UNIVERSITÀ
DEGLI STUDI DI TRIESTE

Dipartimento di Scienze della Vita

Tesi di Dottorato
in
Biomedicina Molecolare
XXX CICLO

An integrated approach identifies mediators of local recurrence in Head and Neck Squamous Cell Carcinoma

RELATORE
Dott. Gustavo Baldassarre

DOTTORANDA
Dott.ssa Francesca Citron

CORRELATORE
Dott.ssa Barbara Belletti

COORDINATORE
Prof. Licio Collavin

Anno Accademico 2016-2017

The main part of the results included in this PhD thesis has been published on "An integrated approach identified mediators of local recurrence in Head and Neck Squamous Cell Carcinoma", Citron F. et al, Clin Cancer Res 23, 3769 (2017).

TABLE OF CONTENTS

ABSTRACT

Head and Neck Squamous Cell Carcinoma (HNSCC) represents the sixth most common cancer worldwide, with about 550,000 new cases/year.

Despite the development of precise and accurate surgery followed by radio- and/or chemo-therapy, in the last decade patients' overall survival was only slightly improved. In particular, only 40-50% of patients with advanced HNSCC will survive for 5 years. The main worse prognostic event is considered the onset of loco-regional and distant recurrences, that is also accepted as surrogate markers of patients' overall survival. HNSCC is a heterogeneous disease and, although many studies have been conducted to clarify the molecular mechanisms behind the development of local or distant metastasis, no clear molecular mediators of recurrence formation have been identified and no valid biomarkers exist to identify and treat patients at high-risk of recurrence.

We aimed to fill this gap taking advantage of an unbiased approach in which we evaluated microRNA (miRNA) expression profile in a cohort of HNSCC primary tumours from recurrent and non-recurrent patients.

We identified and validated a four-miRNA signature, composed by miR-1, miR-9, miR-133a and miR-150 that could be used as biomarker of recurrence. All these miRNAs are not only well-known molecular modulators of cell plasticity in different tumours, but intriguingly in our HNSCC model they are also able to classify HNSCC patients at high- or low-risk of recurrence formation.

We demonstrated that these miRNAs collectively impinge on the epithelial-mesenchymal transition process. *In silico* and wet lab approaches showed that miR-9, expressed at high levels in recurrent HNSCC, targets the epithelial genes SASH1 and KRT13, while miR-1, miR-133a and miR-150, expressed at low levels in recurrent HNSCC, collectively target SP1 and TGFβ pathways.

In vivo, a six-genes signature comprising the above targets, consistently predicts low progression free- and overall-survival in different panels of HNSCC samples. This is of particular clinical relevance, since the detection of this gene signature in tumours specimens could identify patients with poor prognosis who may benefit for a more accurate targeted therapy.

In a preclinical model of HNSCC recurrence*,* the combined pharmacological inhibition of SP1 and TGFβ pathways, when timely administered, induced HNSCC cell death and prevented recurrence formation.

By integrating different experimental approaches, we identified critical mediators of recurrence formation in advanced HNSCC. Since both SP1 and TGFβ pathways could be pharmacologically targeted, their combined inhibition may merit to be considered for future clinical development.

1. INTRODUCTION

1.1 Head and Neck Squamous Cell Carcinoma

Head and Neck Squamous Cell Carcinoma (HNSCC) originates from the mucosal epithelia of the upper aero-digestive tract and encompasses a variety of solid tumours originating in the squamous epithelium of lip, oral cavity, hypopharynx, oropharynx, nasopharynx and larynx[1]. HNSCC represents the sixth most common cancer worldwide, with about 550,000 new cases/year worldwide, and only the 40-50% of patients will survive for 5 years[2].

HNSCC can be stratified in HPV-negative or HPV-positive malignancies on the basis of the etiological factor(s), this is of extremely clinical importance since these two subclasses differ not only for pathological and molecular features, but also for the clinical outcome[3].

HPV-negative cluster comprehends the 80% of HNSCCs and are generally more frequent in the adult (>60 years) male population. The most important risk factors so far identified are tobacco use and alcohol consumption, which seem to have synergistic effect. In the western world, the incidence of this subgroup of HNSCC has been slowly declining during the last years, which can be attributed to a decrease in the prevalence of risk factors, most notably smoking[1,2].

HPV-positive carcinoma, the minor subclass of HNSCC for incidence, is caused by Human Papilloma Virus (HPV) infection and is frequently associated with high-risk sexual behaviours. This subgroup affects in prevalence tongue and oropharynx and has the same frequency in men and women, with an insurgence usually under 60 years[4].

Even if the 95% of head and neck neoplasia are squamous cell carcinoma, recent insights have revealed that HNSCC is unexpectedly heterogeneous. It has been demonstrated that among HPV-negative early lesions there is a marked inter and intra-tumour genomic variability.

In this context, TP53 mutations are the most frequent genetic alteration, detectable in more than 60-80% of HPV-negative HNSCC and playing an uncontested role in the HNSCC pathogenesis, even though it could not be considered a univocal prognostic marker[5]. Many efforts were made in order to validate other genetic alteration that correlate with a poor prognosis. Among these mutations, CCND1, EGFR and MET amplifications are considered frequent events in early lesions, as well as inactivation of important oncosuppressor genes, such as CDKN2A, RB1 and PTEN[6]. In the late stages of HPV-negative diseases, it has been demonstrated increased alterations in pathways belonging to E-Cadherin, VEGF, PDGF and more importantly TGFβ signalling, contributing to a worst prognosis[3,7].

As regarding HPV-positive lesions, there is no correlation neither with TP53 mutations, since they are almost all TP53 wild type, or with other risk factors above mentioned. These malignancies are

usually less heterogenic and display a particular favourable prognosis respect to HPV-negative subclass.

HPV is the pathogenic agent and the viral genome has been observed in about 72% of the cases. The most frequent HPV serotype is HPV-16, which encode E6 and E7 proteins that are able to bind respectively p53 and Rb proteins, inhibiting their oncosuppressor functions with dramatic perturbation of cell cycle in the infected cells[8] (Figure 1).

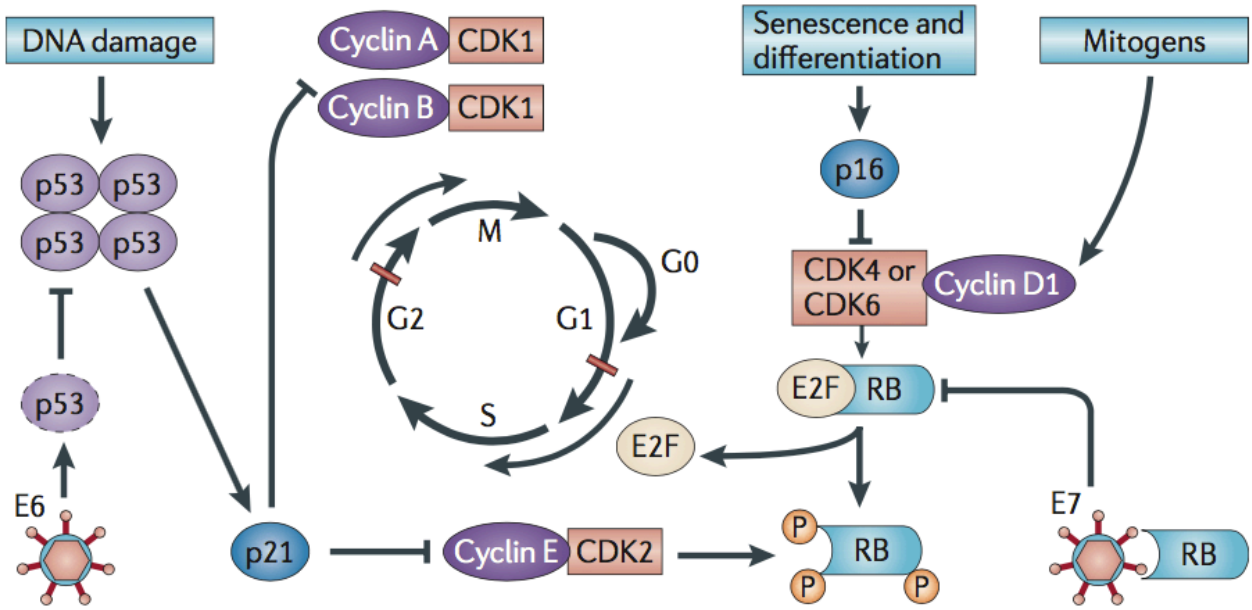

Figure 1. Cell Cycle dysregulation by HPV infection.
This schematic representation of cell cycle principal effectors shows, in red bars, the G1/S and the G2/M checkpoints. These two important steps are mainly controlled by RB pocket proteins and p53 respectively. RB proteins normally bind to and inactivate E2Fs transcription factors, which induce the expression of S phase genes. The key protein p53, once activated by phosphorylation, acts as a stress-induced transcription factor leading to $p21^{CIP}$ expression which in turn inhibits several cyclin-CDK complexes to cause cell cycle arrest. HPV genome contains various early and late open reading frames and encodes two viral onco-proteins: E6 and E7. The E6 protein binds p53 and targets the protein for degradation, whereas the E7 protein binds to inactivate RB pocket proteins. The molecular consequences of these viral onco-proteins expression are cell cycle entry and inhibition of p53-mediated apoptosis, which allows the virus to replicate. In a "oncogenic infection", the expression of E6 and E7 is not confined to the differentiating layer of the squamous epithelium, but it is detectable also the basal layer, where stem cells reside, and causes abrogation of the cell cycle checkpoints and virions production. Figure has been adapted from "The Molecular Biology of Head and Neck Cancer", Leemans et al[9].

The prognosis of HNSCC patients is largely determined by the stage at presentation, which comprehends an evaluation of the extent of the tumour, as well as the presence of lymph-node invasion and distant metastases. The tumour stage of HNSCC is defined by clinical examination, with the integration of different disciplines, including imaging, cytology of lymph-nodes and tumour histopathology after surgery (such as radicality and extra-nodal spread)[9].

Recently, HPV status and tobacco use have also been shown to be of significant prognostic

importance, possibly outweighing the traditional tumour, node, metastasis (TNM) staging system in oropharyngeal tumours[4].

Early-stage tumours (stage I/II), representing the minor part of the cases, are treated with surgery or radiotherapy alone and have a favourable prognosis. The mainstays of treatment for advanced tumours (stage III/IV) are surgery combined with postoperative chemo and/or radiotherapy[10].

The standard chemotherapy for advanced patients (stage III or IV) is the administration of Cisplatin, as radio-sensitizing agent, 100 mg/m^2 once every three weeks[11,12]. To overcome the frequent and highly toxic side effects, this schedule could be adapted to a decreased dose administrated once a week, or preferring the use of Carboplatin or Paclitaxel[13].

A recent meta-analysis demonstrated that the combination of chemo- and radio-therapy is a valid, although highly toxic, therapeutic option and despite this aggressive schedule, the 5-year survival of HNSCC patients ranges from 35 to 55% and has not markedly improved in recent decades, mostly due to the aggressive and invasive growth pattern as well as high resistance against available therapies, leading to loco-regional relapse and/or distant metastasis[13].

As mentioned before, independently from the etiological agents, HNSCC is relative heterogeneous disease and its progression is a stepwise process, resulting from the accumulation of molecular alterations in the squamous epithelium, which eventually drive the progression from premalignant lesions to invasive disease[14,15] (Figure 2).

The limited information available on the molecular carcinogenesis of HNSCC, and the genetic and biological heterogeneity of the disease has hampered the development of new therapeutic strategies. While inactivation of p53 and RB pathways are considered an early, nearly universal event in HNSCC progression, either through somatic mutation/inactivation of critical tumour suppressor genes (*e.g.* TP53 and CDKN1A) or through HPV infection, less is known about the subsequent molecular events necessary for the progression of HNSCC to invasive, metastatic carcinomas[16,17].

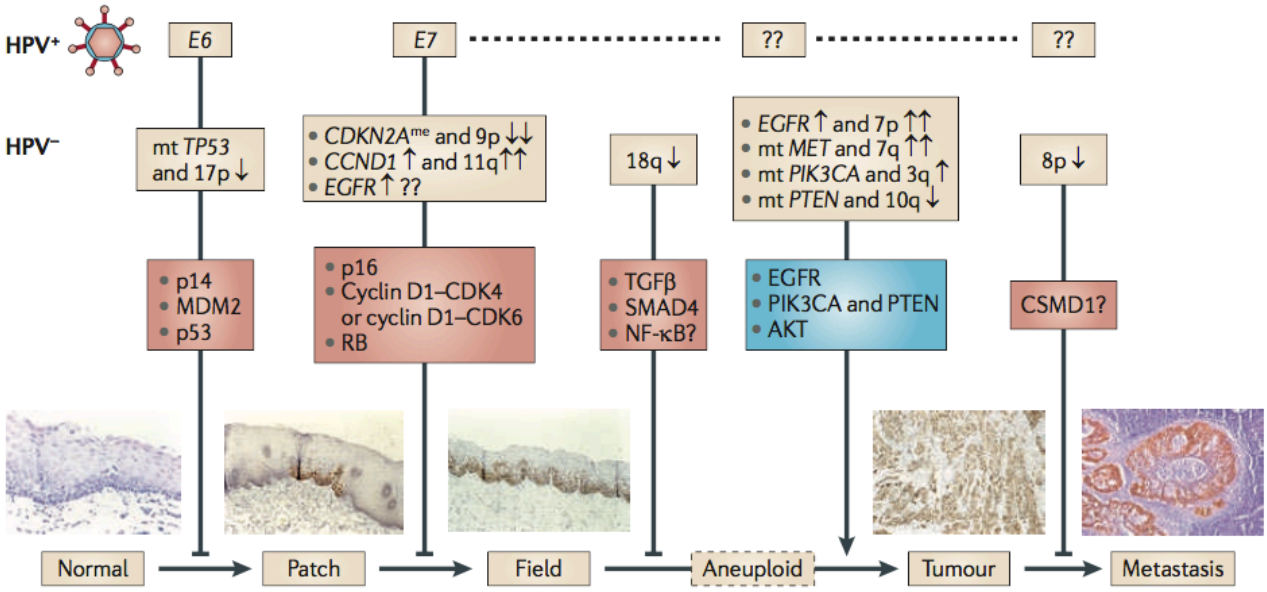

Figure 2. Schematic representation of molecular carcinogenesis of HNSCC.
A progenitor or adult cell in the normal mucosa acquire one (or more) genetic alteration, including TP53, either by mutation (HPV-negative) or by inhibition (HPV-positive). This mutated cell forms a patch containing genetically altered daughter cells, without proper growth control, that develops an expanding field. Successive and eventual transforming events or the accumulation of other mutations turn the field to an overt carcinoma, displaying invasive growth and metastasis capabilities. Figure has been adapted from "The Molecular Biology of Head and Neck Cancer", Leemans et al[9].

