeling by using cell-type-specific gene sets across all samples[106]. For `scGate`, I selected marker-based models optimized for cells within the TME, ensuring the robust identification of major cell types such as fibroblasts, and various immune cell types. These models are particularly useful in TME datasets due to their focus on distinguishing cell types that share complex interactions and closely related transcriptional profiles.

By running the `scGate()` function with model specifications targeted to TME contexts and utilizing four processing cores, I achieved accurate and homogeneous annotations across the datasets. This approach allowed a reliable cell-type identification by aligning

each cell to known marker genes, providing a consistent foundation for downstream analysis.

Tumor Cell Annotation Using the Ikarus Method

For the accurate annotation of tumor cells, I employed a supervised approach using the `Ikarus` method, a model-based classifier designed to detect tumor cells with high specificity across complex cellular landscapes[107]. Unlike semi-supervised approaches, which utilize both labeled and unlabeled data for prediction, Ikarus operates on a fully supervised framework. It applies pre-trained models on fully labeled data to distinguish tumor cells from non-tumor cells by utilizing gene signatures known to characterize tumor cell profiles.

The Ikarus method incorporates an automated tumor cell detection model that uses pre-labeled gene sets to identify cells with expression profiles specific to tumor biology. For this analysis, I used a core model and gene signature files from the `Ikarus` repository, which provide comprehensive, pre-established gene sets specific to tumor cell markers and was already validated in HNSCC. To prepare our integrated Seurat object for compatibility with Ikarus in Python, I streamlined the data using `DietSeurat()` and then converted it into the H5AD format using the `zellkonverter::SCE2AnnData` function.

Once in the H5AD format, I ran `Ikarus` in Python to perform automatic tumor cell identification across our HNSCC datasets. This fully supervised method enabled accurate detection by matching cells against pre-trained tumor cell signatures. To enhance specificity even further, I manually integrated the Ikarus-generated tumor cell labels with our pre-existing tumor cell annotations. This combined approach maximized both the reliability and precision of our tumor cell annotations, ensuring robust identification across the heterogeneous cellular landscape in HNSCC.

Subclassification of Immune Cells and Fibroblasts

Following the main cell-type annotations, I applied detailed subclassification to further characterize immune cells and fibroblasts within the TME. For immune cells, I utilized the ProjectTILS framework, which references pre-labeled immune cell maps to achieve detailed subtype classification of key immune cell types, including dendritic cells, CD4 T cells, and CD8 T cells[138]. Using `ProjecTILs.classifier()`, I matched each cell to reference immune cell signatures, achieving a refined view of immune cell diversity relevant to TME interactions.

For fibroblasts, I created a custom reference within ProjectTILS, incorporating well-defined markers for distinct fibroblast subtypes to capture the fibroblast heterogeneity that is particularly relevant in the HNSCC context. By implementing this customized reference, I achieved detailed annotation of fibroblast subtypes, informed by insights from recent studies on fibroblast roles in tumorigenesis[138,139]. This detailed subclassification of both immune cells and fibroblasts provided a robust framework to study cell-cell interactions and cellular dynamics within the TME, helping to reveal how these components vary across diverse clinical conditions.

3.2.2 Sex Chromosome Dosage Classification and Patient Stratification

To explore the impact of sex chromosome dosage on intercellular communication and TME composition, I employed a dual-level classification system: at the cellular level (XY/XØ cells) and the patient level (XX/XØ/XY patients).

Cellular Classification: XY and XØ Cells

Individual cells were classified based on their expression of Y chromosome genes. For each cell, I calculated the percentage of Y chromosome gene set expression relative to the total gene expression (denoted as pct_chrY). Cells with zero expression across all Y chromosome genes (pct_chrY = 0%) were classified as XØ cells, indicating an absence of active Y chromosome gene expression. Conversely, cells with any detectable Y chromosome gene expression (pct_chrY > 0%) were classified as XY cells. This binary

classification provided a robust framework for identifying and comparing cells within and across patients.

Patient Stratification: XØ, Intermediate, and XY Groups

To classify patients, we focused on tumor cells, which displayed the most pronounced Y chromosomal expression differences compared to other cell types in the TME. For each patient, tumor cells were aggregated to generate a pseudobulk expression profile, effectively summarizing the gene expression of the entire tumor cell population. Using these pseudobulk profiles, we performed GSVA on the Y chromosome gene set to calculate a chromosomal dosage score for each patient. This score provided a continuous measure of Y chromosome gene expression across the cohort.

To avoid confounding effects from HPV status, the analysis was performed independently for HPV-negative and HPV-positive patients, ensuring that stratification was not biased by differences related to viral etiology. Within each HPV group, male patients were ranked based on their GSVA scores and divided into three stratification groups to reflect varying levels of Y chromosome expression. The top tertile of patients, characterized by the highest GSVA scores, were designated as XY patients, indicative of a strong Y chromosome expression. The bottom tertile, with the lowest GSVA scores, represented XØ patients, reflecting minimal or extreme downregulation of Y chromosome expression. Patients in the middle tertile were classified as the Intermediate Group, exhibiting moderate levels of Y chromosome expression.

3.2.3 Analysis of Chromosomal Instability Using inferCNV

I used the R package `inferCNV` to evaluate chromosomal instability across sex chromosome dosage groups[140]. This tool estimates chromosomal copy number variations (CNVs) from scRNAseq data by comparing transcriptional signals to a reference population of normal cells. This approach allowed me to detect large-scale chromosomal gains and losses indicative of genomic instability.

I annotated normal stromal cells (e.g., fibroblasts and endothelial cells) as the reference population to establish a stable baseline for comparison. To prepare the data for inferCNV, I extracted raw gene expression matrices using a custom function, `as_matrix()`, to convert sparse matrices into dense format. I created annotation files that mapped each cell to its respective group (e.g., XX, XY, XØ, or reference) and used the `CreateInfercnvObject()` function to initialize the inferCNV analysis. I used a gene order file based on the hg38 genome assembly to map genes to their chromosomal locations.

I ran inferCNV with a cutoff of 0.1, enabling detection of subtle CNV signals typical of 10x Genomics data. To improve signal clarity, I applied denoising and Hidden Markov Model (HMM) filtering. Due to computational limitations, I analyzed each dataset separately but ensured consistency by using standardized annotations and processing steps across all datasets. For visualization, I used the `plot_cnv()` function to generate heatmaps of CNV profiles.

3.2.4 Cell-Cell Communication Analysis with CellChat

To investigate cell-cell communication in the TME, I employed the R packagge `CellChat`, a computational tool that predicts ligand-receptor interactions and infers communication networks from scRNAseq data[117]. My analysis focused on interactions between the homogeneously annotated cell types, with comparisons across sex chromosome dosage groups (XX, XY, and XØ).

The analysis began by preparing the scRNAseq data for input into CellChat. The dataset, preprocessed in Seurat, was split into subsets by chromosomal dosage using the `SplitObject()` function. The resulting subsets (XX, XY, and XØ) were analyzed separately to construct chromosomal group-specific communication networks. For this step, the RNA expression matrix was extracted using the `GetAssayData()` function, and metadata, including cell type annotations, was included. Using CellChat's default human ligand-receptor database (`CellChatDB.human`), I initialized communication

network inference with the `createCellChat()` function. The analysis involved identifying overexpressed genes (`identifyOverExpressedGenes()`) and ligand-receptor pairs (`identifyOverExpressedInteractions()`) within each chromosomal group. These steps were followed by projecting the gene expression data onto the human protein-protein interaction (PPI) network using `projectData()`. To refine the network, I filtered low-confidence interactions using `filterCommunication()`, with a minimum threshold of 10 cells contributing to each interaction.

The core of the analysis focused on quantifying pathway-specific communication. Using the `computeCommunProb()` and `computeCommunProbPathway()` functions, I calculated communication probabilities and aggregated these at the pathway level with `aggregateNet()`. Centrality scores, including outgoing (secretion activity) and incoming (reception activity), were computed for each pathway using `netAnalysis_computeCentrality()`. These scores provided a quantitative measure of pathway influence in mediating communication within the TME.

To compare communication patterns across chromosomal dosage groups, I merged the networks using the `mergeCellChat()` function and performed differential interaction analysis with `netVisual_diffInteraction()`. Visualization of results included heatmaps (`netVisual_heatmap()`), bubble plots (`netVisual_bubble()`), and circular diagrams (`netVisual_circle()`). To further investigate the functional relevance of these pathways, I performed differential expression analysis on ligand-receptor pairs using `identifyOverExpressedGenes()` and mapped these to the inferred networks with `netMappingDEG()`. Ligand-receptor pairs with significant differential expression were extracted using `subsetCommunication()`, enabling the identification of signaling interactions that were upregulated or downregulated in each chromosomal dosage group.

In addition to the overall cell type interactions, I conducted a fibroblast-specific analysis by focusing on subtypes, including iCAFs, myofibroblasts (mCAFs), and pericytes. Using

the `subsetCellChat()` function, I isolated these cell types and evaluated their communication patterns with tumor cells. Centrality and interaction strengths for fibroblast-specific pathways were calculated and compared across dosage groups.

Pathway Selection and Ranking

For each pathway, differences in centrality scores across sex chromosome dosage groups (XX, XY, and XØ) were evaluated to highlight differential activity. Centrality scores, were computed using the `netAnalysis_computeCentrality()` function in CellChat, as mentioned above. These scores represent both outgoing (signals sent by a cell type) and incoming (signals received by a cell type) communication contributions.

To identify pathways with significant differences, a custom scoring method was developed to rank their relative importance across groups. First, the absolute differences in centrality scores were calculated for each pathway between all pairwise group comparisons (XX vs. XY, XX vs. XØ, and XY vs. XØ). These differences were then combined into a single metric by taking the maximum fold change across the comparisons. This approach ensured that pathways with notable variation in at least one comparison were prioritized.

The pathways were ranked based on their maximum fold change scores, and the top 20 pathways exhibiting the most pronounced differences were selected for detailed analysis. Heatmaps of these pathways were generated using the `pheatmap` R package, clustering them based on their centrality scores[141]. This clustering allowed the identification of shared and unique signaling patterns, providing insights into how specific pathways drive interactions between tumor cells and other cell types, such as iCAFs.

Code availability

All scripts used for the analyses presented in this thesis are available at the following GitHub repository: https://github.com/DKFZ-E220/SexBias. This includes code for data

preprocessing, integration, annotation, and downstream analyses performed on both bulk and single-cell datasets.

4. Results

4.1 Analysis of Bulk Transcriptomic Data

My study started with an investigation of sex-based differences in gene expression using RNA-seq data from The Cancer Genome Atlas (TCGA)[126]. I conducted a DEA comparing samples clinically classified as male with those classified as female. To ensure the robustness of our findings, I used three distinct DEA methods: Limma, EdgeR, and DESeq[108–110]. Recognizing the variability in the stringency of these methods, I focused on the genes that were consistently identified across all three approaches. This stringent filtering resulted in 46 differentially expressed genes, which were subsequently analyzed in greater detail (Figure 4.A).

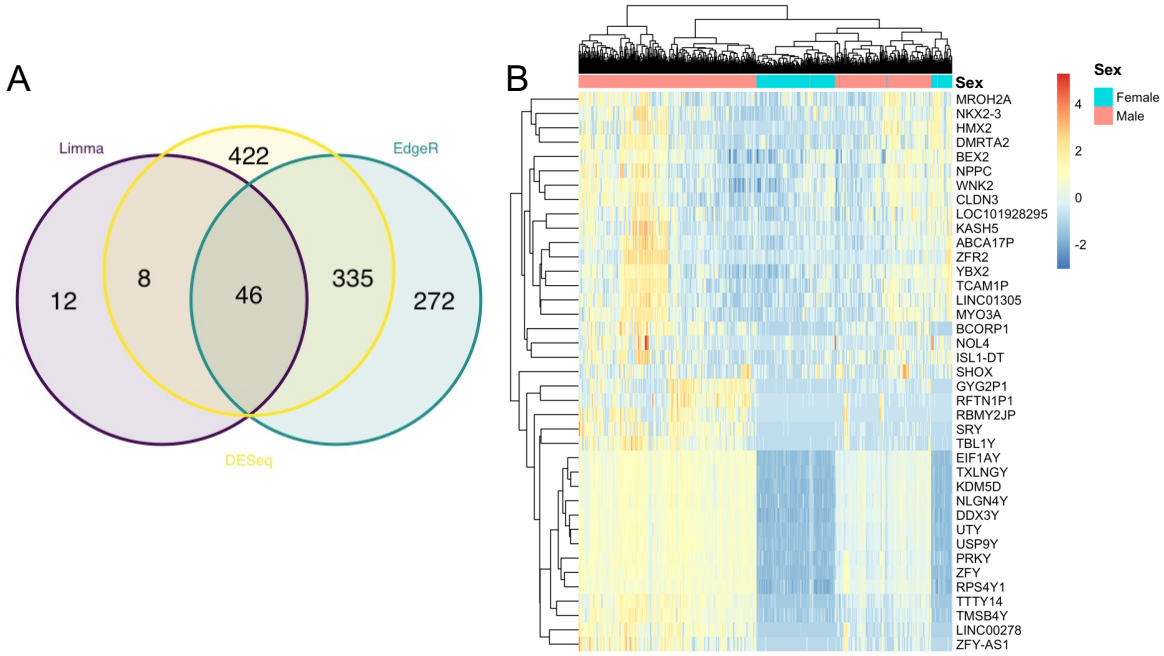


Figure 4. Identification of sex-biased genes and unsupervised clustering based on sex-linked expression profiles. *(A)* Venn diagram showing the overlap of differentially expressed genes (DEGs) identified by Limma, DESeq2, and EdgeR when comparing male and female samples in the TCGA HNSCC cohort. A total of 46 DEGs were consistently identified across all three methods and selected for further analysis. *(B)* Heatmap of DEGs across all samples, stratified by clinical sex. Unsupervised hierarchical clustering revealed a clear separation between male and female samples based on gene expression patterns, although a subset of male patients clustered with the female group, characterized by reduced expression of Y chromosome-linked genes.

An unsupervised hierarchical clustering of the values for the 46 selected genes across samples, as shown in the heatmap (Figure 4.B), revealed significant differences in gene expression between male and female patients. Interestingly, a subset of male patients clustered alongside the female cohort. When I examined it closer, I found that these males exhibited lower expression levels of Y chromosome-linked genes compared to those who did not cluster with the females. This observation suggests that sex chromosome dosage may play a critical role in modulating the expression of autosomal genes, particularly in the context of sex differences in cancer biology.

4.1.1 Copy Number Index Patterns in Male and Female Patients Reveals Loss of Y Chromosome in Males

Building on these transcriptomic findings, I made use of the genomic data available in the TCGA cohort to explore potential structural alterations in the Y chromosome and if this presented any association to the lower expression of the Y chromosome-linked genes. For each patient, I calculated the Copy Number Index (CNI) for each chromosome, including the Y chromosome, using the segmentation data obtained from the TCGA platform. The CNI was computed utilizing a method proposed by Hollows et al., this method takes the weighted average of the copy number segments per chromosome, where the weights correspond to the size of each segment[50]. Specifically, for each chromosome, the total copy number weighted by segment size was divided by the total size of all segments for that chromosome.

For autosomal chromosomes (1 to 22), the expected normal copy number is 2, corresponding to the two copies of each chromosome. For sex chromosomes, the expected copy number depends on the patient's biological sex: females typically have two X chromosomes (CNI of 2 for chromosome X and no Y chromosome), while males have one X and one Y chromosome (CNI of 1 for both X and Y chromosomes).

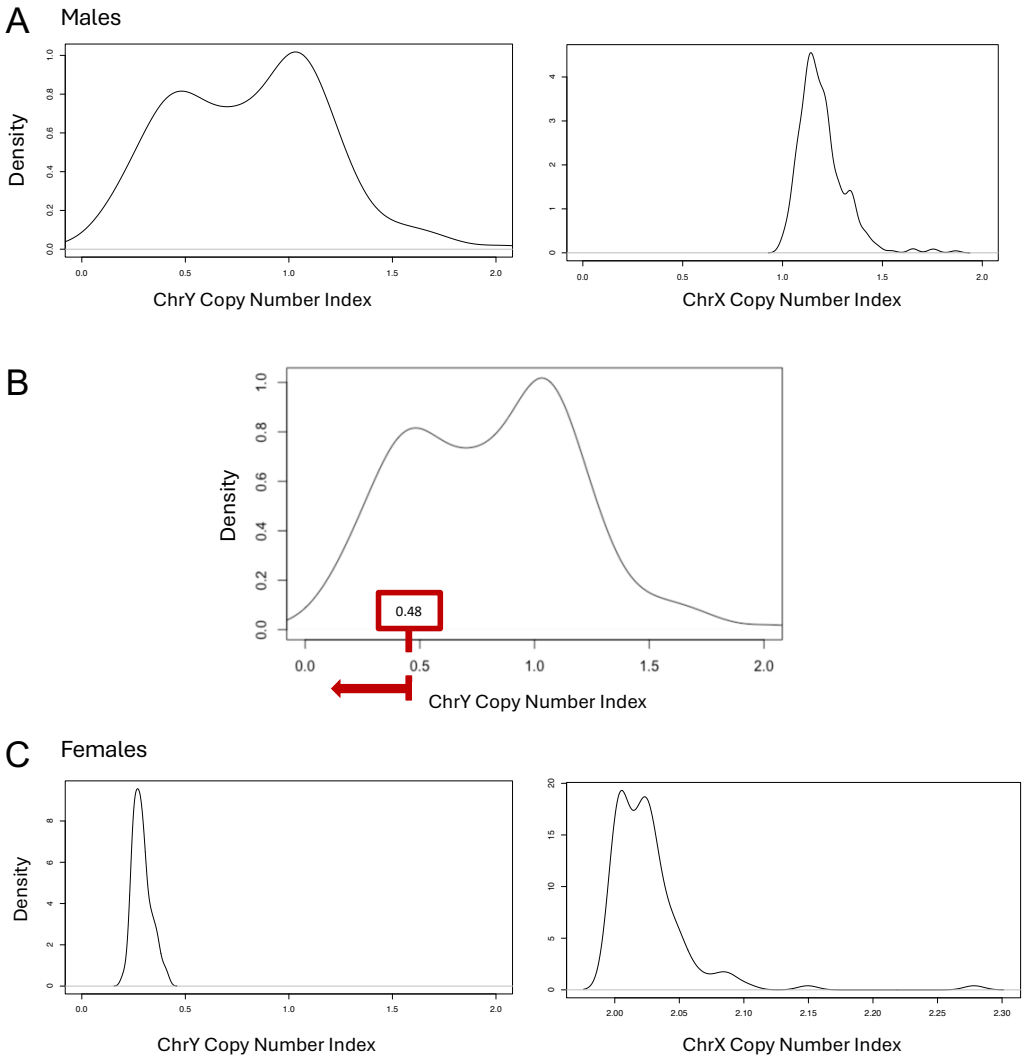


Figure 5. Copy Number Index (CNI) analysis reveals Y chromosome loss in male HNSCC patients. *(A)* Distribution of CNI values for chromosomes X and Y in male patients. The Y chromosome exhibits a bimodal pattern, with peaks near 1 and ~0.48, while the X chromosome shows a unimodal peak at the expected value of 1. *(B)* Classification of male patients based on Y chromosome CNI values. Patients with CNI values below 0.48 (indicated by the red threshold) were classified as experiencing LoY. *(C)* Distribution of CNI values in female patients. The Y chromosome shows a low residual signal (peak ~0.27), likely reflecting artifacts due to X-Y homology. X chromosome CNI values range from 2 to 2.3, consistent with the presence of two X chromosomes.

In male patients, the CNI distribution for the Y chromosome exhibited a bimodal pattern (Figure 5.A). One peak corresponded to the expected CNI value of 1, indicative of a single Y chromosome, while a second peak, centered around 0.48, suggested a subset of males with reduced Y chromosome copy numbers. Patients with CNI values below this secondary peak were classified as experiencing Loss of Y chromosome (LoY) (Figure 5.B).

In contrast, the CNI distribution for the X chromosome in male patients showed no abnormal deviations, aligning with the expected values (Figure 5.A).

For female patients, the CNI for the Y chromosome consistently remained below 0.45, with a peak at 0.27, confirming the expected near absence of the Y chromosome (Figure 5.C). This residual signal is likely due to homologous regions shared between the X and Y chromosomes, which may have been inaccurately classified during the segmentation process. As such, it is anticipated that the CNI for the Y chromosome in female patients is not strictly zero, but reflects these classification artifacts. The X chromosome, on the other hand, displayed a CNI slightly elevated above the expected value of 2, ranging from 2 to 2.3, but without significant abnormalities (Figure 5.C). These findings suggest that the primary sex chromosomal aberrations in this cohort happened predominantly in male patients, especially regarding the partial or complete loss of the Y chromosome.

4.1.2 Y Chromosome Gene Expression Classification Identifies Extreme Downregulation in Males

To keep the consistency with our genomic analysis, I next classified patients based on the expression of Y chromosome-linked genes. This approach allowed us to explore the relationship between genomic LoY and the expression of key Y-linked genes, SRY, RPS4Y1, ZFY, TBL1Y, USP9Y, DDX3Y, UTY, TMSB4Y, NLGN4Y, KDM5D, and EIF1AY, that were expressed in the cancerous tissues of male patients. These were the only Y chromosome genes exhibiting detectable expression levels above zero in these tissues. I applied GSVA scores to both male and female patients, facilitating a direct comparison of Y chromosome gene expression across sexes (Figure 6.A). GSVA is a method that estimates pathway activity by scoring gene sets within each sample, providing insight into differences at the gene set level across groups.

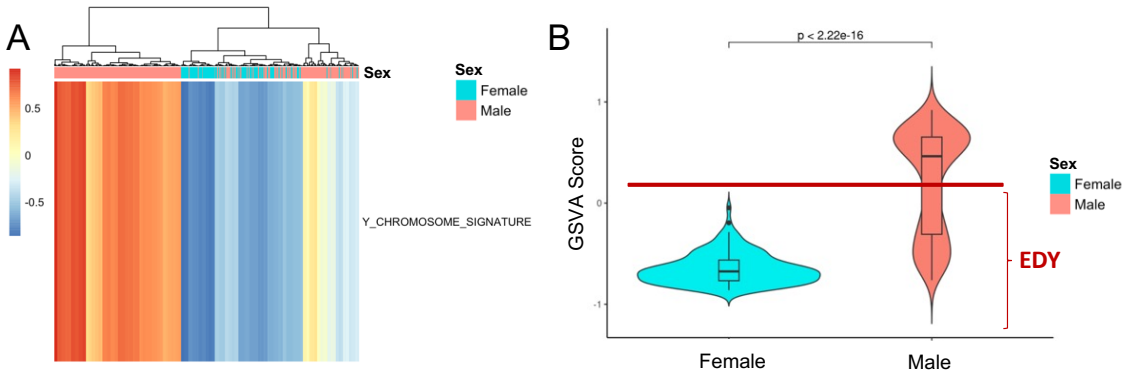


Figure 6. Classification of patients based on transcriptomic Y chromosome gene expression. *(A)* Heatmap showing GSVA scores for the Y chromosome gene signature across male and female patients in the TCGA HNSCC cohort. Unsupervised hierarchical clustering revealed clear sex-based separation, with a subset of male patients exhibiting expression profiles more similar to those of female patients. *(B)* Violin plot comparing GSVA scores for the Y chromosome gene signature between female and male patients. A red line indicates the highest GSVA score observed among females, which was used as a threshold for defining EDY. Male patients with scores equal to or below this threshold were classified as EDY.

Given that females exhibit a certain level of Y-linked gene expression, likely due to homologies between X and Y chromosome genes (as also noted in the probe classification of the segmentation data), I utilized Y-linked gene expression levels in female patients as a threshold for determining low Y chromosome expression in males (Figure 6.B). Male patients with GSVA scores equal to or lower than those observed in females were classified as Extreme Downregulators of the Y chromosome (EDY). This classification allowed me to investigate the correlation between genomic LoY status and transcriptomic EDY status in this patient cohort.

4.1.3 EDY is Closely Associated with LoY in Male Patients

To identify potential causal factors underlying the downregulation of Y chromosome genes, I aimed to overlap the LoY and EDY classifications for each male patient, examining how these phenomena interact at both the genomic and transcriptomic levels. Using a violin plot (Figure 7.A), I compared GSVA scores across three groups: female patients, male patients with LoY, and male patients without LoY. The GSVA

scores were significantly higher in males without LoY compared to those with LoY, indicating that the EDY phenomenon is primarily driven by the physical loss of the Y chromosome in LoY patients.

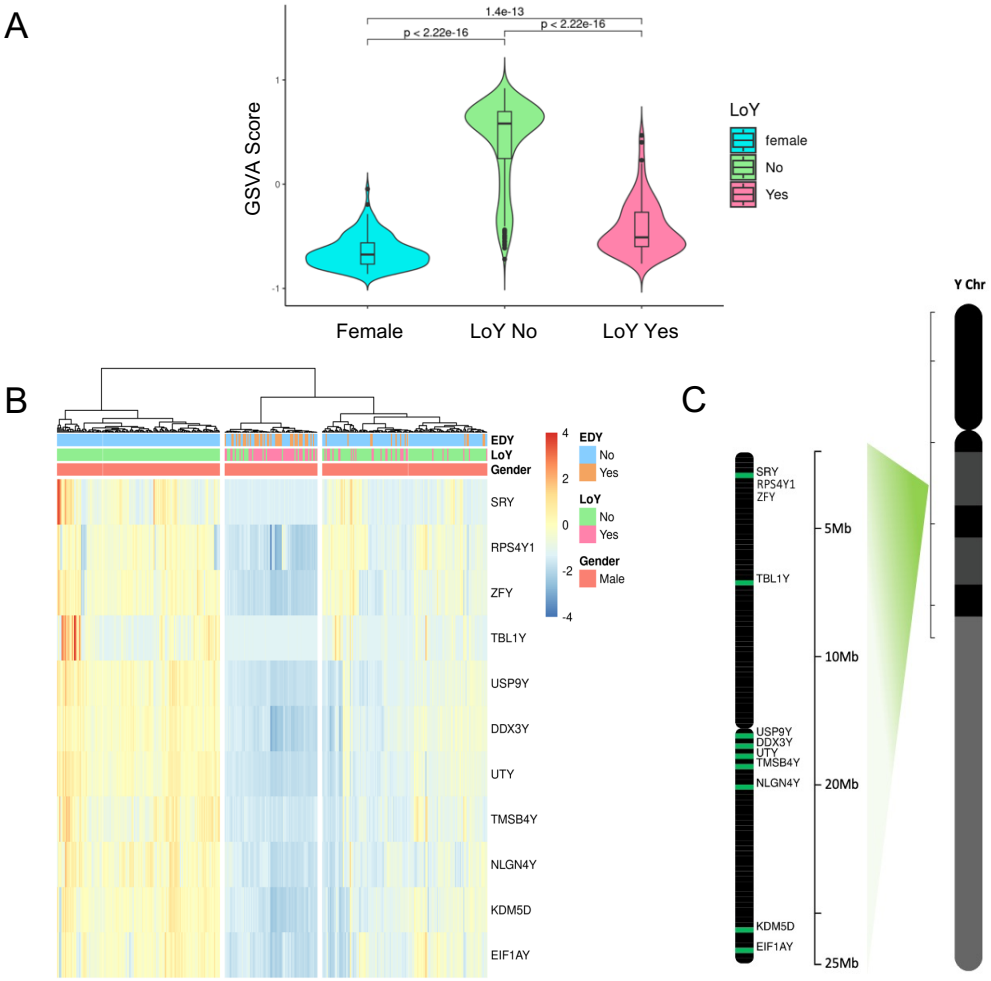


Figure 7. Integration of genomic LoY and transcriptomic EDY reveals their strong association in male HNSCC patients. *(A)* Violin plot comparing GSVA scores for the Y chromosome gene signature across three groups: female patients (cyan), male patients without LoY (green), and male patients with LoY (pink). Males without LoY exhibited significantly higher GSVA scores than those with LoY, indicating that EDY is primarily driven by Y chromosome loss. *(B)* Heatmap showing unsupervised hierarchical clustering of male patients based on the expression of 11 Y-linked genes. Patients with both LoY and EDY predominantly cluster in the low-expression group, while those without LoY/EDY tend to show higher expression. This clustering highlights the correlation between genomic and transcriptomic alterations and allows for classification into high, intermediate, and low Y gene expressors. *(C)* Genomic positions of the 11 Y-linked genes analyzed, plotted on the Y chromosome ideogram. Genes are distributed across the male-specific region of the Y chromosome (MSY), supporting the interpretation that structural alterations in this region may underlie the observed expression differences.

Further evidence for this connection is evidenced in the heatmap (Figure 7.B) based on unsupervised hierarchical clustering of male patients according to the expression of the 11 Y-linked genes. Males classified as both EDY and LoY predominantly clustered in the group of lower expressors, reinforcing the conclusion that LoY and EDY are closely linked. This heatmap also revealed that male patients could be categorized into High, Intermediate, and Low expressors of Y chromosome genes, based on RNA expression levels. Interestingly, patients in the Intermediate group tended to show selective downregulation of certain Y-linked genes, while still expressing others. To explore whether this partial downregulation followed any spatial pattern along the Y chromosome, I visualized the genomic positions of the 11 Y-linked genes (Figure 7.C). These genes are distributed across different regions of the male-specific Y chromosome, and no clear clustering of expression loss within a specific chromosomal region was observed. This suggests that partial downregulation in the Intermediate group does not result from focal deletions but may instead involve heterogeneous regulatory mechanisms affecting subsets of Y-linked genes.

Since some samples exhibiting EDY were not associated with LoY, I explored the possibility that epigenetic mechanisms, such as DNA methylation, might be responsible for the reduction in Y chromosome gene expression. To investigate this, I analyzed DNA methylation patterns in patients with EDY who did not display LoY. However, our analysis did not reveal a correlation between methylation and Y-linked gene expression in these cases, suggesting that other regulatory mechanisms may contribute to the downregulation of Y chromosome genes in this subset of males.

Additionally, these findings were validated across multiple datasets, including other HNSCC cohorts like the CPTAC and HIPO cohorts and HNSCC cell lines from the CCLE (Suppl. Fig. 1-2). This validation supports the conclusion that a subset of male patients consistently exhibits Extreme Downregulation of the Y chromosome 11 gene set, with the majority of these cases overlapping with LoY, the physical loss of the Y chromosome.

4.1.4 EDY and LoY Are Present and Overlapping Events Across Multiple Cancer Types

To extend our understanding of the relationship between EDY and LoY, I analyzed RNA-seq and genomic data from all TCGA cohorts with more than 50 male patients, incorporating a multi-project analysis. Our findings show clear patterns that support LoY as a major driver of EDY and reveal how its presence varies across different cancer types. Some illustrative examples of Y chromosome gene set expression and its overlap with LoY status are shown in Suppl. Fig. 3, where male patient samples from ESCA, LSCC, and STAD display transcriptional patterns similar to those observed in HNSCC.

When I examined the GSVA scores for Y chromosome-linked gene expression across projects, I observed that female patients consistently displayed lower scores, as expected due to the absence of Y chromosome expression (Figure 8.A). In contrast, male patients exhibited a broader range of GSVA scores, which varied across cancer types. Certain projects showed considerable overlap in the GSVA distributions between male and female patients, while in others there is almost no overlap, meaning that there are no male patients having a very low expression of the Y chromosome. This variability points to the need for further investigation into cancer type specific drivers influencing Y chromosome-linked gene expression.

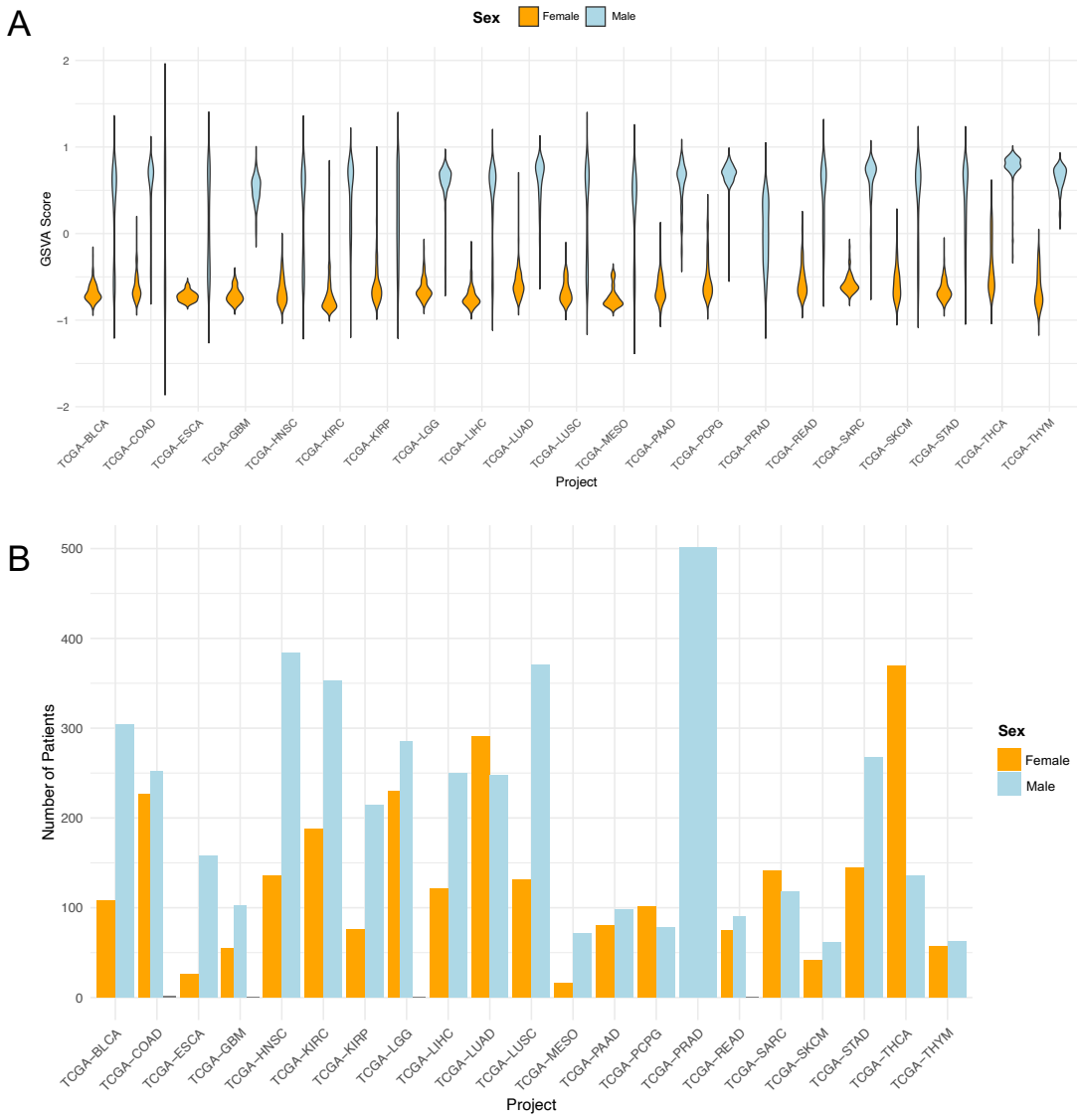


Figure 8. Expression of Y chromosome genes and sex distribution across TCGA cohorts. *(A)* GSVA scores for the Y chromosome gene set across all TCGA cancer types with more than 50 male patients. Female patients (orange) consistently show low scores, while male patients (light blue) display variable expression levels depending on the cancer type. Some cohorts show considerable overlap in GSVA scores between sexes, while others display distinct separation, suggesting variability in Y chromosome gene downregulation across tumor types. *(B)* Number of male and female patients per TCGA project. Sex representation varies across cohorts, with some (e.g., TCGA-HNSC, TCGA-LUSC) being male-dominated, which may influence the observed frequency of EDY events. Projects with a more balanced sex distribution provide additional context for interpreting expression-based classifications.

I analyzed the number of male and female patients per project since it can provide context to our EDY findings (Figure 8.B). I observed that some cohorts, such as TCGA-

LGG and TCGA-COAD, had a more balanced representation of sexes, whereas others, including TCGA-HNSCC and TCGA-LUSC, exhibited a male-dominated composition. This imbalance is significant because projects with more male representation often showed a higher number of EDY cases.

The percentage of EDY males, defined by those with GSVA scores lower than the highest female score when corrected for outliers within the same project, varied significantly (Figure 9.A). I found that in many cancer types, a considerable fraction of the male cohort exhibited EDY, showing that is widespread event. In projects where EDY was more prevalent, I often observed a much larger number of male patients compared to females, suggesting a potential link between cohort composition and the observed EDY rates.