Recently, it has been demonstrated that in about 90% of HNSCC early lesions an overexpression of EGFR could be detected by IHC[18]. This observation, together with the notion that a positive regulatory loop exists between EGFR and Cyclin D1 transcription[19], supported the hypothesis to develop new therapeutic strategies involving the combined treatment with targeted- and standard chemo- or chemo/radio-radiotherapy[17,20]. In this context, several studies have been conducted to validate the efficacy of Cetuximab, a chimeric immunoglobulin which selectively targets an epitope of the EGFR ligand-binding domain, combined with radio or radio/chemo-therapy in advanced HNSCC[21,22]. A phase III multicenter trial randomized showed encouraging results comparing the efficacy of definitive radiotherapy and radiotherapy plus Cetuximab. In this trial, the administration of Cetuximab 400 mg/m^2 loading dose followed by 250 mg/m^2 weekly for 8 planned doses, designed to overlap the planned ~7 weeks of radiotherapy or 7 weeks of radiotherapy alone, demonstrated a significant improvement for the addition of Cetuximab[23].

Although the toxicity of Cetuximab does not overlap with those of Cisplatin, the addition of Cetuximab to standard radio/chemo-therapy did not exert expected results, possibly due to the exacerbation of side effects in advanced HNSCC patients[24].

This observation, together with the remarkably pathway alteration and genomic heterogeneity of HNSCC, implies that identifying patients that will recur could be extremely beneficial for the

management of HNSCC patients, to avoid unnecessary toxicity and improve patients' survival. To date no validated biomarkers exist to identify HNSCC patients with higher probability to develop recurrences and that, therefore, may merit a closer follow-up or different therapeutic approaches.

1.2 Epithelial To Mesenchymal Transition

Epithelial-to-Mesenchymal Transition (EMT) is a complex physiological process in which differentiated epithelial cells lose their characteristic features and acquire a mesenchymal-like phenotype. Typical phenotypic hallmarks of EMT are the loss of cell-cell junctions, loss of apical-basal polarity and acquisition of migratory and/or invasive properties. The contrary process, defined as Mesenchymal-to-Epithelial Transition (MET), is characterized by the loss of these migratory and invasive features and the acquisition of an organized cell polarity[25].

These two transitions, that recapitulate the concept of cell plasticity, and their dynamic phenotypes resemble fundamental processes during embryonic development and wound healing[26–28].

Many studies identified as "master EMT-regulators" a large range of developmental and growth factor signals (principally TGFβ, Notch1, Wnt, EGF, HGF, FGF and IGF) able to drive EMT by triggering genetic and epigenetic programs (*i.e.* post-transcriptional control, alternative splicing, protein stabilization and/or localization), which are controlled by or regulate a specific core set of transcription factors (EMT-TF) belonging to different families, including SNAI1/2, TWIST1, and ZEB1/2, among others (Figure 3)[28–30].

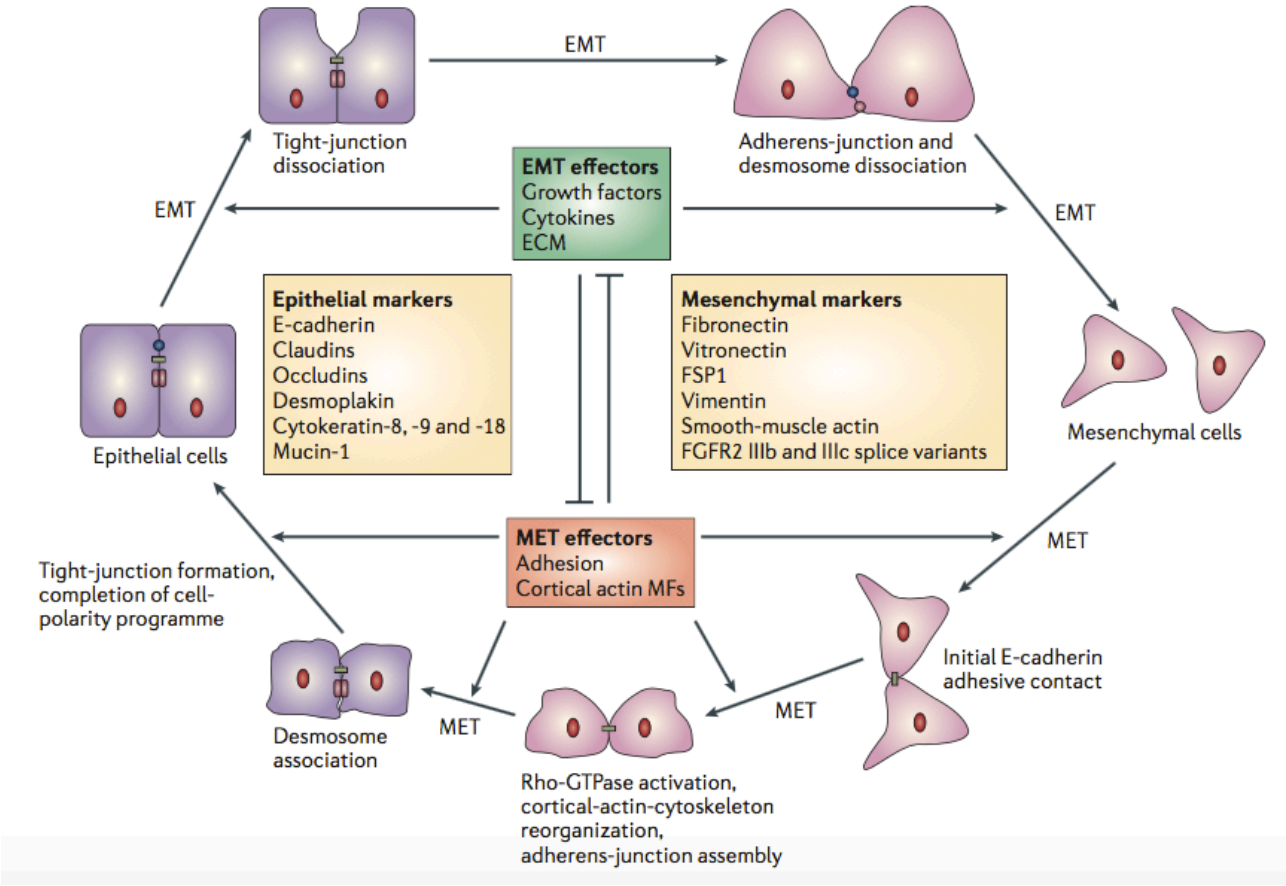

Figure 3. Schematic representation of epithelial cell plasticity.
The diagram shows the main events occurring during the switch from epithelial into mesenchymal cells. As mentioned, several events and effectors influence each other during the progression of EMT and *viceversa* MET process. Figure has been adapted from "Complex networks orchestrate epithelial-mesenchymal transitions", Thiery and Sleeman[31].

The mechanism is the result of sophisticate interactions between microenvironment and EMT-associated pathways whose orchestrate, through direct or indirect interactions in a tissue and cell-specific manner, the modulation of cell plasticity[25,32]. This in turn regulates many hallmarks of cancer, such as apoptosis, cell-cycle, metabolism and, more recently, EMT has been proposed also as an escape strategy mediating also chemo- or radio-resistance. Nevertheless, emerging evidence suggests that EMT is fundamental to gaining "stemness" feature or the transition of neoplastic cells to becoming cancer stem cells (CSCs), that as a fountainhead for tumours, could sustain the remaining population of tumour cells, and eventually contribute to treatment resistance[33–35].

For this reason, in the last decade, EMT has becoming of increased pathophysiologic interest, since cancer cells are able to reactivate the EMT program to gain new and more aggressive properties.

In the past, most studies relied on morphological alterations complemented by the detection of epithelial (*i.e.,* E-cadherin) and mesenchymal markers (*i.e.,* vimentin, fibronectin, N-cadherin) to assess an epithelial or a mesenchymal state as a simple binary decision[30,36–38]. However, more recent

evidence has advanced and broadened the definition of EMT as a program with dynamic transitional states characterized by metastable intermediates, a sort of hybrid neoplastic cells[39]. This concept is in line with the high degree of tumour cell plasticity, which has been described for most tumour entities (including HNSCC), and might be the main driver of intrinsic or acquired resistance to established treatment regimens like platinum-based chemotherapy or radiation[40].

Epithelial plasticity and mesenchymal conversion is often seen in cancer cells as they leave the primary tumour and disseminate to other parts of the body to colonize distant organs and form metastases, which is responsible for the vast majority of cancer-related deaths. However, EMT-related invasion in combination with tumour cell dissemination is initiating, but is not sufficient for completion of the metastatic cascade. For metastatic colonization, the reversal MET is necessary to support tumour cell expansion, an important prerequisite for metastatic growth. In this context, hybrid epithelial/mesenchymal cells are able to detach from the primary tumours as a cluster and maintain the migratory and invasive features to give rise to the metastatic cascade. In contrast to EMT, extracellular signals or cell intrinsic programs involved in the induction of MET have not been well characterized. It is also worth noting that tumour cell dissemination and metastatic colonization does not exclusively rely on changes in cell identity related to the EMT program and its reversal, and other scenarios are also reasonable[32,39].

In cancer research, EMT mechanism represents an important clinical challenge for the peculiar aggressiveness acquired by cancer cells, in term of relapse potential, distant dissemination and chemio/radio-resistance[33,40]. There are many studies which addressed the possibility to validated EMT-regulators as possible therapeutic targets or prognostic markers, but no satisfactory results have been obtained mainly because of the intrinsic heterogeneity of tumour cells and microenvironment. As an example, in the same tumour mass a kaleidoscopic range of different epithelial/mesenchymal phenotypes can be distinguished, not only for gene expression profile, but also for cell localization: the external neoplastic cells tend to express classical mesenchymal markers, as detectable from N-Cadherin or ZEB1, whereas the inner tumour cells express typical epithelial marker, such as E-cadherin and cells with hybrid phenotypes arrange in the middle part of the tumour mass, as result of their EMT-MET balance[41].

1.3 TGF-β pathway in HNSCC

TGF-β, as bioactive dimer, brings together TGF-β receptor I and II (TGFBR-I and TGFBR-II). On binding TGF-β, the type II receptor phosphorylates and activates TGFBR-I, that then propagate the signals by phosphorylating the SMADs 2 and 3 proteins. This activated complex recruits SMAD4 to translocate into the nucleus. The SMAD2/3-SMAD4 complex together with other adaptive proteins and cofactors, is able to bind with high affinity and specificity several selected target genes to enhance their transcription[42].

SMAD2/3/4 complex is very suitable for many interactions, this observation introduces the high pleiotropy of TGF-β functions, closely dependent also from the context (microenvironment, tissue, cell type) (Figure 4).

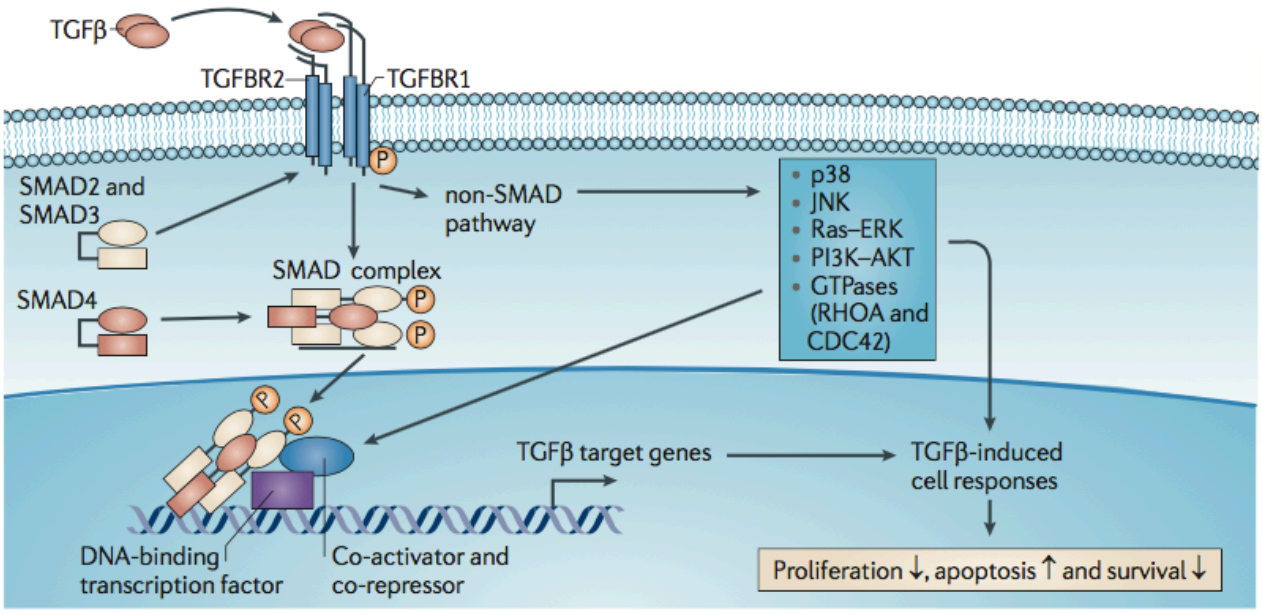

Figure 4. Depiction of TGF-β signalling pathway.
The canonical TGF-β signalling relayed to SMADs pathway seems to play the most important role in HNSCC. This observation reflects the frequent down-modulations of TGFBR-II and SMAD4 described in HNSCC that enable neoplastic cell to respond to the anti-proliferative signals activated by TGF-β cascade. Figure has been adapted from "The Molecular Biology of Head and Neck Cancer", Leemans et al[9]

TGF-β is an important paracrine cytokine that has both growth-promoting and -suppressive effects on cells. Typically, it regulates diverse cellular activities such as inhibition of mitosis, proliferation and cell survival or promotion of differentiation and apoptosis. It is deeply involved in neural development, maintenance of immune surveillance and wound healing process.