Finally, the overlap analysis between EDY and LoY (Figure 9.B) indicated that while the majority of male patients did not exhibit either event, more than half of those identified as EDY also presented LoY. This supports the hypothesis that LoY, characterized by the physical loss of the Y chromosome, may be a significant contributor to EDY. However, as seen before, the EDY cases without LoY points towards additional regulatory mechanisms influencing Y chromosome gene expression beyond chromosomal loss.

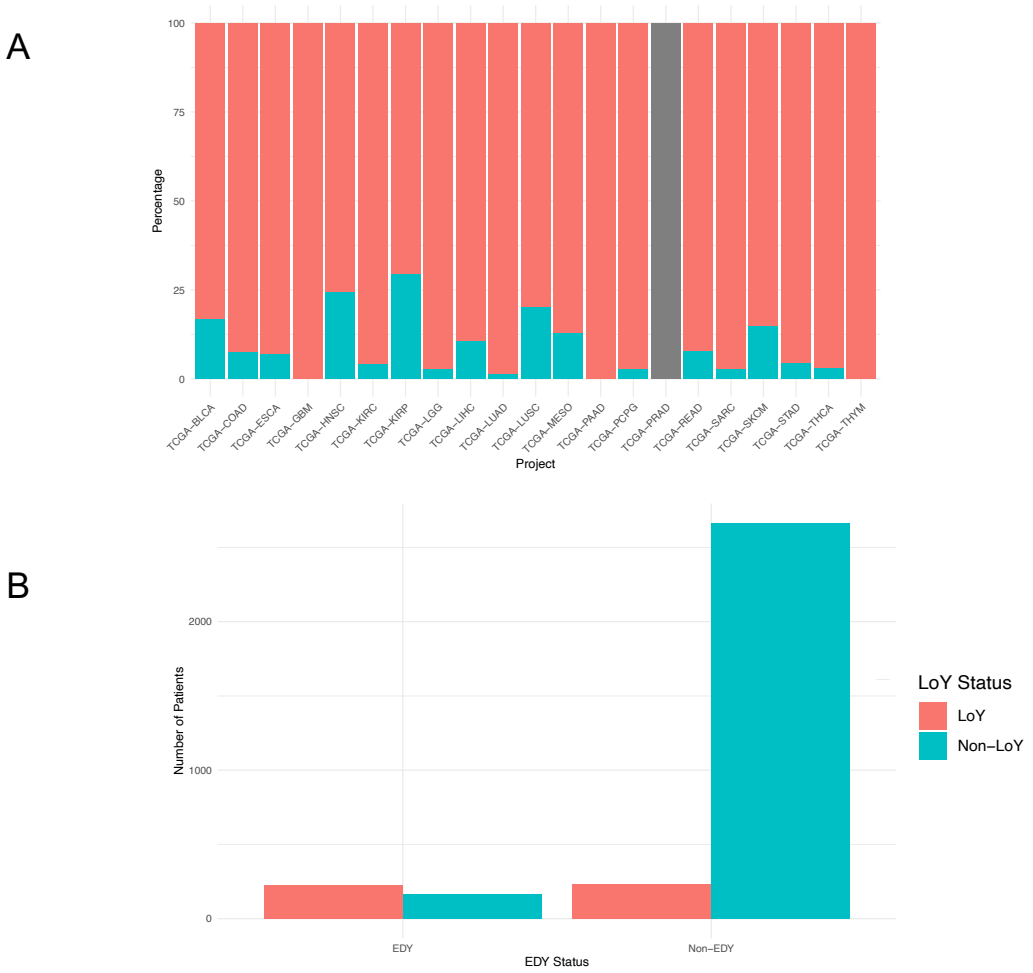


Figure 9. Prevalence and overlap of EDY and LoY events across TCGA cancer types. *(A)* Percentage of male patients classified as EDY across TCGA cohorts with more than 50 male patients. EDY was defined using cohort-specific thresholds based on female GSVA scores. A substantial proportion of males exhibited EDY in several cancer types, suggesting this is a widespread event. In many cases, higher EDY prevalence was observed in male-dominated cohorts, indicating that sample composition may influence observed EDY rates. *(B)* Overlap analysis between EDY and LoY classifications across all cohorts. Most male patients were classified as non-EDY and non-LoY. However, among those identified as EDY, more than half also exhibited LoY, supporting the link between genomic Y chromosome loss and transcriptomic downregulation. The remaining EDY cases without LoY suggest the involvement of additional regulatory mechanisms beyond chromosomal loss.

These findings highlight the presence of EDY across various cancer types and emphasize its significant role in Y chromosome gene regulation. While our analysis confirms that LoY is a major contributor to EDY, it is not the only factor involved. This points towards the importance of EDY as a critical feature in cancer biology and to the need for further research into additional regulatory mechanisms influencing Y chromosome expression.

4.1.5 LoY and EDY Are Predominantly Found in HPV-Negative Patients, with Higher Y Chromosome Expression in HPV-Positive Cases

Another important finding in our study is the observed prevalence of LoY/EDY in HPV-negative patients, with these phenomena being almost absent in HPV-positive cases. To validate this observation, I compared Y chromosome gene expression levels between HPV-negative and HPV-positive male patients using GSVA scores. This comparison revealed a significantly higher expression of Y chromosome genes in HPV-positive cases (t-test, $p = 0.0000225$), supporting the idea that Y expression is largely preserved in these tumors (Figure 10.A). To explore Y chromosome expression patterns, I divided the male patients into three groups, High, Intermediate, and Low Y chromosome expressors, using unsupervised hierarchical clustering based on Y-linked gene expression. This stratification is visualized in the heatmap (Figure 10.C), where patient HPV status is indicated in the annotations, and the enrichment of HPV-positive patients in the High expressor group is evident. In parallel, I evaluated the distribution of HPV status across the Y expression stratification groups. While HPV-negative patients were more evenly distributed across the three categories, HPV-positive patients were strongly enriched in the High expressor group. This association was statistically significant (Fisher's exact test, $p = 0.0000331$), further reinforcing the link between HPV status and Y chromosome gene expression levels (Figure 10.B).

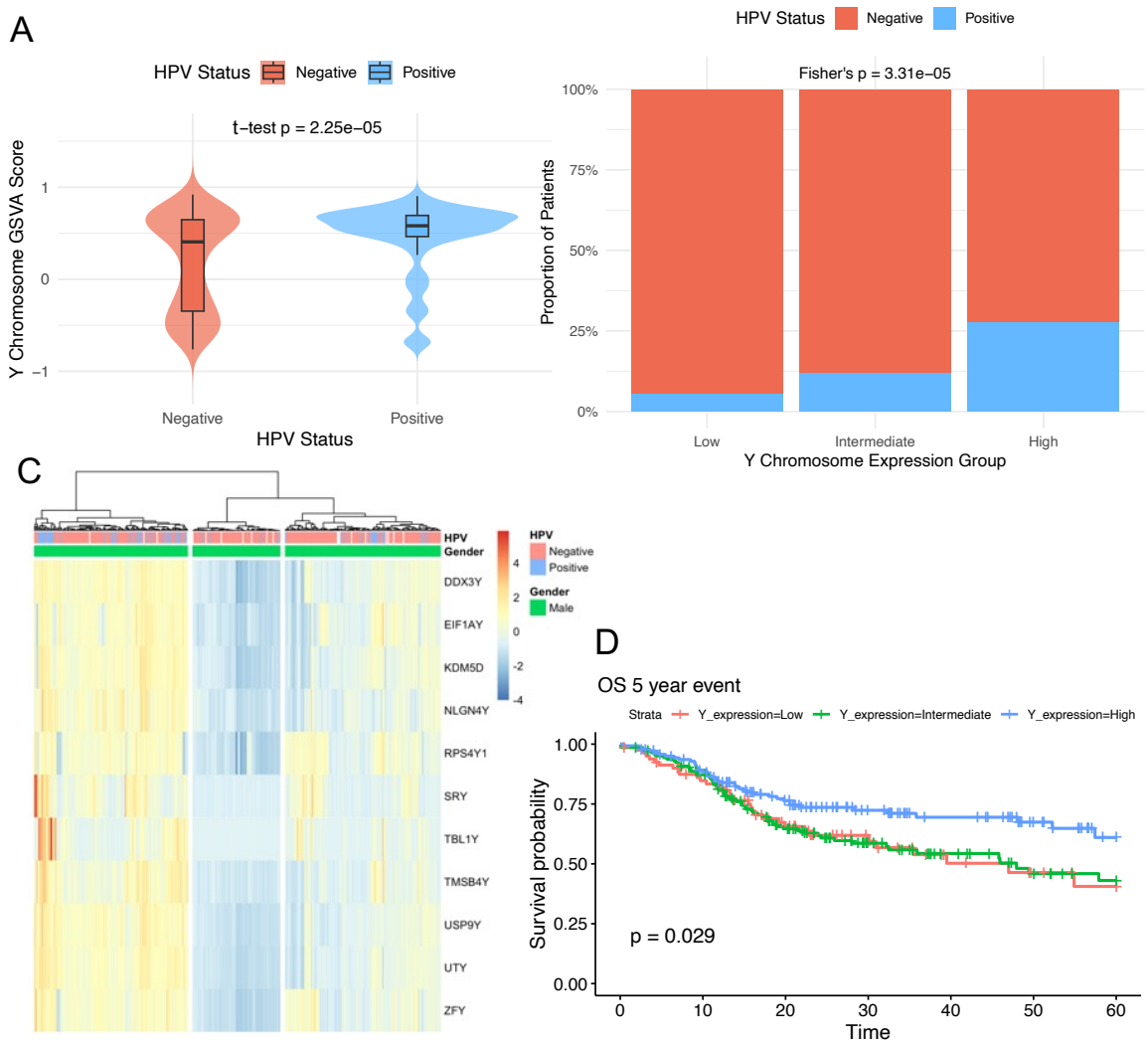


Figure 10. Y chromosome expression is enriched in HPV-positive tumors and is associated with improved survival. *(A)* Violin plot comparing GSVA scores for Y chromosome gene expression between HPV-negative (coral) and HPV-positive (steel blue) male patients. Y chromosome expression was significantly higher in HPV-positive cases (t-test, *p* = 2.25e-5). *(B)* Stacked bar plot showing the distribution of HPV status across Y expression groups. HPV-positive patients were predominantly classified as High expressors, while HPV-negative patients were more evenly distributed (Fisher's exact test, *p* = 3.31e-5). *(C)* Heatmap of Y chromosome gene expression in male patients, clustered into three groups (High, Intermediate, Low) using unsupervised hierarchical clustering. Annotation bars indicate HPV status, revealing enrichment of HPV-positive cases in the High expression group. *(D)* Kaplan–Meier survival analysis comparing overall survival among Y expression groups. Patients in the High expression group showed significantly better survival outcomes (p = 0.029).

After establishing these groups, I performed a survival analysis and discovered that patients classified as high expressors of Y chromosome genes had significantly better

survival outcomes (Figure 10.D). Upon further investigation, I observed that the high expressor group was predominantly enriched with HPV-positive patients. While I could not definitively conclude that high Y chromosome expression directly contributes to improved survival, our results indicate a clear trend: HPV-positive patients tend to exhibit higher Y chromosome expression levels.

I then extended our analysis to the CPTAC cohort, which consists exclusively of HPV-negative cases. In this cohort, the differences in survival across the three groups (High, Intermediate, and Low) were no longer apparent (Figure 11.A-B). Similarly, when I restricted our analysis to only HPV-negative patients within the TCGA cohort, the difference in survival became nonsignificant among groups, reinforcing the idea that the presence of HPV plays a key role in the expression patterns of Y chromosome genes.

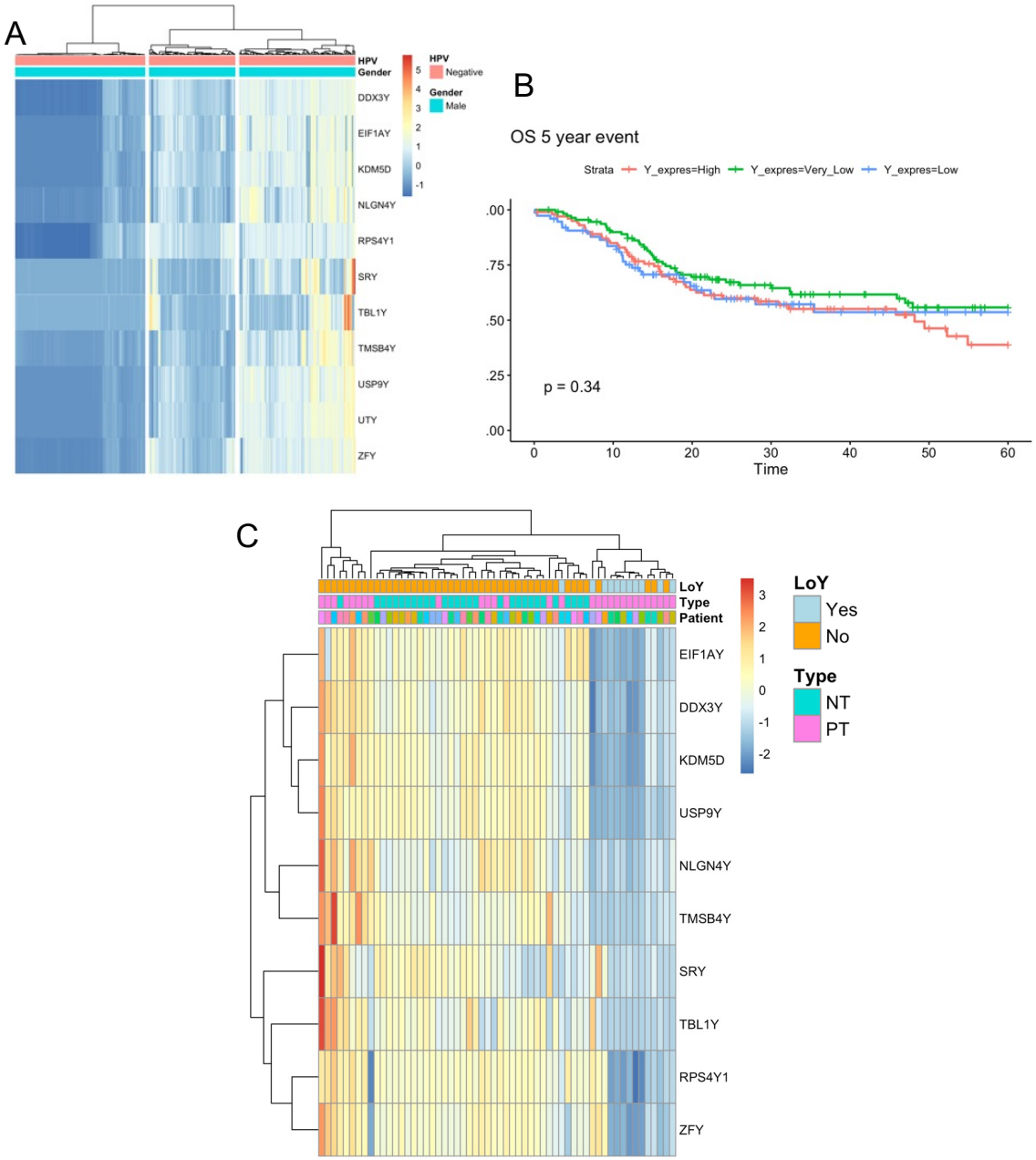


Figure 11. Y chromosome expression does not impact survival in HPV-negative cohorts and is predominantly altered in tumor tissue. *(A)* Heatmap of Y chromosome gene expression in the CPTAC cohort (HPV-negative only), stratified into High, Intermediate, and Low expressor groups via unsupervised clustering. *(B)* Kaplan–Meier survival analysis for the CPTAC cohort comparing overall survival across Y expression groups. No significant differences in survival were observed (log-rank p = 0.34). *(C)* Heatmap of Y-linked gene expression in the TCGA cohort comparing matched primary tumor (PT) and adjacent normal tissue (NT) from the same patients. Y-linked gene expression was consistently lower in tumor tissue compared to adjacent normal samples, suggesting that EDY is a tumor-specific event.

Additionally, I examined the expression of Y-linked genes in adjacent normal tissue and matched primary tumors (Figure 11.C). In these comparisons, I observed that adjacent normal tissues exhibited a lower incidence of EDY compared to the primary tumors, even in patients who showed significant EDY in their tumor samples. This suggests that the downregulation of Y chromosome genes is primarily a tumor-specific phenomenon and is less pronounced in normal tissue.

4.1.6 Patients Can Be Stratified by Sex Chromosome Dosage to Reveal Distinct Biological Profiles

Based on the findings from our analyses, it became clear that classifying patients according to their sex chromosome status, rather than relying only on clinical annotations of biological sex, provided a more accurate framework for exploring sex-related differences in HNSCC. This approach allowed us to refine the stratification of patients beyond traditional classifications based on clinical sex, providing a more biologically representative model of sex. To improve the clarity of these distinctions, I chose to focus on more extreme groups by excluding the intermediate expressors of Y chromosome genes. Although this decision resulted in a reduction in the number of patients included in the final analysis, it allowed for sharper contrasts between the defined groups, providing a clearer understanding of the underlying phenomena.

With this approach, I developed a classification system that stratifies patients according to their sex chromosome dosage state. This classification includes female patients, expected to show no Y chromosome expression, grouped as XX. Male patients with high expression of Y chromosome genes, are classified as XY. Finally, male patients exhibiting low Y chromosome gene expression, characterized by Extreme Downregulation of Y chromosome genes (EDY), are grouped as XØ.

Our findings consistently demonstrated that EDY and LoY are closely related events, with LoY serving as a primary driver of Y chromosome downregulation in the majority of cases. Additionally, I found that EDY is predominantly present in HPV-negative patients, while

HPV-positive patients tend to show a higher Y chromosome expression, further highlighting the complex interplay between viral oncogenesis and sex chromosome dynamics. Importantly, validation of these results across multiple cohorts, including the CPTAC and CCLE datasets, reinforces the robustness of our conclusions. By adopting this classification system and narrowing the focus to extreme expressors, I now have a precise tool to explore the biological implications of Y chromosome loss or downregulation.

4.1.7 Cell Composition Varies by Sex Chromosome Dosage

To investigate how sex chromosome dosage influences cell composition, I analyzed xCell-derived scores and performed Kruskal-Wallis tests to assess statistical significance between groups. I identified several immune cell types that differed significantly ($p < 0.05$) across sex chromosome dosage groups and examined the directionality of these changes (Figure 12). A summary of the mean enrichment scores for each group is provided in Suppl. Table 1.

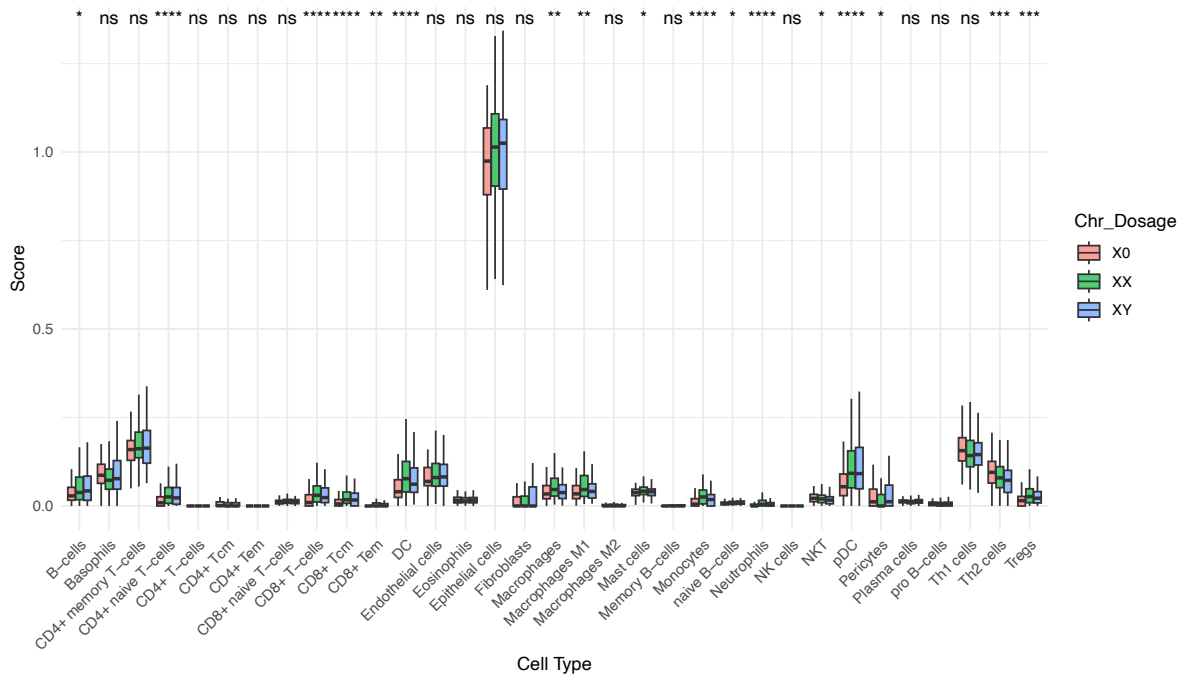


Figure 12. Immune composition varies across sex chromosome dosage groups.

Boxplot of xCell-derived cell type enrichment scores across XØ, XY, and XX groups. Several immune and stromal cell types showed significant differences in abundance based on sex chromosome dosage (Kruskal–Wallis test). Asterisks indicate significance levels: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****).

I observed that B cells (p = 0.0419) followed a progressive increase in abundance from X0 (lowest) to XY (highest), suggesting a dosage-dependent effect. Similarly, CD4+ naïve T cells (p = 0.000031) were most abundant in XX individuals, followed by XY, and showed the lowest scores in X0 individuals, supporting the idea that X chromosome dosage influences T cell differentiation and function. I also found significant differences in CD8+ T cell subsets, where CD8+ naïve T cells, CD8 central memory T cells (CD8Tcm), and CD8 effector memory T cells (CD8Tem) varied across sex chromosome dosage groups. XX individuals showed the highest proportions across all CD8+ populations, while X0 males consistently displayed the lowest levels. These differences suggest a trend in CD8+ T cell abundance that may be influenced by sex chromosome dosage.

Beyond T and B cell populations, I observed notable myeloid lineage differences, with monocytes, neutrophils, dendritic cells (DCs), and macrophages (global and M1 subsets) displaying significant variability across sex chromosome dosage groups. XX

individuals showed the highest proportions across these myeloid populations, while X0 males consistently exhibited the lowest levels. Additionally, I found significant differences in regulatory T cells (Tregs), natural killer T cells (NKT), and T-helper 2 (Th2) cells, further emphasizing the impact of sex chromosome dosage on immune composition. Tregs were most enriched in XX individuals, aligning with prior observations that higher X chromosome dosage correlates with increased regulatory immune functions. In contrast, NKT and Th2 cells showed the highest proportions in X0 individuals, suggesting a distinct immune profile in this group.

Overall, my findings highlight distinct immune composition patterns associated with sex chromosome dosage, suggesting that chromosomal dosage differences influence immune regulation and TME dynamics.

4.2 Analysis of Single Cell Transcriptomic Data

Single-cell transcriptomics has become an essential tool in our research, given that sex chromosome expression occurs at the level of individual cells. Single-cell transcriptomics allows us to extrapolate the concept of sex chromosome dosage at the level of individual cells, allowing us to distinguish between different cell types and assess how variations in sex chromosome expression manifest within the TME.

4.2.1 HNSCC Atlas Construction

Our study involves constructing a comprehensive single-cell atlas of HNSCC by integrating data from multiple single-cell experiments. By pooling data from various studies, I increase statistical power and improve the robustness of our analysis, enabling the detection of non-abundant cell populations and subtle transcriptional changes that might be missed in isolated datasets. This approach addresses the limitations of individual datasets by combining findings into a unified ‘master atlas,’ providing a more comprehensive understanding of the cellular heterogeneity in HNSCC. A manuscript detailing the construction and analysis of this atlas has been submitted to *Communications Medicine*.

To achieve this, I conducted an extensive search for publicly available single-cell RNA-seq datasets that include samples from HNSCC. All identified HNSCC datasets with single-cell RNA-seq data are listed in Supplementary Table 2**,** extracted from Conde-Lopez et al. (2024)[85]. From this pool, I selected four datasets, GSE234933, GSE182227, GSE164690, and GSE181919, based on the presence of appropriate clinical annotations, including sex, HPV status, and the origin of the sample (whether normal tissue, primary tumor, or metastasis). Prior to integration, I conducted a meticulous review and validation of all clinical and experimental annotations to ensure consistency and homogeneity across the datasets. Any datasets that did not meet my inclusion criteria were excluded, establishing a solid foundation for accurate and reliable dataset integration.

I integrated these datasets to gain a more detailed understanding of the TME, capturing cellular phenotypes with greater resolution and exploring variability between datasets (Figure 13). This strategy allowed us to robustly cross-validate our findings across a diverse range of patient samples and experimental conditions. I used the Seurat pipeline for data normalization and initial integration, followed by STACAS, a semi-supervised data integration tool, to make use of the existing labels in the datasets[92,103]. By using pre-assigned labels, I guide the integration process, preserving biologically relevant differences and enhancing both accuracy and interpretability.

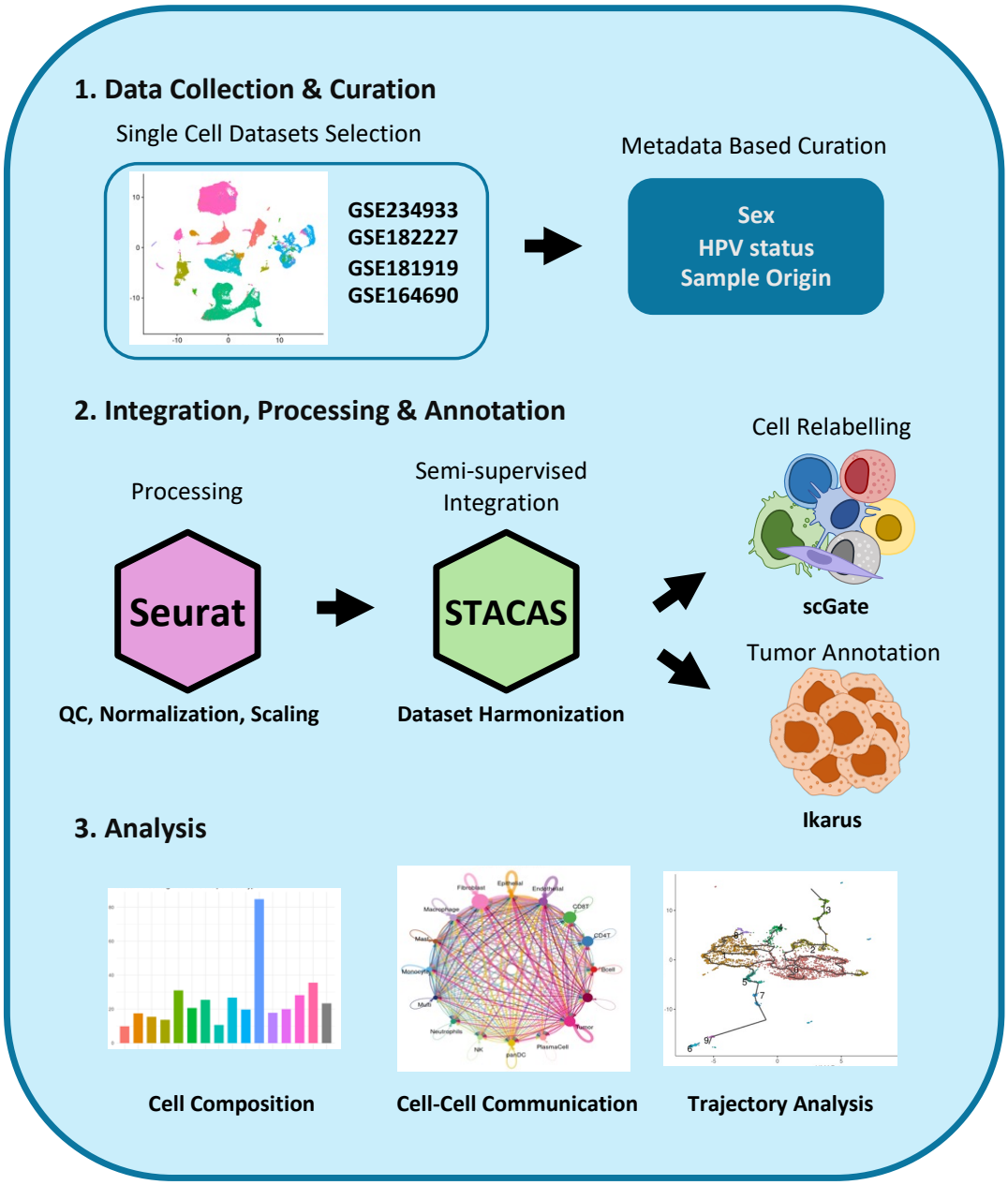


Figure 13. Workflow for creating the HNSCC atlas. The workflow involves three main stages. First, single-cell RNA-seq datasets are collected and curated based on key metadata, including sex, HPV status, and sample origin. Next, data processing is performed with Seurat for quality control, normalization, and scaling, followed by dataset harmonization using STACAS for semi-supervised integration. Cell relabeling is refined with scGate, and tumor cells are specifically annotated using Ikarus. Finally, the harmonized atlas allows comprehensive downstream analyses, for example cell composition profiling, cell-cell communication inference, and trajectory analysis. *This figure is adapted from my manuscript currently under review: Conde Lopez et al.,* A Unified Single-Cell Atlas of HNSCC: Uncovering HPV and Sex Variability in the Tumor Microenvironment, *submitted 2025*.

For cell type annotation, I applied scGate, which uses pre-defined gating models to ensure consistent and homogeneous identification of cell types across our normalized and integrated datasets. This approach allowed us to maintain reliability in cell type classification across different studies, ensuring consistency in our analysis. Together, these tools enabled us to effectively integrate and analyze single-cell data in HNSCC, enhancing our ability to characterize the complex cellular landscape of the TME.

To accurately identify tumor cells in our integrated atlas, I employed a combination of approaches due to the limitations of scGate for this task. I applied the Ikarus algorithm, which is a machine learning-based tool specifically designed for cell type classification in single-cell RNAseq data. Ikarus follows a supervised learning approach, allowing us to train the model using pre-annotated datasets from HNSCC and other cancer types and then apply the trained model to predict tumor cells in our HNSCC datasets[107]. This method is particularly effective for distinguishing tumor cells based on their gene expression profiles. In parallel, I utilized the existing tumor cell annotations provided by each individual study. By combining the output from Ikarus with these pre-defined labels, I ensured a more robust and accurate identification of tumor cells across our dataset. This dual approach allowed us to improve tumor cell classifications, reducing potential inconsistencies that could come from relying on a single method or dataset.

In addition to the broad cell type annotations, I further expanded the classification of immune cells and fibroblasts within the TME. For immune cells, I utilized the ProjectTILS framework, which applies reference-based labeling for key immune subsets such as dendritic cells, $CD4^+$ T cells, and $CD8^+$ T cells[138]. This provided a more detailed subclassification of immune populations across our integrated datasets, ensuring greater specificity in cell type identification. For fibroblasts, I applied the classification reference established by Cords et al.[139]. Using the data provided in this study, I implemented their classification into the ProjectTILS framework, enabling accurate identification of fibroblast subtypes. This combination of immune and fibroblast subclassification allowed for a clearer characterization of the TME across our datasets.

4.2.2 Validation of the Integrated Atlas Through Comparison with Original Labelling

I evaluated the reliability of our integrated atlas by comparing the original cell type annotations with those obtained after reclassification using scGate. This comparison, visualized through UMAP plots (Figure 14.A-B) and further quantified in the mosaic plot (Figure 14.C), underscores the improvements in cell type homogeneity across the HNSCC datasets.

The UMAP plots provide a clear visual distinction: Figure 13.A illustrates the original labelling, while Figure 14.B highlights the relabelled classifications by scGate. The relabelling process significantly enhances the segregation of critical cell types, particularly improving the distinction between T cells and myeloid populations. Notably, scGate's gating models have improved the representation of previously underrepresented or misclassified populations such as endothelial cells and fibroblasts.

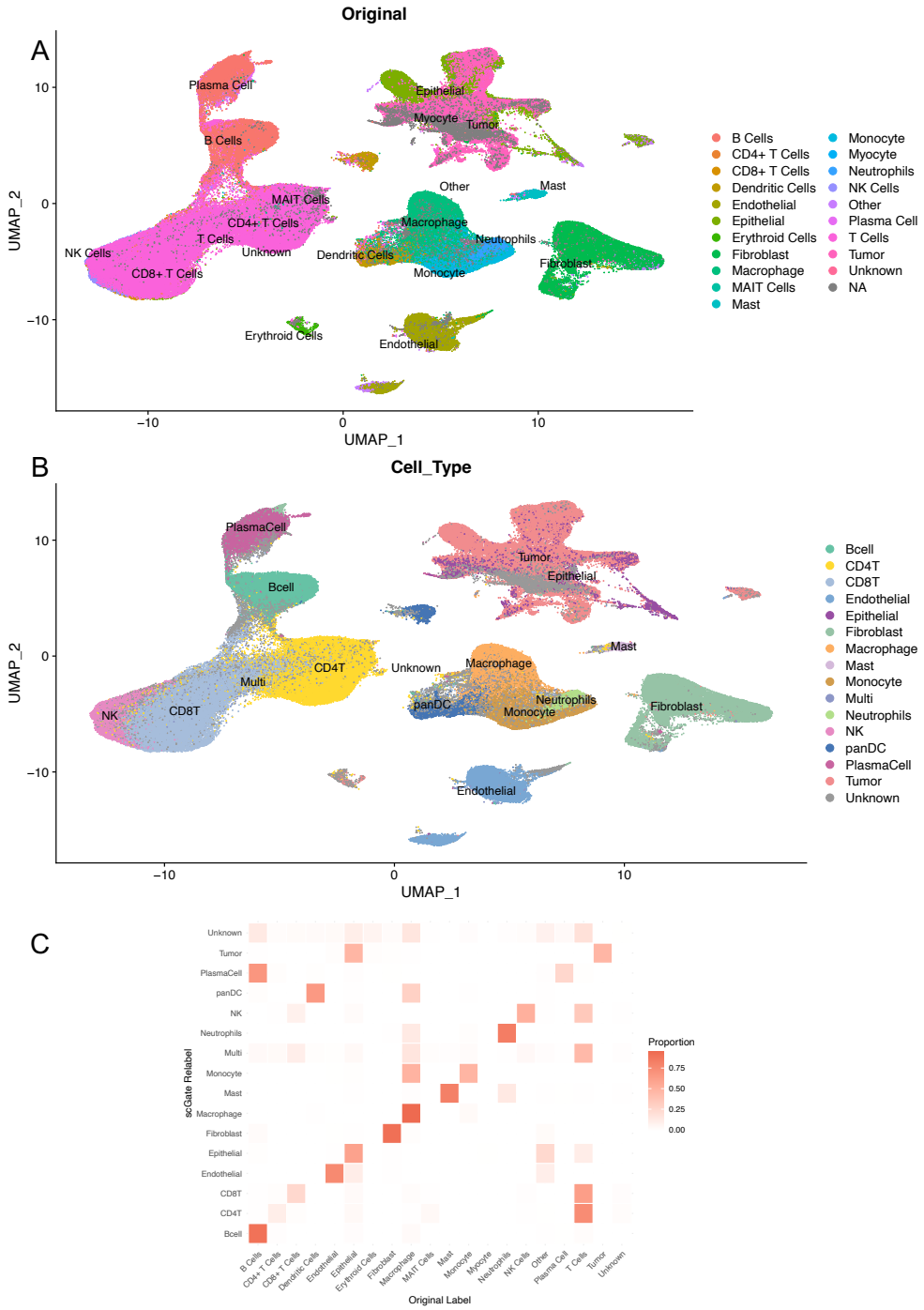


Figure 14. Cell type integration and relabelling in the HNSCC atlas. *(A)* UMAP plot displaying original cell type annotations across the integrated single-cell RNA-seq datasets. (B) UMAP visualization showing updated cell labels after applying scGate-based relabelling, complemented by tumor cell classification using Ikarus. *(C)* Mosaic plot comparing the distribution of original and relabelled cell types, highlighting major shifts in classification. *This figure includes content adapted from the submitted manuscript: "A Unified Single-Cell Atlas of HNSCC: Uncovering HPV and Sex Variability in the Tumor Microenvironment" by Conde Lopez et al., 2025.*

The mosaic plot (Figure 14.C) quantifies these improvements, displaying the frequency and distribution of original versus scGate-relabelled cell types. It reveals a shift towards greater coherence, especially in the classification of T cells, while reducing the number of mislabeled myeloid cells. Additionally, tumor-labeled cells identified using Ikarus and previous annotation methods were included in this analysis. A large proportion of cells previously labelled as epithelial were now reclassified as tumor cells in the final labelling. These results demonstrate the effectiveness of automated tools like scGate or Ikarus in correcting inconsistencies and enhancing the accuracy of cell type annotation across diverse technical backgrounds.

After constructing the single-cell atlas of HNSCC, I sought to explore how key clinical variables, such as HPV status and sex, influence the cellular composition and heterogeneity within the TME. This comparative analysis allowed us to gain a deeper understanding of how these variables shape the immune and stromal profiles of HNSCC patients.

4.2.3 Comparative Analysis of the Tumor Microenvironment by HPV Status and Clinical Sex in HNSCC Atlas

As previously described, HPV status and sex are known to significantly influence the TME in HNSCC. HPV-positive tumors typically display a more active immune response, while sex differences can affect both immune and stromal cell composition. By integrating these variables into our single-cell atlas, I was able to examine their contributions to cellular heterogeneity more closely. Our analysis revealed distinct immune and stromal profiles that varied across HPV status and sex, offering new insights into the complexity of the TME in HNSCC.

In our analysis of the stromal compartment, comprising endothelial cells, epithelial cells, and fibroblasts, I observed notable sex-dependent differences in the composition of HPV-positive patients. Females displayed a substantially higher stromal proportion, accounting for 36.9% of their TME, compared to only 14.8% in males (Figure 15.A, Suppl.

Table 3). Conversely, in the HPV-negative group, this trend reversed, with males exhibiting a higher stromal composition (23%) relative to females (19.9%). However, these differences between sexes in the HPV-negative group were less pronounced than those observed in HPV-positive individuals.

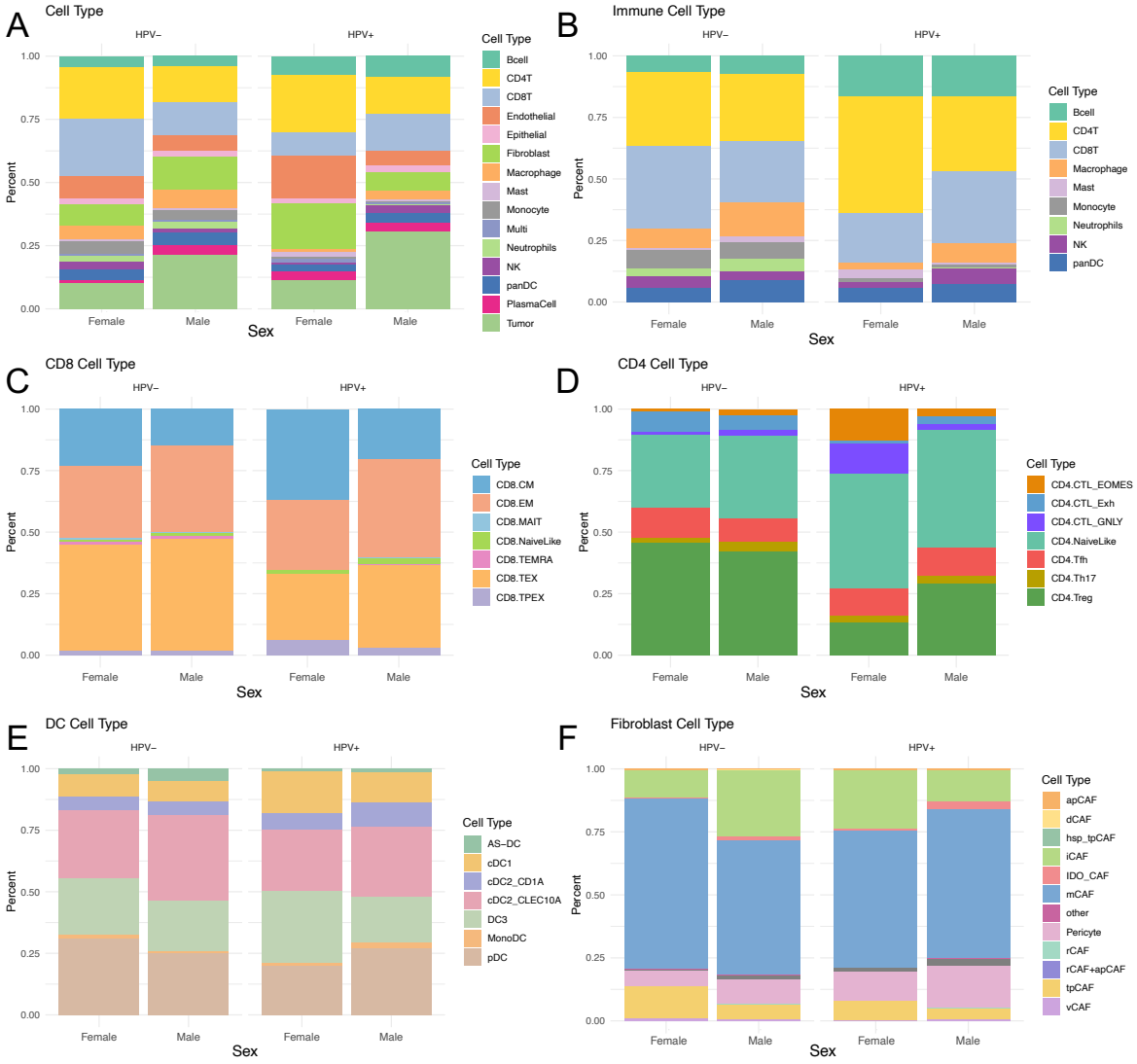


Figure 15. Composition of major cell types in HNSCC tumors by sex and HPV status. *(A)* Stacked bar plot displaying the relative abundance of major cell types in male and female patients, separated by HPV-positive and HPV-negative groups. *(B)* Distribution of overall immune cell populations across the same groups, showing notable shifts between sexes depending on HPV status. *(C)* $CD4^+$ T cell subtype composition across male and female patients, stratified by HPV status. *(D)* $CD8^+$ T cell subtype distribution across the same groups. *(E)* Composition of dendritic cell (DC) subsets by sex and HPV status. *(F)* Distribution of fibroblast subtypes. *This figure incorporates panels adapted from the submitted manuscript: "A Unified Single-Cell Atlas of HNSCC: Uncovering HPV and Sex Variability in the Tumor Microenvironment" by Conde Lopez et al., 2025.*

Regarding tumor cells, I found that males consistently showed a higher proportion of tumor cell populations across both HPV-positive and HPV-negative patients, where tumor cells constituted over 20% of the TME in males in both contexts.

In the immune compartment, our findings confirmed and expanded upon previous research, particularly emphasizing differences in immune cell distribution by HPV status (Figure 15.B, Suppl. Table 4). HPV-positive patients exhibited a significant increase in adaptive immune cells, such as B cells, plasma cells, TH1 and TH2 CD4+ T cells, NK cells, and CD8+ T cells. These findings align with studies indicating robust immune activation in HPV-positive HNSCC[142]. In contrast, HPV-negative patients showed higher levels of innate immune cells, monocytes, macrophages, NK T cells, and neutrophils, corresponding to a more immunosuppressive, or "cold," immune environment[142,143].

When analyzing immune cell distributions by sex, I observed that females exhibited a higher proportion of immune infiltrates compared to males, suggesting a more active immune profile. This finding is consistent with previous studies that report males having a more exhausted CD8+ T cell profile[46,48]. In our study, this phenomenon was particularly evident in HPV-positive males, while HPV-negative cases showed relatively constant T cell proportions between sexes. Notably, HPV-negative females displayed a higher proportion of neutrophils than their male counterparts. Further investigation of intratumoral immune infiltration revealed that males with HPV-positive HNSCC had significantly higher proportions of B cells compared to females.

Our analysis highlights specific patterns shaped by both HPV status and sex. For instance, HPV-positive females displayed a higher percentage of cytotoxic CD8+ T cells and lower T cell exhaustion compared to males, potentially influencing their response to immunotherapies targeting exhausted T cells (Figure 15.C-E, Suppl. Table 5-7). In contrast, males, particularly those with HPV-negative status, showed lower levels of immune infiltration, suggesting they may derive less benefit from conventional immunotherapy strategies. The definitions and functional annotations of the immune cell subtypes discussed can be found in Supplementary Table 8. These findings highlight

the importance of considering both viral and sex-related factors when developing tailored therapeutic approaches for HNSCC, as these variables significantly influence the immunological landscape and potential treatment efficacy.

The comparative analysis of HPV status and sex in HNSCC provided key insights into the cellular composition of the TME. Given the significant influence of sex on immune and stromal cell populations, I next sought to explore the role of sex chromosome dosage (XX, XY, XØ) in shaping the TME. This step allowed us to further investigate how sex chromosome dosage contributes to the observed variations in cellular composition and immune infiltration across the TME.

4.2.4 Y Chromosome Downregulation in Single-Cell Data Reveals Tumor Cells as the Primary Source of Dosage Variability

Since the single-cell data available for our analysis consists solely of transcriptomic data, the first step was to translate the concept of EDY into a framework applicable to individual cells. In bulk data, EDY is defined at the patient level, with a score assigned based on the overall expression of Y chromosome genes. However, in the single-cell context, I assign this classification at the cell level, identifying XØ cells as those with no detectable Y chromosome gene expression. This key difference makes the single-cell analysis not directly comparable to the bulk data, as the former reflects Y chromosome dosage variability at the resolution of individual cells rather than providing an aggregated score per patient.

To select XØ cells, I evaluated the expression of all detectable Y chromosome genes within the single-cell transcriptomic data. I calculated the percentage of expression for the Y chromosome gene set over the expression of all genes for each cell. Any cell with a score equal to zero across all Y chromosome genes was classified as an XØ cell. This process was conducted for both male and female patients. As expected, female patients exhibited no Y chromosome gene expression (Figure 16.A). In the visualization of the percentage of Y chromosome expression, I observed a small number of outliers among

the female samples, these outliers might be attributed to the homology between certain Y and X chromosome genes. Nevertheless, the general trend toward zero Y chromosome expression is clearly maintained, reinforcing the reliability of our classification. Importantly, XØ cells were observed in all male patients, making the presence of XØ cells a homogenous event across the cohort (Figure 16.B)

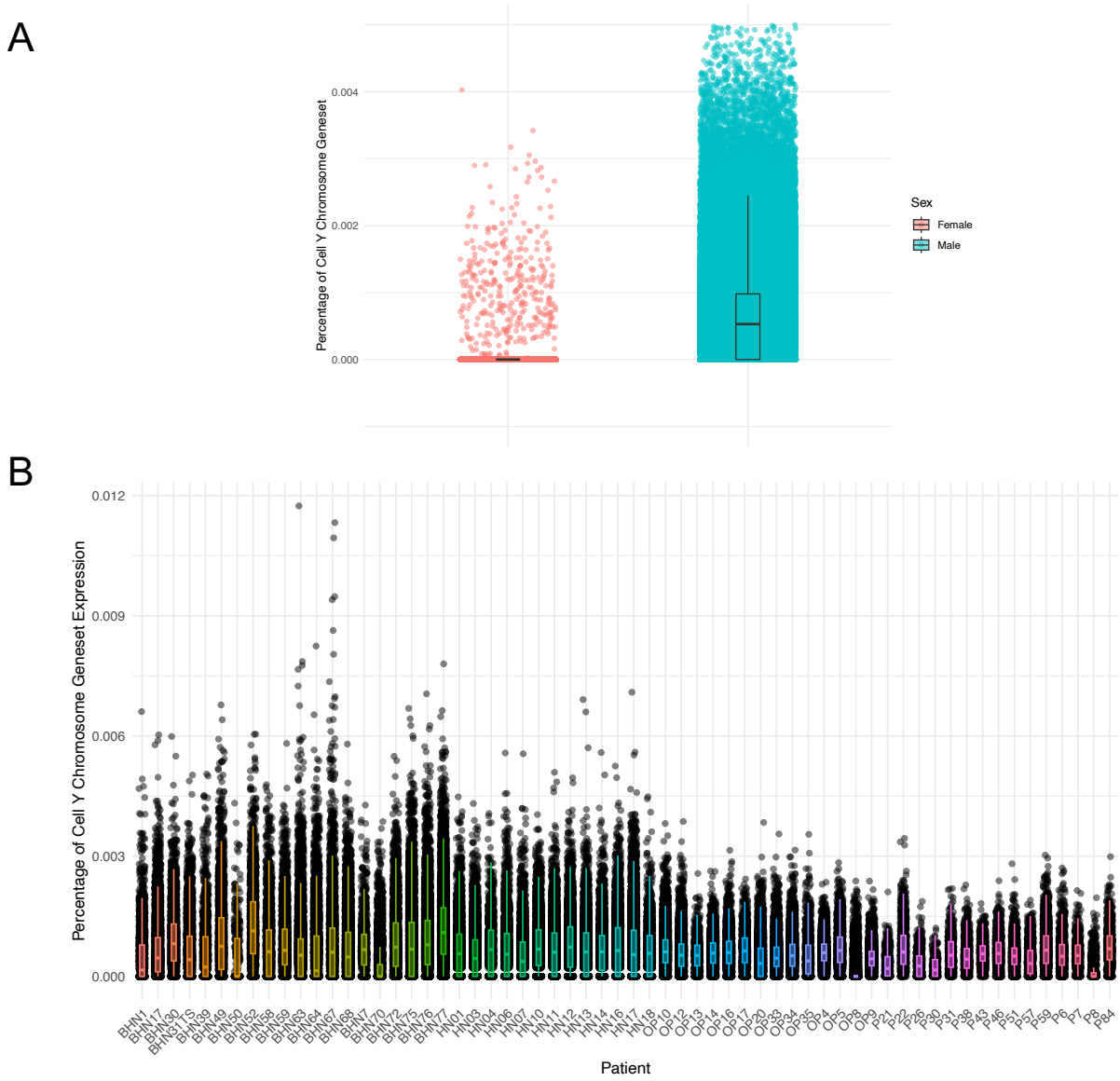


Figure 16. Single-cell quantification of Y chromosome expression highlights dosage variability across sex and cell types. *(A)* Distribution of Y chromosome gene set expression per cell, shown as the percentage of Y chromosome gene expression over total gene expression. Female samples show near-zero expression. In contrast, male samples display a wide range of Y chromosome expression. *(B)* Per-patient distribution of Y chromosome expression across male samples. Despite variability in expression levels, all male patients have a proportion of XØ cells, supporting the notion that Y chromosome downregulation is a widespread and consistent phenomenon across the cohort.

Having classified XØ cells within our single-cell data, I next aimed to investigate whether specific cell types were driving the observed differences between XØ and XY males in the bulk data. My hypothesis was that, while some cell types might exhibit a consistent, homogenous proportion of XØ cells across patients, others could show greater heterogeneity. Identifying the cell types that display this variability could provide insight into the specific cell populations responsible for the sex chromosome dosage differences observed in the bulk analysis.

To test this, I examined the percentage of XØ cells across various cell types, assessing whether these proportions remained stable across patients or if certain cell types were more heterogeneous in their distribution. I began by looking at the overall percentages of XØ cells in each cell type, correcting for the representation of each population to ensure meaningful comparisons (Figure 17.A). Although neutrophils showed a relatively high percentage of XØ cells, their representation in the data was minimal and inconsistent across patients (Figure 17.B), making them less relevant for driving the bulk-level differences.

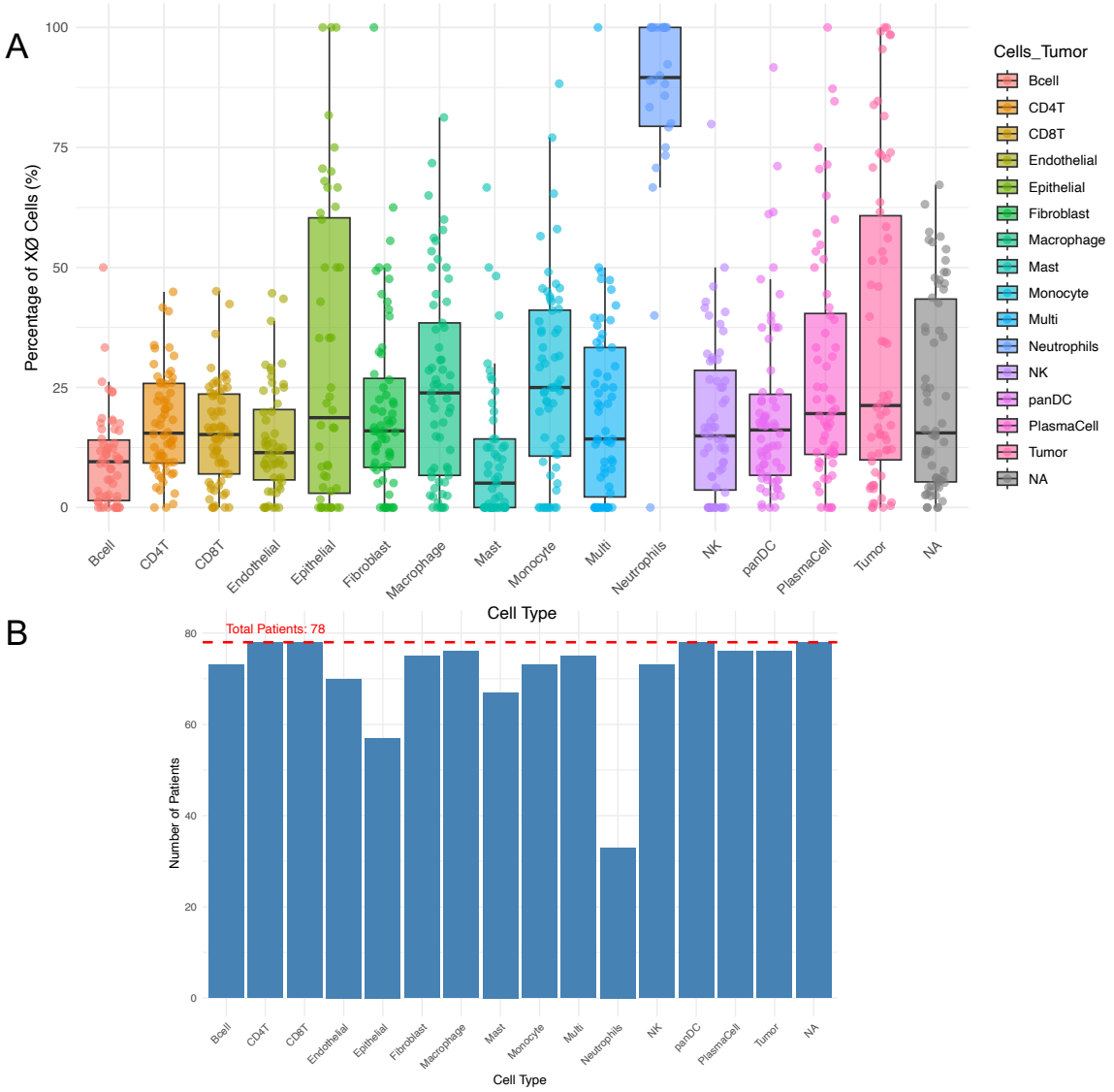


Figure 16. Cell type-specific variability in the proportion of XØ cells across male patients. *(A)* Distribution of the percentage of XØ cells across different cell types, normalized per patient. Each point represents an individual patient's percentage of XØ cells for a given cell type, enabling visualization of both central tendency and variability. *(B)* Number of male patients in which each cell type was detected. This information ensures that conclusions drawn about XØ percentages are supported by adequate patient coverage. Although neutrophils appear to have high XØ percentages, they are inconsistently represented across the cohort and thus likely do not contribute substantially to the bulk-level observations.

The most relevant findings emerged from my analysis of tumor cells. Not only did tumor cells exhibit the highest percentage of XØ cells when corrected for representation, but they were also the most abundant cell type in males (Figure 18.A). Additionally, the distribution of XØ cells within tumor cells varied significantly among patients (Figure 18.B). This combination of high percentage, abundance, and variability made tumor cells

the ideal choice for classifying patients. This finding highlights tumor cells as a key factor in the observed differences between XØ and XY males. Understanding the enrichment of XØ cells in this population may provide deeper insights into the biological implications of Y chromosome downregulation and loss in HNSCC.

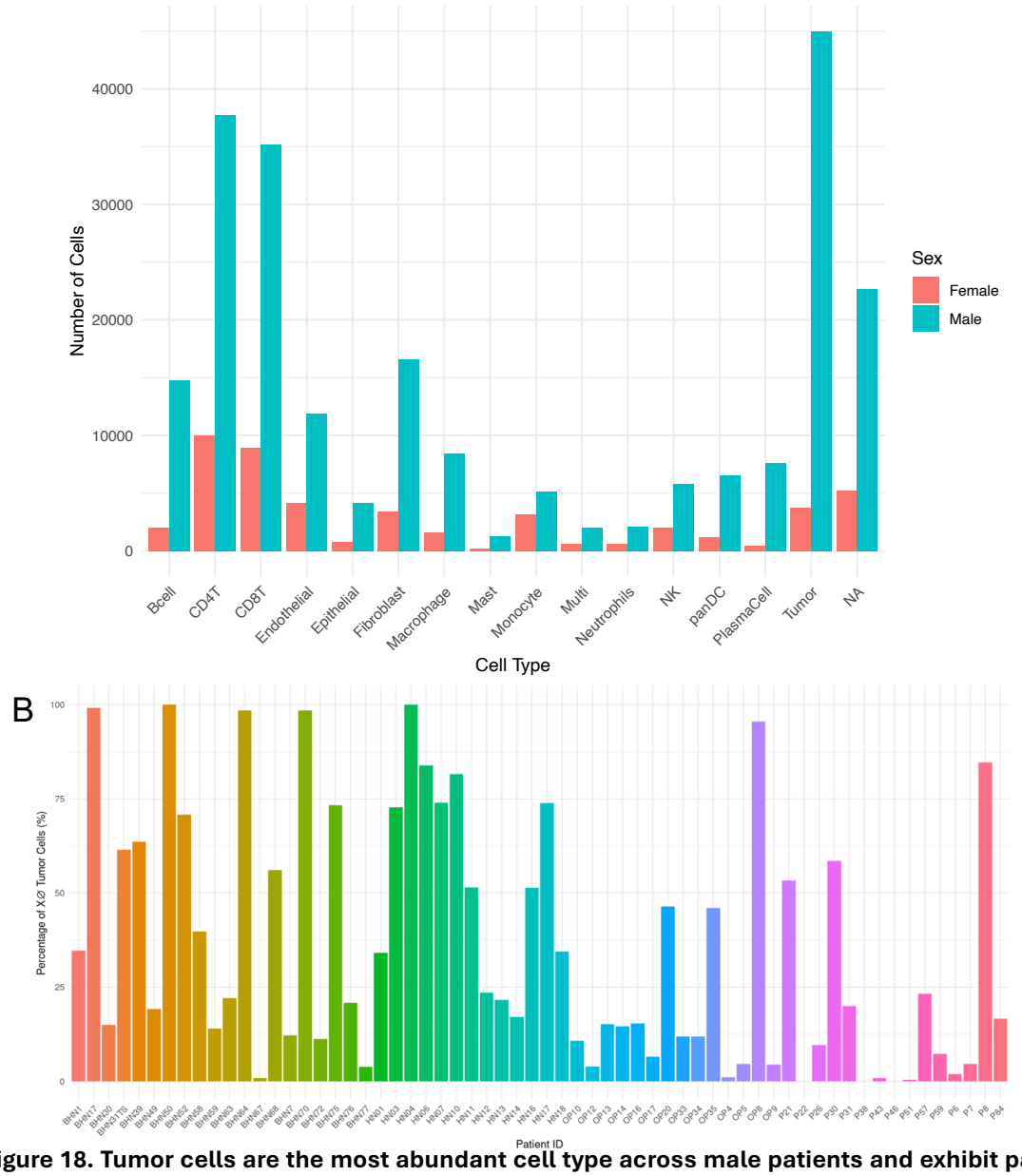


Figure 18. Tumor cells are the most abundant cell type across male patients and exhibit patient-specific variability in XØ cell proportions. *(A)* Absolute number of cells per cell type in male and female patients. Tumor cells are the most represented population in males, supporting their relevance for downstream analyses involving sex chromosome dosage. *(B)* Percentage of XØ tumor cells per male patient. While tumor cells consistently appear across patients, the fraction of XØ cells within them varies considerably, indicating inter-patient heterogeneity in Y chromosome expression among tumor cells.

4.2.5 Patient Classification Based on Tumor Cell XØ Characteristics

While the previous section focused on detecting individual XØ cells and examining their distribution across different cell types and patients, the next step required me to classify patients according to their XØ characteristics, translating the concept of sex chromosome dosage that I had established in bulk data in order to allow to directly compare my single-cell results with bulk data, where classifications were made at the patient level.

To achieve this, I needed a consistent framework for patient classification, and given our findings from the previous section, I chose to base this classification on tumor cells. Tumor cells had emerged as the cell type of primary interest due to their high percentage of XØ cells, abundance in the data, and variability across patients. I hypothesized that focusing on tumor cells would allow us to more accurately track the source of sex chromosome dosage variability and draw clearer parallels to the bulk data results.

I pseudobulked the transcriptomes of tumor cells from each patient and performed a GSVA analysis using the predefined set of 11 Y chromosome genes that I had established in bulk data (Results 4.2.1). After calculating the GSVA scores for the Y chromosome gene set, I observed that, as expected, males had higher scores than females except for some that showed lower expression which marks them as possible XØ patients (Figure 19.A). When focusing on male patients and examining their HPV status (Figure 19.B), I observed a clear trend similar to what was seen in the bulk data: HPV-positive patients tend to have significantly higher GSVA scores for the Y chromosome gene set compared to HPV-negative patients. These findings further strengthen our previous observations regarding the relationship between HPV status and Y chromosome expression.

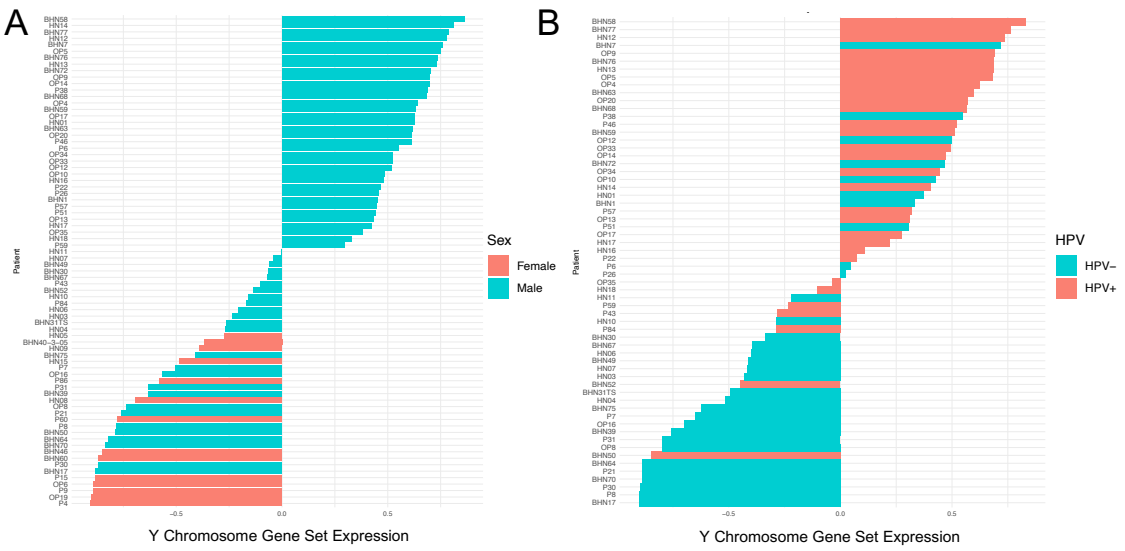


Figure 19. Y chromosome gene expression in tumor cells mirrors trends observed in bulk data, with HPV status influencing dosage levels. *(A)* GSVA scores for the Y chromosome gene set calculated from pseudobulked tumor cells per patient, grouped by sex. As expected, male patients showin general higher expression of Y chromosome genes compared to females. *(B)* Focusing on male patients, Y chromosome gene expression is stratified by HPV status. HPV-positive males exhibit significantly higher Y chromosome expression than HPV-negative counterparts, consistent with patterns observed in bulk transcriptomic data.

To classify male patients more effectively, I decided to calculate GSVA scores separately for the HPV-positive and HPV-negative cohorts. Given that GSVA scores are dataset-dependent, this separation was crucial to avoid HPV status acting as a confounding factor. By independently calculating the scores for each cohort, I could classify the highest-scoring patients as XY (high expressors) and the lowest-scoring patients as XØ (low expressors), without skewing the results in favour of HPV-positive patients (Figure 20). Without this separation, all highest-scoring patients would have been HPV-positive, and all lowest-scoring patients would have been HPV-negative, making it difficult to compare the two groups meaningfully.

This approach allowed for a more balanced classification across dosage groups: among HPV-negative and HPV-positive patients, respectively, there were 10 and 10 classified as intermediate, 10 and 10 as XØ, 14 and 2 as XX, and 12 and 11 as XY. To further sharpen the contrast between high and low expressors, I decided to exclude the intermediate

group. While this meant sacrificing some patients from the analysis, it allowed us to draw more distinct comparisons, which I expect will yield more informative insights.

In the current analysis (Figure 20.B), I plotted the proportion of XØ and XY tumor cells per patient, ordered by their Y chromosome GSVA score. As seen in the graph, patients with HPV-negative status tend to have much lower GSVA scores compared to those with HPV-positive status. Furthermore, the distribution of XØ and XY cells reveals a clear relationship between GSVA score and Y chromosome expression (Figure 20.B). Patients with lower GSVA scores exhibit a higher proportion of XØ tumor cells, indicating a complete lack of Y chromosome gene expression in these cells. Conversely, patients with higher GSVA scores demonstrate a much lower percentage of XØ cells and a correspondingly higher proportion of XY cells.

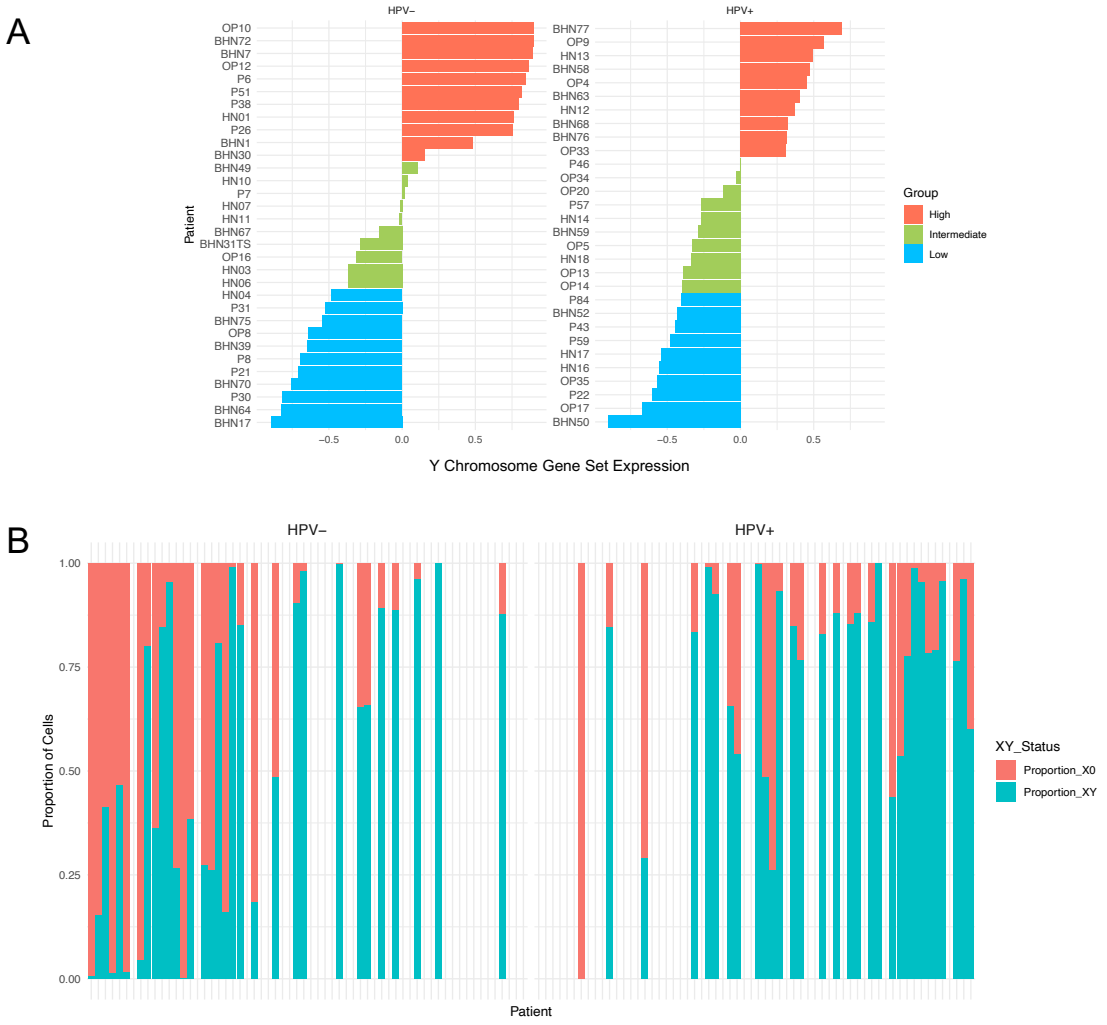


Figure 20. GSVA-based classification of male patients reveals distinct XØ/XY tumor cell profiles linked to HPV status. *(A)* GSVA scores for Y chromosome gene expression in tumor cells of male patients were computed separately for HPV-negative (left) and HPV-positive (right) cohorts to prevent HPV status from acting as a confounder. Patients were classified into XØ (low expressors, blue), XY (high expressors, red), and Intermediate (green) groups. This separation allowed for more accurate inter-cohort comparisons by mitigating bias introduced by the global score shift observed between HPV statuses. *(B)* Proportions of XØ (blue) and XY (red) tumor cells per patient, stratified by HPV status and ordered by their Y chromosome GSVA score. A strong correlation is observed: patients with lower GSVA scores tend to exhibit a higher proportion of XØ tumor cells, while those with higher scores predominantly harbor XY cells. This confirms the relevance of Y chromosome downregulation as a biologically meaningful feature across HPV contexts.

This visualization confirms my earlier hypothesis: patients with lower Y chromosome expression (as reflected by their GSVA scores) tend to harbor a higher fraction of XØ tumor cells, while higher Y chromosome expression is associated with a predominance

of XY cells. The combination of these two independent metrics—GSVA scores and XØ/XY cell proportions—provides further insight into the biological relevance of Y chromosome downregulation, especially in the context of HPV status.

In this section, I used single-cell transcriptomic data to investigate sex chromosome dosage variability, focusing on the identification of XØ tumor cells, which lack detectable Y chromosome gene expression. My analysis revealed distinct differences between HPV-positive and HPV-negative patients, with the former showing higher GSVA scores for Y chromosome genes and a predominance of XY cells, while the latter exhibited lower scores and a higher proportion of XØ cells. To further define my analysis, I classified male patients into XØ (low expressors) and XY (high expressors) groups based on their Y chromosome expression. However, due to the complexity of disentangling the effects of HPV from sex chromosome dosage and the potential confounding influence of the virus, I decided to focus exclusively on HPV-negative patients in the subsequent analyses. This decision enabled a more accurate classification of sex chromosome dosage, ensuring that I could better interpret the differences between the of XØ and XY patients without interference from viral factors, and offering clearer insights into the biological variability of sex chromosome dosage in HNSCC.

4.2.6 Chromosomal Instability is Present Across Sex Chromosome Dosage Groups but Does Not Correlate with Y Chromosome Expression Loss

To explore whether the loss of the Y chromosome expression in XØ cells could be linked to broader chromosomal instability, I conducted an analysis using inferCNV, a tool that estimates chromosomal CNVs from single-cell RNA sequencing data. inferCNV infers chromosomal ploidy by assessing transcriptional signals relative to a reference population of normal cells, allowing us to detect large-scale chromosomal gains and losses that may indicate broader genomic instability. This analysis was critical to determine whether the observed loss of the Y chromosome was an isolated phenomenon or part of a larger pattern of genomic instability. If the chromosomal anomalies extended beyond the Y chromosome, it would suggest that the observed loss

of the Y chromosome was simply one aspect of a more generalized genomic instability, making it a less specific event.