Nowadays, its controversial role is under investigation, since it is well understood that acts as a tumour-suppressor in early tumorigenesis, and becomes an enhancer of malignant phenotype in advanced phases. Deregulation of TGF-β pathway in malignancy occurs through several mechanisms,

including loss of response to its ligand, defects in the transduction pathway, and others[42–44]. In particular, HNSCC has been shown to be resistant to the suppressive effects of TGF-β, through a frequent down-modulation of TGFBR-II or SMAD4, which elicits for lethal TGF-β-driven EMT and correlates with a poor prognosis[45–48].

Intriguingly, TGF-β as primary factor triggering EMT, is able to activate, as well as other inflammatory cytokines such as TNF-α or IL-6, STAT3 signalling which converge in the activation of diverse EMT-associated transcription factor: Twist, Snail, and Slug (Snail2), that exacerbate the EMT phenotype, not only in HNSCC[29,35].

As mentioned above, although EMT is of potential clinical impact, the pharmaceutical inhibition of "master EMT regulators" presents some grave limitations: considering the TGF-β pathway in pancreatic cancer, the abrogation of its signalling limits its pro-apoptotic function and nevertheless the EMT inhibition, depending on cell type, could also revert the phenotype of hybrid tumour cells, leading to a reverse MET, able to establish an epithelial cell expansion and subsequent metastasis formation[49]. In this scenario, the major biological challenge is to defined and identify EMT-evolution markers and translate them into the clinical practice, since advanced HNSCC patients, stratified for tumour isotype, could benefit for a timely administration of EMT-targeted therapy.

As regard to locally advanced HNSCC, different studies have demonstrated a strong and frequent alteration of the TGF-β pathway and demonstrated its pivotal and controversial role in the establishment of an EMT phenotype, with dramatic consequence concerning recurrence and/or metastasis formation in HNSCC[45–47].

1.4 microRNA in HNSCC

microRNA (miRNAs) are small non-coding RNA molecules (17-25 nucleotides) that regulate post-transcriptional gene expression by relatively non-specific binding to the 3'-untranslated region (3'-UTR) of target mRNA[50]. Their biogenesis has been clarified and seems to be more complex than other genes. miRNAs sequences are located within various genomic context: the majority of canonical miRNAs are encoded by introns of non-coding or coding transcripts, and also by exon regions. Some miRNA genes reside in the introns of protein-coding genes and, thus, share the promoter of the host gene. However, it has been noted that miRNA genes often have multiple transcription start sites and that the promoter of intronic miRNAs are sometimes distinct from the promoter of their host gene[51]. The transcription of primary miRNA (pri-miRNA) is carried out by RNA-polymerase-II, and its associated co-factors, that leads to the formation of about 1 kb pri-miRNA, containing different stem-loop sequence in which the mature miRNA sequence is

embedded[52]. Following the transcription, the first step of miRNA maturation is initiated by nuclear RNA-III, Drosha. This enzyme crops the stem-loop to release a small hairpin-shaped RNA of about 65-75 nucleotides in length (pre-miRNA)[53]. Following the cleavage, pre-miRNA is exported into the cytoplasm by Exportin 5 (EXP5), where the maturation can be completed. Once in the cytosol, pre-miRNA is cleaved by Dicer, that release a small double strand RNA (dsRNA)[54]. This small RNA duplex is loaded onto AGO protein to form an effector complex, named RNA-induced silencing complex (RISC), that rapidly remove the passenger strand releasing the mature single strand miRNA[55–57] (Figure 5).

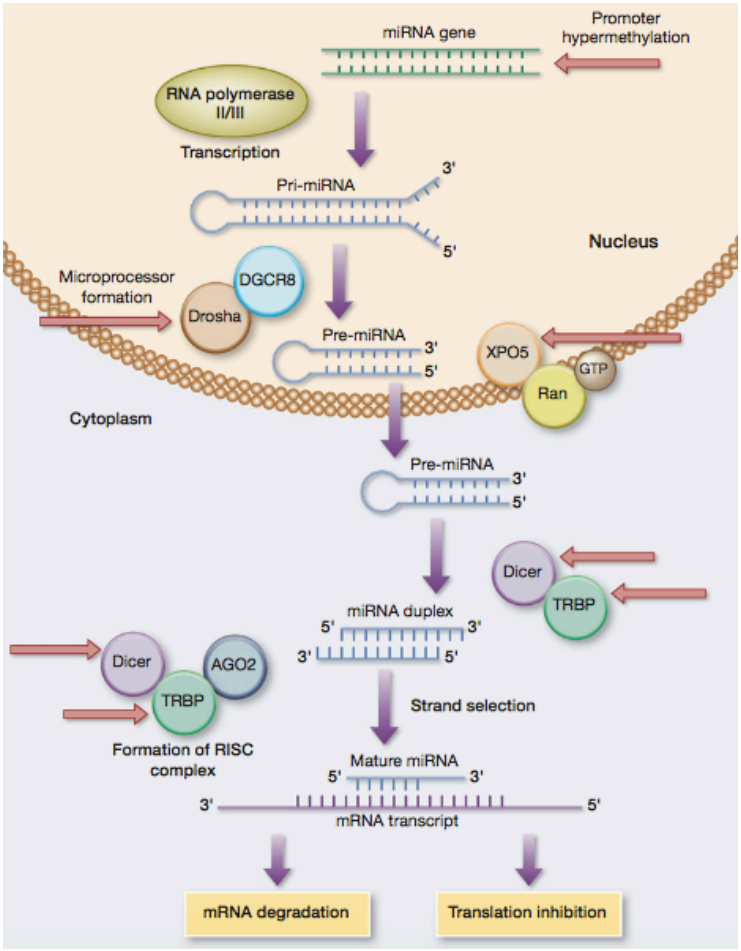

Figure 5. microRNA biogenesis pathway. Schematic representation of miRNA biogenesis with the principal effectors of transcription and maturation processes. The complex regulation of miRNA expression is perturbed in almost all biogenesis step (as indicated with the red arrows) by genetic, and more frequently, epigenetic factors. Figure has been adapted from "miRNA Dysregulation in Breast Cancer" Mulrane et al[58].

The imperfect base-pairing between a single miRNA and the target mRNA leads to a potential and dynamic regulation of several gene expression. It has been proposed that miRNAs regulate approximately 30% of human genes[59].

It has been demonstrated that miRNAs are specifically express in diverse tissue and biological fluids, and particular miRNAs pattern are characteristic of different cell type and function. Moreover, they are involved in essential biological activities and cellular processes, including the regulation of cell cycle, proliferation, differentiation, development and apoptosis[60].

Due to their pleiotropic features and the amount of essential cell activities in which they are involved, it is not surprising that an unbalanced miRNA expression could play a role in cell plasticity and cancer development.

In this context, the interesting in the exploration of miRNA expression and their role in the molecular biology of human diseases has increased in last decades. The relative easy detection of miRNAs in different tissue and biological fluids has led to intensive investigations with innovative technologies with the aim to identify specific biomarkers, useful in characterizing cancers biology and discovering novel therapeutic targets[61,62].

Since, cancer cell plasticity is regulated epigenetically, it has been hypothesized that the regulation of miRNAs could play a primary role also in HNSCC cell plasticity[63].

After the seminal demonstration that a reciprocal feedback loop exists between the miRNA-200 family and the ZEB transcription factors to tightly control EMT, the number of miRNAs that has been directly or indirectly associated with EMT is becoming an extensive list[64].

The role of miRNA expression in HNSCC has been widely investigated and several differentially expressed miRNAs in normal/peritumoural mucosa *versus* primary tumours have been identified. However, no study has specifically compared the expression profile of miRNAs in primary tumours from patients who have recurred *versus* patients that have not[65–67].

To further improve patient outcomes with more effective therapeutic approach, we investigate and identify a 4-miRNAs signature able to identify HNSCC patients at high risk of recurrence and describe the mechanism whereby they orchestrate the expression of genes regulating cancer cell plasticity *via* EMT modulation.

2. AIMS

Head and Neck Squamous Cell Carcinoma represents the sixth most common cancer worldwide, with about 550,000 new cases/year. Independently from the etiological risk factor(s), HNSCC is an unexpected heterogeneous disease, possibly due to the accumulation in genetic mutation occurring during the tumour staging.

Most patients are diagnosed with a locally advanced disease and are treated with surgery in combination with radiotherapy and/or chemotherapy. This highly toxic schedule is curative only in about half of the cases, and despite this aggressive treatment about 40-50% of patients will survive for 5 years. On one side, for advanced HNSCC patients, surgery alone is not sufficient for the whole excision of tumour mass mainly due to the invasive burden close to physiological complex structure for which the maintenance of functionality is mandatory and this allows the growth of residual tumour cells and their local or distant invasion. On the other side the combined addiction of radio- or chemotherapy is not well tolerated due to the grave side effects, and for this reason, in many cases, the effective dose could be decreased and adapted to each patient, leading to a high percentage of resistance to standard therapy.

The development of local or distant recurrences remains the major clinical challenge, since recurrent patients do not have effective salvage therapies. Therefore, there is the urgency to identify and validate solid biomarkers able to classify patients at high risk that may benefit for specific targeted approaches. This PhD project tackled these two unmet clinical needs and aimed to identify a microRNAs signature of loco-regional recurrence in patients with HNSCC.

3. MATERIALS AND METHODS

3.1 Patient samples

Specimens from primary HNSCC were collected from patients who underwent surgery at our institution and at Santa Maria degli Angeli Hospital, Pordenone, Italy. HNSCC specimens were immediately frozen and stored at -80°C. The study was approved by the Internal Review Board of the Centro di Riferimento Oncologico (CRO) of Aviano (#IRB-08/2013) and all patients provided written informed consent.

3.2 Bioinformatic analyses

Computational analysis includes two parts: A) Uni-Variant Statistical Significance Testing of miRNA-1, miRNA-9, miRNA-133a, and miRNA-150 (one miRNA at a time)[68]. B) Testing Sample Classification based on molecular signatures using models constructed from combinations of miRNAs. More specifically, we test the above 4 miRNAs together, and every combination of 2 or 3 miRNAs listed above[69,70].

3.2.1 Uni-variant Significance Test by The Permutation Test

This established test was used to calculate the statistical significance of each of the four miRNAs (miRNA-1, miRNA-9, miRNA-133a, and miRNA-150) individually. A low p-value for the miRNA indicates higher confidence in how well it separates recurrence from non-recurrence patients. For this, we used the permutation test to estimate the p-values for each miRNA. The permutation test works by first measuring the observed difference between the two patient groups (recurrence vs. non-recurrence) for the given gene, then, it randomly shuffles the samples across the two groups k times and it counts how often it encounters a difference greater or equal to the observed difference. When K is large (we used 10000000), the p-value is the ratio between how often a difference greater or equal to the observed difference is measured randomly (due to chance) out of K. The permutation test is advantageous because it is parametric, it makes no assumptions, and it can be used with small sample size.

3.2.2 Testing of Sample Classification

This second test was used to build and test a computational prediction model that predicts recurrence based on the miRNAs expressions. However, the accuracy of prediction relies on the process of classifier construction. In this process, we address challenges mentioned earlier to improve prediction accuracy. Therefore, we adopt the under-sampling technique described by W. Klement et al[70], to

counter the effects of class imbalance and the potential over-fitting. The idea is based on selecting equal portions from both patient groups used to train the classifier. For illustration, if the training data has 9 recurrence cases and 29 non-recurrence cases, we select all 9 recurrence cases, plus randomly selected (without replacement) 9 non-recurrence cases to a total of 18 cases for use in training the classifier. In other words, the under-sampling policy is to include the entire minority class cases and randomly subsample an equal proportion of cases in the majority class. This process is then repeated and the average performance is determined. The repeated training and testing demonstrates the reliability and robustness of predictions. For method of classification, we report our results using two very different methods: Naïve Bayes and Logistic Regression methods. While the former is a simple probabilistic classifier based on applying Bayes' theorem, the latter constructs a regression model for binary classification. Classifier testing policy is based on repeated 10-fold cross-validation. This strategy follows: we first partition the data randomly into 10 stratified folds. The stratification ensures that every fold i of the 10 folds has approximately the same proportions of positive (recurrence) samples and negative (non-recurrence) samples. For every fold i, we use the remaining 9 folds for training a classifier, and then, we test the resulting classifier on samples in fold i. This process is repeated for all 10 folds to produce predictions for all samples. Then, we calculate the performance metrics for evaluation (they are described below). This entire testing strategy (starting with the random data partitioning) is repeated 10000 times and the average and standard deviations are calculated. The reason for the repetitions is to eliminate the potential systematic overlap of 8 folds in testing two consecutive folds, a consequence of 10-fold cross validation strategy.