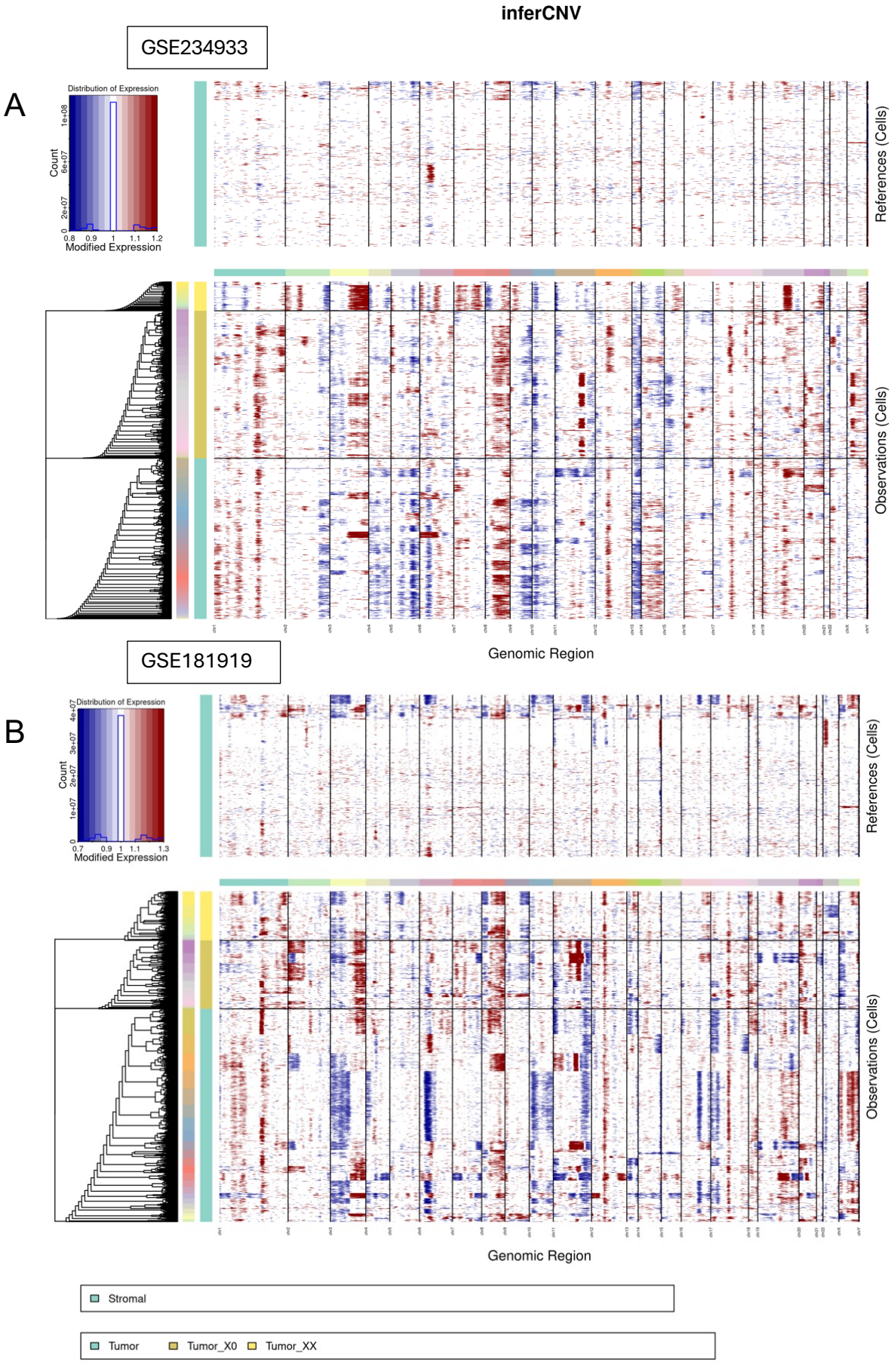
Due to computational limitations, I applied inferCNV separately to each dataset, but because the cell types were homogeneously annotated across datasets, this did not affect the consistency of our analysis. Normal stromal cells served as the reference population, providing a stable baseline for comparison. I clustered the tumor cells based on their sex chromosome dosage status (XX, XY, and XØ) to investigate chromosomal instability within these distinct groups.

The results from inferCNV (Figure 21) revealed that all groups, regardless of their sex chromosome dosage, exhibited a higher degree of chromosomal instability compared to the normal stromal reference cells. Each group showed patterns of chromosomal anomalies, with certain genomic regions demonstrating clear deviations from the baseline reference. Interestingly, while chromosomal instability was present across all groups, the specific patterns of gains and losses were not consistent between XX, XY, and XØ cells. Each group displayed distinct patterns of chromosomal aberrations, suggesting that while instability was a shared characteristic, the particular genomic regions affected varied between the sex chromosome dosage groups.

Moreover, despite the variability in chromosomal anomalies, all three groups (XX, XY, and XØ) showed evidence of instability, with no single group exhibiting a significantly higher level of chromosomal aberrations than the others. This observation indicates that chromosomal instability is a common feature in all tumor cell groups, but it does not disproportionately affect XØ cells. As such, the loss of the Y chromosome in XØ cells does not appear to be a result of an overall increase in chromosomal instability. Rather, it remains a specific event, separate from the general chromosomal anomalies observed across other chromosomes.

These findings reinforce the biological relevance of categorizing cells into XX, XY, and XØ groups. The Y chromosome downregulation observed in XØ cells appears to be a distinct event, rather than a manifestation of broader genomic instability. This highlights the

importance of continuing to investigate the implications of sex chromosome dosage as a specific factor influencing the tumor microenvironment in HNSCC.


inferCNV
GSE234933
A
Distribution of Expression
Count
Modified Expression
References (Cells)
Observations (Cells)
Genomic Region
GSE181919
B
Distribution of Expression
Count
Modified Expression
References (Cells)
Observations (Cells)
Genomic Region
Stromal
Tumor
Tumor_X0
Tumor_XX

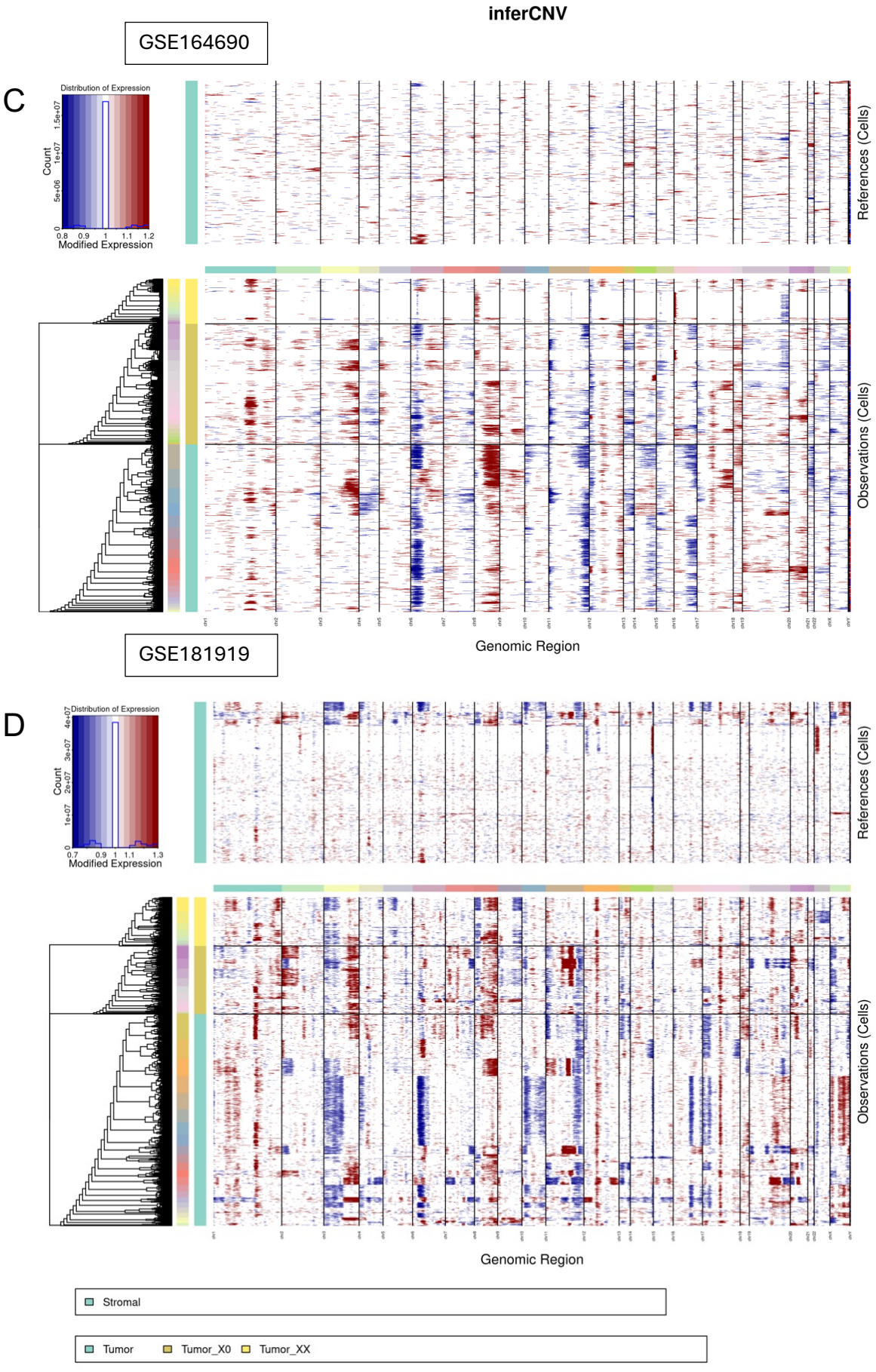

inferCNV
GSE164690
C
Distribution of Expression
Count
Modified Expression
References (Cells)
Observations (Cells)
Genomic Region
GSE181919
D
Distribution of Expression
Count
Modified Expression
References (Cells)
Observations (Cells)
Genomic Region
Stromal
Tumor
Tumor_X0
Tumor_XX

Figure 20. Chromosomal instability analysis using inferCNV across tumor cells with different sex chromosome dosage statuses. To determine whether the loss of Y chromosome expression in XØ cells was associated with broader chromosomal instability, we performed inferCNV analysis using stromal cells as the reference (upper panel). Tumor cells were clustered by sex chromosome dosage groups (XY (Tumor), XX (Tumor_XX), and XØ (Tumor_X0)) and analyzed separately for each dataset (lower panel). inferCNV identifies large-scale chromosomal gains and losses based on gene expression patterns along the genome. *(A)* inferCNV results from dataset **GSE234933.** *(B)* inferCNV results from dataset **GSE18227**. *(C)* inferCNV results from dataset **GSE164690**. *(D)* inferCNV results from dataset **GSE181919**.

4.2.7 Analysis of Cell Type Proportions and Tumor Microenvironment Composition Among Sex Chromosome Dosage Groups in HPV-Negative Patients

To further investigate the impact of sex chromosome dosage on the TME, I analyzed the cell type proportions across XX, XY, and XØ groups, focusing exclusively on HPV-negative patients. My analysis revealed differences in the composition of the stromal and immune compartments between these sex chromosome dosage groups.

The stromal compartment was notably more pronounced in XY individuals, followed by XX, compared to those in the XØ group (Figure 22). This enrichment was particularly evident for fibroblasts, which constituted a substantial proportion of the TME in XY patients. These observations suggest that the presence of a functional Y chromosome may contribute to the expansion or maintenance of the stromal compartment. In contrast, tumor cells were the dominant population in XØ males, indicating a potentially more aggressive tumor profile with reduced immune infiltration in this group.

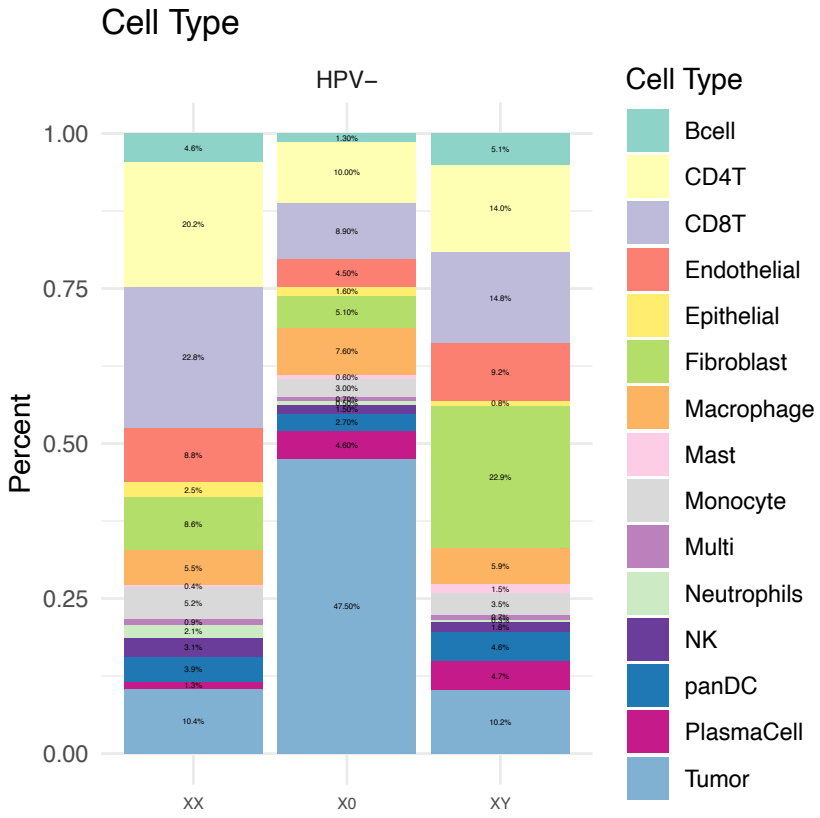


Figure 22. Cell type composition across sex chromosome dosage groups in HPV-negative patients. Bar plot displaying the aggregated proportions of all annotated cell types across XX, XY, and XØ groups, considering only HPV-negative tumors. Cell type annotations were derived from the establisand grouped by sex chromosome dosage category. Proportions were calculated across all cells within each group to allow comparative visualization of the TME composition. This approach enables the assessment of differences in stromal, immune, and tumor compartments between sex chromosome dosage groups.

For the immune compartment (Figure 22, Suppl. Table.9) B cells were most abundant in XY tumors, followed by XX tumors, and were least prevalent in XØ tumors. This distribution suggests a potentially stronger humoral immune response in XY individuals, while XØ tumors, with their lower B cell representation, may experience reduced antibody-mediated immunity. Natural Killer (NK) cells followed a similar trend, being more abundant in XX tumors compared to XY and XØ tumors. The higher presence of NK cells in XX tumors may indicate a more robust innate immune surveillance, whereas the diminished NK cell proportions in XØ tumors suggest a weakened innate immune response in this group. For the myeloid populations, macrophages were more prevalent in XØ tumors compared to XX and XY tumors, indicating a more immunosuppressive

environment that may contribute to tumor progression in XØ individuals. Monocytes, on the other hand, were most abundant in XX tumors, followed by XY tumors, and were least common in XØ tumors. When considering all immune cell types together, XX tumors exhibited the highest overall immune cell content, suggesting a more immune-active microenvironment. In contrast, XØ tumors had the lowest immune cell infiltration, pointing to a more immunosuppressive TME in these individuals.

With respect to T cell composition, XY tumors demonstrated a higher proportion of $CD8^+$ T exhausted and $CD4^+$ T naive cells compared to XØ and XX tumors, aligning with the hypothesis that XY individuals exhibit a less robust immune infiltration in their primary tumors (Figure 23.A-B, Suppl. Table. 10-11). In contrast, XX female tumors were characterized by a higher overall proportion of immune cells, particularly enriched in follicular helper T cells, cytotoxic $CD4^+$ T cells, and more active $CD8^+$ T cells. This is consistent with studies that suggest females have a more active immune phenotype than males[45,46,48]. Interestingly, XØ tumors showed the highest levels of regulatory T cells and slightly lower proportions of naive $CD4^+$ T cells compared to their XY counterparts, further supporting the idea that sex chromosome dosage not only influences the quantity of immune infiltration but also the functional state of T cell populations within the TME.

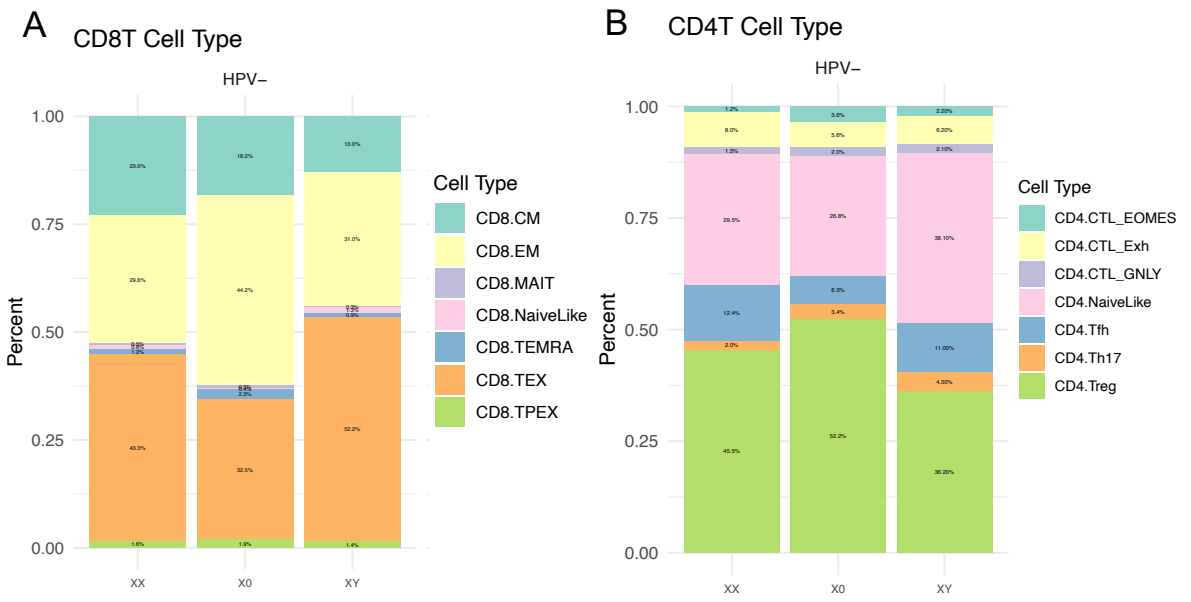


Figure 23. Distribution of T cell subtypes across sex chromosome dosage groups in HPV-negative tumors. *(A)* Relative proportions of $CD8^+$ T cell subtypes, including exhausted and cytotoxic populations, in XX, XY, and XØ tumors. *(B)* Relative proportions of $CD4^+$ T cell subtypes, including naïve, regulatory, follicular helper, and cytotoxic $CD4^+$ T cells across the same groups.

To further extend our investigation into the TME in HPV-negative patients, I focused on the fibroblast compartment, exploring the distribution of fibroblast subtypes across the sex chromosome dosage groups (XX, XY, and XØ). Using the classification system established by Cords et al., I identified several distinct fibroblast subtypes, each characterized by unique gene expression profiles and functional roles[139]. These subtypes include Matrix CAFs (mCAFs), Inflammatory CAFs (iCAFs), Vascular CAFs (vCAFs), Tumor-like CAFs (tCAFs), and others, with each subtype contributing differently to the tumor stroma.

The analysis of fibroblast subtype proportions across the sex chromosome dosage groups is illustrated in Figure 24, Suppl. Table 12. The most striking observation is the dominance of mCAFs across all groups, with this subtype being particularly prevalent in XY and XØ tumors. mCAFs are known for their involvement in matrix remodeling and are heavily enriched in genes related to extracellular matrix production, such as collagen-encoding mRNAs. Their higher representation in XY tumors may indicate more extensive matrix remodeling processes, which could be linked to increased tumor progression in

this group. Interestingly, while XØ tumors also showed a significant proportion of mCAFs, other fibroblast subtypes were less represented in this group, suggesting that matrix remodeling may play a more central role in XØ tumors than in those of XX or XY individuals.

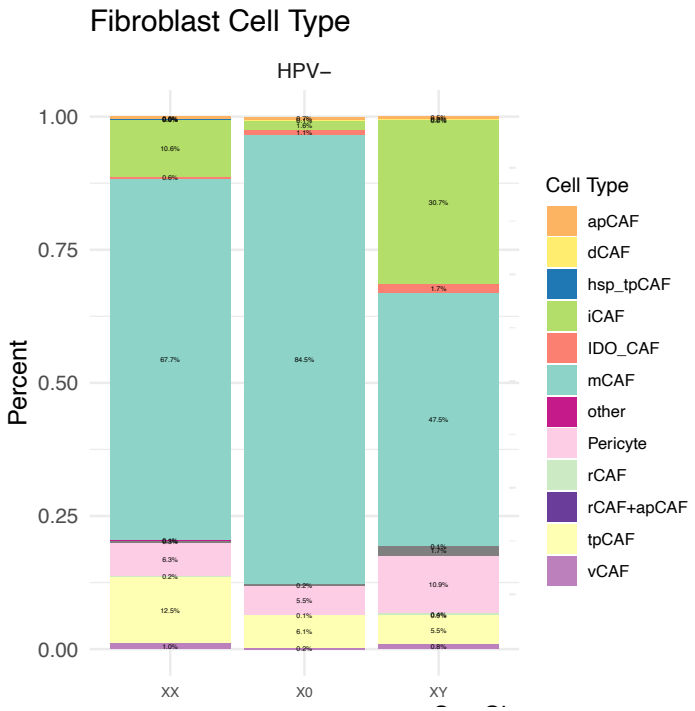


Figure 23. Distribution of fibroblast subtypes across sex chromosome dosage groups in HPV-negative tumors. Relative proportions of annotated fibroblast subtypes—including matrix CAFs (mCAFs), inflammatory CAFs (iCAFs), vascular CAFs (vCAFs), tumor-like CAFs (tCAFs), and others—across XX, XY, and XØ groups. Fibroblasts were classified using the reference framework described by Cords et al. (2023), which distinguishes subtypes based on distinct transcriptional and functional profiles.

iCAFs, characterized by their expression of pro-inflammatory cytokines like IL6 and CXCL12, were more prominent in XY tumors than in XØ or XX tumors. The increased presence of iCAFs in XY individuals suggests an environment with higher inflammatory signalling, which may influence immune cell recruitment and activation within the tumor. In contrast, XØ tumors had a much lower proportion of iCAFs, potentially reflecting a different inflammatory landscape, with implications for immune modulation and tumor-immune interactions in these patients.

Interestingly, XX tumors exhibited a higher proportion of tumor-like CAFs (tCAFs), characterized by the expression of genes associated with stress response, proliferation, and metastasis, such as PDPN, MME, and TMEM158. The increased presence of tCAFs, alongside the higher proportion of vCAFs in XX tumors, points to a distinct stromal composition in female patients. These differences suggest that the stroma in XX tumors might have unique functional properties compared to XY and XØ tumors, although further investigation is needed to explore how these characteristics might influence tumor behaviour and progression.

Taken together, these findings highlight the complex interactions between fibroblast subtypes and sex chromosome dosage within the TME. The differences in fibroblast composition across the XX, XY, and XØ groups suggest that sex chromosome dosage plays a significant role in shaping the stromal environment of HPV-negative tumors. Specifically, the higher prevalence of matrix remodelling and inflammatory fibroblast subtypes in XY tumors may contribute to more aggressive tumor behaviour, while the reduced presence of iCAFs in XØ tumors could point to a less inflamed, but still highly matrix-remodelled, microenvironment.

In summary, our compositional analysis of the TME in HPV-negative patients, stratified by sex chromosome dosage (XX, XY, and XØ), reveals distinct differences in both immune and stromal cell populations. XX tumors exhibited a higher overall immune cell infiltration, including a more active T cell profile, enriched in follicular helper T cells and cytotoxic $CD4^+$ T cells. In contrast, XY tumors displayed a more exhausted T cell state with increased $CD8^+$ T cell exhaustion, while XØ tumors showed the lowest immune infiltration and higher levels of immunosuppressive regulatory T cells. Additionally, the fibroblast compartment showed significant variability, with mCAFs more prominent in XY and XØ tumors, and tumor-like CAFs tCAFs and vCAFs more abundant in XX tumors.

4.2.8 Fibroblast-Mediated Communication Drives Differential Tumor Microenvironment Dynamics Across Sex Chromosome Dosage Groups

Cell-to-cell communication plays a pivotal role in the coordination of various biological processes within the TME. In this section, I used the CellChat tool to investigate how communication patterns vary among the XX, XY, and XØ sex chromosome dosage groups in HPV-negative patients[117]. CellChat provides a comprehensive framework for analyzing intercellular communication by inferring ligand-receptor interactions based on single-cell transcriptomic data. It assesses the signaling pathways and their interactions, taking into account both the number of interactions (based on ligand-receptor pairs) and their overall strength, which reflects the expression levels of the receptors and ligands in each interacting cell type. The tool integrates both the frequency of interactions and the extent of expression, providing insight into which cell types are sending and receiving signals most actively.

In the analysis, I focused on the differential interaction strength between the XX, XY, and XØ groups. As defined by CellChat, interaction strength is not simply the count of interactions but a combinative measure that factors in the expression level of the interacting receptors and ligands, scaled according to the number of cells in each group. This allows us to visualize and compare how strongly certain cell types communicate with one another across the sex chromosome dosage groups.

The results of our cell-cell communication analysis underscored the significant role of the stromal compartment, particularly fibroblasts, in mediating interactions within the TME (Figure 25.A-B). In XY tumors, fibroblasts were particularly active in outgoing signaling, especially in communication with tumor cells, suggesting their prominent role in tumor-stromal interactions in XY individuals. Conversely, XØ tumors showed increased communication between NK cells and fibroblasts, indicating a stronger NK cell-mediated immune response in this group. The interaction strength between fibroblasts and other cell types was weakest in XX tumors, reflecting reduced stromal communication compared to XY and XØ groups. Additionally, in both XX and XY tumors, there was a higher interaction strength between $CD8^+$ T cells and neutrophils, a pattern absent in XØ tumors. These findings reveal distinct immune and stromal interaction dynamics across sex chromosome dosage groups.

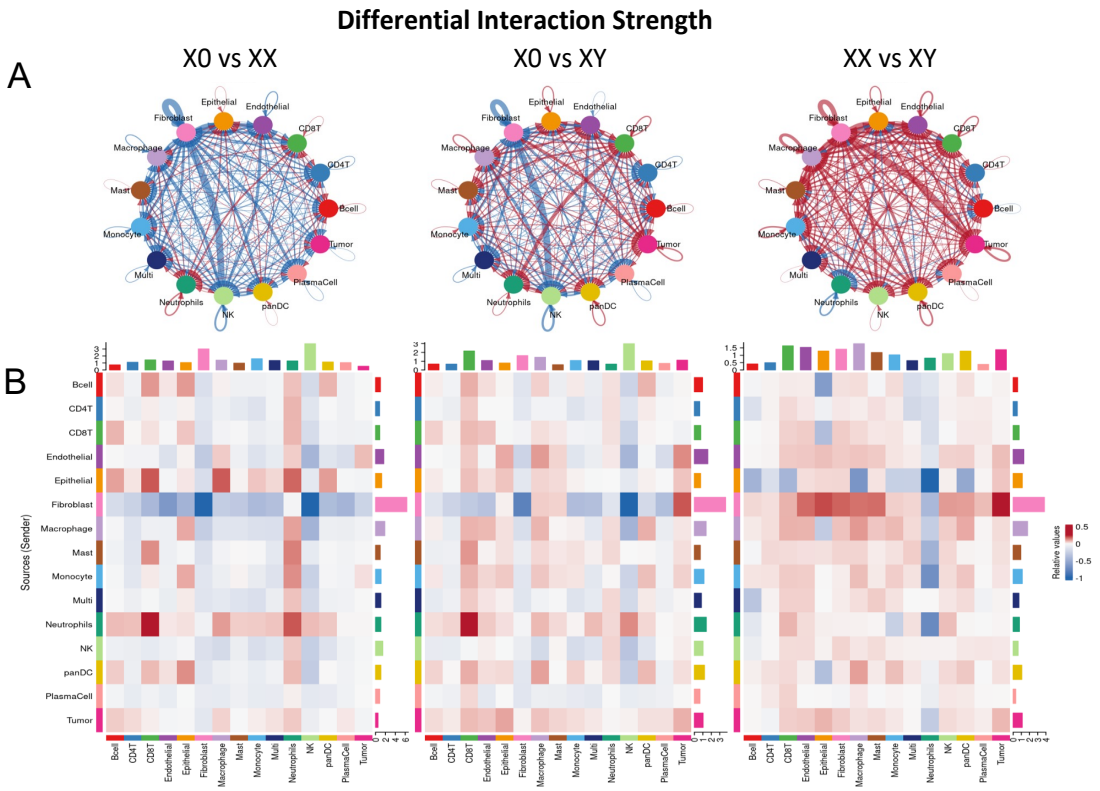


Figure 25. Differential cell-cell communication strength across sex chromosome dosage groups in HPV-negative tumors. The analysis visualizes differential interaction strength, as defined by CellChat, between cell types in XØ, XX, and XY tumors. Red indicates stronger signaling in the second group of the comparison, while blue reflects stronger signaling in the first group. *(A)* Chord diagrams depict global intercellular communication networks for each pairwise comparison (XØ vs XX, XØ vs XY, and XX vs XY), with edge thickness corresponding to the relative strength of interactions. *(B)* Heatmaps quantify the relative interaction strength between sender (rows) and receiver (columns) cell types, highlighting specific interaction patterns across groups. Fibroblasts emerge as major stromal signaling hubs, consistently exhibiting the highest differential interaction strength across all comparisons. This effect is particularly pronounced in XY tumors, where fibroblasts show enhanced outgoing communication toward tumor cells, highlighting their pivotal role in shaping the tumor–stroma interface.

Incorporating fibroblast subtypes into the analysis provided additional insights into their role in the TME. mCAFs and iCAFs emerged as key players across all sex chromosome dosage groups. These fibroblast subtypes demonstrated the strongest outgoing signals, further highlighting their central role in coordinating communication within the TME (Figure 26.A). However, when comparing the differential interaction strength between the groups, I found that mCAFs and iCAFs displayed the most variation in their communication behavior between the XØ, XX, and XY groups (Figure 26.B). This suggests

that the functional role of these fibroblast subtypes may differ based on the sex chromosome dosage, potentially influencing tumor progression in distinct ways across the different groups.

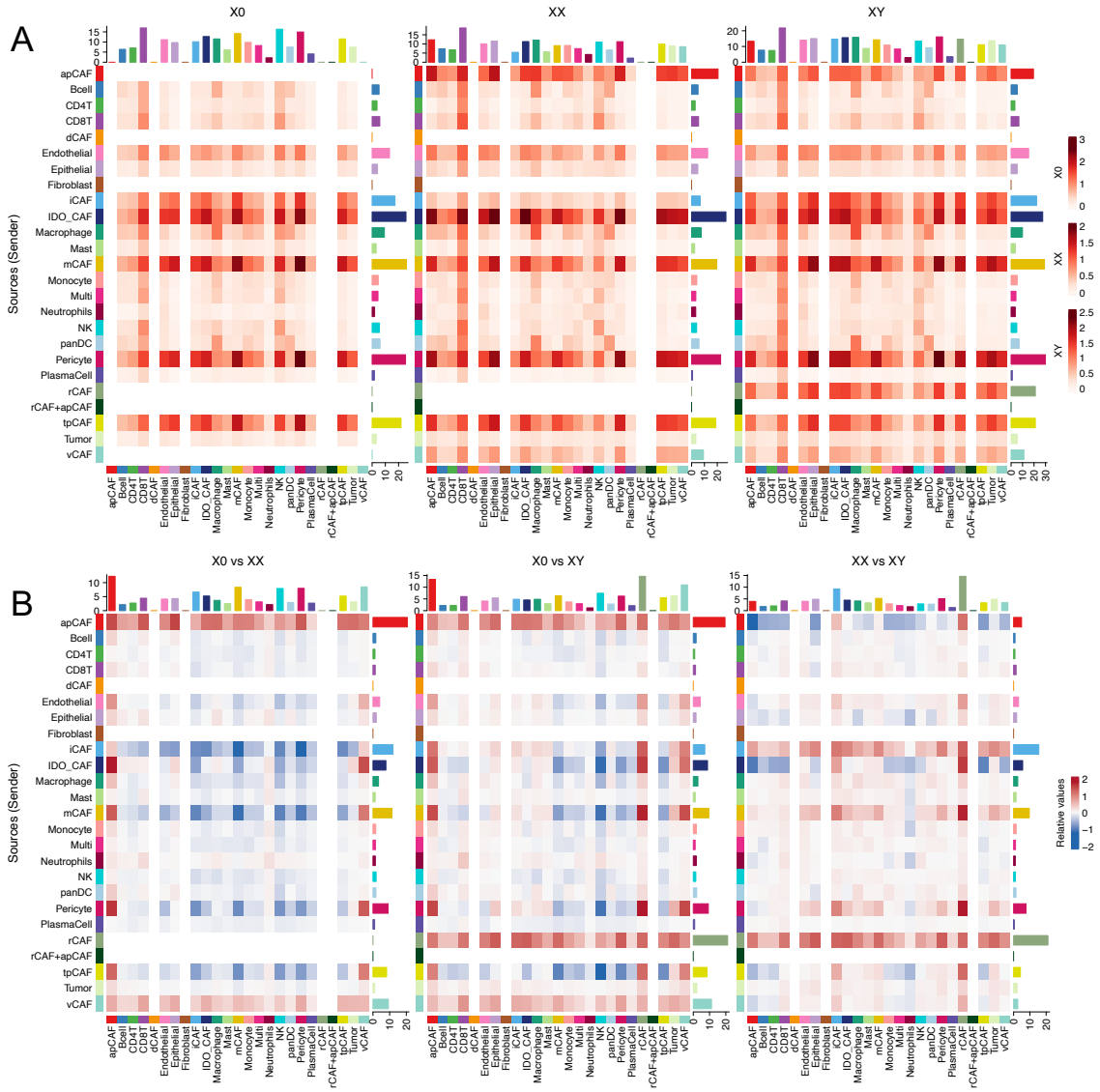


Figure 26. Differential signaling profiles of fibroblast subtypes across sex chromosome dosage groups. Cell–cell communication analysis was extended by incorporating fibroblast subtypes, providing a more granular view of stromal interactions in the TME. *(A)* The upper panel shows the absolute interaction strength of each sender–receiver cell type pair within the XØ, XX, and XY groups. *(B)* The lower panel presents the differential interaction strength between groups (XØ vs XX, XØ vs XY, and XX vs XY). In this analysis, red indicates stronger interactions in the second group of the comparison, while blue indicates stronger interactions in the first. mCAFs and iCAFs emerged again as the fibroblast subtypes with pronounced differences in communication across sex chromosome dosage groups.

Together, these findings highlight the importance of cell-cell communication in shaping the TME and reveal how differences in sex chromosome dosage can alter the signaling landscape, particularly through the activity of fibroblasts. By integrating these communication patterns into our broader understanding of the TME, I gain further insight into how cellular interactions contribute to the distinct biological behaviors of tumors across XX, XY, and XØ individuals.

4.2.9 Patient-Specific Analysis Reveals Distinct Patterns of Tumor-iCAF Interactions by Sex Chromosome Dosage

To ensure the observed cell-to-cell communications reflect interactions occurring within individual tumors, I performed a patient-specific analysis focusing on interactions between tumor cells and iCAFs, given that this was the most differential interaction across sex chromosome dosage groups. By examining interactions on a per-patient basis within XX, XY, and XØ groups, I aimed to distinguish true intra-tumoral signaling from potential inter-patient artifacts in pooled datasets.

For each patient, I visualized the interaction strength using heatmaps categorized by the presence and activity of iCAFs: (1) iCAFs detected and interacting with tumor cells, (2) iCAFs detected but without interaction with tumor cells, and (3) patients with no iCAFs detected, and example of each type can be found in Figure 27. These visualizations allowed me to assess variations in tumor-iCAF signaling dynamics across individual patients within each sex chromosome dosage group, highlighting not only the presence of specific iCAF populations but also the direct communication channels active within each patient's TME.

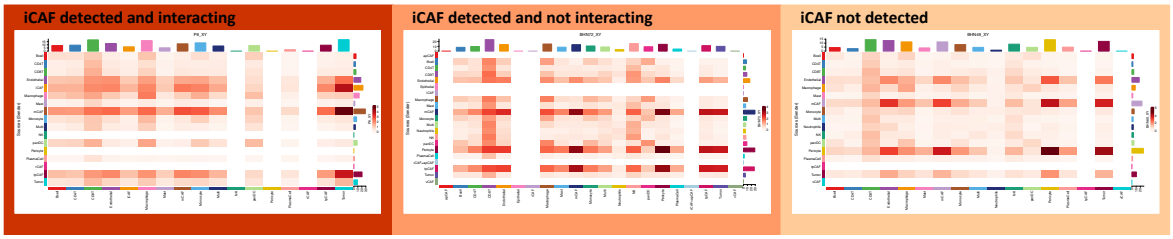


Figure 27. Patient-level variation in iCAF–tumor interactions across the TME. To investigate inter-patient variability in iCAF activity, heatmaps were generated to visualize cell–cell communication strength, stratified by iCAF presence and interaction status. Each heatmap represents an example patient from one of three categories: *(Left) iCAF detected and interacting with tumor cells*: strong communication signals between iCAFs and tumor cells suggest active stromal involvement in the TME. *(Middle) iCAF detected but not interacting with tumor cells*: iCAFs are present but show no direct signaling toward tumor cells, indicating a potential functional disconnect or niche restriction. *(Right) iCAF not detected*: no iCAF population was observed in the TME.