3.2.3 Network analysis

After downloading predicted miRNA–gene interactions for the genes in our network from the miRNADIP portal ver. 1 (http://ophid.utoronto.ca/mirDIP)[71], which integrates 12 microRNA prediction datasets, we kept only those interactions that were identified in at least three independent datasets. We then integrated the analysis using genes up- or down-regulated in head neck recurrences from CDIP, the Cancer Data Integration Portal ver. 1 (http://ophid.utoronto.ca/cdip), a collection of gene expression data from published studies. We also used a list of genes associated with recurrence formation in HNSCC from a published cohort[72].

Next, we uploaded this list of gene IDs into NAViGaTOR 3 as our network visualization tool (http://ophid.utoronto.ca/navigator)[73] and retrieved known, publicly available human physical protein interaction using the I2D 2.2 portal (http://ophid.utoronto.ca/i2d)[74]. Our goal was to explore the relationships between our 4 miRNAs and genes known to regulate recurrence formation in HNSCC.

Network nodes represent miRNAs and proteins respectively, while edges represent physical protein-protein interactions (PPIs) and microRNA:gene regulation. We downloaded the list of the 4 miRNAs targets from the miRNADIP database ver. 1 (http://ophid.utoronto.ca/miRNADIP/) and then we looked at the intersection among genes that control recurrence formation, genes associated to a pro-survival signature and the 4 miRNAs targets[72].

3.2.4 *PathDIP analyses*

We analyzed 56 genes, involved in recurrence formation[72], as potential miRNA-9 targets. These pathway analyses were conducted using pathDIP Ver. 2.4.3.12 (http://ophid.utoronto.ca/pathDIP)[75].

3.2.5 Analysis of the TCGA dataset

Correlation analysis using HNSCC TCGA RNAseq and microRNA-seq data for SASH1, KRT13 and hsa-miRNA-9 were performed using Spearman correlation[76]. All statistical analyses were performed using [R] (https://www.r-project.org/). KRT13/SASH1 correlation analysis using HNSCC TCGA RNAseq data were performed using cBioPortal for Cancer Genomics (http://www.cbioportal.org/)[77]. Clustering of RNAseq values was performed using Ward linkage. This approach identified cluster 1 with high expression of KRT13/SASH1 and low expression of TGFβR1/2, SMAD3 and SP1, while cluster 2 is characterized by high expression of TGFβR1/2, SMAD3, SP1 and low expression of SASH1/KRT13. The associations of the 2 clusters with survival was evaluated with the log-rank test using the survival package in R. HNSCC TCGA patients were stratified either by high mRNA expression of KRT13/SASH1 and low SMAD3 (above median of mRNA expression, below median of mRNA expression, respectively) or by high mRNA expression of SMAD3 and low mRNA expression of KRT13/SASH1. Cox proportional hazards regression analysis based on the *KMsurv* package in R was used to assess the hazard ratios. All HNSCC TCGA data were downloaded using the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/) and the cBioPortal for Cancer Genomics (http://www.cbioportal.org/)

To validate prognostic properties of the four-miRNA signature, we used SurvMicro v.0.9 (http://bioinformatica.mty.itesm.mx:8080/Biomatec/Survmicro.jsp)[78]. Signature was validated on TCGA LUAD Ilumina HiSeq dataset, comprising 311 patient samples. All settings were used as default.

3.3 Cell Biology Experiments

3.3.1 Cell culture

FaDu, Cal27, UMSCC1 and UMSCC74b cells were cultured in Dulbecco modified Eagle medium (DMEM, Sigma) supplemented with 10% fetal bovine serum (FBS, Gibco). SCC9, SCC15 and SCC25 cells were cultured in a 1:1 mixture of Dulbecco's modified Eagle's medium (Sigma) and Ham's F12 medium (Sigma) containing 1.2 g/L sodium bicarbonate (Sigma), 2.5 mM L-glutamine (Sigma), 15 mM HEPES (Sigma) and 0.5 mM sodium pyruvate (Sigma) supplemented with 400 ng/ml hydrocortisone (Sigma) and FBS 10%. All cell lines were authenticated by BMR Genomics srl Padova, Italia, according to Cell ID ^{TM} System (Promega, USA) protocol and using Genemapper ID Ver 3.2.1, to identify DNA STR profiles. UMSCC74b and UMSCC1 cells were kindly provided by Dr. Thomas Carey (University of Michigan, Ann Arbor, MI). All other head and neck squamous cell lines were obtained from ATCC (LGC Standards).

All *in vitro* studies were performed in triplicate, unless otherwise specified.

3.3.2. Lentiviral trasduction of FaDu and SCC9.

FaDu and SCC9 cells were transduced with Lentiviruses expressing anti-miRNA-9-5p (MISSION® Lenti miRNANA inhibitor human has-miRNA-9-5p, HLTUD0946, Sigma) according to the manufacturer protocol. Cells were transduced with anti-miRNA-9 (or control) lentivirus and selected in 1.0-1.5 μg/ml puromycin.

3.3.3 miRNA-mimic or inhibitor transfection.

Anti-miRNA-9 (MISSION® Synthetic miRNA inhibitor Sigma) were used to transiently knockdown miRNA-9 in FaDu and SCC9 cells. Hsa-miRNA-1, -133a, -150 (mission® miRNA mimic) were used to overexpress alone or in combination these microRNAs. miRNA inhibitor, mimic or control were transfected into different HNSCC cells using Oligofectamine Transfection Reagent (LifeTechnologies) according to manufacturer protocol.

3.3.4 Cell viability and IC50 drugs calculation.

HNSCC cells were seeded in 96-well culture plates (2-4x10^3 cells/well) and after 24 hours treated with increasing doses of the different drug (MTA, SB43, SB52, Erlotinib, Gefitinib and Cetuximab) for 72 hours. Combination treatments using MTA plus SB52 or SB43 were performed as above using MTA at the IC50 concentration, calculated for each cell line, and increasing doses of SB52 or SB43

for 72 hours of treatment. Cell viability was determined at the end of treatment using the CellTiter^{96}AQueous cell proliferation assay kit (Promega).

For CDDP treatments, FaDu or SCC9 control or stably expressing antimiR-9 cells were plated as above and treated with increasing dose of CDDP for 16 hours. Cell viability was determined as above 24 hours after the end of the treatment.

3.3.5 Luciferase assay.

Luciferase assay was used to validate miRNA-9, miRNA-1, miRNA-133a and miRNA-150 putative target sites on SP1 and SASH1 3' UTR. We amplified sequences surrounding the putative miRNA-1, -133a, -150 and -9 binding sites from FaDu cells genomic DNA using specific primers carrying restriction endonuclease sequence XbaI. PCR products were cloned in pGL3 control vector (Promega) digested with XbaI (Promega), at the 3' of the luciferase gene, which is under the regulation of SV40 promoter. Primers used to amplify SASH1 and SP1 sequences are reported below:

Primer	***Sequence 5'- 3'***	***GeneBank Accession***
3' UTR SASH1 A FW	TCTAGATGGACAAGAGCCACCCTTTC	NM_015278
3' UTR SASH1 A Rev	TCTAGAAGACAGAAGAGCAAGGGCAC	NM_015278
3' UTR SASH1 B FW	TCTAGATGGCTGCGAATGCTCTATCT	NM_015278
3' UTR SASH1 B Rev	TCTAGACTGGTCGAGCCAGATGTTCTA	NM_015278
3' UTR SASH1 C FW	TCTAGATGTTTTACAACTGATTTCAGCACA	NM_015278
3' UTR SASH1 C Rev	TCTAGATCAAGGGCTTGTGGTCAAGG	NM_015278
3' UTR SP1 A FW	TCTAGATTCTCTTCTCAGCTCTTCCATGA	NM_138473
3' UTR SP1 A Rev	TCTAGAAGTCGAAGAAGCTGATCCCAA	NM_138473
3' UTR SP1 B FW	TCTAGAGTTCTTGGGTGGTTCCTAAGGG	NM_138473
3' UTR SP1 B Rev	TCTAGAACAATGCTTTTATGGCTGGGC	NM_138473
3' UTR SP1 C FW	TCTAGATGCTGGTTTACCCTCAACCC	NM_138473
3' UTR SP1 C Rev	TCTAGACTGCCCTCGAAGTTTTGGTC	NM_138473
3' UTR SP1 D FW	TCTAGACATAGCTCTCCTTCCCCCTCA	NM_138473
3' UTR SP1 D Rev	TCTAGATTGGGAAGGCAGACAAAGAGA	NM_138473

For the luciferase assays, 293T17, CAL27 or FaDu were co-transfected with 500 ng of the reporter construct and 50 ng of pRL-TK vector (internal control) in 24-well plate using FuGENE® HD Transfection Reagent (Promega) according to manufacturer's recommendations. After transfection cell lysates were assayed for luciferase activity using the Dual-Luciferase reporter assay system (Promega). Values were normalized using Renilla luciferase.

3.3.6 Irradiation and clonogenic survival assay

Irradiations were performed using a ^{137}Cs irradiator (Nordion) at a dose rate of ~1 Gy/minute or a Clinac 600 C (Varian Medical Systems, Palo Alto, CA) linear accelerator (LINAC) for external beam radiation therapy, at ambient oxygen concentrations and in cell adhesion conditions. In the case of irradiation with the LINAC, cell plates were positioned at the center of the radiation field of 40 × 40 cm^2 size, with LINAC gantry at 180°, between two 5 cm layers of solid water. The dose delivered to cell plates was 2 or 5 Gy at a dose rate of ~ 2.5 Gy/minute, as calculated from measurements with radiochromic films in the same setup of irradiation.

For the clonogenic assay the number of cells to seed was calculated using the following formula:

N° cell = N° optimal counting colonies/plating efficiency in standard conditions/likelihood of predicted survival.

Cells were seeded in 6-well plates or 60 mm dishes (two dilutions, in triplicate) and let adhere to the plates. Cells were then irradiated and maintained at 37 °C and 5% CO_2 for 10-15 days, refreshing the medium every 3-4 days. Colonies were then fixed and stained with 1% (w/v) methylene blue in 50% ethanol or with 0.5 mg/ml crystal violet in 20% methanol. Colonies with more than approximately 50 cells were counted manually and clonogenic survival fraction (SF) was expressed as the relative plating efficiencies of the irradiated to the control cells. At least three independent experiments were performed for every condition. Where indicated, the clonogenic assays were performed without radiation treatment. Briefly, cells were seeded in 6-well plates (two dilutions, in triplicate) and let adhere. Cells were then incubated for 10–15 days with appropriate culture medium changes every 3-4 days. Past 10-15 days, growth was blocked and the colonies stained and counted, as described above.

3.3.7 Anchorage independent soft agar assay

To evaluate the anchorage-independent cell growth, $1.5x10^4$ FaDu or SCC9 control or stably transduced with antimiR-9 were re-suspended in 2 ml top agar medium (appropriate cell medium + 10% FBS, 0.4% Low Melting Agarose, SIGMA) and quickly overlaid on a previously gelified 0.6% bottom agar medium (appropriate cell medium + 10% FBS, 0.6% Low Melting Agarose, SIGMA). The experiments were performed in six-well tissue culture plates, in triplicate. Fresh medium was added to the wells twice a week as a feeder layer. After three weeks, the number of colonies was counted in 10 randomly chosen fields, at 10X magnification.

3.3.8 Reagents

Mithramycin A (MTA), an aureolic acid compound that specifically inhibits SP1 activity, SB-525334 (SB52) and SB431542 (SB43), two potent and selective inhibitors of TGFβR1. All compounds were purchased from Sigma and used for *in vitro* and/or *in vivo* experiments.

Cisplatin (CDDP) was purchased from TEVA *Italia*.

EGFR inhibitors, Gefitinib and Erlotinib, were purchase from Selleck Chemicals, whereas Cetuximab (Erbitux) from Merck Serono.

3.4 Molecular Biology Experiments

3.4.1 DNA extraction and evaluation of TP53 mutational status and HPV DNA status

Total DNA was isolated from patient-derived primary tumors using QIAamp DNA Mini Kit (QIAGEN, Germany) according to manufacturer protocol. Total DNA was quantified using QuantiFluor® dsDNA System (Promega).

HPV genotyping was determined in all samples using the INNO-LiPA HPV Genotyping Extra assay (Innogenetics, Belgium) and following the manufacturer's instructions. The INNO-LiPA HPV assay is a PCR-based hybridization assay that includes a cocktail of biotinylated consensus primers (SPF 10) to amplify a 65 bp region within the L1 ORF (open reading frame) of multiple HPV types. After DNA extraction, all specimens were subjected to PCR amplification (40 cycles) using the Inno-LiPA HPV Genotyping Extra Amp. The PCR product was then denatured, and a 10-µl aliquot was hybridized onto nitrocellulose strips where the HPV type-specific oligonucleotides were already bound. After 60 min at 49°C, the PCR product bound to a specific probe was detected by an alkaline phosphatase-streptavidin conjugate and colorimetric detection.

TP53 mutational status was assessed by next generation sequencing (NGS) with an amplicon-based strategy to cover with at least a 1000X coverage and 5% of sensitivity the whole *TP53* sequence.

TP53 NGS primer list is available upon request. PCR products were generated using a high fidelity Taq polymerase (Phusion High-Fidelity DNA Polymerase; Thermo Scientific, Waltham, MA) and subjected to NGS on a MiSeq sequencer (Illumina, San Diego, CA). Data were analyzed with MiSeq reporter (Illumina) and IGV software against human genome assembly hg19. Results were expressed as percentage of mutated DNA.