My patient-specific analysis of iCAF presence and interactions with tumor cells reveals distinct patterns across the XX, XY, and XØ groups, underscoring the heterogeneity in stromal-tumor communication linked to sex chromosome dosage (Figure 28).

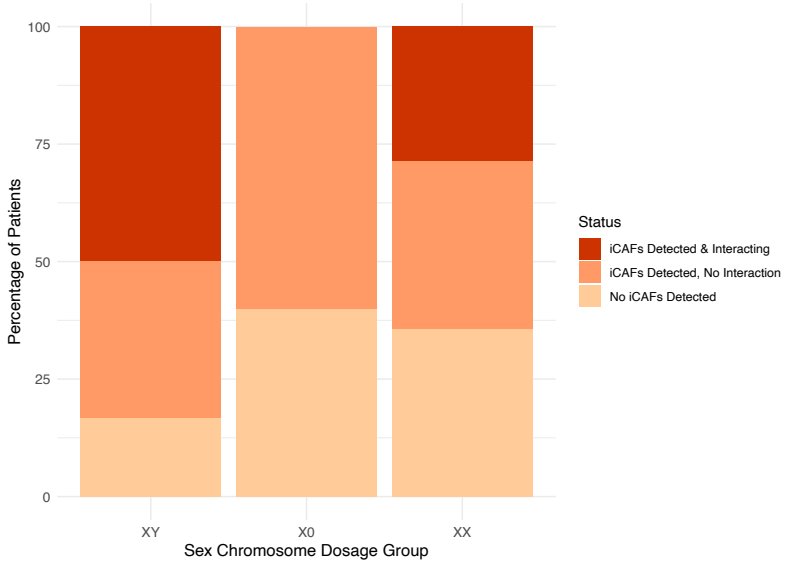


Figure 28. Proportion of patients with different iCAF–tumor interaction states across sex chromosome dosage groups. Bar plot showing the percentage of patients within each sex chromosome dosage group (XY, XØ, XX) classified into three categories based on iCAF presence and interaction with tumor cells: (1) iCAF detected and interacting with tumor cells (dark red), (2) iCAF detected but not interacting (orange), and (3) no iCAF detected (peach). XY patients showed the highest proportion of iCAF–tumor interactions, while XØ patients had the lowest, highlighting sex chromosome dosage–linked differences in stromal engagement.

In the XY group, the majority of patients exhibited iCAF presence, with six out of twelve patients displaying active interactions between iCAFs and tumor cells. This relatively high frequency of tumor-iCAF communication suggests a prominent role for inflammatory signaling in XY individuals, potentially facilitating tumor progression and influencing immune cell recruitment within the TME. In four additional XY patients, iCAFs were present but not actively engaging with tumor cells, and only two patients had no detectable iCAFs. These findings imply that the inflammatory signaling potential, while generally high, varies among XY individuals, reflecting a degree of intra-group heterogeneity.

For XØ patients, no cases showed active interactions between iCAFs and tumor cells, with six patients displaying iCAFs without interaction and four patients having no detectable iCAFs. The absence of direct iCAF-tumor cell interactions across XØ patients could indicate a less inflamed and more immunosuppressive TME in this group, which aligns with findings of reduced immune activity previously observed in XØ tumors. This

absence of interaction suggests that, in XØ patients, iCAFs may play a more limited or indirect role in modulating tumor behavior.

In the XX group, the variability in iCAF-tumor interactions was notable. While four patients showed active iCAF-tumor interactions, these varied in strength, with two patients exhibiting high interaction strength and two displaying low strength. This diversity, coupled with five patients showing iCAFs without active tumor interactions and five with no iCAFs, suggests a more variable inflammatory landscape in XX tumors. The range of interaction strengths and presence of non-interacting iCAFs may indicate that, in XX patients, iCAF activity and tumor engagement are more context-dependent, potentially influenced by additional factors such as immune infiltration and local TME composition.

Importantly, Fisher's exact test revealed a statistically significant difference in the distribution of iCAF–tumor interaction patterns between XY and XØ patients ($p = 0.035$), underscoring a potential biological divergence in the role of iCAFs across sex chromosome dosage groups.

4.2.10 Tumor-Stroma Crosstalk in HNSCC Reveals That XY Fibroblasts Promote Tumor Survival, Stemness, and Collective Invasion

To investigate potential functional differences among iCAFs across chromosome dosage groups, I first analyzed Hallmark pathway enrichment related to fibroblast activity (Figure 29.A). The analysis revealed that XY iCAFs exhibited the highest scores across most fibroblast-associated pathways, including Epithelial-Mesenchymal Transition (EMT), TGF-beta signaling, hypoxia and angiogenesis, suggesting a metabolic and signaling profile distinct from XØ and XX iCAFs. Conversely, XØ and XX fibroblasts displayed relatively higher scores in specific pathways, such as inflammatory response, glycolysis and Wnt-Beta Catenin signaling, indicating potential differences in immune modulation and stromal maintenance.

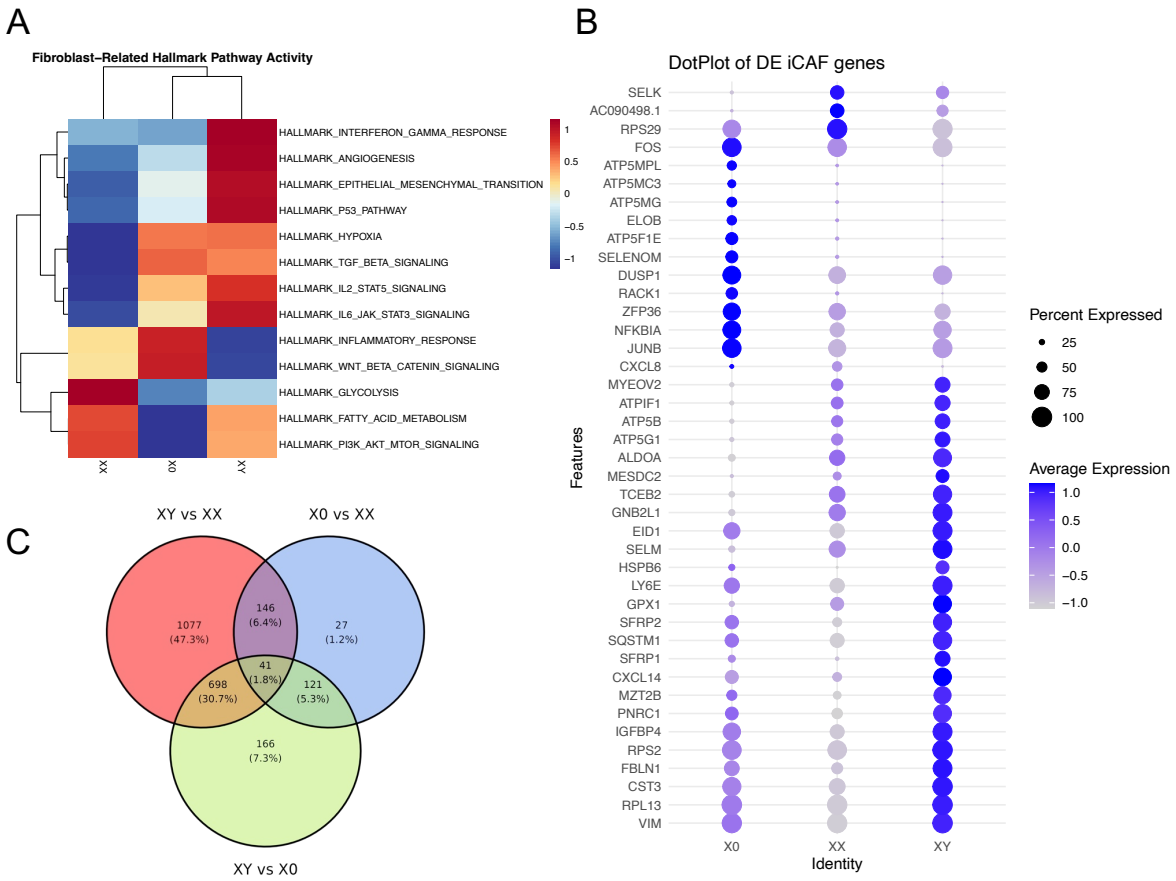


Figure 29. Comparative analysis of iCAFs across sex chromosome dosage groups.

(A) Heatmap displaying Hallmark pathway enrichment scores for iCAFs stratified by sex chromosome dosage (XØ, XX, XY), highlighting variation in fibroblast-related signaling activity across groups. *(B)* Dot plot showing scaled expression values and detection frequency of selected differentially expressed genes in iCAFs across XØ, XX, and XY tumors. The size of the dots reflects gene detection rate, while the color scale indicates average expression level. *(C)* Venn diagram summarizing the overlap of differentially expressed genes identified in pairwise comparisons between XØ, XX, and XY iCAFs, with 41 genes shared across all three comparisons.

To further explore molecular differences, I performed a DE analysis across the three fibroblast groups. This analysis identified a set of 41 genes (Figure 29.C) consistently differentially expressed across all comparisons. Several of these genes are functionally relevant to fibroblasts and HNSCC, including VIM (Vimentin), a key EMT marker; CXCL8 and CXCL14, chemokines involved in immune modulation; and FBLN1 (Fibulin-1), an extracellular matrix protein with roles in fibroblast function and tumor-stroma interactions. Additionally, SFRP1 and SFRP2, known regulators of Wnt signaling, were differentially expressed, supporting a role for fibroblast-associated pathways in tumor

progression. Notably, JUNB and FOS, transcription factors involved in fibroblast activation, as well as NFKBIA and CXCL8, regulators of inflammatory signaling, exhibited stronger expression in XØ fibroblasts, while VIM, CXCL14, FBLN1, and SFRP family members were more highly expressed in XY fibroblasts, suggesting distinct activation states between these groups (Figure 29.B).

To gain insight into the biological processes underlying these transcriptional differences, I performed Gene Ontology (GO) enrichment analysis separately for each chromosome dosage group (Figure 30). XØ iCAFs were enriched for pathways associated with tumor necrosis factor (TNF) response, lipopolysaccharide (LPS) signaling, and ATP biosynthesis, suggesting an inflammatory and metabolically active fibroblast phenotype. In contrast, XY iCAFs exhibited significant enrichment in extrinsic apoptotic signaling regulation, stem cell fate commitment, and planar cell polarity, indicating a distinct role in tumor-stromal interactions and fibroblast differentiation. Notably, XX iCAFs showed limited pathway enrichment, with only cytoplasmic translation emerging as a significantly enriched term, suggesting a more general fibroblast phenotype compared to XØ and XY groups.

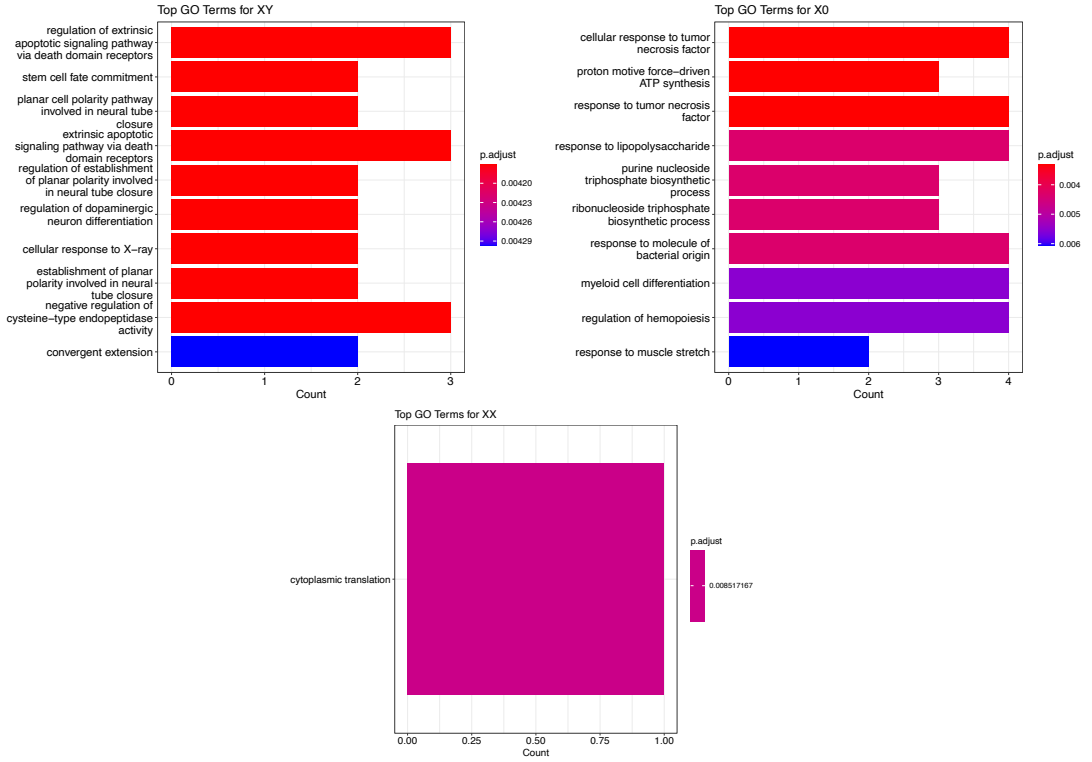


Figure 30. Gene Ontology (GO) enrichment analysis of iCAFs by sex chromosome dosage. Bar plots display the top significantly enriched GO biological processes for differentially expressed genes in iCAFs from (top left) XY, (top right) XØ, and (bottom) XX tumors. Bar length indicates gene count per term, while the color gradient reflects the adjusted p-value (FDR), with darker hues denoting higher significance. The XY and XØ groups show enrichment in distinct pathways, while the XX group presents limited enrichment, with only a single significant term identified.

Together, these results indicate that fibroblasts within the HNSCC TME exhibit distinct functional states depending on their chromosome dosage status, with XY fibroblasts demonstrating a more pro-tumorigenic role, contributing to tumor-stroma interactions through enhanced survival signaling, stemness promotion, and facilitation of collective invasion, while XØ fibroblasts are characterized by a heightened inflammatory response, and XX fibroblasts present a less specialized transcriptional program.

4.2.11 Deconvolution of TCGA Data Reveals Differences in mCAFs and iCAFs Proportions Across Sex Chromosome Dosage Groups

To further investigate if the cellular composition differences observed in my single-cell analysis can also be traced back into the bulk RNA-seq data, I used BayesPrism to

deconvolute the TCGA dataset. BayesPrism is a deconvolution tool designed for bulk transcriptomic data, using probabilistic modeling to assign proportions of various cell types based on single-cell RNA-seq reference profiles. This approach allows us to infer the cellular composition in bulk datasets, using information from our detailed single-cell atlas to identify and quantify specific cell populations within the TCGA data. By doing so, I aimed to check if specific cell type distributions observed in the single-cell setting across XX, XY, and XØ sex chromosome dosage groups are also detectable in the bulk data.

The box plots (Figure 31.A-B) illustrate the estimated proportions of each cell type across the three dosage groups. Kruskal-Wallis tests were conducted to identify significant differences among XX, XY, and XØ groups for each cell type. Notably, key fibroblast populations, including mCAFs and iCAFs, showed significant differences among the sex chromosome dosage groups (Suppl. Table 13). These findings suggest that these fibroblast subtypes may have distinct roles in the TME across different sex chromosome configurations. Importantly, the directionality of these differences aligns with our single-cell findings. Specifically, the deconvolution results indicate that fibroblast populations are most abundant in XY tumors, followed by XØ, and least abundant in XX tumors (Table X). This trend mirrors our single-cell analysis, reinforcing the notion that fibroblast presence, particularly mCAFs and iCAFs, is enriched in XY tumors. These observations further support the idea that fibroblast-driven stromal remodeling may be influenced by sex chromosome dosage, potentially shaping the tumor microenvironment in a dosage-dependent manner.

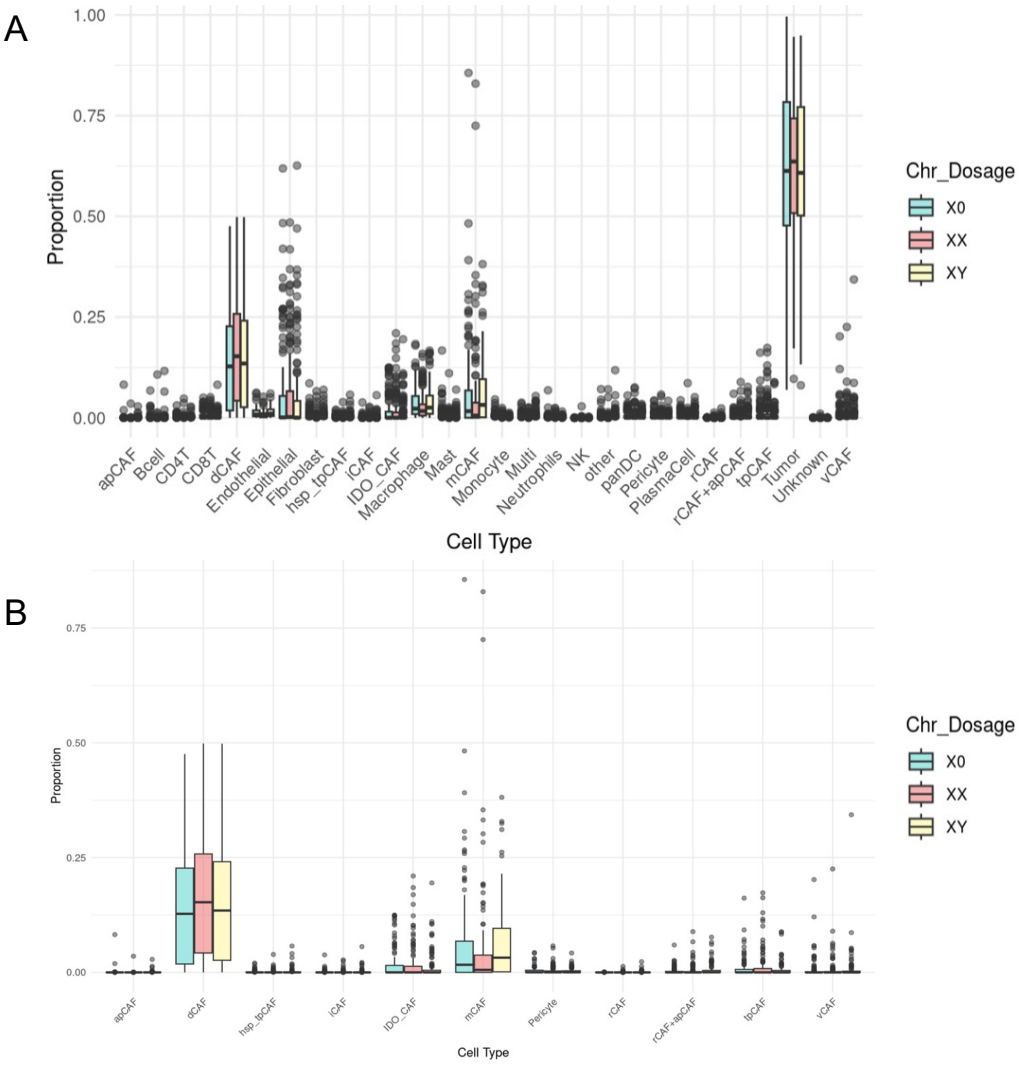


Figure 31. Cell type deconvolution of TCGA bulk RNA-seq data using BayesPrism and single-cell reference profiles. Box plots display the estimated proportions of cell types (*(A)* all TME cell types, *(B)* fibroblast subtypes) in bulk tumor samples across XX, XY, and XØ sex chromosome dosage groups. Inference was performed using BayesPrism, with reference signatures derived from my single-cell HNSCC atlas. Each dot represents an individual patient, and boxes show interquartile range and median for each group. This approach enables the comparison of bulk-level cellular composition across dosage groups and highlights differences in fibroblast subtype abundance detectable at the bulk transcriptome level.

These results strengthen our single-cell findings, showing that some of the differences in cell type composition can also be seen in the bulk data. The alignment between single-cell and bulk RNA-seq data for key cell populations, especially iCAFs and mCAFs, highlights the important role these fibroblast subtypes play in shaping the unique TME characteristics of the XX, XY, and XØ groups.

4.2.12 Enrichment of COX2 and AR Expression in iCAFs and mCAFs Highlights Potential Tumor-Stromal Interactions Specific to Sex Chromosome Dosage Groups

Building on our initial exploration into COX2's role as a driver of inflammation in cancer and its differences in expression among the sexes in HNSCC patient stainings, I revisited its impact within the TME by examining PTGS2 (COX2) and androgen receptor (AR) expression across the sex chromosome dosage groups (XX, XY, and XØ). COX2, known for promoting inflammatory signaling, has been associated with processes that favor tumor progression, such as immune cell recruitment, angiogenesis, and cell proliferation[144]. Earlier studies pointed to COX2 overexpression as a potential marker of more aggressive tumor behavior, leading us to explore its function within specific cell types in the TME. Likewise, AR, though mainly studied in hormone-responsive cancers, has shown signs of influencing cellular behaviors in a broader range of contexts[47].

With the single-cell RNA sequencing atlas now available, I aimed to see if COX2 and AR have specific expression patterns within distinct cell types that might help explain the differences in the TME linked to each sex chromosome dosage group. By examining these expression patterns, I aimed to determine whether and how COX2 and AR contribute to shaping the stromal and immune composition and behavior in ways that correspond with the observed TME variations across these groups.

In our analysis of PTGS2 (COX2) expression across general cell types, I observed distribution across multiple cell populations, which led me to examine fibroblasts more closely given their established role as a key cell type of interest (Figure 32.A). When I specifically analyzed fibroblast subtypes, I found that COX2 expression was particularly elevated in iCAFs and mCAFs within the XY group, with comparatively lower expression in XX and XØ fibroblasts (Figure 32.C). In contrast, AR expression was largely confined to fibroblasts across the general cell types, but it still showed higher levels in the XY group compared to XX and XØ (Figure 32.B). Within fibroblast subtypes, I again observed that iCAFs and mCAFs in XY patients displayed the most prominent AR expression (Figure

32.D). While these findings suggest that COX2 and AR may contribute to the distinct stromal environment observed in XY tumors, their restricted expression to select cells within the population highlights the need for further investigation into whether downstream inflammatory and androgen-related pathways are broadly activated within the fibroblast compartment.

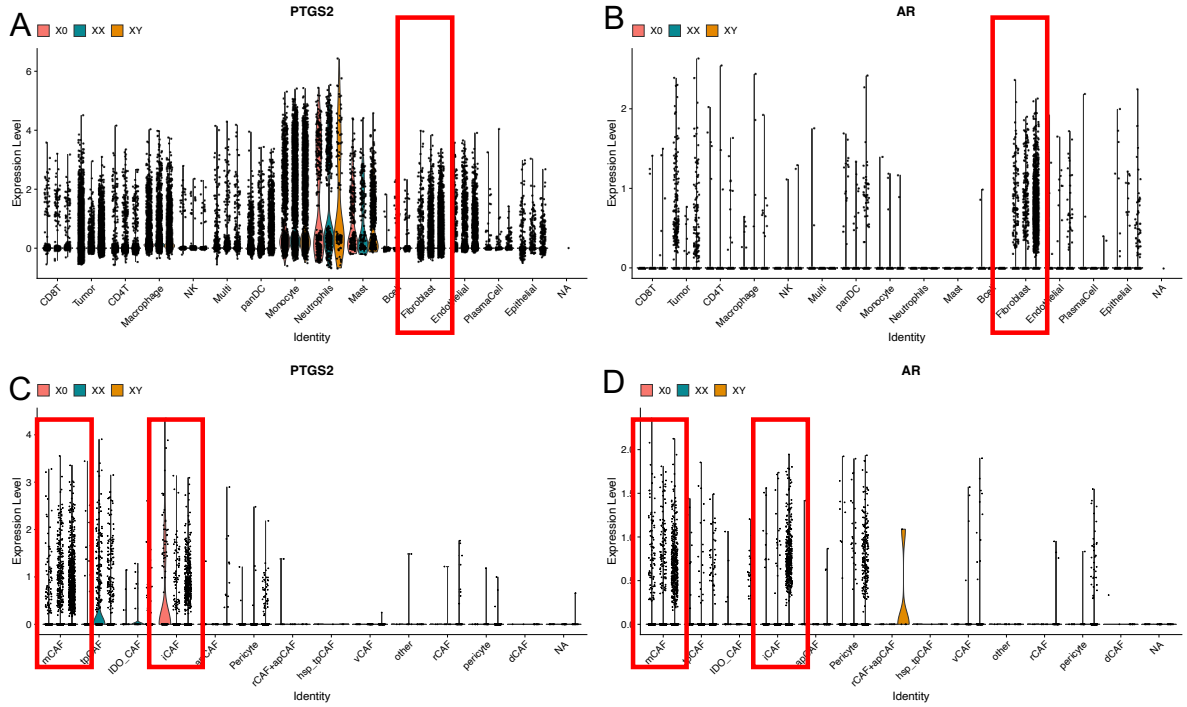


Figure 32. Expression patterns of *PTGS2* (COX2) and *AR* across general cell types and fibroblast subtypes in the HNSCC single-cell atlas. *(A, B)* Violin plots showing the expression levels of *PTGS2 (A)* and *AR (B)* across major cell types, with fibroblasts highlighted in red boxes. *PTGS2* shows widespread distribution, while *AR* expression is primarily restricted to fibroblasts. *(C, D)* Expression of *PTGS2 (C)* and *AR (D)* within annotated fibroblast subtypes. Both genes show elevated expression in inflammatory (iCAF) and matrix (mCAF) fibroblasts, with highest levels in the XY group, suggesting sex chromosome dosage–linked modulation of stromal signaling within the TME.

Further exploration of the co-expression of COX2 and AR revealed that in the XY group, iCAFs demonstrated the highest levels of simultaneous expression of these two markers, a pattern not observed in the XX or XØ groups (Figure 32). This co-expression within iCAFs suggests a possible interaction between inflammatory and androgen receptor-mediated pathways that could modulate the TME in XY patients. The presence of co-expressing iCAFs in XY tumors might indicate a unique fibroblast-mediated mechanism contributing

to the tumor's immune and stromal landscape, potentially facilitating tumor-stromal communication and supporting tumor growth.

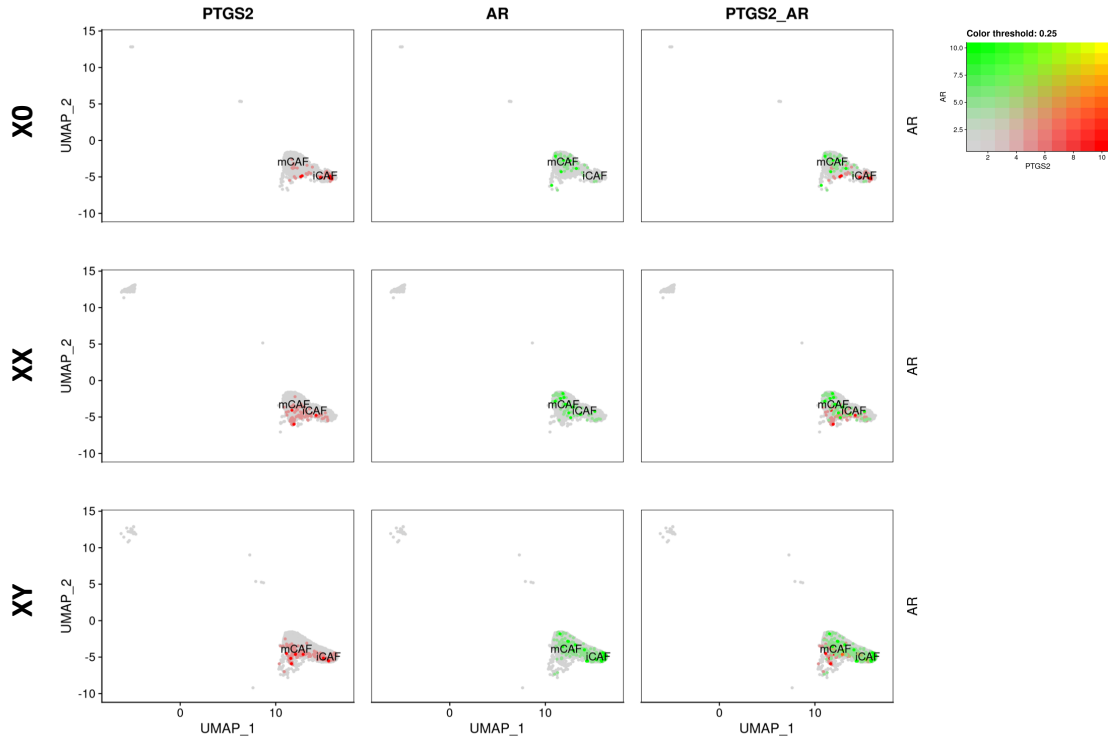


Figure 33. UMAP visualization of PTGS2 (COX2) and AR co-expression in fibroblast subtypes across sex chromosome dosage groups. UMAPs display expression levels of PTGS2 (left), AR (middle), and their combined co-expression (right) in mCAFs and iCAFs for XØ (top row), XX (middle row), and XY (bottom row) fibroblast populations. The composite color scale (top right) represents joint expression, with green indicating high AR expression, red indicating high PTGS2, and yellow indicating co-expression of both markers. iCAFs in XY tumors exhibit the most prominent co-expression of PTGS2 and AR, a pattern absent in XX and XØ groups.

4.2.13 Differential Incoming/Outgoing Communication Pathways Between Tumor Cells and iCAFs Across Sex Chromosome Dosage Groups

To explore the communication landscape between tumor cells and iCAFs across sex chromosome dosage groups (XX, XY, and XØ), I investigated signaling pathways using the SC atlas data processed through the CellChat framework. I used the calculated pathway communication scores to get a clearer picture of how much and how strongly tumor cells and iCAFs interact in the TME. By focusing on these pathways, I wanted to see if

differences in sex chromosome dosage affect the way these cells communicate, especially in pathways involved in immune response and cell growth.

I calculated pathway communication scores for each sex chromosome dosage group and selected interactions between tumor cells and iCAFs. This was done calculating centrality scores, which measure how much influence each pathway has on communication by looking at both signals sent out and received by each cell type. This let us quantify how important specific pathways are for communication in each group. To improve the comparability across groups, I identified a shared set of pathways and designed a ranking metric which highlighted pathways showing differences between the XX, XY, and XØ groups. From this ranking, I selected the top 20 pathways, as these showed the most differences and were therefore the most relevant and I clustered the pathways to find pathways specific for each sex chromosome dosage group. For this I did it from both perspectives, incoming and outgoing signals for iCAFs and incoming and outgoing singnals for tumor.

For the analysis of incoming and outgoing pathways in tumor cells (Figure 34.A-B), the XX group exhibited stronger scores across all pathways, suggesting that tumor cells in this group might have a more active signaling behavior. I observed that pathways such as Apo A, FASLG, IL1, and NPR2 were exclusively present in the XØ group for incoming signals. There was limited overlap between XØ and XY groups, while more pathways were shared between XY and XX, including Activin, SEMA3, and BMP. Unique pathways for the XX group included TGFβ, IL4, and PARs, highlighting a more active communication role in this dosage group. For outgoing pathways, the XØ group displayed activation of pathways such as VEGF, MPZ, CD40, and JAM, which were shared with XX. Notably, IL1 was strong but specific to XØ, while pathways like KIT, SEMA4, and VTN were stronger in XY and shared with XX.

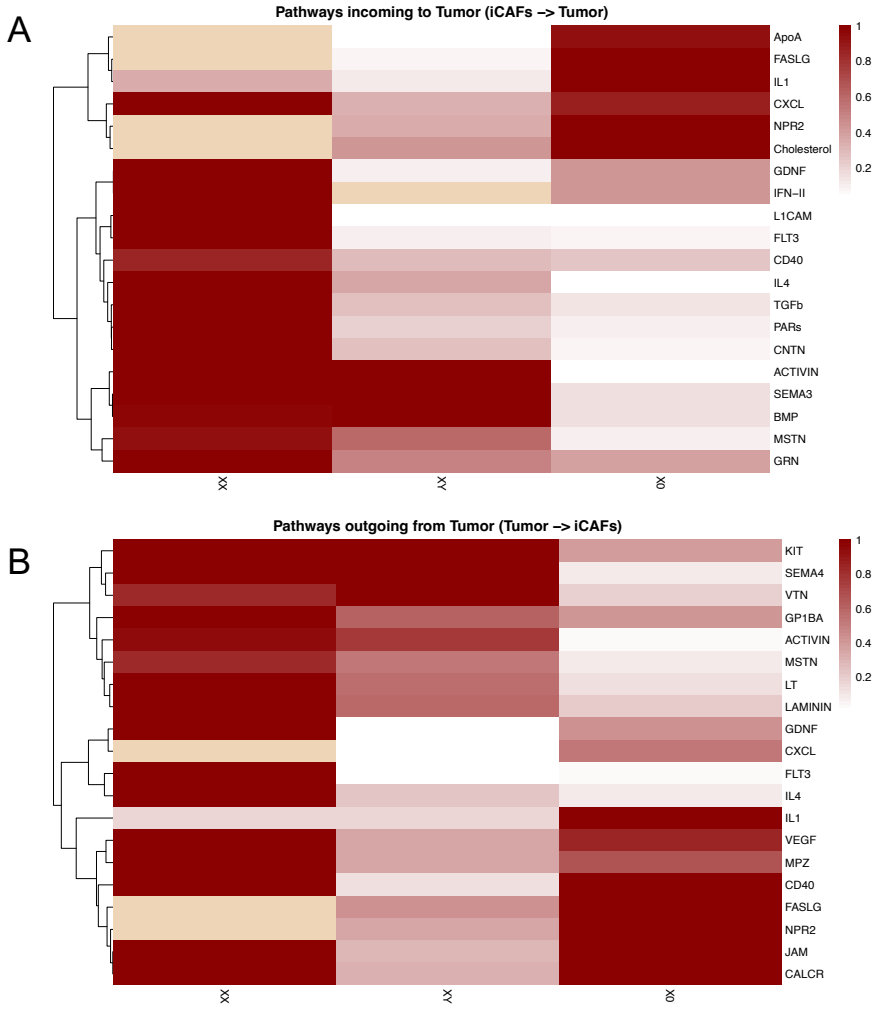


Figure 34. Heatmap visualization of incoming and outgoing signaling pathway activity in tumor cells across sex chromosome dosage groups. *(A)* Incoming signaling pathways show the relative strength of pathway activity received by tumor cells from iCAFs in XØ, XX, and XY groups. (B) Outgoing signaling pathways show the relative strength of pathway activity sent from tumor cells to iCAFs in XØ, XX, and XY groups.

In contrast, the signaling in iCAFs revealed a different pattern (Figure 35.A-B). Overall, pathway scores were much weaker in the XX group compared to the XY and XØ groups. Incoming signals seem frequently shared between XØ and XY and included IL4, CNTN, IFN-II, and VEGF, whereas pathways like RELN, SEMA3, and ncWNT were more activated only in XØ. NPR2 and NT were more activated in the XY group. The XX group showed activation primarily of PLAU, which was unique to this group, and IL1, which was shared with XY. For outgoing signals, VEGF, IL4, and IFN-II were again shared between XØ and XY, while CD40 and NPR2 were more prominent in XY. Pathways like SEMA4 and BMP

were more exclusive to XØ. In the XX group, IL1 and VTN were more active, with IL1 shared with XY and VTN present in both XY and XØ but primarily active in XØ.

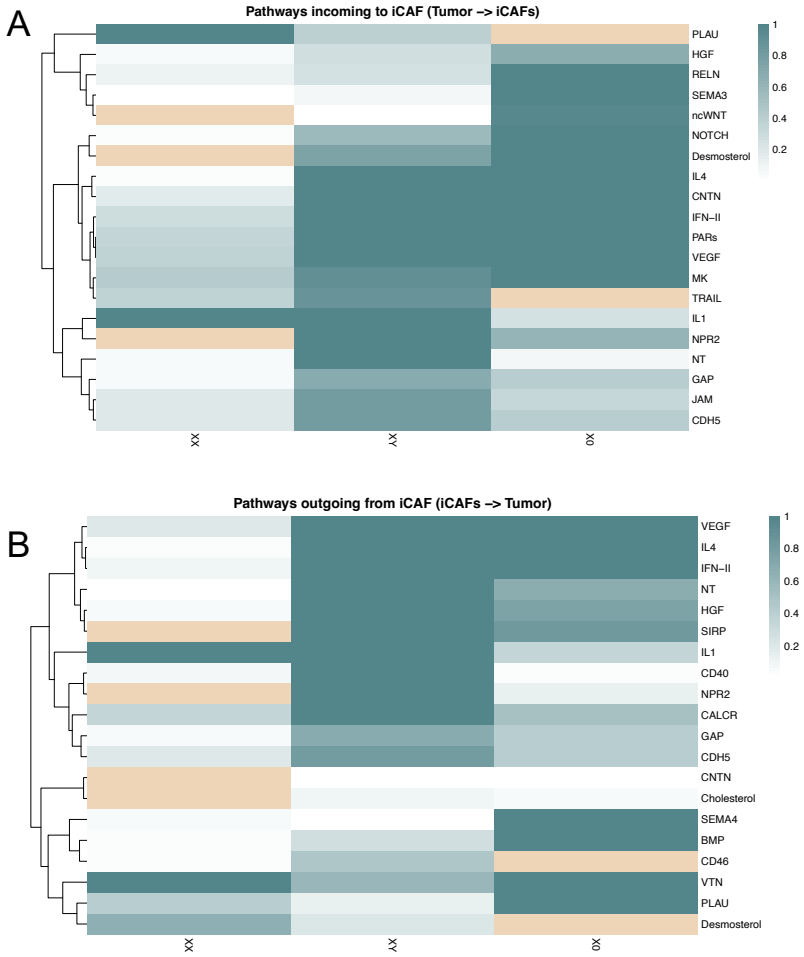


Figure 35. Heatmap visualization of incoming and outgoing signaling pathway activity in iCAFs across sex chromosome dosage groups. *(A)* Incoming signaling pathways represent the relative strength of pathway activity received by iCAFs from tumor cells in XØ, XX, and XY groups. *(B)* Outgoing signaling pathways show the strength of signals sent from iCAFs to tumor cells across dosage groups.

To further understand the shared communication pathways between tumor cells and iCAFs, I extended my analysis by selecting pathways that appeared in both the outgoing signals from one cell type and the incoming signals to the other. This allowed us to focus on pathways that facilitate reciprocal interactions within the TME. I identified these shared pathways by intersecting the top-ranked outgoing and incoming pathways for both tumor cells and iCAFs. This comparison was performed for two sets of pathways: outgoing tumor to incoming iCAF and outgoing iCAF to incoming tumor (Figure 36.A-B).

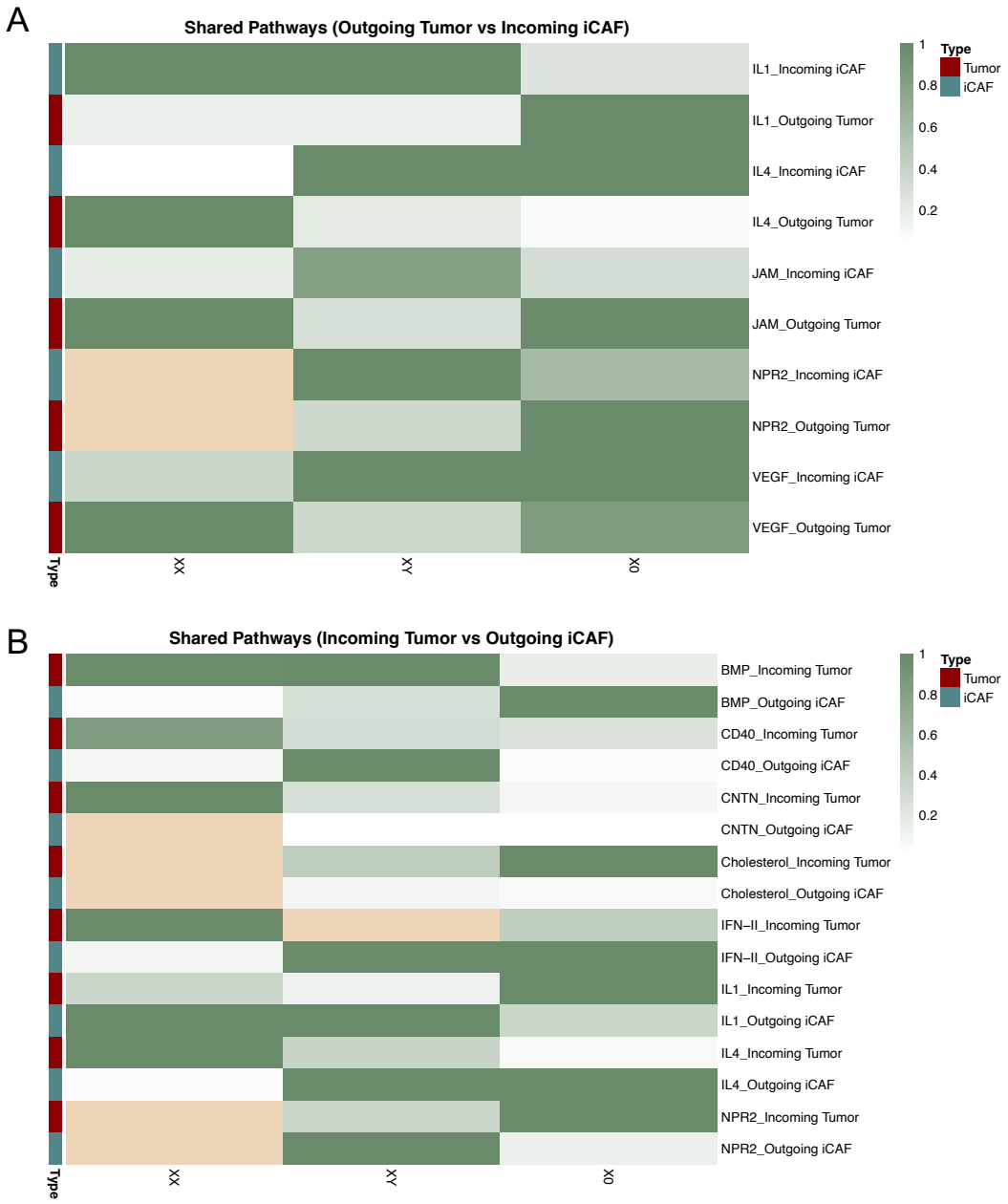


Figure 36. Shared signaling pathways mediating reciprocal communication between tumor cells and iCAFs across sex chromosome dosage groups. *(A)* Heatmap of pathways identified as both outgoing from tumor cells and incoming to iCAFs in each group (XØ, XX, XY), representing potential tumor-to-stroma signaling. *(B)* Heatmap of pathways identified as both outgoing from iCAFs and incoming to tumor cells, indicating potential stroma-to-tumor signaling.

In this analysis IL1 and IL4 had strong outgoing and incoming pathway scores. This finding indicates the importance of these pathways in tumor-iCAF crosstalk, where they may be involved in modulating immune responses. The strength of IL4 in the XY group,

particularly, points to an immune-modulatory role that may be influenced by sex chromosome dosage. BMP and CD40 appeared with notable intensity in incoming tumor and outgoing iCAF signaling, especially within XY and XØ groups, suggesting a potential enhancement in immune-related feedback mechanisms, fibroblast activation states and the regulation of cell proliferation, differentiation, and apoptosis. NPR2 was also present in both directions and showed differences between XX and XØ/XY groups, the lack of receptor and ligand in the XX group made the score be NA. This could indicate that chromosomal dosage could affect pathways involved in matrix remodeling or cellular adhesion, with potential implications for tumor progression and microenvironment interactions.

In summary, this analysis reinforced the critical roles of IL4, IL1, and CD40 as central mediators in tumor-iCAF communication, with distinct activity patterns linked to chromosomal dosage. The observed variations suggest that sex chromosome dosage may significantly influence not only the strength but also the directionality of signaling pathways. This provides insights into potential therapeutic targets, as disrupting these key pathways could impact feedback-rich communication networks and potentially alter tumor behavior and treatment responses.

5. Discussion

This study presents a comprehensive analysis of sex chromosome dosage differences in HNSCC, integrating bulk and single-cell transcriptomic data to investigate their impact on tumor biology and the TME.

5.1 LoY and EDY Are Highly Prevalent in HNSCC and Are More Frequent in HPV-Negative Tumors

I first identified loss of the Y chromosome (LoY) and extreme downregulation of Y chromosome genes (EDY) in bulk RNA-seq data, revealing that these phenomena are highly prevalent among male HNSCC patients, as it was previously described in literature[50,145]. Across multiple cancer types, LoY has been widely documented as a frequent somatic event, particularly in hematological malignancies, bladder cancer, and renal cancer, where it has been implicated in tumor progression and immune evasion[146,147]. However, my analysis suggests that HNSCC is among the tumor types where LoY and EDY occur with the highest frequency, indicating a strong selection pressure for the loss or silencing of Y chromosome genes in these tumors. Additionally, I found that in the majority of cases where LoY was present, EDY was also detected, indicating that extreme downregulation of Y chromosome genes can often be traced to the physical loss of the chromosome. However, a subset of tumors exhibited EDY in the absence of detectable LoY, suggesting the involvement of additional regulatory mechanisms such as epigenetic modifications or transcriptional repression. For instance, transcriptional repression via dysregulated transcription factors or noncoding RNAs may contribute to EDY[145]. These cases highlight the need for further investigation into the non-genetic factors contributing to Y chromosome gene silencing and their potential implications for tumor progression.

A key observation in my study is that LoY is significantly more prevalent in HPV-negative tumors compared to HPV-positive tumors, aligning with the distinct molecular and immune landscapes of these subtypes[50,148]. HPV-negative HNSCC, which is predominantly associated with tobacco and alcohol exposure, exhibited higher rates of LoY, suggesting that this phenomenon may be linked to environmental carcinogens or intrinsic genomic instability associated with HPV-negative disease[26,149]. The lower frequency of LoY in HPV-positive tumors may reflect differences in tumor evolution, immune selection pressures, or the presence of viral-driven oncogenic pathways that preserve Y chromosome gene expression[150].

Additionally, my analysis demonstrated that LoY and EDY are significantly more pronounced in tumor tissue compared to adjacent normal tissue, reinforcing the hypothesis that these events are cancer-driven rather than background somatic alterations[70,145,151]. While previous studies have reported age-related LoY in non-malignant cells, my findings suggest that the selective pressure for Y chromosome loss is not only a byproduct of aging but a specific feature of tumorigenesis in HNSCC[152,153]. The increased occurrence in tumor cells, combined with the relatively lower frequency in normal tissues, highlights the potential functional role of Y chromosome genes in suppressing oncogenic processes.

I examined the association between EDY and patient survival, finding that patients with lower Y chromosome expression had significantly worse overall survival ($p = 0.029$) when analyzing all HPV-positive and HPV-negative cases together. However, when stratified by HPV status, this difference was only significant in the combined cohort, as HPV-positive tumors predominantly fell into the high expressor group. Within the HPV-negative subgroup, EDY did not reach statistical significance, suggesting that while not the sole driver of tumor aggressiveness, it may contribute to molecular vulnerabilities that could be targeted for therapy. A previous study reported significantly shorter overall survival in HPV-negative HNSCC patients with LoY ($p = 0.003$)[50]. This contrasts with my findings, where EDY alone was not a significant prognostic factor in HPV-negative cases. This discrepancy may arise from differences in classification, as LoY reflects chromosomal loss, while EDY represents transcriptional downregulation. Though often linked, these

processes do not always overlap, suggesting distinct biological implications. Further investigation is needed to clarify the functional consequences of Y chromosome inactivation and its role in shaping patient outcomes.

Overall, our findings establish LoY and EDY as recurrent and functionally relevant events in HNSCC, particularly in HPV-negative tumors, where their prevalence suggests a potential role in shaping tumor progression and therapeutic responses. While the exact mechanisms through which LoY and EDY drive tumor aggressiveness remain to be fully elucidated, the strong association with immune exclusion and tumor-intrinsic genomic instability highlights their potential as biomarkers for patient stratification. Moreover, the distinct differences in LoY and EDY frequency between HPV-positive and HPV-negative tumors suggest that these subtypes are driven by distinct oncogenic mechanisms, underscoring the need for treatment strategies tailored to their specific molecular landscapes.

5.2 Sex Chromosome Dosage Influences Tumor Heterogeneity and Immune Evasion in HNSCC

To better capture the role of sex chromosome dosage in tumor biology, I developed a classification framework that moves beyond conventional clinical sex categories. Instead of grouping patients solely by male or female status, I classified tumors into XY (tumors with retained Y chromosome expression), XØ (tumors exhibiting extreme Y chromosome gene downregulation), and XX (female patients). This approach provides a refined perspective on tumor heterogeneity, offering potential applications for prognosis and therapeutic decision-making.

Beyond HNSCC, LoY has been implicated in immune evasion across multiple cancers. Tumors with LoY often exhibit reduced immune infiltration, impaired antigen presentation, and a more immunosuppressive microenvironment[154,155]. In bladder cancer, LoY promotes CD8+ T-cell exhaustion (elevated PD-1, TIM-3) and reduces IFN-γ signaling and dendritic cell activation, driven by the loss of KDM5D and UTY[147]. Similarly,

in lung adenocarcinoma (LUAD), LoY suppresses cancer/testis antigens (CTAs) and reduces HLA class I/II expression, leading to diminished CD4+/CD8+ T-cell infiltration[156]. In line with this, my analysis of bulk TCGA RNA-seq data indicated a reduced presence of several immune cell types across sex chromosome dosage groups. The CD8+ T cell compartment showed the highest proportions in XX patients while XØ males exhibited the lowest levels, suggesting that the loss of Y chromosome expression may be associated with reduced cytotoxic T cell infiltration. Similarly, key myeloid cell types, including monocytes and macrophages, were most abundant in XX individuals and lowest in X0 tumors, reinforcing a potential link between Y chromosome dosage and altered immune cell composition. These findings further support the idea that LoY/EDY may contribute to immune evasion mechanisms in HNSCC, potentially shaping the TME in a way that impacts immune surveillance and anti-tumor immunity.

5.3 Single-Cell Integration Enhances Resolution of the HNSCC TME

Given the complexity of bulk transcriptomic data and its inherent limitations in resolving cellular heterogeneity, I sought to extend this analysis to the single-cell level. A major limitation of bulk RNA-seq is that it averages gene expression across all cells in a tumor sample, potentially obscuring cell type-specific differences. To overcome this, I constructed a comprehensive HNSCC single-cell atlas, integrating multiple datasets into a unified framework that enables a higher-resolution examination of the TME.

Compared to previous single-cell studies in HNSCC, which often relied on smaller sample sizes, manual cell type annotation, or lacked robust integration across datasets, my atlas represents a significant advancement. Prior approaches have typically been constrained by dataset heterogeneity, where technical batch effects and inconsistencies in annotation methods have limited cross-cohort comparability. Compared to previous efforts in integrating single-cell data from HNSCC samples, such as the work by Dai et al., my approach stands out for incorporating an integration strategy and refined annotation methodologies that build upon existing approaches. While Dai et al. integrated five datasets using the standard Seurat pipeline with manual annotation based on cluster markers, I expanded upon this by incorporating STACAS for semi-

supervised integration, leveraging existing labels to guide the harmonization process[157]. In addition, I implemented automated annotation pipelines, including Ikarus and scGate, which improved reproducibility, minimized subjective biases, and ensured consistency across datasets[106,107]. Unlike studies that relied solely on manual clustering and marker-based annotation, these refinements not only enhance the accuracy of cell type identification but also provide a scalable framework that can be applied to future multi-cohort analyses.

Several previously published HNSCC single-cell atlases have provided foundational insights into tumor heterogeneity and immune dynamics, but their scope has often been limited by dataset size, focus, or annotation methodology[85]. The pioneering atlas by Puram et al. profiled ~6,000 cells from 18 patients, primarily characterizing malignant and non-malignant cell types and exploring tumor progression at the single-cell level[158]. Subsequent studies expanded upon these efforts by incorporating larger datasets or focusing on specific disease subtypes, such as Kürten et al., who profiled tumor-infiltrating leukocytes and non-immune cells to investigate inflammation and HPV-associated differences, and Choi et al., who mapped the progression of HNSCC from normal tissue to metastases, highlighting shifts in T cell infiltration and regulatory T cell involvement[32,136]. The largest HNSCC single-cell dataset to date, presented by Bill et al., included 52 tumor samples from 51 patients and provided new insights into tumor-associated macrophage polarization and its influence on the TME[94]. However, while these studies have deepened our understanding of HNSCC at the single-cell level, most have remained limited to single-cohort analyses, potentially restricting their ability to capture broader interpatient variability.

By integrating multiple datasets into a single standardized framework, I build upon these previous efforts while addressing key limitations. My atlas harmonizes data from 79 patients and 274,911 cells across independent cohorts, allowing for a more comprehensive and statistically robust characterization of the HNSCC TME. The integration strategy I implemented not only mitigates dataset-specific biases but also improves the resolution of cellular heterogeneity and enhances cross-cohort comparability. Additionally, the use of STACAS, Ikarus, and scGate ensures greater

reproducibility and scalability, refining tumor and stromal cell classification with increased precision. In doing so, I provide a more detailed and biologically meaningful representation of the TME, particularly in the context of sex chromosome dosage, and establish a high-resolution resource for investigating tumor-immune-stroma interactions in HNSCC. These improvements underscore the importance of large-scale data harmonization in single-cell oncology research and demonstrate the power of integrated approaches in revealing novel insights into tumor biology.

5.4 HPV and Clinical Sex Shape the Immune and Stromal Landscape in HNSCC

Before stratifying tumors by sex chromosome dosage, I first examined how HPV status and clinical sex influence the TME, revealing differences in immune and stromal composition. The analysis showed that HPV-positive tumors from female patients exhibited a higher stromal content than those from males, whereas the opposite trend was observed in HPV-negative tumors, where males had a more prominent stromal compartment. These findings suggest that sex-dependent factors shape the tumor stroma in an HPV-dependent manner. Prior studies across multiple cancer types have reported that females generally exhibit a higher stromal score compared to males, indicating a potential sex-linked regulation of fibroblast activity and extracellular matrix remodeling[159]. However, for HNSCC, the evidence regarding sex-specific differences in stromal composition remains inconclusive, with some studies reporting variability in fibroblast abundance but without a consistent pattern across cohorts[160]. Given that HPV-positive tumors are typically characterized by a more inflammatory microenvironment, the increased stromal content observed in female HPV-positive tumors may reflect a heightened fibroblast response to immune activation[161]. Conversely, the higher stromal content in HPV-negative males could correspond to a more rigid, fibrotic tumor architecture, contributing to immune exclusion and treatment resistance[162,163]. While these findings support a role for sex and viral status in shaping the TME, further investigation is needed to determine whether these differences reflect inherent stromal biology or are secondary to other tumor-intrinsic and extrinsic factors.

Further analysis of the immune compartment revealed that HPV-positive tumors from females harbored a higher proportion of cytotoxic CD8+ T cells and displayed reduced T cell exhaustion compared to their male counterparts, consistent with previous reports indicating stronger immune activation in females[164]. In contrast, HPV-negative tumors from males exhibited a more immunosuppressive immune profile, with increased regulatory T cells and lower cytotoxic activity. These differences in immune composition suggest that both HPV-negative and HPV-positive males, with their higher levels of exhausted CD8+ T cells, may represent a subgroup that is particularly responsive to immune checkpoint blockade. The increased abundance of exhausted T cells in these tumors indicates a pre-existing, albeit dysfunctional, immune response, which could be effectively reinvigorated by anti-PD-1/PD-L1 therapies[165,166].

5.5 Single-Cell Resolution of Sex Chromosome Dosage in HNSCC

To extend the analysis of sex chromosome dosage beyond bulk data, I utilized single-cell RNA sequencing to refine the assessment of Y chromosome downregulation at the cellular level and enable patient stratification to ensure comparability across datasets. To do so, I classified individual tumor cells as XØ if they exhibited no detectable expression of any Y chromosome genes. This approach allowed for the identification of XØ cells across the dataset, revealing that they were present in all male patients and in all cell types, confirming that Y chromosome gene silencing is a widespread phenomenon in HNSCC tumors. Importantly, while XØ cells were detected across various cell populations, tumor cells exhibited the highest proportion, suggesting that Y chromosome downregulation may be particularly enriched in malignant cells rather than being uniformly distributed across the TME.

To extend this analysis to the patient level, I pseudobulked the tumor cell transcriptomes from each patient and calculated a GSVA score for Y chromosome gene expression. Based on these scores, I classified male patients into XØ (low expressors) and XY (high expressors), ensuring that comparisons were made within HPV-specific cohorts to avoid confounding effects. This patient-level stratification revealed a clear relationship between Y chromosome expression and the proportion of XØ tumor cells, with patients

exhibiting lower GSVA scores showing a higher fraction of XØ cells. Notably, HPV-negative tumors exhibited significantly lower Y chromosome expression scores, further reinforcing the link between HPV status and Y chromosome gene regulation.

These findings suggest that Y chromosome gene silencing in tumor cells is not a random event but may be under selective pressure, particularly in HPV-negative tumors where immune infiltration is lower. Prior studies in bladder and lung adenocarcinoma have demonstrated that loss of the Y chromosome in tumor cells is associated with immune evasion, including increased CD8+ T-cell exhaustion and reduced HLA class I/II antigen presentation[69,156]. In these cancers, tumors with LoY exhibit impaired antigen presentation and cytokine signaling, contributing to an immunosuppressive microenvironment. The enrichment of XØ cells in HPV-negative tumors, which are already characterized by lower immune infiltration and reduced adaptive immune activation, raises the possibility that Y chromosome silencing may contribute to immune evasion in HNSCC as well[167–169].

By refining the classification of HNSCC tumors using single-cell transcriptomics and GSVA-based scoring, this study provides a framework for understanding how sex chromosome dosage differences shape tumor behavior and the TME, with potential implications for treatment stratification. The integration of single-cell and bulk transcriptomic data offers a more detailed approach to classifying tumors, moving beyond traditional male-female distinctions and providing a biologically informed perspective on sex chromosome dosage variability in cancer.

5.6 Single-Cell Analysis Reveals TME Differences Across HPV-negative HNSCC Sex Chromosome Dosage Groups

Given that the XØ group appeared more relevant in HPV-negative tumors and that patient groups were better balanced in this subset, I focused the analysis of TME composition on HPV-negative patients. This revealed distinct differences across sex chromosome dosage groups in tumor, immune, and stromal compartments. XØ tumors exhibited a markedly higher tumor cell composition (47.5%) compared to XX (10.4%) and XY (10.2%)

tumors, suggesting a more tumor-dominant microenvironment with reduced stromal and immune cell infiltration. The increased proportion of tumor cells in XØ tumors may contribute to a more aggressive phenotype, potentially enhancing tumor growth, immune evasion, and resistance to therapy. The immune compartment also exhibited notable differences, with XX tumors showing the highest overall immune cell infiltration, particularly enriched in T cells. In contrast, XØ tumors displayed the lowest levels of immune infiltration, reinforcing the hypothesis that extreme Y chromosome downregulation may contribute to immune exclusion. This aligns with previous findings in bladder and lung adenocarcinoma, where tumors with LoY exhibited impaired antigen presentation and cytokine signaling, leading to increased CD8+ T-cell exhaustion and a more suppressive microenvironment[66,69]. XY tumors exhibited intermediate immune infiltration but retained a distinct profile, suggesting that chromosomal dosage influences both immune composition and the extent of immune engagement in the TME. The stromal compartment further reflected these differences, with fibroblasts being most abundant in XY tumors (22.9%), compared to 8.8% in XX and only 5.1% in XØ tumors. The depletion of fibroblasts in XØ tumors suggests a less structured stromal environment, which could impact extracellular matrix remodeling and immune cell recruitment. Together, these findings indicate that sex chromosome dosage shapes the TME, with XX tumors exhibiting a more immune-active landscape, XY tumors retaining stromal complexity, and XØ tumors presenting a tumor-dominant and immune-depleted phenotype.

The lower immune infiltration in HPV-negative XØ and XY tumors, combined with the increased stromal remodeling in XY, suggests that these tumors may be more resistant to immunotherapy, necessitating alternative therapeutic strategies. Some studies show the relation between lower fibroblast infiltrate with a better prognosis[170,171]. One potential avenue for intervention could involve targeting fibroblast-mediated immune suppression, particularly in XY tumors where matrix remodeling is prominent. Previous studies have demonstrated that fibroblast-targeting therapies, such as TGF-β inhibitors, can help disrupt fibroblast-driven immune exclusion and enhance response to checkpoint blockade in immune-cold tumors, raising the possibility that such approaches could be beneficial in XY HNSCC[172,173].

5.7 Fibroblast Heterogeneity and Tumor-Stroma Interactions Across Sex Chromosome Dosage Groups

Given the key role of fibroblasts in mediating tumor-stroma interactions, I further examined the distribution of cancer-associated fibroblasts (CAFs) subtypes across sex chromosome dosage groups. CAFs in HNSCC are highly heterogeneous, with multiple subtypes playing distinct roles in tumor progression and immune modulation[171]. scRNA-seq studies have helped classify CAFs based on their expression of specific markers such as α-SMA, fibroblast activation protein (FAP), and podoplanin (PDPN). Among these, matrix CAFs (mCAFs), inflammatory CAFs (iCAFs), and antigen-presenting CAFs (apCAFs) have been identified as key subtypes, with unique functional roles in shaping the TME[174].

In this study, I used the CAF subtype classification proposed by Cords et al. (2023), which provides a comprehensive classification of CAF subtypes across multiple cancers, including HNSCC[139]. The most notable changes were observed in mCAFs and iCAFs, which displayed distinct enrichment patterns depending on the chromosomal dosage group. XØ tumors showed an increased presence of mCAFs, suggesting a more rigid extracellular matrix that could contribute to immune exclusion by creating physical barriers to immune cell infiltration. Meanwhile, iCAFs were more abundant in XY tumors, but rather than driving a pro-inflammatory response, they likely modulate immune signaling within the TME. iCAFs are characterized by high expression of cytokines such as IL-6, CXCL12, and CXCL14, which play key roles in immune regulation[171,175]. Through the activation of pathways such as IL-6-JAK-STAT3 and KRAS signaling, iCAFs can shape the local immune environment, potentially facilitating immune suppression rather than direct immune activation[176]. Notably, certain subsets of iCAFs, particularly CKS2+ iCAFs, have been associated with glycolytic metabolic activity, supporting tumor survival by providing metabolic intermediates that help sustain tumor cell proliferation under hypoxic conditions[174]. Additionally, iCAFs exhibit strong interactions with tumor cells through ligand-receptor signaling, which may contribute to immune evasion. Their negative correlation with CD8+ T cells and NK cells, along with their positive association with exhausted CD8+ T cells, further suggests a role in promoting T-cell dysfunction[27,177].

These findings indicate that while XY tumors retain a more fibroblast-enriched stroma, the functional properties of iCAFs may ultimately contribute to immune modulation rather than an outright pro-inflammatory response.

Patient-specific analyses revealed distinct patterns of tumor-iCAF interactions across sex chromosome dosage groups. While iCAFs were present in most XY tumors, only a subset showed active interactions with tumor cells, indicating variability in fibroblast-driven signaling within this group. In contrast, XØ tumors largely lacked direct tumor-iCAF interactions, reinforcing a more immunosuppressive TME. XX tumors displayed the greatest variability, with some patients showing strong tumor-iCAF communication while others had minimal engagement, suggesting that iCAF activity in XX tumors may be more context-dependent. Given their involvement in resistance to ICIs, targeting specific iCAF subpopulations, such as CKS2+ iCAFs, may represent a potential therapeutic avenue, particularly in XY tumors where their enrichment is more pronounced.

My analysis of iCAF function across chromosomal dosage groups revealed distinct roles in shaping the TME. In XY tumors, fibroblasts exhibited the highest enrichment in pathways associated with EMT, TGF-β signaling, and angiogenesis, suggesting a more tumor-supportive role through extracellular matrix remodeling, stromal crosstalk, and immune evasion[178–180]. In contrast, in XØ tumors fibroblasts displayed increased activation of inflammatory pathways and Wnt/β-catenin signaling, indicating a microenvironment shaped by chronic inflammation but potentially lacking effective anti-tumor immunity[181,182]. These functional differences were further reflected in transcriptional profiles, as fibroblasts in XY tumors showed upregulation of genes involved in extracellular matrix regulation (VIM, FBLN1, and SFRP family members), while fibroblasts in XØ tumors exhibited higher expression of inflammatory mediators (JUNB, FOS, NFKBIA, and CXCL8), reinforcing their distinct activation states[183–186]. Interestingly, fibroblasts in XX tumors displayed a less specialized transcriptional profile, with limited pathway enrichment, suggesting that they may remain more plastic and adaptable in response to TME cues rather than adopting a fixed functional role. Given the established link between iCAFs and resistance to immune checkpoint blockade, I found that their greater abundance and signaling activity in XY tumors suggest they may serve as a barrier

to effective immunotherapy in this group[187–189]. Conversely, the inflammatory phenotype observed in fibroblasts in XØ tumors may contribute to a dysfunctional immune landscape, where inflammation fails to translate into effective anti-tumor immunity. These findings reinforce the need for tailored therapeutic strategies that specifically target fibroblast-mediated immune modulation, potentially suppressing fibroblast-driven immune evasion in XY tumors while reprogramming dysfunctional inflammatory signaling in XØ tumors.

A deeper analysis of cell-cell communication provided further insight into how tumor cells and iCAFs interact across sex chromosome dosage groups. By leveraging CellChat-based pathway inference, I identified key ligand-receptor interactions that differentiate the TME composition of XX, XY, and XØ tumors. The communication landscape varied notably across groups, with XX tumors exhibiting the highest communication scores in tumor cells, both in outgoing and incoming signaling. This suggests that tumor cells in XX tumors may engage in more active signaling across multiple pathways. Interestingly, while XX tumors had the most active tumor cell signaling, XY tumors exhibited the highest scores for incoming and outgoing signaling specifically in iCAFs. This aligns with the observation that iCAFs are more abundant in XY tumors and suggests that these fibroblasts play a more active role in shaping the TME in this group. XY tumors exhibited stronger activation of KIT, SEMA4, and VTN, pathways that were also shared with XX tumors but showed greater specificity in XY. The activation of KIT, which is involved in cell proliferation and survival, suggests that tumor cells in XY tumors may engage in signaling that supports fibroblast expansion and maintenance[190,191]. SEMA4, a known modulator of immune and stromal interactions, may also play a role in dictating the fibroblast-mediated immune landscape in XY tumors[192,193].

To further examine fibroblast-mediated interactions, I analyzed pathways shared between tumor cell signaling and iCAF responses. IL1 and IL4 emerged as central mediators across all groups, with IL1 being particularly strong in XØ tumors, suggesting a role in chronic inflammation and immune suppression, while IL4 was more prominent in XY tumors, shaping fibroblast-immune interactions[194,195]. BMP and CD40 signaling in both XY and XØ tumors indicate active fibroblast involvement in stromal remodeling and

immune modulation[196]. Overall, tumor-fibroblast communication varied by sex chromosome dosage, with XØ tumors primarily engaging in VEGF, IL1, and CD40 signaling, potentially reinforcing immune exclusion, while XY tumors displayed enrichment for KIT, SEMA4, and VTN, suggesting a role in stromal remodeling. These differences underscore fibroblasts' distinct contributions to the TME across chromosomal dosage groups, highlighting potential therapeutic targets to disrupt tumor-fibroblast signaling and enhance immune responses in immune-cold tumors.

The expression patterns of COX2 (PTGS2) and androgen receptor (AR) in fibroblasts provided additional mechanistic insights into how stromal interactions may be modulated by sex chromosome dosage. COX2 expression was highest in iCAFs and mCAFs within XY tumors, suggesting that inflammatory signaling is more pronounced in these tumors. Prior studies have shown that COX2 overexpression in the TME is associated with increased tumor-associated inflammation and immune suppression, with COX2 inhibitors being explored as a potential strategy to improve responses to immunotherapy[144,197,198]. Similarly, AR expression was enriched in fibroblasts in XY tumors, particularly in iCAFs and mCAFs, raising the possibility that androgen signaling may contribute to stromal remodeling and immune modulation in these tumors. Given that AR inhibitors have been explored in other cancers, including prostate and lung, their potential role in XY HNSCC warrants further investigation[199–201].

These findings underscore the critical role of fibroblasts in shaping the TME across sex chromosome dosage groups, revealing distinct stromal compositions and fibroblast-mediated immune modulation. XY tumors exhibited the highest fibroblast abundance, particularly enriched in iCAFs, which contribute to immune suppression through cytokine signaling and extracellular matrix remodeling. In contrast, XØ tumors displayed a more rigid extracellular matrix with increased mCAF presence, yet lacked significant iCAF-tumor interactions, reinforcing a more immune-excluded phenotype. Meanwhile, XX tumors demonstrated a more variable fibroblast landscape, suggesting a context-dependent stromal role. The differential enrichment of key signaling pathways and gene expression further highlights how fibroblast activity varies by chromosomal dosage, potentially influencing immune responses and tumor progression. Given the established

link between fibroblast-mediated immune suppression and resistance to immunotherapy, these findings open new avenues for therapeutic intervention. Fibroblast-targeting strategies, such as TGF-β inhibitors, could help overcome stromal barriers in XY tumors, while approaches aimed at reprogramming inflammatory fibroblasts in XØ tumors may restore immune competence. These insights provide a framework for integrating fibroblast-targeted strategies into personalized treatment approaches for HNSCC, particularly in the context of sex chromosome dosage-driven tumor heterogeneity.

6. Conclusion and Outlook

This study provides a comprehensive, multi-layered analysis of sex chromosome dosage in HNSCC, revealing how Y chromosome loss or downregulation shapes tumor biology and modulates the tumor microenvironment. Through integration of bulk and single-cell transcriptomic data, I identified that EDY and LoY are frequent events in HNSCC, particularly in HPV-negative tumors. By translating these findings to the single-cell level, I refined patient classification into biologically meaningful XY, XØ, and XX categories, uncovering distinct immune, stromal, and transcriptional profiles that transcend traditional male–female stratifications.

Sex chromosome dosage was shown to influence not only tumor cell-intrinsic features but also the broader architecture of the TME. XØ tumors displayed a tumor-dominant and immune-depleted phenotype with elevated mCAF presence, while XY tumors were enriched in iCAFs and showed a more fibroblast-driven, immunomodulatory environment. XX tumors exhibited greater immune infiltration and a more variable fibroblast profile, potentially reflecting a more adaptable stromal context. These patterns suggest that Y chromosome downregulation contributes to immune exclusion and fibroblast-mediated immune suppression, both of which may impact response to therapy, particularly immunotherapy.

While these findings provide mechanistic insight and establish a novel classification framework, translational validation remains essential. Future efforts could involve spatial profiling approaches, such as multiplexed imagingor spatial transcriptomics, to assess the expression and localization of key stromal, immune, and tumor markers in HNSCC tissue sections. Such experiments would enable validation of transcriptionally inferred patterns, like differential fibroblast abundance, immune exclusion, and cytokine signaling, within their native spatial context. Moreover, spatial resolution would offer the opportunity to explore immune–fibroblast interactions and assess whether sex chromosome dosage shapes gradients of immune suppression within the TME.

Ultimately, this work lays the foundation for incorporating sex chromosome dosage into future HNSCC stratification efforts, with implications for biomarker development and therapeutic targeting. The prospect of tailoring fibroblast- or immune-modulatory therapies, such as TGF-β or COX2 inhibition, in a sex chromosome dosage aware way represents an exciting avenue for precision oncology. As spatial transcriptomic and proteomic technologies continue to evolve, they will be instrumental in validating and expanding these insights, helping bridge the gap between single-cell discovery and clinical application.

Publications

First author:

- Conde-Lopez, C., Marripati, D., Elkabets, M., Hess, J. & Kurth, I. Unravelling the complexity of HNSCC using single-cell transcriptomics. Cancers 16, 3265 (2024)
- Conde Lopez, C., Marripati, D., Besso, M. J., Roscher, M., Han, R., Hadiwikarta, W. W., Elkabets, M., Hess, J. & Kurth, I. A unified single-cell atlas of HNSCC: uncovering HPV and sex variability in the tumor microenvironment. Submitted to *Communications Medicine* (2025).