3.4.2 RNA extraction and qRT-PCR analyses.

Total RNA for miRNA microarray and qRT-PCR analyses was isolated from patient-derived primary tumors or cell cultures using Trizol solution (Roche Applied Science Mannheim, Germany) according to manufacturer protocol. Total RNA was quantified using NanoDrop (Thermo Fisher Scientific Inc., USA). MiRNAs expression profile was performed using the Nanostring™ technology (NanoString nCounter Human miRNA assay (v1.1) that allowed to evaluate the expression of 746 miRNAs (664 Human 82 Viral) along with the one of housekeeping genes (GeneBank GSE89000).

Differentially expressed miRNAs were validated using the TaqMan single tube MicroRNA Assays. All reagents, primers and probes were obtained from Applied Biosystems and Reverse Transcriptase (RT) reactions and qRT-PCR were performed according to the manufacturer instructions (Applied Biosystems, Life Technologies). Normalization of the validation set was performed on the U6 RNA. All RT reactions were run in an Opticon qRT-PCR Thermocycler (Bio-Rad). Comparative qRT-PCR was performed in triplicate, including no-template controls. miRNA levels were quantified using the MyiQ2 (Bio-Rad). Relative expression was calculated using the comparative Ct method.

For gene expression analysis, RNA was retro-transcribed with AMV Reverse Transcriptase to obtain cDNAs, according to provider's instruction (Promega). Absolute quantification of targets was evaluated by qRT-PCR, using SYBR Green dye-containing reaction buffer (Real SG Master Mix 5x, Experteam). Standard curves (10-fold dilution from 10^{1} to 10^{-4} attomoles) were prepared for all the analyzed genes (TGFβR1, TGFβR2, TGFβR3, SASH1, SP1, KRT13, β-catenin, WNT4, WNT5a, JUP, FLG, HPV16 E6 and E7) and housekeeping genes (GAPDH, SDHA and POL2a). All the primers for gene expression analyses were purchased from Sigma-Aldrich and the table below reports all the primer sequences with the corresponding Gene Bank accession ID. Incorporation of the SYBR Green dye into the PCR products was monitored in real time PCR, using the MyiQ2 Two Color Real-time PCR Detection System (Bio-Rad). Ct values were converted into attomoles and normalized expression was evaluated by using GAPDH, SDHA or POL2a as housekeeping genes.

Primer	*Sequence 5' – 3'*	*Gene Bank Accession*
Human TGFβR1 For	CATTTTTCCCAAGTGCCAGT	NM_001306210.1
Human TGFβR1 Rev	ACACCCCTAAGCATGTGGAG	"
Human TGFβR2 For	CCATGTCTCACAGCCAGCTA	NM_001024847.2
Human TGFβR2 Rev	CCAGGAGAAATAAGGGCACA	"
Human TGFβR3 For	CCAAGATGAATGGCACACAC	NM_003243.4
Human TGFβR3 Rev	CCATCTGGCCAACCACTACT	"
Human SASH1 For	CTGTCACCCCCTCAGTGTTT	NM_015278.3
Human SASH1 Rev	GAACAGGGTGGAGTCCGTTA	"
Human Sp1 For	GGTGCCTTTTCACAGGCTC	NM_138473.2
Human Sp1 Rev	CATTGGGTGACTCAATTCTGCT	"
Human KRT13 For	GTCTTCAGCACCCAGAGGAG	NM_153490.2
Human KRT13 Rev	TTCGAGAAAGGCAGGAAACT	"
Human β-catenin For	GAAACGGCTTTCAGTTGAGC	NM_001904.3
Human β-catenin Rev	CTGGCCATATCCACCAGAGT	"
Human WNT4 For	GCTGTGACAGGACAGTCGAT	NM_030761.4
Human WNT4 Rev	GCCTCATTGTTGTGGAGGTT	"
Human WNT5A For	GGGTGGGAACCAAGAAAAAT	NM_003392.4
Human WNT5A Rev	TGGAACCTACCCATCCCATA	"
Human JUP For	GAAAAGCTGCTCTGGACCAC	NM_002230.2
Human JUP Rev	GACGTTGACGTCATCCACAC	"
Human FLG For	GGCAAATCCTGAAGAATCCA	NM_002016.1
Human FLG Rev	TGCTTTCTGTGCTTGTGTCC	"
Human GAPDH For	GAAGGTGAAGGTCGGAGTC	NM_001289746.1
Human GAPDH Rev	GAAGATGGTGATGGGATTTC	"
Human SDHA For	AGAAGCCCTTTGAGGAGCA	NM_004168.3
Human SDHA Rev	CGATTACGGGTCTATATTC	"
Human POL2a For	GCAAATTCACCAAGAGAGACG	NM_000937.4
Human POL2a Rev	CACGTCGACAGGAACATCAG	"
HPV16 E6 1 For	ACAAACCGTTGTGTGATTTGTT	
HPV16 E6 1 Rev	CAGTGGCTTTTGACAGTTAATACA	
HPV16 E6 2 For	GCAAAGACATCTGGACAAAAAG	
HPV16 E6 2 Rev	ACCGACCCCTTATATTATGGAATC	

HPV16 E7 1 For	GAACCGGACAGAGCCCATTA	
HPV16 E7 1 Rev	ACACTTCGAACAAAAGGTTACA	
HPV16 E7 2 For	TAACCTTTTGTTCGAAGTGTGA	
HPV16 E7 2 Rev	TTTGTACGCACAACCGAAGC	

3.4.3 Protein extraction and Western Blot Analyses

For cellular protein lysates, cells were scraped on ice using cold Ripa lysis buffer (150 nM NaCl, 50 mM Tris HCl pH 8, 1% Igepal, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with a protease inhibitor cocktail (CompleteTM, Roche), 1 mM Na_3VO_4 (Sigma), 100 mM NaF (Sigma) and 1 mM DTT (Sigma). Proteins were separated in 4-20% SDS-PAGE (Criterion Precast Gel, Bio-Rad) and transferred to nitrocellulose membranes (GE Healthcare). Membranes were blocked with 5% dried milk in TBS-0·1% Tween 20 or in Odyssey Blocking Buffer (LI-COR, Biosciences) and incubated at 4°C overnight with primary antibodies.

Primary antibodies were from Cell Signaling (rabbit polyclonal Claudin-1 #4933, rabbit polyclonal SMAD2/3 #5678, rabbit monoclonal pS465-7 SMAD2 #3108, rabbit monoclonal β-Actin #8457), Abcam (mouse monoclonal SP1 #ab51513), Bethyl (rabbit polyclonal SASH1 #A302-265), Sigma (rabbit polyclonal KRT13 #SAB2104755, mouse monoclonal β-Tubulin #T8203), Transduction Lab (mouse monoclonal E-cadherin #610404), Santa Cruz (mouse monoclonal PARP-1 #sc8007, goat polyclonal Vinculin N19 #sc7649), Millipore (mouse monoclonal GRB2 MABS89, mouse monoclonal γH2AX JBW301 05-636). Membranes were washed in TBS-0·1% Tween20 and incubated 1 hour at RT with IR-conjugated (AlexaFluor680, Invitrogen or IRDye 800, Rockland) secondary antibodies for infrared detection (Odyssey Infrared Detection System, LI- COR) or with the appropriate horseradish peroxidase-conjugated secondary antibodies (GE Healthcare) for ECL detection (Clarity Western ECL Substrate, BioRad). Band quantification was performed using the Odyssey v1.2 software (LI- COR) or the QuantiONE software (Bio-Rad Laboratories). The Re-Blot Plus Strong Solution (Millipore) was used to strip the membranes, when reblotting was needed.

3.5 Xenograft growth in mouse flanks and treatments

Primary tumors were established by subcutaneous injection of $2x10^6$ FaDu cells bilaterally in the flanks of female athymic nude mice (Harlan, 6 weeks old). Growth of primary tumors was monitored by measuring tumor width (W) and length (L) and calculating tumor volume based on the formula ($W^2xL/2$). When tumors reached a volume of 1000-1200 mm^3, mice pre-anesthetized underwent surgery to remove the primary tumors. To evaluate the effect of SB-525334 (Sigma) and Mithramycin A (Sigma) on the formation of recurrences, the day before surgery, the animals were randomly divided into groups according to experimental design (3 mice/control group, 4 mice/treatment). The administration of the drugs was performed following a "peri-operative schedule", i.e. mice received the treatments the day before, the same day and the day after surgery. Vehicle or Mithramycin A 1 mg/kg, resuspended in DMSO, were administered intraperitoneally for three consecutive days, following the "peri-operative schedule". Vehicle or SB525334, resuspended in PBS, were administered 15 mg/kg with oral gavages following the "peri-operative schedule", and also three and six days after surgery. The appearance of local recurrence was monitored by macroscopic examination of mice over a period of 8 weeks. Unless tumor burden was incompatible with the well-being of the animals, mice were sacrificed at the end of the experiment and pathologically examined for the presence of microscopic recurrences. *In vivo* experiment was carried out in strict accordance with the recommendations contained in the Guide for the Care and Use of Laboratory Animals of the CRO Aviano, National Cancer Institute, Italy. The project was reviewed and approved by the CRO Institutional Animal Care and Use Committee and conducted according to that committee's guidelines. All mice were monitored twice a week and euthanized when required in accordance to the "AVMA guidelines on Euthanasia". All efforts were made to minimize suffering.

3.6 Statistical analyses

All graphs and statistical analyses were performed using PRISM (version 6, GraphPad, Inc.) and R, SAS Software 9.2 and R for statistical analyses. In all experiments, differences were considered significant when p was <0.05. Statistical analyses included paired and un-paired t-tests, Mann-Whitney un-paired t-test and Spearman correlation test, used as appropriate and as specified in each figure. Differences in miRNA expression between patients' groups were evaluated by non-parametric Wilcoxon test (two groups) or Kruskal-Wallis test (three groups). Correlation between Array and PCR quantification of miRNAs was evaluated through Spearman correlation coefficient.

4. RESULTS

4.1 Identification of miRNAs differentially expressed in recurrent and non-recurrent HNSCC.

Primary HNSCC fresh frozen surgical samples collected from 44 patients (Table 1) who experienced (n=11) or not (n=33) local recurrence within 2 years from the first surgery were analyzed by Nanostring™ technology for the expression of 746 human and 82 viral miRNAs.

	Relapsed		Not relapsed		miR-9	miR-1	miR-133a	miR-150
	n	(%)	n	(%)				
Sex								
Men	11	(100.0)	24	(72.7)	319.0	2193.8	396.7	537.7
Women	0	(0.0)	9	(27.3)	180.9	186.5	132.0	1752.8
		p=0.0849			*p=0.6812*	*p=0.4459*	*p=0.9266*	***p=0.0358***
Age								
<60 years	8	(72.7)	15	(46.9)	310.8	512.3	86.2	542.4
≥60 years	3	(27.3)	17	(53.1)	288.2	3777.7	721.0	896.1
		p=0.1753			*p=0.5217*	*p=0.6463*	*p=0.6575*	*p=0.4025*
Cancer site								
Tongue/Oral cavity	5	(45.5)	23	(69.7)	149.9	724.8	110.0	684.6
Oro/Hypopharynx	5	(45.5)	6	(18.2)	329.9	3425.2	662.3	499.5
Larynx	1	(9.0)	4	(12.1)	648.3	78.7	18.4	1452.2
		p=0.2104			*p=0.3178*	*p=0.3296*	*p=0.5829*	*p=0.2347*
cT								
T1-T2	3	(27.3)	21	(65.6)	209.0	941.5	169.3	584.8
T3-T4	8	(72.7)	11	(34.7)	341.2	2365.4	446.6	744.9
		p=0.0383			*p=0.9737*	***p=0.0409***	*p=0.1021*	*p=0.9211*
cN								
0	3	(27.3)	13	(40.6)	321.0	5360.7	1024.5	626.5
1-2	8	(72.7)	19	(59.4)	292.2	432.0	72.5	726.6
		p=0.4942			*p=0.7667*	*p=0.4503*	*p=0.4409*	*p=0.7168*
Adjuvant radio/chemotherapy								
No	6	(54.6)	26	(78.8)	160.0	638.9	126.2	489.1
Yes	5	(45.4)	7	(21.2)	391.6	2763.3	513.9	829.3
		p=0.1387			*p=0.3709*	*p=0.9752*	*p=0.4672*	*p=0.4043*

Table 1. Distribution of relapsed or not relapsed HNSCC patients of the discovery cohort.
Table reports the mean level of miRNA expression according to selected covariates. The significant differences are shown in bold. Abbreviations: cN, clinical evaluation of node status; cT, clinical evaluation of tumour size.

Statistical analyses of normalized miRNA expression demonstrated that 7 miRNAs were significantly different between the two groups (Table 2) of the Discovery Cohort. Validation of these data by qRT-PCR analyses confirmed significant differences for miR-1, miR-133a, miR-150 and miR-9 between recurrent and non-recurrent HNSCC.

miRNA		Chip		Real-time PCR		Correlation between chip and qRT-PCR*
		Fold change	Wilcoxon	Fold change	Wilcoxon	
	Overexpressed					
11	hsa-miR-9	3.239	P=0.0209	1.742	P=0.0136	0.734
	Underexpressed					
9	hsa-miR-1	0.048	P=0.0281	0.031	P=0.0126	0.736
79	hsa-miR-133a	0.077	P=0.0183	0.024	P=0.0073	0.894
91	hsa-miR-142-3p	0.399	P=0.0041	0.435	P=0.1669	0.082
104	hsa-miR-150	0.254	P=0.0034	0.222	P=0.0044	0.566
275	hsa-miR-492	0.258	P=0.0377	4.677	P=0.4440	0.081
329	hsa-miR-520g	0.166	P=0.0048	0.783	P=0.6841	0.423

Table 2. qRT-PCR validation of miRNAs differentially expressed in HNSCC patients. The table reports qRT-PCR validation data of the differentially miRNAs expression among HNSCC who underwent selective surgery that experienced recurrence respect to those that had not. * Speraman correlation: satisfactory value >0.5.