Contributions:

- Prasad, M. et al. MEK1/2 inhibition transiently alters the tumor immune microenvironment to enhance immunotherapy efficacy against head and neck cancer. J. Immunother. Cancer 10, e003917 (2022).
- Novoplansky, O. et al. Activation of the EGFR/PI3K/AKT pathway limits the efficacy of trametinib treatment in head and neck cancer. Mol. Oncol. 17, 2618–2636 (2023).
- Barbosa, S. et al. The role of SOX2 and SOX9 in radioresistance and tumor recurrence. Cancers (Basel) 16, (2024).
- Besso, M. J. et al. Transcriptomic and epigenetic landscape of nimorazole-enhanced radiochemotherapy in head and neck cancer. Radiother. Oncol. 199, 110348 (2024).

List of Abbreviations

apCAF – Antigen-Presenting Cancer-Associated Fibroblast
AR – Androgen Receptor
ATP – Adenosine Triphosphate
BayesPrism – Bayesian probabilistic deconvolution tool for bulk RNA-seq using single-cell references
BMP – Bone Morphogenetic Protein
CAF – Cancer-Associated Fibroblast
CCLE – Cancer Cell Line Encyclopedia
CD40 – Cluster of Differentiation 40
$CD4^+$ – Cluster of Differentiation 4 positive T Cells
CD8Tcm – $CD8^+$ Central Memory T Cells
CD8Tem – $CD8^+$ Effector Memory T Cells
$CD8^+$ – Cluster of Differentiation 8 positive T Cells
CKS2 – CDC28 Protein Kinase Regulatory Subunit 2
CNI – Copy Number Index
CNTN – Contactin
CNV – Copy Number Variant
COX2 (PTGS2) – Cyclooxygenase-2 (Prostaglandin-Endoperoxide Synthase 2)
CPTAC – Clinical Proteomic Tumor Analysis Consortium
CTAs – Cancer/Testis Antigens
CTLA-4 – Cytotoxic T-Lymphocyte-Associated Protein 4
CXCL14 – C-X-C Motif Chemokine Ligand 14
CXCL8 – C-X-C Motif Chemokine Ligand 8
DCs – Dendritic Cells
DE – Differential Expression
DEA – Differential Expression Analysis
DNA – Deoxyribonucleic Acid
EDY – Extreme Downregulation of Y-linked Genes
EIF1AY – Eukaryotic Translation Initiation Factor 1A, Y-linked
EMT – Epithelial-Mesenchymal Transition
FAP – Fibroblast Activation Protein
FBLN1 – Fibulin 1
FOS – FBJ Murine Osteosarcoma Viral Oncogene Homolog
GEO – Gene Expression Omnibus
GO – Gene Ontology
GSEA – Gene Set Enrichment Analysis
GSVA – Gene Set Variation Analysis
HIPO-HNC – Heidelberg Center for Personalized Oncology - Head and Neck Cancer

HLA – Human Leukocyte Antigen
HNSCC – Head and Neck Squamous Cell Carcinoma
HPV – Human Papillomavirus
iCAF – Inflammatory Cancer-Associated Fibroblast
ICI – Immune Checkpoint Inhibitor
IFN-II – Interferon gamma (Type II Interferon)
IL1 – Interleukin 1
IL4 – Interleukin 4
IL6 – Interleukin 6
InferCNV – Infer Copy Number Variation (R package for single-cell CNV inference)
JAK-STAT3 – Janus Kinase - Signal Transducer and Activator of Transcription 3 signaling pathway
JUNB – JunB Proto-Oncogene
KIT – KIT Proto-Oncogene, Receptor Tyrosine Kinase
KRAS – Kirsten Rat Sarcoma Viral Oncogene Homolog
Limma – Linear Models for Microarray Data
logCPM – Log-transformed Counts Per Million
LoY – Loss of the Y chromosome
LPS – Lipopolysaccharide
LUAD – Lung Adenocarcinoma
mCAF – Matrix Cancer-Associated Fibroblast
miRNA – MicroRNA
MPZ – Myelin Protein Zero
ncWNT – Non-canonical Wnt signaling pathway
NFKBIA – Nuclear Factor of Kappa Light Polypeptide Gene Enhancer in B-cells Inhibitor, Alpha
NK cells – Natural Killer Cells
NKT – Natural Killer T Cells
NPR2 – Natriuretic Peptide Receptor 2
NT – Neurotrophin
OS – Overall Survival
PCA – Principal Component Analysis
PD-1 – Programmed Cell Death Protein 1
PD-L1 – Programmed Death-Ligand 1
PDPN – Podoplanin
PLAU – Plasminogen Activator, Urokinase
PPI – Protein–Protein Interaction
PTGS2 – Prostaglandin-Endoperoxide Synthase 2
QC – Quality Control
R – Statistical programming language
RB – Retinoblastoma

RELN – Reelin
RPS4Y1 – Ribosomal Protein S4, Y-linked 1
scGate – Supervised Gating for Single-Cell Classification
scRNA-seq – Single-Cell RNA Sequencing
SEMA3 – Semaphorin 3
SEMA4 – Semaphorin 4
SFRP1/2 – Secreted Frizzled-Related Protein 1 / 2
SRY – Sex Determining Region Y
STACAS – Subtype Anchor Correction for Alignment in Seurat
t-SNE – t-Distributed Stochastic Neighbor Embedding
tCAF – Tumor-like Cancer-Associated Fibroblast
TCGA – The Cancer Genome Atlas
TCGAbiolinks – R/Bioconductor package to access TCGA data
TGFβ – Transforming Growth Factor Beta
TILs – Tumor-Infiltrating Lymphocytes
TME – Tumor Microenvironment
TMSB4Y – Thymosin Beta 4, Y-linked
TNF – Tumor Necrosis Factor
Tregs – Regulatory T Cells
UMAP – Uniform Manifold Approximation and Projection
UMI – Unique Molecular Identifier
USP9Y – Ubiquitin Specific Peptidase 9, Y-linked
UTY – Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-linked
VEGF – Vascular Endothelial Growth Factor
VIM – Vimentin
VTN – Vitronectin
XX – Female cells or patients with no Y chromosome expression
XY – Male cells or patients with retained Y chromosome gene expression
XØ – Male cells or patients with absent Y chromosome gene expression
ZFY – Zinc Finger Protein, Y-linked
α-SMA – Alpha Smooth Muscle Actin

Acknowledgments

First, I would like to thank my supervisors, Ina Kurth and Jochen Hess, for giving me the opportunity to work on this project and for their trust and guidance throughout this entire journey. I'm especially grateful for their support—not only in shaping this thesis but also in helping me grow as a scientist. Thank you for always keeping the patients in mind and for guiding me through the process with a steady hand. To Ina, thank you for being there and for truly listening throughout these years filled with science, meetings, and conferences. Even during the more stressful moments, we always managed to keep a smile. To Jochen, thank you for your invaluable advice, which helped shape this thesis into what it is today. Thanks also to Moshe Elkabets, your scientific input and support throughout our collaboration have been very meaningful, and I'm grateful for the chance to work with you. I would also like to thank the members of my thesis committee, Sergio Acebrón, Frank Zapata, and Frank Lyko, for their time, thoughtful feedback, and support.

To my group: Rose Euer-Lange, Wahyu Wijaya Hadiwikarta, Safayat Mahmud Khan, Julian Schlecker, Mahnaz Bonrouhi, Manuela Dittrich, and Olga Giraldo Pasmin, thank you for the conversations, the laughs, and for making daily lab life feel lighter and friendlier. Special thanks to Rose, Manuela, and Mahnaz for taking over the wet lab experiments; to Julian, for always being ready to help when the numbers weren't adding up; to Safayat and Wahyu for our conversations about science and life (complexity is totally underrated!); and to Olga, for expanding and brightening the Hispanic front with me.

A very special thank you to Mareike, who has been a constant support throughout these years. Whether it was practical help at work, navigating the never-ending German bureaucracy, making phone calls I couldn't manage in German, or simply being there to hang out and talk about life, I truly appreciate your kindness, patience, and presence.

To Maria, thank you for being one of the greatest supports I could have asked for. Your friendship has been a steady presence since the very first day I arrived in Heidelberg. I'm so grateful to have shared this chapter with you, all the trips, dinner dates, and random

plans have made my life here so much brighter. And not only that, but also the hours spent sitting together, going through our science, because yes, we do more than just chusmear! Even if we don't always believe it ourselves.

To all my friends in Heidelberg who helped me feel at home, Marta, Sandra, Maria, Paula, Sonja, Chase, Johannes, David, Nikita, Helena, and everyone else, thank you for your care, your company, and the countless memories that kept me grounded. It has been a rollercoaster, filled with some of the best times, and, at times, moments where we just needed a lloradita y a seguir existiendo. But through it all, you made Heidelberg feel like home, a place I always wanted to return to. I truly believe that this is the most valuable gift anyone can offer.

And finally, to my family:
Gracias por estar siempre ahí. A mi padre, mi madre y mi hermano, de quienes aprendí que con el estómago lleno todo se lleva mejor. Y sí, esta tesis también se ha hecho a base de videollamadas hasta que mamá decia "bueno...", recetas compartidas y alguna caja capaz de destruir las arterias. Gracias por aseguraros de que nunca me faltara comida ni ánimo. A Paula, Mónica y Miguel, por venir a visitarme, por traerme siempre un trocito de casa (a veces literalmente), y por hacer que Heidelberg pareciera un poco menos lejos. A todas las personas que han estado ahí durante tantos años, en especial, Magda, Carmen y Gabriela, gracias por seguir siendo hogar, aunque cada una esté en un sitio distinto. Vuestra amistad ha sido mi red de seguridad en más momentos de los que puedo contar. A todo ellos, gracias por ese entorno perfectamente desordenado donde siempre hay ruido, comida, opiniones y mucho cariño. Puede que no sigamos el método científico, pero claramente funciona: de ahí salió una tesis.

I would like to acknowledge the use of OpenAI's ChatGPT as a writing assistant throughout the preparation of this thesis. This tool was employed exclusively to enhance the clarity, structure, and readability of the text. All scientific content, data analysis, and interpretation were conducted independently. Figures 1, 2, and 3 were created using BioRender (https://BioRender.com), which was used to generate schematic illustrations for improved visual communication of concepts.

Bibliography

1. Johnson, D. E. *et al*. Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Primers* **6**, 92 (2020).
2. Barsouk, A., Aluru, J. S., Rawla, P., Saginala, K. & Barsouk, A. Epidemiology, risk factors, and prevention of head and neck squamous cell carcinoma. *Med. Sci. (Basel)* **11**, (2023).
3. Bray, F. *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin*. **68**, 394–424 (2018).
4. Jawa, Y. *et al*. Current Insights and Advancements in Head and Neck Cancer: Emerging Biomarkers and Therapeutics with Cues from Single Cell and 3D Model Omics Profiling. *Front. Oncol*. **11**, 676948 (2021).
5. Nagel, R. *et al*. Treatment response of HPV-positive and HPV-negative head and neck squamous cell carcinoma cell lines. *Oral Oncol*. **49**, 560–566 (2013).
6. Cillo, A. R. *et al*. Immune landscape of viral- and carcinogen-driven head and neck cancer. *Immunity* **52**, 183-199.e9 (2020).
7. Liu, S., Wang, R. & Fang, J. Exploring the frontiers: tumor immune microenvironment and immunotherapy in head and neck squamous cell carcinoma. *Discov. Oncol*. **15**, 22 (2024).
8. Boffetta, P., Hecht, S., Gray, N., Gupta, P. & Straif, K. Smokeless tobacco and cancer. *Lancet Oncol*. **9**, 667–675 (2008).
9. Hashibe, M. *et al*. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *J. Natl. Cancer Inst*. **99**, 777–789 (2007).
10. Benson, E., Li, R., Eisele, D. & Fakhry, C. The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas. *Oral Oncol*. **50**, 565–574 (2014).
11. de Bakker, T. *et al*. Restoring p53 function in head and neck squamous cell carcinoma to improve treatments. *Front. Oncol*. **11**, 799993 (2021).

12. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
13. Seiwert, T. Y. *et al*. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin. Cancer Res.* **21**, 632–641 (2015).
14. Tabatabaeian, H., Bai, Y., Huang, R., Chaurasia, A. & Darido, C. Navigating therapeutic strategies: HPV classification in head and neck cancer. *Br. J. Cancer* **131**, 220–230 (2024).
15. Joo, Y.-H. *et al*. Characteristics and prognostic implications of high-risk HPV-associated hypopharyngeal cancers. *PLoS One* **8**, e78718 (2013).
16. Sabatini, M. E. & Chiocca, S. Human papillomavirus as a driver of head and neck cancers. *Br. J. Cancer* **122**, 306–314 (2020).
17. Human papillomavirus (HPV): diagnosis and treatment. in *Animal biotechnology* (eds. Verma, A. S. & Singh, A.).
18. Wang, J. *et al*. HPV-positive status associated with inflamed immune microenvironment and improved response to anti-PD-1 therapy in head and neck squamous cell carcinoma. *Sci. Rep.* **9**, 13404 (2019).
19. Wang, H.-F. *et al*. The double-edged sword-how human papillomaviruses interact with immunity in head and neck cancer. *Front. Immunol.* **10**, 653 (2019).
20. Mandal, R. *et al*. The head and neck cancer immune landscape and its immunotherapeutic implications. *JCI Insight* **1**, e89829 (2016).
21. Park, R. & Chung, C. H. Advanced human Papillomavirus-negative head and neck squamous cell carcinoma: Unmet need and emerging therapies. *Mol. Cancer Ther.* **23**, 1717–1730 (2024).
22. Bauwens, L. *et al*. Prevalence and distribution of cervical lymph node metastases in HPV-positive and HPV-negative oropharyngeal squamous cell carcinoma. *Radiother. Oncol.* **157**, 122–129 (2021).
23. Sanguineti, G. Re: Prevalence and distribution of cervical lymph node metastases in HPV-positive and HPV-negative oropharyngeal squamous cell carcinoma. Radiother Oncol, 2021. 157: p. 122-129. *Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology* vol. 161 251–252 (2021).

24. Özcan-Wahlbrink, M., Schifflers, C. & Riemer, A. B. Enhanced radiation sensitivity of human Papillomavirus-driven head and neck cancer: Focus on immunological aspects. *Front. Immunol.* **10**, 2831 (2019).
25. Fakhry, C. *et al.* Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J. Natl. Cancer Inst.* **100**, 261–269 (2008).
26. Thakral, A. *et al.* Smoking and alcohol by HPV status in head and neck cancer: a Mendelian randomization study. *Nat. Commun.* **15**, 7835 (2024).
27. Canning, M. *et al.* Heterogeneity of the Head and Neck Squamous Cell Carcinoma Immune Landscape and Its Impact on Immunotherapy. *Front Cell Dev Biol* **7**, 52 (2019).
28. Mitsuda, J. *et al.* A 14-marker multiplexed imaging panel for prognostic biomarkers and tumor heterogeneity in head and neck squamous cell carcinoma. *Front. Oncol.* **11**, 713561 (2021).
29. Deshmukh, S. & Saini, S. Phenotypic heterogeneity in tumor progression, and its possible role in the onset of cancer. *Front. Genet.* **11**, 604528 (2020).
30. Weber, P. *et al.* Therapy-related transcriptional subtypes in matched primary and recurrent head and neck cancer. *Clin. Cancer Res.* **28**, 1038–1052 (2022).
31. Wu, L. *et al.* Single-cell RNA sequencing and traditional RNA sequencing reveals the role of cancer-associated fibroblasts in oral squamous cell carcinoma cohort. *Front. Oncol.* **13**, 1195520 (2023).
32. Choi, J.-H. *et al.* Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. *Nat. Commun.* **14**, 1055 (2023).
33. Zhang, Q. *et al.* Predicting local persistence/recurrence after radiation therapy for head and neck cancer from PET/CT using a multi-objective, multi-classifier radiomics model. *Front. Oncol.* **12**, 955712 (2022).
34. Hutchinson, M.-K. N. D., Mierzwa, M. & D'Silva, N. J. Radiation resistance in head and neck squamous cell carcinoma: dire need for an appropriate sensitizer. *Oncogene* **39**, 3638–3649 (2020).
35. Sindhu, S. K. & Bauman, J. E. Current concepts in chemotherapy for head and neck cancer. *Oral Maxillofac. Surg. Clin. North Am.* **31**, 145–154 (2019).

36. Goel, B. *et al*. Therapeutic approaches for the treatment of head and neck squamous cell carcinoma-An update on clinical trials. *Transl. Oncol.* **21**, 101426 (2022).
37. Tsai, C. J. *et al*. Evaluation of substantial reduction in elective radiotherapy dose and field in patients with human Papillomavirus-associated oropharyngeal carcinoma treated with definitive chemoradiotherapy. *JAMA Oncol.* **8**, 364–372 (2022).
38. Mensour, E. A. *et al*. What is the future of treatment de-escalation for HPV-positive oropharyngeal cancer? A review of ongoing clinical trials. *Front. Oncol.* **12**, 1067321 (2022).
39. Clocchiatti, A., Cora, E., Zhang, Y. & Dotto, G. P. Sexual dimorphism in cancer. *Nat. Rev. Cancer* **16**, 330–339 (2016).
40. Arnold, A. P. & Disteche, C. M. Sexual inequality in the cancer cell. *Cancer research* vol. 78 5504–5505 (2018).
41. Xiao, T. *et al*. Hallmarks of sex bias in immuno-oncology: mechanisms and therapeutic implications. *Nat. Rev. Cancer* **24**, 338–355 (2024).
42. Park, J.-O. *et al*. Sex differences in the prevalence of head and neck cancers: A 10-year follow-up study of 10 million healthy people. *Cancers (Basel)* **14**, 2521 (2022).
43. Lin, C. Y. *et al*. Sex differences in using systemic inflammatory markers to prognosticate patients with head and neck squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **27**, 1176–1185 (2018).
44. Langevin, S. M., Grandis, J. R. & Taioli, E. Female hormonal and reproductive factors and head and neck squamous cell carcinoma risk. *Cancer Lett.* **310**, 216–221 (2011).
45. Lee, J. *et al*. Sex-Biased T-cell Exhaustion Drives Differential Immune Responses in Glioblastoma. *Cancer Discov.* **13**, 2090–2105 (2023).
46. Ning, K. *et al*. Sex differences in renal cell carcinoma: a single-cell analysis reveals exhausted CD8+ T-cells highly infiltrated in males. *Biol. Sex Differ.* **14**, 58 (2023).
47. Zhao, J., Wang, Q., Tan, A. F., Loh, C. J. L. & Toh, H. C. Sex differences in cancer and immunotherapy outcomes: the role of androgen receptor. *Front. Immunol.* **15**, 1416941 (2024).

48. Yan, H., Huang, J., Li, Y. & Zhao, B. Sex disparities revealed by single-cell and bulk sequencing and their impacts on the efficacy of immunotherapy in esophageal cancer. *Biol. Sex Differ.* **15**, 22 (2024).
49. Yuan, Y. *et al.* Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. *Cancer Cell* **29**, 711–722 (2016).
50. Hollows, R. *et al.* Association between loss of Y chromosome and poor prognosis in male head and neck squamous cell carcinoma. *Head Neck* **41**, 993–1006 (2019).
51. Čonkaš, J., Sabol, M. & Ozretić, P. "toxic masculinity": What is known about the role of androgen receptors in head and neck squamous cell carcinoma. *Int. J. Mol. Sci.* **24**, (2023).
52. Wang, S., Pan, W., Mi, W.-X. & Wang, S.-H. Sex-specific gene expression patterns in head and neck squamous cell carcinomas. *Heliyon* **9**, e14890 (2023).
53. Lee, Y.-G. *et al.* Comparison of treatment patterns and clinical outcomes by gender in locally advanced head and neck squamous cell carcinoma (KCSG HN13-01). *Cancers (Basel)* **15**, 471 (2023).
54. Hu, W. *et al.* Sex - a potential factor affecting immune checkpoint inhibitor therapy for cancers. *Front. Immunol.* **13**, 1024112 (2022).
55. Grassadonia, A. *et al.* Effect of gender on the outcome of patients receiving immune checkpoint inhibitors for advanced cancer: A systematic review and meta-analysis of phase III randomized clinical trials. *J. Clin. Med.* **7**, 542 (2018).
56. Ahmed, A. *et al.* Sex differences in the systemic and local immune response of pancreatic cancer patients. *Cancers (Basel)* **15**, (2023).
57. Brown, D. W. & Machiela, M. J. Why Y? Downregulation of chromosome Y genes potentially contributes to elevated cancer risk. *Journal of the National Cancer Institute* vol. 112 871–872 (2020).
58. Qi, M., Pang, J., Mitsiades, I., Lane, A. A. & Rheinbay, E. Loss of chromosome Y in primary tumors. *Cell* (2023) doi:10.1016/j.cell.2023.06.006.
59. Forsberg, L. A. *et al.* Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat. Genet.* **46**, 624–628 (2014).
60. Wright, D. J. *et al.* Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. *Nat. Genet.* **49**, 674–679 (2017).

61. Noveski, P. *et al*. Loss of Y chromosome in peripheral blood of colorectal and prostate cancer patients. *PLoS One* **11**, e0146264 (2016).
62. Gutiérrez-Hurtado, I. A. *et al*. Loss of the Y chromosome: A review of molecular mechanisms, age inference, and implications for men's health. *Int. J. Mol. Sci.* **25**, (2024).
63. Bérgamo, N. A. *et al*. Classic and molecular cytogenetic analyses reveal chromosomal gains and losses correlated with survival in head and neck cancer patients. *Clin. Cancer Res.* **11**, 621–631 (2005).
64. Forsberg, L. A. *et al*. Mosaic loss of chromosome Y in leukocytes matters. *Nature genetics* vol. 51 4–7 (2019).
65. Caceres, A., Jene, A., Esko, T., Perez-Jurado, L. A. & Gonzalez, J. R. Extreme downregulation of chromosome Y and Alzheimer's disease in men. *Neurobiol. Aging* **90**, 150.e1-150.e4 (2020).
66. Loftfield, E. *et al*. Mosaic Y loss is moderately associated with solid tumor risk. *Cancer Res.* **79**, 461–466 (2019).
67. Dumanski, J. P. *et al*. Immune cells lacking Y chromosome show dysregulation of autosomal gene expression. *Cell. Mol. Life Sci.* **78**, 4019–4033 (2021).
68. Case, L. K. *et al*. The Y chromosome as a regulatory element shaping immune cell transcriptomes and susceptibility to autoimmune disease. *Genome Res.* **23**, 1474–1485 (2013).
69. Abdel-Hafiz, H. A. *et al*. Y chromosome loss in cancer drives growth by evasion of adaptive immunity. *Nature* **619**, 624–631 (2023).
70. Müller, P. *et al*. Why loss of Y? A pan-cancer genome analysis of tumors with loss of Y chromosome. *Comput. Struct. Biotechnol. J.* **21**, 1573–1583 (2023).
71. Bruhn-Olszewska, B. *et al*. The effects of loss of Y chromosome on male health. *Nat. Rev. Genet.* (2025) doi:10.1038/s41576-024-00805-y.
72. Chakraborty, S., Hosen, M. I., Ahmed, M. & Shekhar, H. U. Onco-multi-OMICS approach: A new frontier in cancer research. *Biomed Res. Int.* **2018**, 9836256 (2018).
73. Heo, Y. J., Hwa, C., Lee, G.-H., Park, J.-M. & An, J.-Y. Integrative multi-omics approaches in cancer research: From biological networks to clinical subtypes. *Mol. Cells* **44**, 433–443 (2021).

74. Hu, Z.-Z. *et al*. Omics-based molecular target and biomarker identification. *Methods Mol. Biol.* **719**, 547–571 (2011).
75. Yang, J., Chen, Y., Luo, H. & Cai, H. The landscape of somatic copy number alterations in head and neck squamous cell carcinoma. *Front. Oncol.* **10**, 321 (2020).
76. Hoesli, R. C. *et al.* Genomic sequencing and precision medicine in head and neck cancers. *Eur. J. Surg. Oncol.* **43**, 884–892 (2017).
77. Li, Q., Tie, Y., Alu, A., Ma, X. & Shi, H. Targeted therapy for head and neck cancer: signaling pathways and clinical studies. *Signal Transduct. Target. Ther.* **8**, 31 (2023).
78. Chi, L.-H., Wu, A. T. H., Hsiao, M. & Li, Y.-C. J. A transcriptomic analysis of head and neck squamous cell carcinomas for prognostic indications. *J. Pers. Med.* **11**, 782 (2021).
79. Dankó, B. *et al.* Metabolic pathway-based subtypes associate glycan biosynthesis and treatment response in head and neck cancer. *NPJ Precis. Oncol.* **8**, 116 (2024).
80. Le, J. M., Squarize, C. H. & Castilho, R. M. Histone modifications: Targeting head and neck cancer stem cells. *World J. Stem Cells* **6**, 511–525 (2014).
81. Callahan, S. C. *et al.* High enhancer activity is an epigenetic feature of HPV negative atypical head and neck squamous cell carcinoma. *Front. Cell Dev. Biol.* **10**, 936168 (2022).
82. De Cecco, L. *et al.* Head and neck cancer subtypes with biological and clinical relevance: Meta-analysis of gene-expression data. *Oncotarget* **6**, 9627–9642 (2015).
83. Zhang, P. *et al.* Characterization of molecular subtypes in head and neck squamous cell carcinoma with distinct prognosis and treatment responsiveness. *Front. Cell Dev. Biol.* **9**, 711348 (2021).
84. Walter, V. *et al.* Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. *PLoS One* **8**, e56823 (2013).
85. Conde-Lopez, C., Marripati, D., Elkabets, M., Hess, J. & Kurth, I. Unravelling the complexity of HNSCC using single-cell transcriptomics. *Cancers (Basel)* **16**, (2024).
86. Jia, Q., Chu, H., Jin, Z., Long, H. & Zhu, B. High-throughput single-cell sequencing in cancer research. *Signal Transduct. Target. Ther.* **7**, 145 (2022).

87. Chen, S. *et al*. Single-cell analysis technologies for cancer research: from tumor-specific single cell discovery to cancer therapy. *Front. Genet.* **14**, 1276959 (2023).
88. Hawsawi, Y. M. *et al*. Recent progress and applications of single-cell sequencing technology in breast cancer. *Front. Genet.* **15**, 1417415 (2024).
89. Arora, R. *et al*. Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat. Commun.* **14**, 5029 (2023).
90. Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput. Biol.* **14**, e1006245 (2018).
91. Heumos, L. *et al*. Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* 1–23 (2023).
92. Hao, Y. *et al*. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
93. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
94. Bill, R. *et al*. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. *Science* **381**, 515–524 (2023).
95. Liu, S., Lian, M., Han, B., Fang, J. & Wang, Z. Single-cell integrated transcriptomics reveals the role of keratinocytes in head and neck squamous cell carcinoma. *J. Appl. Genet.* **65**, 727–745 (2024).
96. Park, Y. M. & Lin, D.-C. Moving closer towards a comprehensive view of tumor biology and microarchitecture using spatial transcriptomics. *Nat. Commun.* **14**, 7017 (2023).
97. Lytal, N., Ran, D. & An, L. Normalization methods on single-cell RNA-seq data: An empirical survey. *Front. Genet.* **11**, 41 (2020).
98. Sheng, J. & Li, W. V. Selecting gene features for unsupervised analysis of single-cell gene expression data. *Brief. Bioinform.* **22**, (2021).
99. Linderman, G. C. Dimensionality reduction of single-cell RNA-seq data. *Methods Mol. Biol.* **2284**, 331–342 (2021).
100. Zhou, B. & Jin, W. Visualization of single cell RNA-seq data using t-SNE in R. *Methods Mol. Biol.* **2117**, 159–167 (2020).

101. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).
102. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
103. Andreatta, M. *et al.* Semi-supervised integration of single-cell transcriptomics data. *Nat. Commun.* **15**, 872 (2024).
104. Jia, Y., Ma, P. & Yao, Q. CellMarkerPipe: cell marker identification and evaluation pipeline in single cell transcriptomes. *Sci. Rep.* **14**, 13151 (2024).
105. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
106. Andreatta, M., Berenstein, A. J. & Carmona, S. J. scGate: marker-based purification of cell types from heterogeneous single-cell RNA-seq datasets. *Bioinformatics* **38**, 2642–2644 (2022).
107. Dohmen, J. *et al.* Identifying tumor cells at the single-cell level using machine learning. *Genome Biol.* **23**, 123 (2022).
108. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
109. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
110. Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L. & Smyth, G. K. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res.* **53**, (2025).
111. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
112. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
113. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb.)* **2**, 100141 (2021).
114. Xu, S. *et al.* Using clusterProfiler to characterize multiomics data. *Nat. Protoc.* **19**, 3292–3320 (2024).

115. Xie, Z. *et al.* Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
116. Dimitrov, D. *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224 (2022).
117. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
118. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
119. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
120. Gao, R. *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **39**, 599–608 (2021).
121. Huang, R. *et al.* Robust analysis of allele-specific copy number alterations from scRNA-seq data with XClone. *bioRxiv* (2023) doi:10.1101/2023.04.03.535352.
122. Mahdipour-Shirayeh, A., Erdmann, N., Leung-Hagesteijn, C. & Tiedemann, R. E. sciCNV: high-throughput paired profiling of transcriptomes and DNA copy number variations at single-cell resolution. *Brief. Bioinform.* **23**, (2022).
123. Ivich, A. *et al.* Missing cell types in single-cell references impact deconvolution of bulk data but are detectable. *bioRxiv* (2024) doi:10.1101/2024.04.25.590992.
124. Xu, X. *et al.* Cell-type deconvolution for bulk RNA-seq data using single-cell reference: a comparative analysis and recommendation guideline. *Brief. Bioinform.* **26**, (2024).
125. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat. Cancer* **3**, 505–517 (2022).
126. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
127. Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
128. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

129. Therneau, T. M. *A Package for Survival Analysis in R*. https://CRAN.R-project.org/package=survival (2024).
130. Kassambara, A., Kosinski, M. & Biecek, P. survminer: Drawing Survival Curves using "ggplot2." *CRAN: Contributed Packages* The R Foundation https://doi.org/10.32614/cran.package.survminer (2016).
131. Edwards, N. J. *et al*. The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *J. Proteome Res*. **14**, 2707–2713 (2015).
132. Ghandi, M. *et al*. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
133. Mock, A. *et al*. EGFR and PI3K pathway activities might guide drug repurposing in HPV-negative head and neck cancers. *Front. Oncol.* **11**, 678966 (2021).
134. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
135. Puram, S. V. *et al*. Cellular states are coupled to genomic and viral heterogeneity in HPV-related oropharyngeal carcinoma. *Nat. Genet.* **55**, 640–650 (2023).
136. Kürten, C. H. L. *et al*. Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat. Commun.* **12**, 7338 (2021).
137. Wickham, H. The split-apply-combine strategy for data analysis. *J. Stat. Softw.* **40**, (2011).
138. Andreatta, M. *et al*. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat. Commun.* **12**, 2965 (2021).
139. Cords, L. *et al*. Cancer-associated fibroblast classification in single-cell and spatial proteomics data. *Nat. Commun.* **14**, 1–13 (2023).
140. Tickle T, Tirosh I, Georgescu C, Brown M, Haas B. *InferCNV of the Trinity CTAT Project*. https://github.com/broadinstitute/inferCNV (2019).
141. Kolde, R. *Pheatmap: Pretty Heatmaps*. https://github.com/raivokolde/pheatmap (2018).
142. Ruffin, A. T. *et al*. Improving head and neck cancer therapies by immunomodulation of the tumour microenvironment. *Nat. Rev. Cancer* **23**, 173–188 (2023).

143. Mito, I. *et al*. Comprehensive analysis of immune cell enrichment in the tumor microenvironment of head and neck squamous cell carcinoma. *Sci. Rep*. **11**, 16134 (2021).

144. Pelly, V. S. *et al*. Anti-inflammatory drugs remodel the tumor immune environment to enhance immune checkpoint blockade efficacy. *Cancer Discov*. **11**, 2602–2619 (2021).

145. Cáceres, A., Jene, A., Esko, T., Pérez-Jurado, L. A. & González, J. R. Extreme Downregulation of Chromosome Y and Cancer Risk in Men. *J. Natl. Cancer Inst*. **112**, 913–920 (2020).

146. Büscheck, F. *et al*. Y-chromosome loss is frequent in male renal tumors. *Ann. Transl. Med*. **9**, 209 (2021).

147. Abdel-Hafiz, H. A. *et al*. Y chromosome loss in cancer drives growth by evasion of adaptive immunity. *Nature* **619**, 624–631 (2023).

148. Leemans, C. R., Snijders, P. J. F. & Brakenhoff, R. H. The molecular landscape of head and neck cancer. *Nat. Rev. Cancer* **18**, 269–282 (2018).

149. Miranda-Galvis, M., Loveless, R., Kowalski, L. P. & Teng, Y. Impacts of environmental factors on head and neck cancer pathogenesis and progression. *Cells* **10**, 389 (2021).

150. Yu, X. *et al*. Comprehensive analysis of the carcinogenic process, tumor microenvironment, and drug response in HPV-positive cancers. *Front. Oncol*. **12**, 842060 (2022).

151. Veiga, L. C. S., Bérgamo, N. A., Reis, P. P., Kowalski, L. P. & Rogatto, S. R. Loss of Y-chromosome does not correlate with age at onset of head and neck carcinoma: a case-control study. *Braz. J. Med. Biol. Res*. **45**, 172–178 (2012).

152. Thompson, D. J. *et al*. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).

153. Orta, A. H. *et al*. Rats exhibit age-related mosaic loss of chromosome Y. *Commun. Biol*. **4**, 1418 (2021).

154. Wójcik, M. *et al*. Loss of Y in regulatory T lymphocytes in the tumor micro-environment of primary colorectal cancers and liver metastases. *medRxiv* (2023) doi:10.1101/2023.06.17.23289722.

155. Wang, Y. & Sano, S. Why Y matters? The implication of loss of Y chromosome in blood and cancer. *Cancer Sci.* **115**, 706–714 (2024).

156. Fischer, J. *et al.* Selective loss of Y chromosomes in lung adenocarcinoma modulates the tumor immune environment through cancer/testis antigens. *bioRxivorg* (2024) doi:10.1101/2024.09.19.613876.

157. Dai, Y. *et al.* Integrative single-cell and bulk transcriptomes analyses identify intrinsic HNSCC subtypes with distinct prognoses and therapeutic vulnerabilities. *Clin. Cancer Res.* **29**, 2845–2858 (2023).

158. Puram, S. V. *et al.* Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611-1624.e24 (2017).

159. Han, J. *et al.* Comprehensive characterization of tumor microenvironment differences in cancer between male and female patients. *SSRN Electron. J.* (2021) doi:10.2139/ssrn.3894380.

160. Han, J. *et al.* Pan-cancer analysis reveals sex-specific signatures in the tumor microenvironment. *Mol. Oncol.* **16**, 2153–2173 (2022).

161. Gameiro, S. F., Evans, A. M. & Mymryk, J. S. The tumor immune microenvironments of HPV+ and HPV- head and neck cancers. *WIREs Mech Dis* **14**, e1539 (2022).

162. Tharp, K. M. *et al.* Myeloid mechano-metabolic programming restricts anti-tumor immunity. *bioRxiv* (2022) doi:10.1101/2022.07.14.499764.

163. Khalaf, K. *et al.* Aspects of the tumor microenvironment involved in immune resistance and drug resistance. *Front. Immunol.* **12**, 656364 (2021).

164. Ray, A. L. *et al.* The role of sex in the innate and adaptive immune environment of metastatic colorectal cancer. *Br. J. Cancer* **123**, 624–632 (2020).

165. Conforti, F. *et al.* Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. *Lancet Oncol.* **19**, 737–746 (2018).

166. Botticelli, A. *et al.* The sexist behaviour of immune checkpoint inhibitors in cancer therapy? *Oncotarget* **8**, 99336–99346 (2017).

167. Wang, Z. *et al.* Characterization of immune microenvironment in patients with HPV-positive and negative head and neck cancer. *Sci. Data* **10**, 694 (2023).

168. Elmusrati, A., Wang, J. & Wang, C.-Y. Tumor microenvironment and immune evasion in head and neck squamous cell carcinoma. *Int. J. Oral Sci.* **13**, 24 (2021).