In particular, univariant significance testing confirmed that miRNA-9 was upregulated, while miR-1, miR-133a and miR-150 were downregulated in tumours from recurrent patients (Fig. 1). No significant associations were found between expression of these miRNAs and other clinical and biological variables of the tumours, including the presence of TP53 mutation and the positivity for HPV infection (Table 1 and 3). Accordingly, miR-9 upregulation and miR-1, miR-133a and miR-150 downregulation in recurrent tumours was confirmed also when only TP53 mutant or HPV-negative cases were considered (Fig. 1). Bioinformatic validation by data reiteration confirmed the significance of this interaction (Fig. 2). Using the classifier testing, we calculated the Area Under the Curve (AUC) to estimate the ability of the four miRNAs, each one alone or in combination, to predict recurrence. Models 1-4, built using a Naïve Bayes or a Logistic Regression models, in which the classifier combines miR-133a with miR-150, with or without miRNA-9, achieved a high AUC (80-81%), high sensitivity (82-88%) and low false positive rates (29-35%), which translated into 65-71% specificity (Fig. 2). Our analyses also suggested that the addition of miR-1 does little to improve the classification accuracy (Fig. 3A).

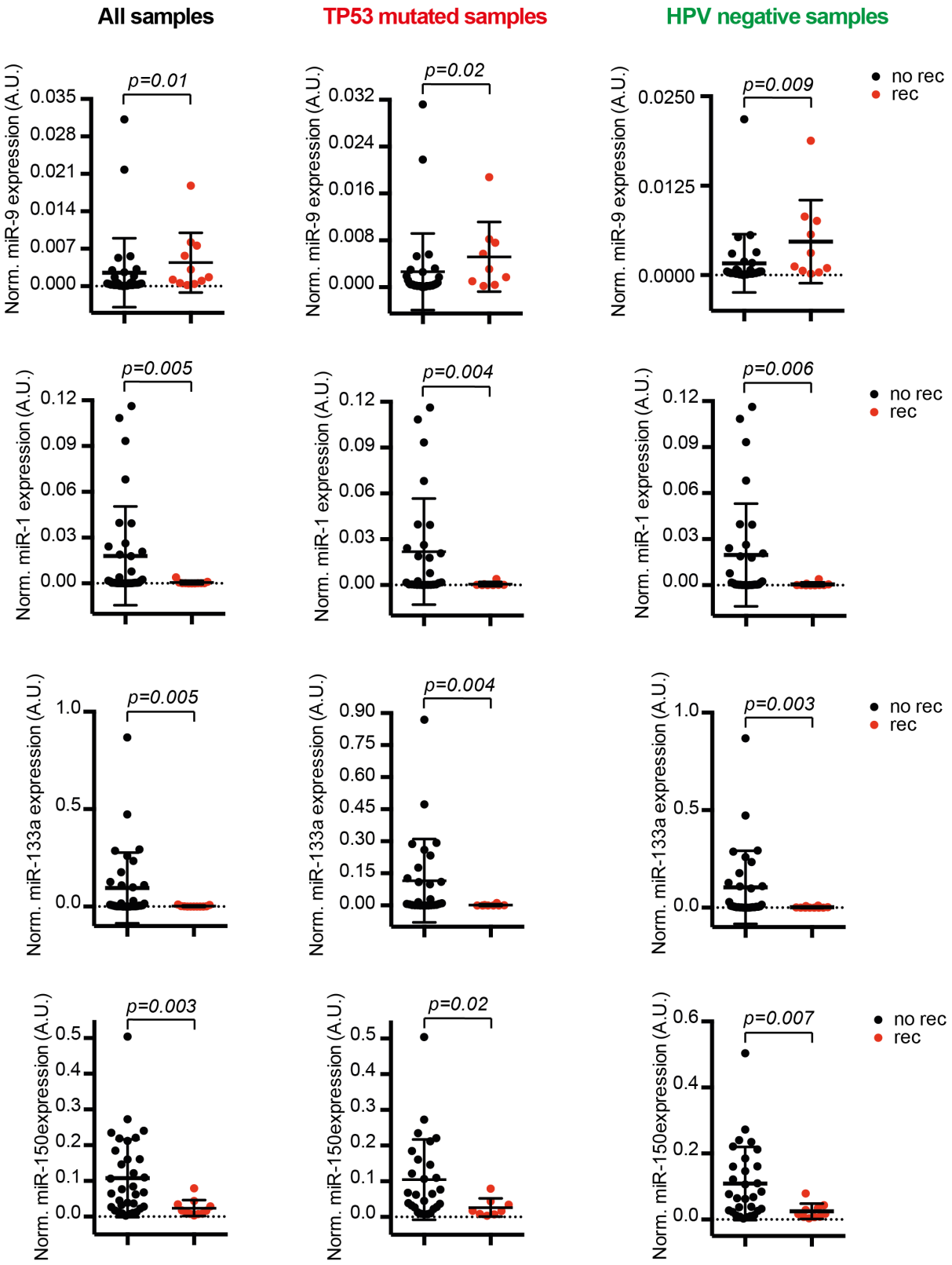

Figure 1. Validation of differentially expressed miRNAs in the Discovery Dataset.
Dot plots show the normalized expression of miR-1, -9, -133a and -150 in samples from primary HNSCC from recurrent (red) and non-recurrent patients (black). Bars indicate the mean and the 95% confidence interval. In the left column dot plots show the normalized miRNA expression of all 44 HNSCC samples. In the central column only the TP53 mutated samples and in left column only the HPV-negative samples were included in the analyses. Statistical significance is reported in each graph and was calculated with Mann-Whitney test.

ID	SEX	AGE	CANCER SITE	TREATMENT	HISTOLOGY	cT	cN	cM	pT	pN	GRADING	RECURRENCE	TP53	TYPE of TP53 mutation	HPV (DNA)	HPV (RNA)
1	M	43	tongue	Su+RT	SCC	3	2A	N.A.	N.A.	N.A.	G3	YES	MUT	Deletion	16	NEG
2	M	N.A.	tongue	Su+RT/CT	SCC	2	1	M0	2	0	G2	NO	MUT	Missense	NEG	NEG
4	F	50	oropharynx	Su+RT	SCC	2	1	M0	2	1	G3	NO	MUT	Missense	16	NEG
7	M	50	tongue	Su+RT	SCC	2	0	M0	2	1	G2	NO	MUT	Splice donor variant	NEG	NEG
8	F	69	tongue	Su	SCC	2	0	N.A.	1	0	G3	NO	MUT	Missense	6	NEG
15	M	82	hypopharynx	Su+RT	SCC	4	1	N.A.	2A	2B	G3	YES	MUT	Deletion + Missense	16	NEG
18	F	66	oropharynx	Su+RT/CT	SCC	3	1	M0	3	2B	G3	NO	MUT	Nonsense	16	NEG
19	M	60	hypopharynx	Su+RT	UndC	4A	0	M0	4	0	G4	YES	MUT	Missense	16	NEG
20	M	46	oral cavity	Su+RT/CT	SCC	2	1	M0	2	2B	G3	NO	MUT	Nonsense	NEG	NEG
21	M	59	tongue	Su+RT/CT	SCC	1	2B	M0	1	2B	G2	NO	WT	WT	16	E6 E7+
22	F	63	oral cavity	Su+RT/CT	SCC	2	0	N.A.	4A	0	G3	NO	MUT	Insertion	NEG	NEG
23	M	60	oropharynx	Su	SCC	2	0	N.A.	2	0	G2	NO	MUT	Missense	16	NEG
24	M	49	tongue	Su+RT/CT	SCC	1	1	M0	1	2B	G2	NO	MUT	Missense + Splice accep var	NEG	NEG
25	M	50	larynx	Su+RT/CT	SCC	4	2B	M0	4A	2B	G3	YES	MUT	Deletion	16	NEG
26	M	61	tongue	Su+RT/CT	SCC	3	2B	N.A.	3	2B	G2	NO	MUT	Insertion	18	NEG
31	M	55	oropharynx	Su+RT	SCC	2	2B	M0	2	2B	G2	NO	MUT	Deletion	16	NEG
32	M	70	tongue	Su+RT	SCC	2	0	M0	2	2B	G3	YES	N.A.	N.A.	N.A.	NEG
33	F	58	oropharynx	Su	SCC	1	0	M0	1	0	G2	NO	MUT	Missense	16	E6 E7+
34	M	49	hypopharynx	Su+RT/CT	SCC	4	0	Mx	4A	2C	G1	YES	MUT	Missense	16	NEG
36	M	73	tongue	Su+RT	SCC	2	1	M0	1	0	G2	NO	MUT	Nonsense	NEG	NEG
37	F	80	oral cavity	Su+RT	SCC	4	1	M0	4A	1	G2	NO	MUT	Missense	NEG	NEG
38	M	67	tongue	Su	SCC	2	1	N.A.	2	0	G2	NO	MUT	Splice acceptor variant	NEG	NEG
40	M	56	larynx	Su	SCC	3	0	M0	2	0	N.A.	NO	MUT	Deletion + Missense	NEG	NEG
42	M	55	oral cavity	Su+RT	SCC	4	2B	M0	2	2B	G2	NO	MUT	Splice acceptor variant	16	NEG
44	M	69	tongue	Su	SCC	3	0	N.A.	2	0	G2	NO	MUT	Missense	NEG	NEG
45	M	58	larynx	Su	SCC	2	2C	M0	1	0	G1	NO	MUT	Missense	NEG	NEG
46	M	65	oral cavity	Su+RT	SCC	3	1	M0	4A	0	G2	NO	WT	WT	NEG	NEG
49	M	53	tongue	Su+RT/CT	SCC	4	2B	M0	3	2C	G3	YES	MUT	Missense	NEG	NEG
50	M	56	tongue	Su+RT/CT	N.A.	2	1	M0	2	1	G3	YES	WT	WT	16	E6 E7+
52	F	53	tongue	Su	SCC	1	0	M0	1	0	G2	NO	MUT	Missense	16	NEG
55	F	52	tongue	Su+RT	SCC	2	1	M0	2	2A	G2	NO	WT	WT	16	E6 E7+
56	M	53	oropharynx	Su+RT	SCC	3	2A	M0	3	2	G2	YES	MUT	Deletion	NEG	NEG
61	M	67	tongue	Su+RT	SCC	3	1	Mx	2	2B	G3	NO	MUT	Deletion	NEG	NEG
62	M	62	oral cavity	Su+RT/CT	SCC	3	1	M0	3	1	G3	NO	MUT	Splice acceptor variant	16	NEG
63	M	57	oropharynx	Su+RT/CT	SCC	2	1	M0	2	2B	G2	YES	MUT	Deletion + Missense	16	NEG
65	M	47	tongue	Su+RT	SCC	1	2C	N.A.	1	0	N.A.	NO	WT	WT	16	NEG
66	M	59	oral cavity	Su+RT	SCC	4	1	M0	2	0	G2	YES	MUT	Missense	NEG	NEG
69	M	62	larynx	Su	SCC	1	0	N.A.	1	2	G2	NO	MUT	Missense	16	NEG
72	M	50	oral cavity	Su	SCC	4	0	M0	2	0	G2	NO	MUT	Missense	NEG	NEG
73	F	75	oral cavity	Su+RT/CT	SCC	4	0	N.A.	2	0	G2	NO	MUT	Nonsense	16	NEG
74	M	70	larynx	Su	SCC	2	0	N.A.	2	0	G2	NO	WT	WT	NEG	NEG
75	M	73	oropharynx	Su	SCC	2	0	Mx	2	0	G1	NO	MUT	Deletion	NEG	NEG
76	M	54	tongue	Su	SCC	N.A.	N.A.	N.A.	1	0	G2	NO	N.A.	N.A.	N.A.	N.A.
78	M	65	tongue	Su	SCC	2	1	M0	1	1	G2	NO	N.A.	N.A.	N.A.	N.A.

Legend to Table 3

N.A. = Not Available
Su = Primary Surgery with Curative Intent
RT = Radiotherapy
CT = Chemotherapy
CT = Chemotherapy
SCC = Squamous Cell Carcinoma
UndC = Undifferentiated Carcinoma
cT = Clinical Tumor Size
cN = Clinical Node Status
pT = Pathological Tumor Size
pN = Pathological Node Status

Table 3. Histological and clinical features of HNSCC patients included in the Discovery Set.
Table reports all the different covariates analysed in HNSCC patients of the Discovery Cohort, including TP53 mutations and HPV status.

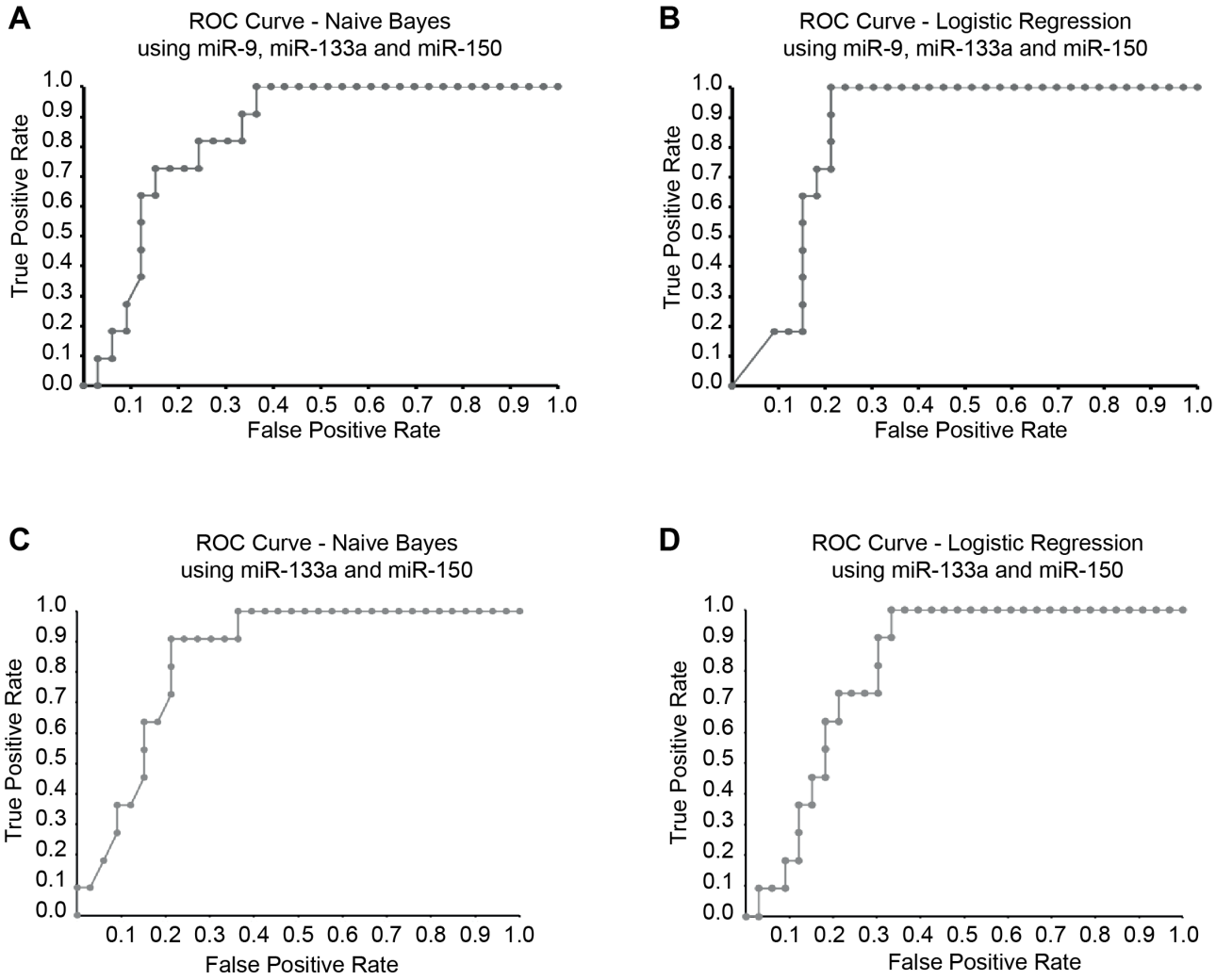

Figure 2. Design of a recurrence prediction model in HNSCC based on miRNA expression using a bioinformatic approach.
A/B. ROC curve prediction recurrence formation using miRNA-9, -133a and -150 data iteration applying the Naïve Bayes (**A**) or Logistic Regression (**B**) models. The AUC is 81.3% (Sensitivity 878% and Specificity 75%) in **A** and 80.3% (Sensitivity 82% and Specificity 71%) in **B**. **C/D.** ROC curve prediction recurrences formation using miRNA-133a and -150 data iteration applying the Naïve Bayes (**C**) or Logistic Regression (**D**) model. The AUC is 81% (Sensitivity 88% and Specificity 65%) in **C** and 81% (Sensitivity 83% and Specificity 69%) in **D**.

These computational analyses are in accord with the notion that miR-1 and miR-133a belongs to the same cluster[79] and, consequently, their expression highly correlates, as we observed in our samples set (Spearman correlation value r=0.9621 and p<0.0001). Correlation analyses also indicated that miRNA-133a expression directly correlates with miR-150 (Spearman correlation value r=0.2984 and p=0.049) and that miR-9 expression inversely correlates with both miRNA-133a (Spearman correlation value r=-0.1715, p=n.s.) and miR-150 (Spearman correlation value r=-0.4757 and p=0.0011).

Accordingly, using the HNSCC TCGA dataset[76] we confirmed that i) miR-1 and 133a expression strongly correlate (R= 0.79; p<0.0001); ii) miR-9 inversely correlates with both miR-1 (R= -0.27;

p<0.0001) and miR-133a (R= -0.34; p<0.0001) and iii) the expression of miR-1, -133a, -150 and -9 classify patients at high risk of relapse (Fig. 3B).

A

Classification		Signature				TPR	FPR	AUC
	Model	miR-1	miR-9	miR-133a	miR-150	(Thr = 0.5)	(Thr = 0.5)	(All Thr)
#1	**Naïve Bayes**	-	x	x	x	0.87±0.07	0.35±0.03	**0.81±0.04**
#2	**Naïve Bayes**	-	-	x	x	0.88±0.04	0.35±0.02	**0.81±0.03**
#3	**Logistic Regression**	-	-	x	x	0.83±0.07	0.31±0.03	**0.81±0.03**
#4	**Logistic Regression**	-	x	x	x	0.82±0.10	0.29±0.04	**0.80±0.04**
#5	Naïve Bayes	x	x	x	x	0.87±0.08	0.38±0.03	0.80±0.04
#6	Logistic Regression	x	x	-	x	0.80±0.10	0.29±0.05	0.79±0.05
#7	Naïve Bayes	x	-	x	x	0.89±0.06	0.38±0.02	0.79±0.03
#8	Naïve Bayes	x	-	-	x	0.82±0.01	0.37±0.03	0.79±0.03
#9	Naïve Bayes	x	x	-	x	0.83±0.08	0.37±0.03	0.79±0.04
#10	Logistic Regression	x	-	-	-	0.83±0.04	0.32±0.04	0.79±0.04
#11	Logistic Regression	x	x	x	x	0.76±0.1	0.32±0.05	0.76±0.05
#12	Logistic Regression	x	-	x	x	0,76±0.06	0.34±0.05	0.76±0.05
#13	Logistic Regression	-	x	x	-	0.77±0.10	0.40±0.06	0.74±0.05
#14	Logistic Regression	-	x	-	x	0.77±0.10	0.37±0.04	0.74±0.04
#15	Naïve Bayes	-	x	-	x	0.77±0.10	0.40±0.05	0.73±0.05
#16	Logistic Regression	x	x	-	-	0.75±0.10	0.41±0.07	0.73±0.05
#17	Naïve Bayes	-	x	x	-	0.92±0.07	0.61±0.04	0.71±0.06
#18	Logistic Regression	x	x	x	-	0.72±0.10	0.40±0.06	0.71±0.06
#19	Logistic Regression	x	-	x	-	0.78±0.06	0.46±0.05	0.70±0.04
#20	Naïve Bayes	x	x	x	-	0.89±0.07	0.61±0.03	0.69±0.06
#21	Naïve Bayes	x	x	-	-	0.88±0.07	0.63±0.04	0.68±0.07
#22	Naïve Bayes	x	-	x	-	0.91±0.02	0.61±0.03	0.65±0.03

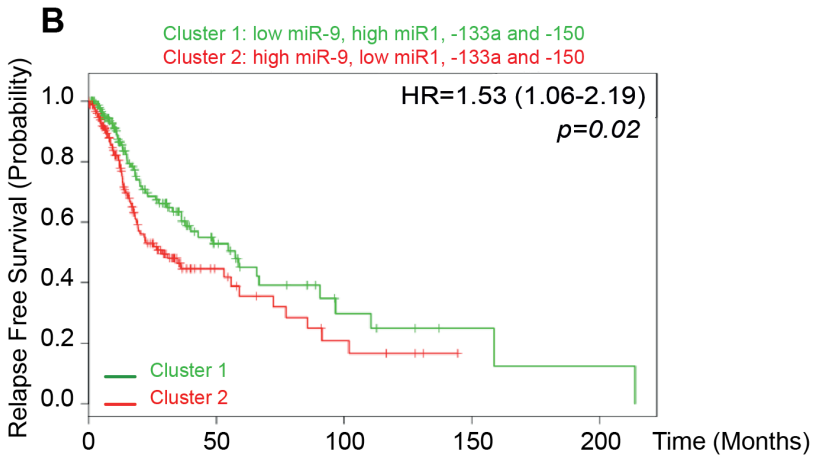

Figure 3. Differentially miRNA expressions stratified HNSCC TCGA dataset with diverse prognosis. **A.** Table reporting the predictive value of miRNA combination tested using Logistic Regression or the Naïve Bayes models. TPR=True Positive Value (Sensitivity); FPR= False Positive Ratio (Sensitivity); AUC=Area Under the Curve. Thr=Threshold. **B.** Kaplan–Meier curve evaluating progression-free survival of patients with HNSCC clustered on the basis of the expression of miR-1, -9, -133a, and -150.

4.2 Bioinformatic analyses identified miRNA targets involved in the regulation of cell plasticity.

Confirmed targets of the identified 4-miRNAs impinge on EMT process. Specifically, miR-9, promotes EMT by targeting E-Cadherin[80] and miR-1, -133 and -150, act as EMT suppressors by targeting SLUG[81], SNAIL[82] and ZEB1[83], respectively.

Since EMT plays a pivotal role in HNSCC progression and recurrence formation[13,84], we applied the miRNADIP[71] and NAVIGaTOR[73] bioinformatic tools to integrate miRNA-target predictions with experimentally determined proteins-protein interactions (PPIs) from I2D as described[74,85], focusing on genes regulating EMT (Fig.4).

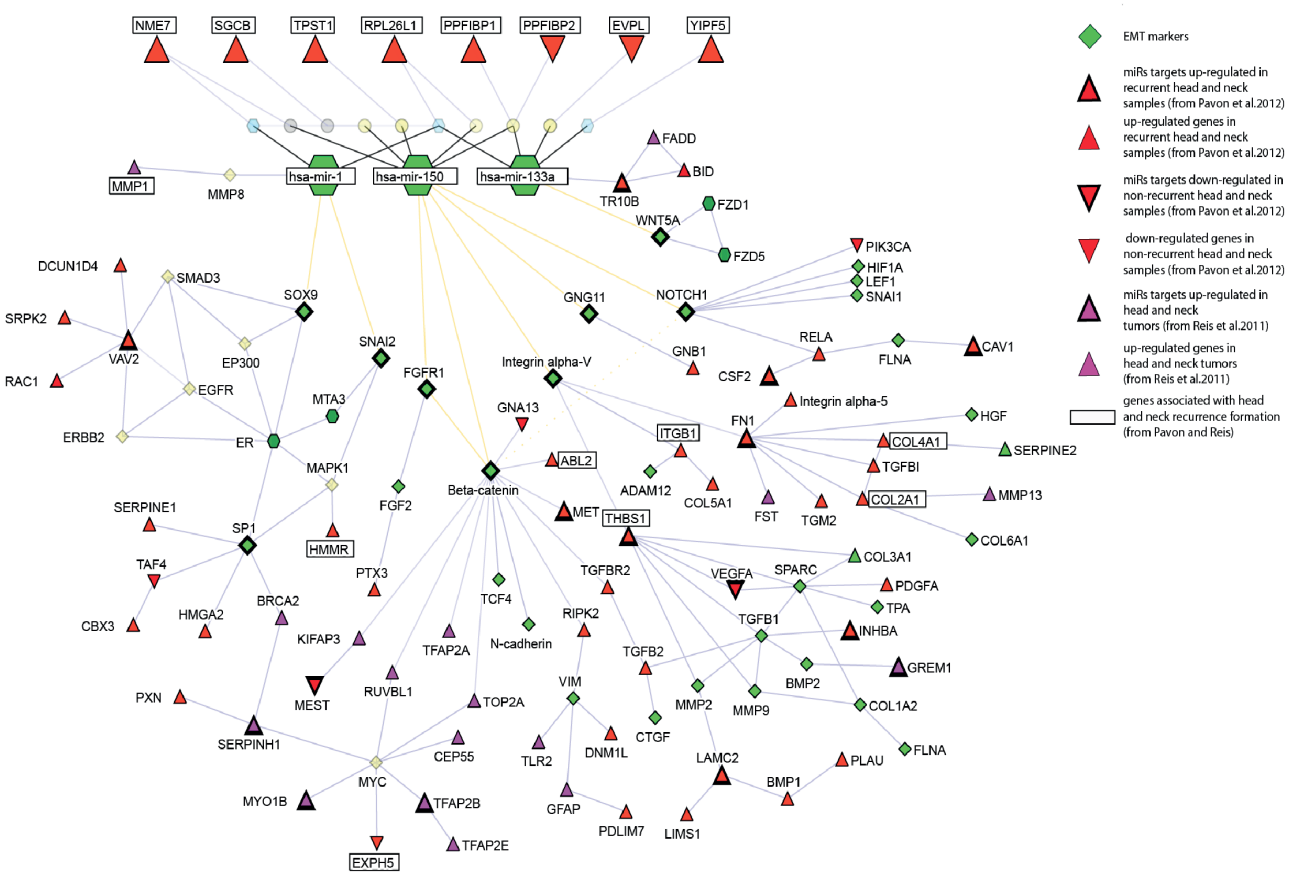

Figure 4. Bioinformatic analysis of miR-1, -133a and -150 targets.
Bioinformatic analyses, using Navigator portal, highlighted miR-1, -133a and -150 putative targets among genes known to be markers of EMT, up-regulated in HNSCC recurrent samples, down-regulated in non-recurrent HNSCC or up-regualted in HNSCC tumors as evaluated using Reis et al and Pavón et al datasets. These analyses showed that miR-1, -133a and -150 collectively targets SP1, β-Catenin and TGFβ pathways.