169. Fialová, A., Koucký, V., Hajdušková, M., Hladíková, K. & Špíšek, R. Immunological network in head and neck squamous cell carcinoma-A prognostic tool beyond HPV status. *Front. Oncol.* **10**, 1701 (2020).
170. Wang, B., Zhang, S., Tong, F., Wang, Y. & Wei, L. HPV+ HNSCC-derived exosomal miR-9-5p inhibits TGF-β signaling-mediated fibroblast phenotypic transformation through NOX4. *Cancer Sci.* **113**, 1475–1487 (2022).
171. Li, X., González-Maroto, C. & Tavassoli, M. Crosstalk between CAFs and tumour cells in head and neck cancer. *Cell Death Discov.* **10**, 303 (2024).
172. Knops, A. M. *et al.* Cancer-associated fibroblast density, prognostic characteristics, and recurrence in head and neck squamous cell carcinoma: A meta-analysis. *Front. Oncol.* **10**, 565306 (2020).
173. Kieffer, Y. *et al.* Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Discov.* **10**, 1330–1351 (2020).
174. Hu, C., Zhang, Y., Wu, C. & Huang, Q. Heterogeneity of cancer-associated fibroblasts in head and neck squamous cell carcinoma: opportunities and challenges. *Cell Death Discov* **9**, 124 (2023).
175. Custódio, M., Biddle, A. & Tavassoli, M. Portrait of a CAF: The story of cancer-associated fibroblasts in head and neck cancer. *Oral Oncol.* **110**, 104972 (2020).
176. Thuya, W. L. *et al.* Insights into IL-6/JAK/STAT3 signaling in the tumor microenvironment: Implications for cancer therapy. *Cytokine Growth Factor Rev.* (2025) doi:10.1016/j.cytogfr.2025.01.003.
177. Freeman, P. & Mielgo, A. Cancer-associated fibroblast mediated inhibition of CD8+ cytotoxic T cell accumulation in tumours: Mechanisms and therapeutic opportunities. *Cancers (Basel)* **12**, 2687 (2020).
178. Erdogan, B. & Webb, D. J. Cancer-associated fibroblasts modulate growth factor signaling and extracellular matrix remodeling to regulate tumor metastasis. *Biochem. Soc. Trans.* **45**, 229–236 (2017).
179. Ghahremanifard, P., Chanda, A., Bonni, S. & Bose, P. TGF-β mediated immune evasion in cancer-spotlight on cancer-associated fibroblasts. *Cancers (Basel)* **12**, 3650 (2020).
180. Calon, A., Tauriello, D. V. F. & Batlle, E. TGF-beta in CAF-mediated tumor growth and metastasis. *Semin. Cancer Biol.* **25**, 15–22 (2014).

181. Huang, T. *et al*. Wnt inhibition sensitizes PD-L1 blockade therapy by overcoming bone marrow-derived myofibroblasts-mediated immune resistance in tumors. *Front. Immunol.* **12**, 619209 (2021).

182. Ershaid, N. *et al*. NLRP3 inflammasome in fibroblasts links tissue damage with inflammation in breast cancer progression and metastasis. *Nat. Commun.* **10**, 4375 (2019).

183. Cole, M. A., Quan, T., Voorhees, J. J. & Fisher, G. J. Extracellular matrix regulation of fibroblast function: redefining our perspective on skin aging. *J. Cell Commun. Signal.* **12**, 35–43 (2018).

184. Kay, E. J., Koulouras, G. & Zanivan, S. Regulation of extracellular matrix production in activated fibroblasts: Roles of amino acid metabolism in collagen synthesis. *Front. Oncol.* **11**, 719922 (2021).

185. Dunbar, A. *et al*. CXCL8/CXCR2 signaling mediates bone marrow fibrosis and represents a therapeutic target in myelofibrosis. *bioRxiv* (2021) doi:10.1101/2021.12.08.471791.

186. Guo, Q. *et al*. NF-κB in biology and targeted therapy: new insights and translational implications. *Signal Transduct. Target. Ther.* **9**, 53 (2024).

187. Saw, P. E., Chen, J. & Song, E. Targeting CAFs to overcome anticancer therapeutic resistance. *Trends Cancer* **8**, 527–555 (2022).

188. Nicolas, A. M. *et al*. Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer. *Cancer Cell* **40**, 168-184.e13 (2022).

189. Piwocka, O., Piotrowski, I., Suchorska, W. M. & Kulcenty, K. Dynamic interactions in the tumor niche: how the cross-talk between CAFs and the tumor microenvironment impacts resistance to therapy. *Front. Mol. Biosci.* **11**, 1343523 (2024).

190. Tao, L., Huang, G., Song, H., Chen, Y. & Chen, L. Cancer associated fibroblasts: An essential role in the tumor microenvironment. *Oncol. Lett.* **14**, 2611–2620 (2017).

191. Tokhanbigli, S., Haghi, M., Dua, K. & Oliver, B. G. G. Cancer-associated fibroblast cell surface markers as potential biomarkers or therapeutic targets in lung cancer. *Canc. Drug Resist.* **7**, 32 (2024).

192. Xuan, Z. *et al*. PLXNB1/SEMA4D signals mediate interactions between malignant epithelial and immune cells to promote colorectal cancer liver metastasis. *J. Cell. Mol. Med.* **28**, e70142 (2024).

193. Jiang, J. *et al*. Semaphorins as potential immune therapeutic targets for cancer. *Front. Oncol.* **12**, 793805 (2022).

194. Biffi, G. *et al*. IL1-induced JAK/STAT signaling is antagonized by TGFβ to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Discov.* **9**, 282–301 (2019).

195. Kwaśniak, K. *et al*. Scientific reports concerning the impact of interleukin 4, interleukin 10 and transforming growth factor β on cancer cells. *Cent. Eur. J. Immunol.* **44**, 190–200 (2019).

196. Luheshi, N., Davies, G. & Legg, J. Understanding the influence of the tumor microenvironment on macrophage responses to CD40 agonists. *Oncoimmunology* **3**, e27615 (2014).

197. Wang, S.-J. *et al*. Effect of cyclo-oxygenase inhibitor use during checkpoint blockade immunotherapy in patients with metastatic melanoma and non-small cell lung cancer. *J. Immunother. Cancer* **8**, e000889 (2020).

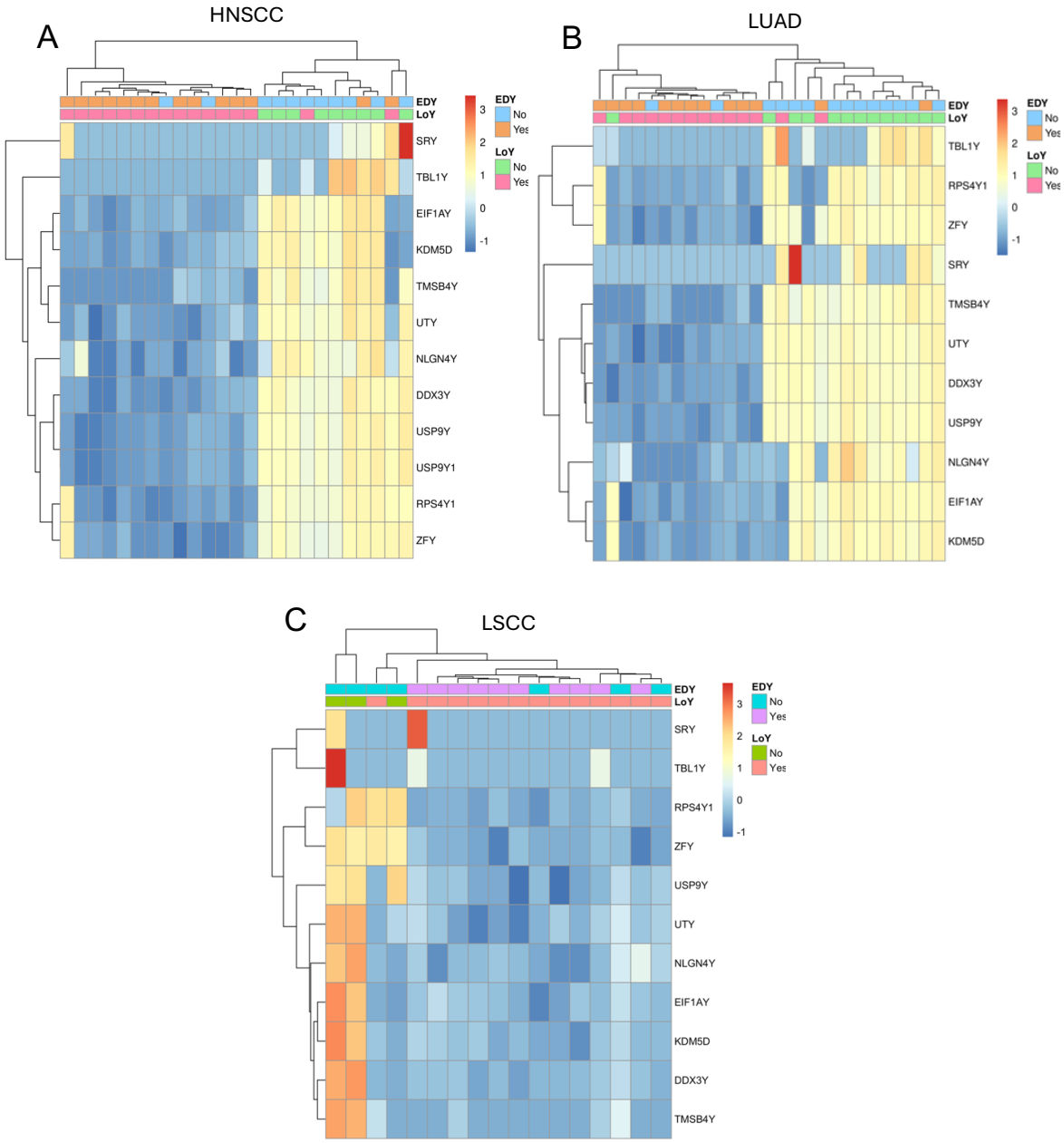
198. Jin, K., Qian, C., Lin, J. & Liu, B. Cyclooxygenase-2-Prostaglandin E2 pathway: A key player in tumor-associated immune cells. *Front. Oncol.* **13**, 1099811 (2023).

199. Jacob, A., Raj, R., Allison, D. B. & Myint, Z. W. Androgen receptor signaling in prostate cancer and therapeutic strategies. *Cancers (Basel)* **13**, 5417 (2021).

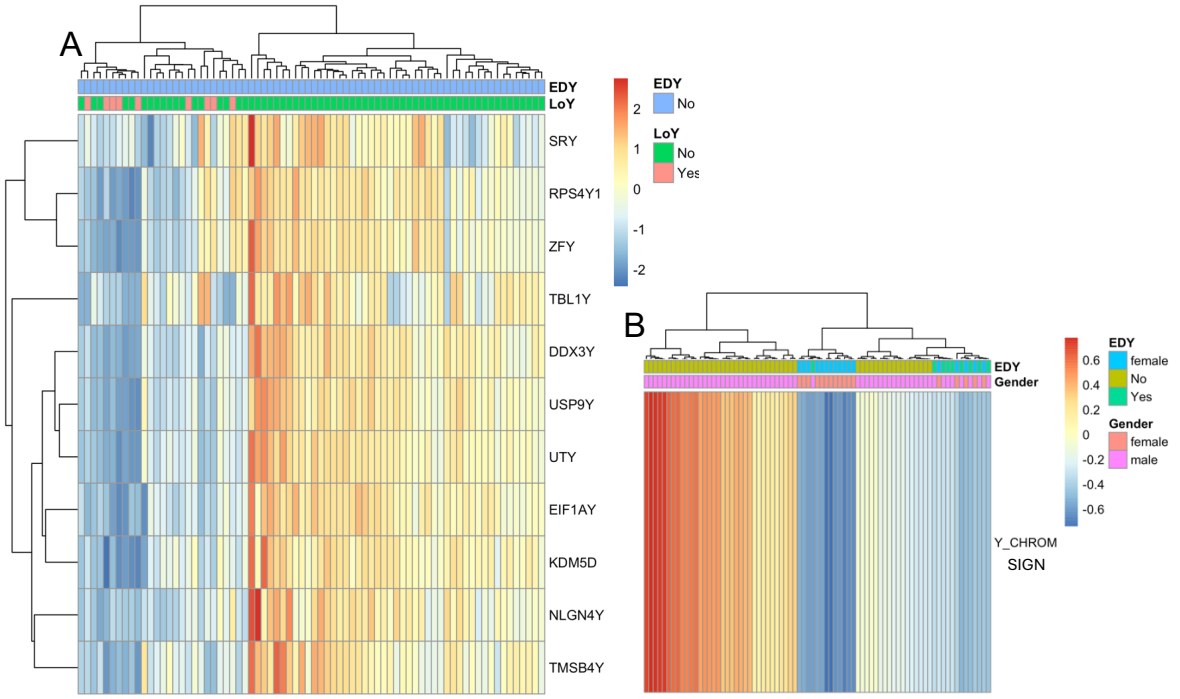
200. Chen, Y. & Lan, T. N-terminal domain of androgen receptor is a major therapeutic barrier and potential pharmacological target for treating castration resistant prostate cancer: a comprehensive review. *Front. Pharmacol.* **15**, 1451957 (2024).

201. Liu, S. *et al*. Androgen receptor suppresses lung cancer invasion and increases cisplatin response via decreasing TPD52 expression. *Int. J. Biol. Sci.* **19**, 3709–3725 (2023).

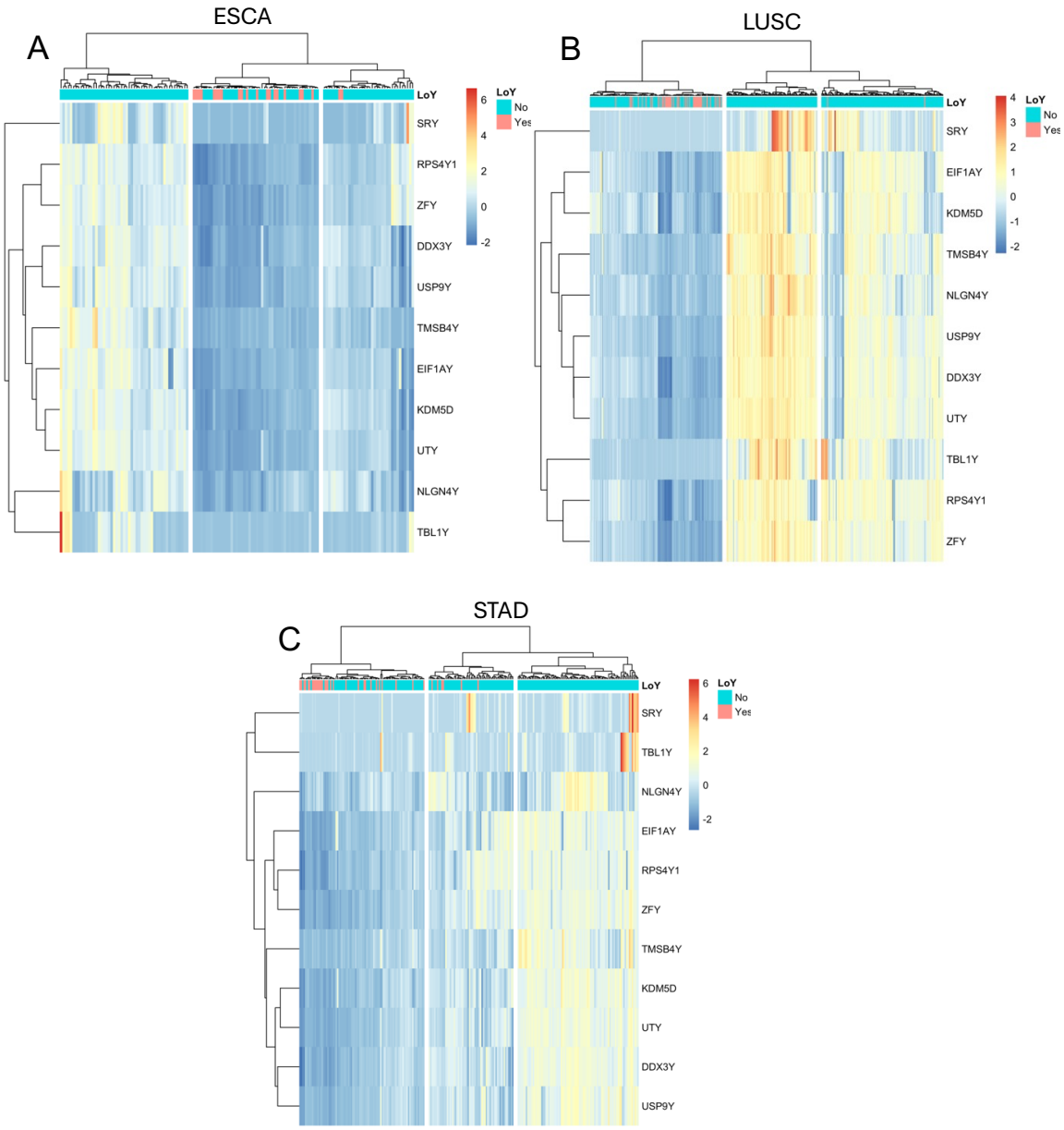
Supplementary Materials



Supplementary Figure 1. Heatmap showing Y chromosome gene expression in male cell lines from the Cancer Cell Line Encyclopedia (CCLE) for HNSCC (A), LUAD (B), and LSCC (C). The heatmaps include 11 Y chromosome genes from the HNSCC gene set. Each column represents a male cell line, with annotations for LoY and EDY. Expression values are scaled, and genes and samples are clustered by similarity. A clear downregulation pattern is visible in EDY samples, consistent with what we see in patient data.



Supplementary Figure 2. Y chromosome gene expression and GSVA-based clustering across HNSCC patient bulk datasets. *(A)* Heatmap of scaled expression for 11 Y chromosome genes in the CPTAC HNSC cohort. Only male samples are shown, annotated for LoY. ***(B)*** Heatmap showing the GSVA enrichment score of the Y chromosome gene set in the HIPO cohort. Both male and female samples are included, annotated by sex and EDY status. Some male samples cluster together with females, indicating reduced Y chromosome activity in those cases.



Supplementary Figure 3. Expression of Y chromosome gene set across male patient samples from three additional TCGA cancer types. Heatmaps show scaled expression of the 11 Y chromosome genes in *(A)* Esophageal Carcinoma (ESCA), *(B)* Lung Squamous Cell Carcinoma (LSCC), and *(C)* Stomach Adenocarcinoma (STAD). Samples are annotated for LoY. A similar transcriptional downregulation pattern is observed in LoY-classified samples across all three cancer types, consistent with findings in HNSCC.

Supplementary Table 1. Mean xCell-derived enrichment scores for each cell type across sex chromosome dosage groups (X0, XX, and XY). To investigate the influence of sex chromosome dosage on cell composition, enrichment scores were computed using xCell and compared between groups using Kruskal–Wallis tests. Several immune cell types showed significant differences ($p < 0.05$), as described in Figure 11. This table provides a detailed summary of the mean scores per group.

Cell_Type	X0	XX	XY
Bcell	0.00180792048982	0.00111845236062	0.00198960499867
CD4T	0.00130213008617	0.00123661094076	0.00110940286231
CD8T	0.00533037156298	0.00506523075327	0.00385897000729
Endothelial	0.01237560741828	0.00817258588318	0.01413446530205
Epithelial	0.05727558644811	0.05836944749523	0.05313055922096
Fibroblast	0.00403467294506	0.00243849624142	0.00396174460154
IDO_CAF	0.01675392352409	0.01703127795680	0.01112690167099
Macrophage	0.03772692240728	0.02808252630797	0.03883320514009
Mast	0.00630571382764	0.00246436541151	0.00285374179025
Monocyte	0.00245853873497	0.00118040021371	0.00087564273447
Multi	0.00245243939455	0.00135731080218	0.00366643819432
NK	3,62E+09	0.00037796679362	7,36E+09
Neutrophils	0.00161288772601	0.00109782562344	0.00086878692315
Pericyte	0.00377072004144	0.00367867901840	0.00303867650096
PlasmaCell	0.00389026528807	0.00273358576945	0.00401886807404
Tumor	0.60583950518437	0.62332323382870	0.61241897405784
Unknown	8,39E+09	0.00018585396906	9,31E+09

apCAF	0.00092915478160	0.00032421463611	0.00065089915753
dCAF	0.14753871751695	0.16355179918619	0.14947328330123
hsp_tpCAF	0.00104134966799	0.00111993996276	0.00214445608988
iCAF	0.00091318140449	0.00090527762374	0.00145881840844
mCAF	0.06109814765637	0.04641448870211	0.06297348508529
other	0.00253785825958	0.00161801330286	0.00344259077818
panDC	0.00500297576270	0.00466713030394	0.00387840362183
rCAF	9,01E+09	0.00032162717189	0.00065199581718
rCAF+apCAF	0.00289187320592	0.00393845431483	0.00579612724865
tpCAF	0.00811624409517	0.01255097727271	0.00568322393209
vCAF	0.00678312899904	0.00667422815341	0.00779401814278

Supplementary Table 2. Summary of publicly available human HNSCC single-cell RNA-seq datasets.This table provides an overview of HNSCC single-cell transcriptomic datasets deposited in the Gene Expression Omnibus (GEO), including accession numbers, publication year, first authors, data type, tissue types analyzed, number of samples, and number of cells. This table was extracted from: Conde-Lopez, C., Marripati, D., Elkabets, M., Hess, J., & Kurth, I. *Unravelling the complexity of HNSCC using single-cell transcriptomics*. Cancers **16**, 3265 (2024).

Accession Number	Year	Publication	Data type	Tissue types	Sample Number	Cell Number
GSE234933	2023	Bill R, Wirapati P, Messemaker M, Roh W et al. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. Science 2023 Aug 4;381(6657):515-524	scRNAseq	Primary tumor Local recurrence Distant metastasis	52	87399
GSE182227	2022	Puram SV, Mints M, Pal A, Qi Z et al. Cellular states are coupled to genomic and viral heterogeneity in HPV-related oropharyngeal carcinoma. Nat Genet2023 Apr;55(4):640-650	scRNAseq	Primary tumor Normal tissue	24	70970
GSE139324	2019	Cillo AR, Kürten CHL, Tabib T, Qi Z et al. Immune Landscape of Viral- and Carcinogen-Driven Head and Neck Cancer. Immunity 2020 Jan 14;52(1):183-199.e9.	scRNAseq	Peripheral/Intra-tumoral CD45+ populations	63	131224
GSE164690	2021	Kürten CHL, Kulkarni A, Cillo AR, Santos PM et al. Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. Nat Commun 2021 Dec 17;12(1):7338	scRNAseq	Primary tumor Pheripheral Blood Leucocytes	51	134606
GSE103322	2017	Puram SV, Tirosh I, Parikh AS, Patel AP et al. Single-Cell Transcriptomic Analysis of Primary and	scRNAseq	Primary tumor	18	5902

		Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell 2017 Dec 14;171(7):1611-1624.e24.				
GSE181919	2022	Choi JH, Lee BS, Jang JY, Lee YS et al. Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. Nat Commun 2023 Feb 24;14(1):1055.	scRNAseq	Primary tumor Normal tissue Leukoplakia Lymph-node metastasis	37	54239
GSE173647	2022	---	scRNAseq	Primary tumor	2	13903
GSE195832	2022	Obradovic A, Graves D, Korrer M, Wang Y et al. Immunostimulatory Cancer-Associated Fibroblast Subpopulations Can Predict Immunotherapy Response in Head and Neck Cancer. Clin Cancer Res 2022 May 13;28(10):2094-2109.	scRNAseq	Primary tumor	8	22906
GSE140042	2021	---	scRNAseq	Primary tumor Lymph-node metastasis	9	---
GSE200996	2022	Luoma AM, Suo S, Wang Y, Gunasti L et al. Tissue-resident memory and circulating T cells are early responders to pre-surgical cancer immunotherapy. Cell 2022 Aug 4;185(16):2918-2935.e29.	scRNAseq + scTCR	Peripheral/Intra-tumoral CD45+ populations	204	74557
GSE153559	2020	Wieland A, Patel MR, Cardenas MA, Eberhardt CS et al. Defining HPV-specific B cell responses in patients with head and neck cancer. Nature 2021 Sep;597(7875):274-278.	scRNAseq	B cells Primary tumor Lymph-node metastasis Pheriphery	7	8271
GSE180268	2021	Eberhardt CS, Kissick HT, Patel MR, Cardenas MA et al. Functional HPV-specific PD-1(+) stem-like CD8 T cells in head and neck cancer. Nature 2021 Sep;597(7875):279-284.	scRNAseq	TILs Primary tumor/Lymph-node metastasis	39	---
GSE162025	2020	Liu Y, He S, Wang XL, Peng W et al. Tumour heterogeneity and	scRNAseq + scTCR	Primary tumor/Pheripheral Blood Leucocytes	40	176447

		intercellular networks of nasopharyngeal carcinoma at single cell resolution. Nat Commun2021 Feb 2;12(1):741				
GSE150321	2020	Song L, Zhang S, Yu S, Ma F et al. Cellular heterogeneity landscape in laryngeal squamous cell carcinoma. Int J Cancer 2020 Nov 15;147(10):2879-2890.	scRNAseq	Primary tumor	2	12985
GSE213047	2022	Lin M, Sade-Feldman M, Wirth L, Lawrence MS et al. Single-cell transcriptomic profiling for inferring tumor origin and mechanisms of therapeutic resistance. NPJ Precis Oncol 2022 Oct 10;6(1):71.	scRNAseq	Primary tumor/Normal tissue/Lymph-node metastasis	3	11470
GSE172577	2021	Peng Y, Xiao L, Rong H, Ou Z et al. Single-cell profiling of tumor-infiltrating TCF1/TCF7(+) T cells reveals a T lymphocyte subset associated with tertiary lymphoid structures/organs and a superior prognosis in oral cancer. Oral Oncol2021 Aug;119:105348.	scRNAseq	Primary tumor	6	---

Supplementary Tables 3–7. Cell type composition in HNSCC tumors by sex, stratified by HPV status. These tables provide a detailed overview of cell type proportions across female and male patients, separately for HPV-positive and HPV-negative tumors: (3) Proportions of major cell types across sex groups by HPV status. (4) Proportions of major immune cell types across sex groups by HPV status. (5) Proportions of $CD8^+$ T cell subtypes across sex groups by HPV status. (6) Proportions of $CD4^+$ T cell subtypes across sex groups by HPV status. (7) Proportions of DC subtypes across sex groups by HPV status.These tables were adapted from: Conde Lopez, C., Marripati, D., Besso, M. J., Roscher, M., Han, R., Hadiwikarta, W. W., Elkabets, M., Hess, J., & Kurth, I. *A unified single-cell atlas of HNSCC: uncovering HPV and sex variability in the tumor microenvironment*. Submitted to *Communications Medicine* (2025).

3.

Percentage All Cell Types

	HPV -		HPV +	
	Female	Male	Female	Male
Bcell	4.4%	3.6%	7.7%	10.20%
CD4T	23.9%	18.7%	22.5%	20.00%
CD8T	21.7%	17.2%	9.5%	16.30%
Endothelial	8.9%	6.8%	17.0%	4.90%
Epithelial	1.6%	2.0%	1.6%	2.00%
Fibroblast	6.7%	10.9%	18.3%	5.50%
Macrophage	4.2%	5.9%	1.3%	2.90%
Mast	0.5%	0.9%	1.7%	0.40%
Monocyte	7.8%	4.2%	0.8%	0.80%
Multi	1.2%	0.9%	1.4%	0.90%
NK	5.5%	3.1%	1.0%	2.70%
Neutrophils	1.6%	2.2%	0.0%	0.00%
PlasmaCell	0.9%	3.0%	3.5%	4.60%
Tumor	8.4%	16.9%	11.2%	26.10%
panDC	2.9%	3.7%	2.7%	2.90%

4.

Percentage Immune Cell Types

	HPV -		HPV +	
	Female	Male	Female	Male
Bcell	6.0%	6.0%	16.3%	18.2%

CD4T	33.0%	31.4%	47.7%	35.6%
CD8T	29.9%	28.8%	20.1%	29.0%
Macrophage	5.9%	9.9%	2.8%	5.2%
Mast	0.7%	1.5%	3.7%	0.7%
Monocyte	10.7%	7.1%	1.6%	1.4%
NK	7.5%	5.2%	2.1%	4.8%
Neutrophils	2.3%	3.8%	0.0%	0.1%
panDC	4.0%	6.3%	5.7%	5.1%

5.

Percentage CD8 Cell Types

	HPV -		HPV +	
	Female	Male	Female	Male
CD8.CM	23.0%	14.8%	37.1%	20.3%
CD8.EM	29.6%	35.4%	28.5%	40.2%
CD8.MAIT	0.5%	0.4%	0.0%	0.4%
CD8.NaiveLike	0.8%	1.0%	1.3%	2.2%
CD8.TEMRA	1.2%	1.3%	0.3%	0.4%
CD8.TEX	43.3%	45.5%	26.6%	33.6%
CD8.TPEX	1.6%	1.6%	6.2%	3.0%

6.

Percentage CD4 Cell Types

	HPV -		HPV +	
	Female	Male	Female	Male
CD4.CTL_EOMES	1.2%	2.7%	12.8%	2.9%
CD4.CTL_Exh	8.0%	6.0%	1.2%	3.5%
CD4.CTL_GNLY	1.5%	2.0%	12.5%	2.1%
CD4.NaiveLike	29.5%	33.8%	46.3%	47.9%
CD4.Tfh	12.4%	9.3%	11.2%	11.5%
CD4.Th17	2.0%	3.9%	2.8%	3.2%
CD4.Treg	45.5%	42.2%	13.1%	28.9%

7.

Percentage DC Cell Types

	HPV -		HPV +	
	Female	Male	Female	Male
AS-DC	2.3%	5.0%	1.0%	1.5%
DC3	23.1%	20.5%	29.5%	18.9%
MonoDC	1.7%	0.9%	1.0%	2.4%

cDC1	9.2%	8.4%	17.1%	12.1%
cDC2_CD1A	5.4%	5.7%	6.7%	10.1%
cDC2_CLEC10A	27.4%	34.6%	24.8%	28.1%
pDC	30.9%	25.0%	20.0%	27.0%

Supplementary Table 8. Overview of $CD8^+$, $CD4^+$, and dendritic cell subtypes. This table summarizes the major subtypes of $CD8^+$ T cells, $CD4^+$ T cells, and dendritic cells (DCs), providing definitions and outlining their functional roles in immune regulation. It serves as a reference framework for the classification and functional annotation of these immune subsets in HNSCC tumors, supporting cross-study comparisons and interpretability of single-cell data.This table was adapted from: Conde Lopez, C., Marripati, D., Besso, M. J., Roscher, M., Han, R., Hadiwikarta, W. W., Elkabets, M., Hess, J., & Kurth, I. *A unified single-cell atlas of HNSCC: uncovering HPV and sex variability in the tumor microenvironment*. Submitted to *Communications Medicine* (2025).

Cell Type	**Description**
CD8.NaiveLike	Antigen-naive T cells
CD8.CM	Central Memory T cells
CD8.EM	Effector Memory T cells
CD8.TEMRA	Effector Memory cells re-expressing CD45RA. Sometimes called Short Lived Effectors (SLEC), or Cytotoxic effectors
CD8.TPEX	Progenitor exhausted T cells
CD8.TEX	Exhausted T cells
CD8.MAIT	Mucosal-associated invariant T cells, innate-like T cells defined by their semi-invariant αβ T cell receptor
CD4.NaiveLike	T cells with naive-like phenotype
CD4.Tfh	T follicular helper cells
CD4.Th17	Th17 helper cells
CD4.Treg	T regulatory cells
CD4.CTL_EOMES	Cytotoxic CD4 T cells expressing EOMES and GZMK
CD4.CTL_GNLY	Cytotoxic CD4 T cells expressing GNLY
CD4.CTL_Exh	Cytotoxic CD4 T cells with exhaustion phenotype
AS-DC	AXL+ SIGLEC6+ Dendritic Cells, also referred to as DC5 or Pre-DCs
cDC1	Conventional Dendritic Cells type 1, specialized in antigen cross-presentation and CD8+ T cell activation
cDC2_CD1A	Conventional Dendritic Cells type 2 subset expressing CD1A, involved in CD4+ T cell activation
cDC2_CLEC10A	Conventional Dendritic Cells type 2 subset expressing CLEC10A, functionally distinct from CD1A-expressing cDC2

DC3	Tissue-resident DCs, lacking a direct counterpart in circulation, potentially derived from cDC1 or cDC2/MoDC
MonoDC	Monocyte-derived Dendritic Cells, inflammation-induced, with similarities to cDC2
pDC	Plasmacytoid Dendritic Cells, key producers of type I IFNs, fostering antitumor immunity

Supplementary Tables 9–12. Cell type composition in HNSCC tumors by sex chromosome dosage. These tables provide a detailed overview of cell type proportions across XX, XØ and XY patients (9) Proportions of major cell types across sex chromosome dosage groups. (10) Proportions of $CD8^+$ T cell subtypes across groups. (11) Proportions of $CD4^+$ T cell subtypes across groups. (12) Proportions of fibroblast subtypes across groups.

9.

Percentage All Cell Types			
	XX	XØ	XY
Bcell	4.6%	1.3%	5.1%
CD4T	20.2%	10.0%	14.0%
CD8T	22.8%	8.9%	14.8%
Endothelial	8.8%	4.5%	9.2%
Epithelial	2.5%	1.6%	0.8%
Fibroblast	8.6%	5.1%	22.9%
Macrophage	5.5%	7.6%	5.9%
Mast	0.4%	0.6%	1.5%
Monocyte	5.2%	3.0%	3.5%
Multi	0.9%	0.7%	0.7%
Neutrophils	2.1%	0.5%	0.3
NK	3.1%	1.5%	1.8%
panDC	3.9%	2.7%	4.6%
PlasmaCell	1.3%	4.6%	4.7%
Tumor	10.4%	47.5%	10.2%

10.

Percentage CD8 Cell Types			
	XX	XØ	XY
CD8.CM	23.0%	18.2%	13.0%
CD8.EM	29.6%	44.2%	31.0%
CD8.MAIT	0.5%	0.5%	0.3%

CD8.NaiveLike	0.8%	0.4%	1.3%
CD8.TEMRA	1.2%	2.3%	0.9%
CD8.TEX	43.3%	32.5%	52.2%
CD8.TPEX	1.6%	1.9%	1.4%

11.

Percentage CD4 Cell Types			
	XX	XØ	XY
CD4.CTL_EOMES	1.2%	3.6%	2.2%
CD4.CTL_Exh	8.0%	5.6%	6.2%
CD4.CTL_GNLY	1.5%	2.0%	2.1%
CD4.NaiveLike	29.5%	26.8%	38.1%
CD4.Tfh	12.4%	6.3%	11.0%
CD4.Th17	2.0%	3.4%	4.3%
CD4.Treg	45.5%	52.2%	36.2%

12.

Percentage Fibroblast Cell Types			
	XX	XØ	XY
apCAF	0.6%	0.7%	0.5%
dCAF	0.0%	0.1%	0.3%
hsp-tpCAF	0.0%	0.0%	0.0%
iCAF	10.6%	1.6%	30.7%
IDO-CAF	0.6%	1.1%	1.7%
mCAF	67.7%	84.5%	47.5%
other	0.1%	0.0%	0.1%
Pericyte	6.3%	5.5%	10.9%
rCAF	0.2%	0.1%	0.4%
rCAF+apCAF	0.0%	0.0%	0.0%
tpCAF	12.5%	6.1%	5.5%
vCAF	1.0%	0.2%	0.8%

Supplementary Table 13. Statistical comparison of BayesPrism-inferred cell type proportions across sex chromosome dosage groups in TCGA HNSCC tumors. This table reports the results of Kruskal–Wallis tests assessing differences in estimated cell type proportions across XX, XY, and XØ groups in bulk RNA-seq data deconvoluted with

BayesPrism. For each cell type, the Kruskal–Wallis p-value is provided along with a summary of pairwise significance. Several stromal and immune cell populations, including mCAFs, iCAFs, apCAFs, and macrophages, showed significant differences between dosage groups, supporting findings from the single-cell analysis (Figure 30.A–B).

Cell_Type	KW_p_value	Pairwise_significant_pairs
Tumor	0,808446	Non-significant
CD8T	0,9221	Non-significant
CD4T	0,586958	Non-significant
Macrophage	**0,020996**	**Significant**
Multi	**0,018843**	**Significant**
NK	0,068868	Non-significant
Unknown	0,20964	Non-significant
Monocyte	0,374392	Non-significant
panDC	0,432993	Non-significant
Neutrophils	0,646379	Non-significant
Mast	0,164652	Non-significant
Bcell	0,114387	Non-significant
mCAF	**0,000354**	**Significant**
tpCAF	0,24684	Non-significant
Endothelial	**0,00017**	**Significant**
PlasmaCell	0,501215	Non-significant
Epithelial	0,459779	Non-significant
IDO_CAF	0,730948	Non-significant
iCAF	**0,008151**	**Significant**
apCAF	**0,000318**	**Significant**
Pericyte	0,607022	Non-significant
rCAF+apCAF	**0,003405**	**Significant**
hsp_tpCAF	0,13569	Non-significant
vCAF	0,20199	Non-significant
other	**0,004678**	**Significant**
rCAF	**0,0053**	**Significant**
Fibroblast	0,025889	Significant
dCAF	**0,432031**	**Non-significant**