We first identified 16 possible common targets of miR-1, -133a and 150 (Fig. 5A). Then, we built an interaction network by integrating these analyses with the genes upregulated in HNSCC, as determined by the TCGA consortium. These analyses led us to the identification of SP1 as a common target of these 3 miRNAs (Fig. 5B). Refinement of these results, using two other available HNSCC datasets of coding gene signatures that could predict recurrences in HNSCC[72,86], also identified SP1 (Fig. 5C) together with other genes belonging to the TGFβ and β-Catenin pathways, as potential targets of miR-1, -133a and -150 (Fig. 5C).

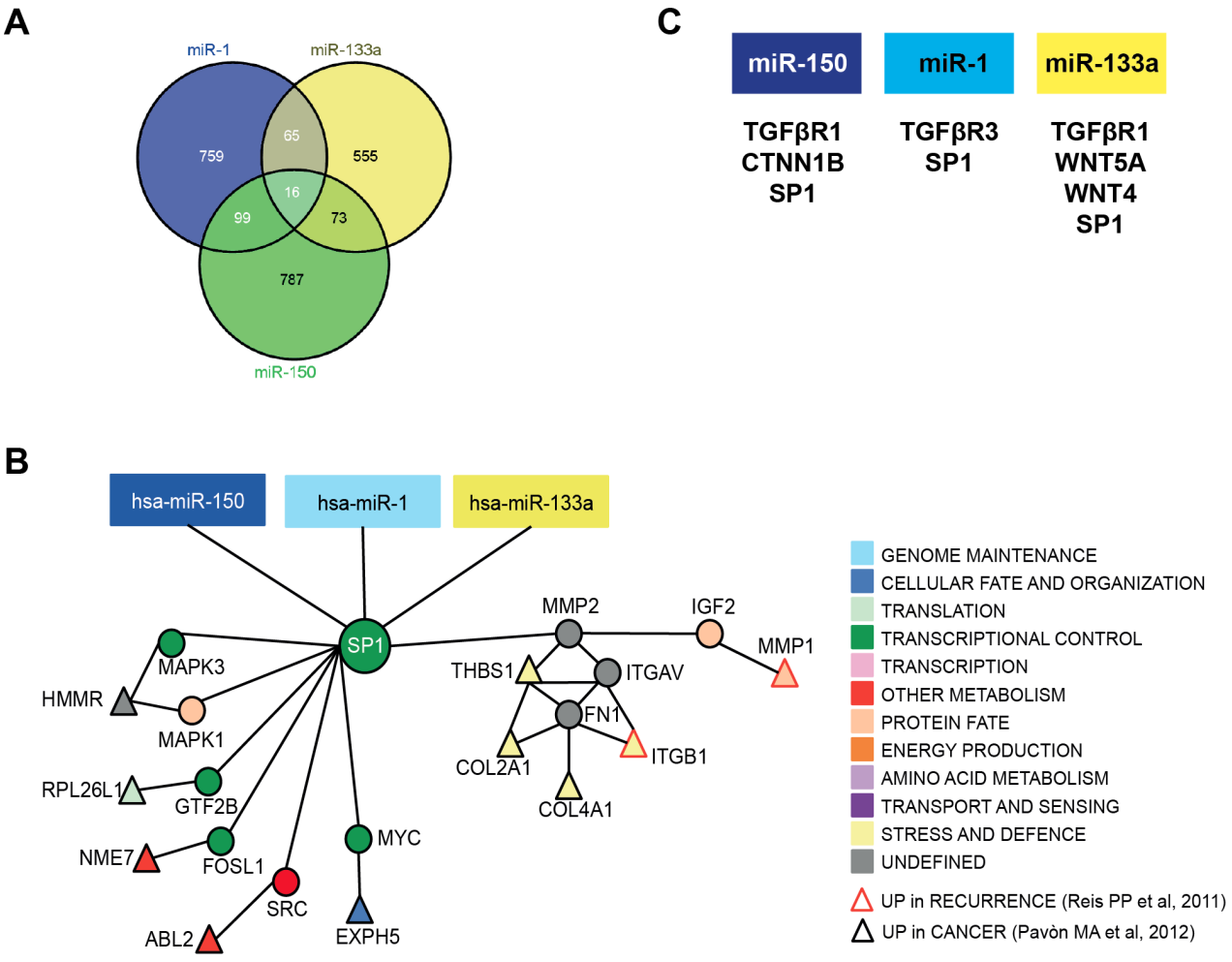

Figure 5. Identification of miR-1, 133a and -150 putative targets.
A. Venn diagram showing the number of miR-1, -133a, and -150 common potential targets, among the EMT genes, altered in HNSCC. **B.** Visualization of miR-1, -133a, and -150 and SP1 network in HNSCC. Direct interactions are shown by edges. Red-border triangles identify genes upregulated in recurrent HNSCC and black-border triangles identify genes upregulated in HNSCC primary tumours. **C.** List of miR-1, -133a, and -150 targets belonging to the TGFβ and WNT pathways.

A similar approach was used to screen the 1611 potential miR-9 targets. By intersecting these targets with genes associated with EMT negative regulators and HNSCC recurrence formation and/or patients' survival, we defined a complex interaction network that highlighted potential targets altered in HNSCC (Fig. 6A). Further refinement of these analyses led to the identification of Plakoglobin (JUP1), SASH1, Keratin 13 (KRT13) and Filaggrin (FLG), as the genes with the highest probability to represent miRNA-9 targets in recurrent HNSCC (Fig. 6B).

Pathway enrichment analyses using pathDIP[75], to integrate miR-1, -133a, and -150 with miR-9 networks, identified significantly enriched pathways (Supplementary Table S2 ref. Citron et al[87]). The most frequently occurring genes belonged to Signal Transduction (30 genes; $p<0.05$), EGFR1 (29 genes; $p=0.01$), Immune System (29 genes; $p<0.02$), Integrin $\alpha 6\beta 4$ (26 genes; $p=0.001$), TNFα (23 genes; $p=0.05$), Developmental Biology (22 genes; $p=0.02$), and TGFβ Signaling (22 genes; $p=0.04$) pathways. Interestingly, four prognostic genes from the work of Reiss and collegues[86], are predicted to be regulated by miR-9-5p and miR-150-5p (using an updated miRNADIP 3.0.1; http://ophid.utoronto.ca/mirDIP)[71].

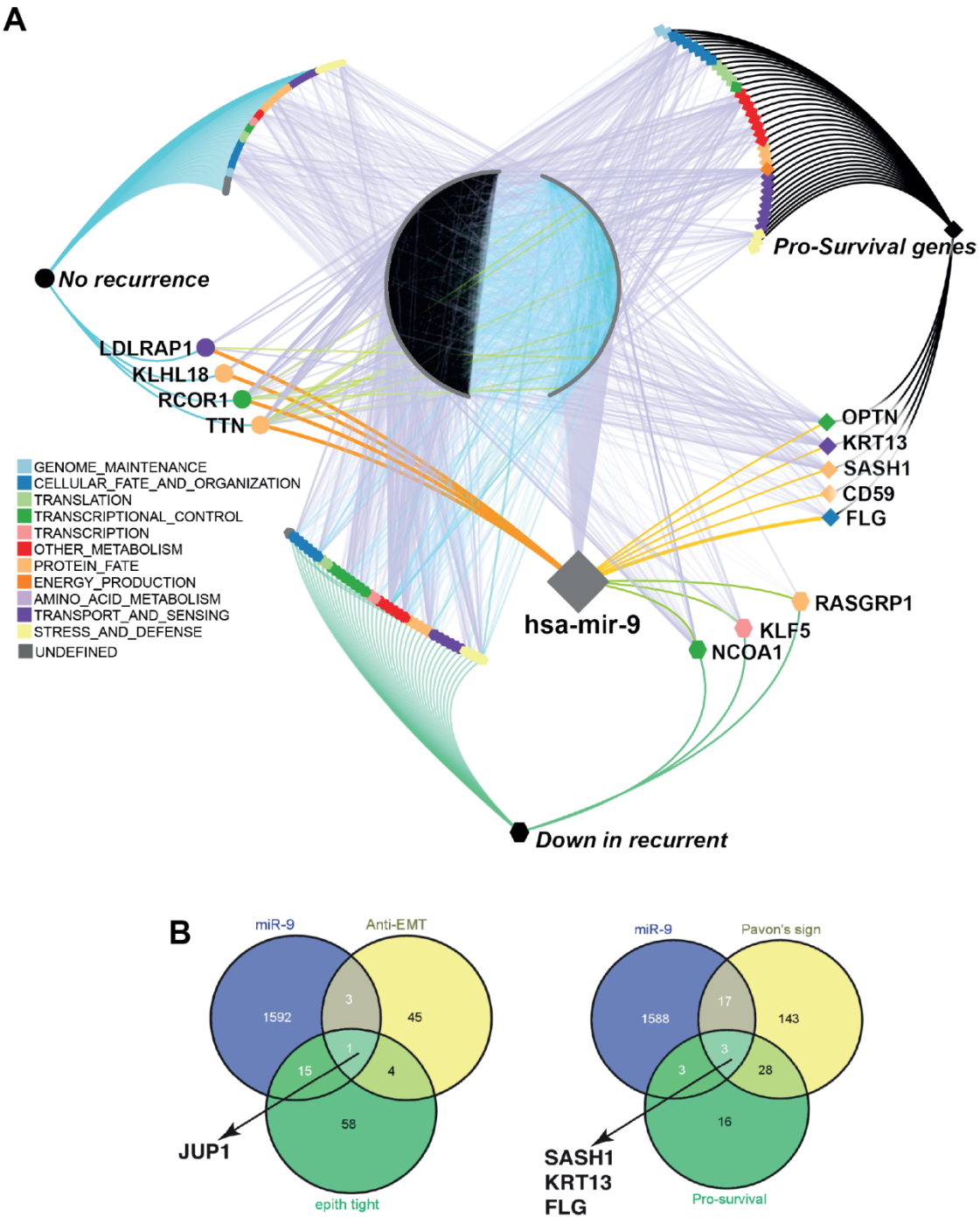

Figure 6. Identification of miR-9 putative targets in HNSCC.
A. Bioinformatic analyses, using pathDIP portal, highlighted miR-9 putative targets among genes known to be markers of EMT and altered in recurrent HNSCC. Direct interactions are shown by edges. Diamonds identify genes belonging to the pro-survival signature, hexagons are genes downregulated in recurrent tumors and circles represent genes upregulated in non-recurrent HNSCC as determined using the *Pavón* dataset. **B.** Venn diagram showing the number of miR-9 potential targets, among the anti-EMT genes and among the genes altered in HNSCC.

4.3 Experimental validation of identified miRNAs targets involved in the regulation of EMT.

Overall these analyses suggested that the relative expression of miR-1, -133a, -150 and miR-9 could play a functional role in HNSCC progression possibly by regulating EMT.

To experimentally validate miRNAs targets identified *in silico*, we screened a panel of HNSCC-derived cell lines for endogenous miRNAs levels to choose the most appropriate *in vitro* model. Generally, all cell lines expressed low levels of miR-1, -133a and -150 and higher levels of miR-9 (Fig.7A-B).

Intriguingly, FaDu and SCC9 cells expressing highest level of miR-9, display low levels of SASH1, KRT13 and of the epithelial marker E-Cadherin (Fig.7A-B).

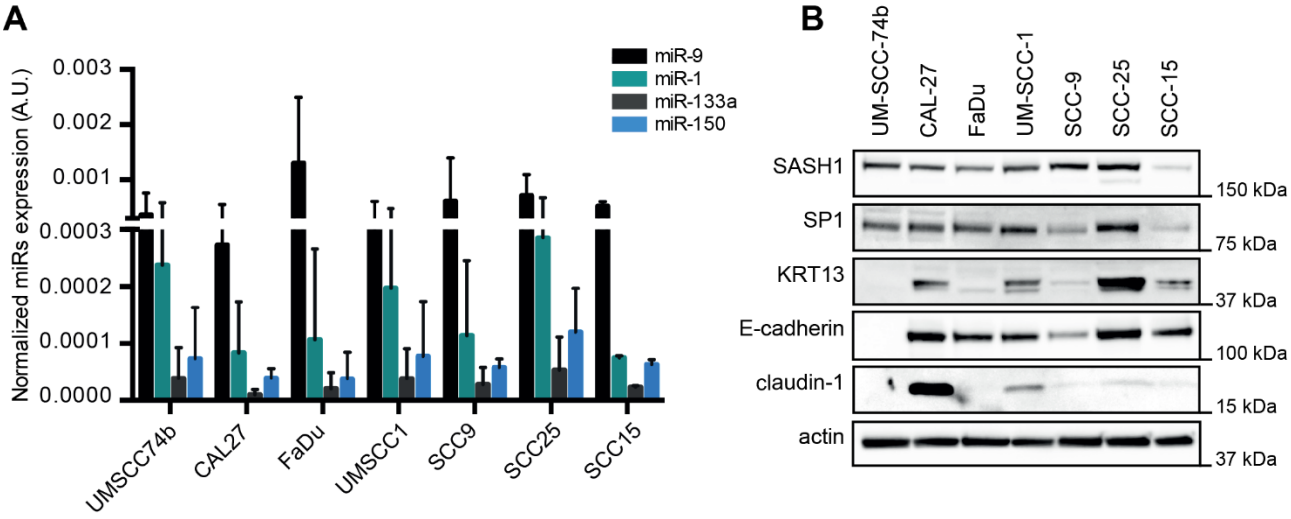

Figure 7. Characterization of miRNAs expression and related putative targets in a panel of HNSCC cells. A. Graph showing the normalized expression of miR-1, -9, 133a and 150 in a panel of HNSCC cell lines. Data represent the mean (±S.D.) of 3 independent experiments performed in duplicate. **B.** Western blot analysis of the indicated proteins in a panel of HNSCC cells. Actin was used as loading control.

First, we focused on the four most probable miR-9 targets, namely JUP1, SASH1, KRT13 and FLG (Fig. 6B). In both cell lines, miR-9 knock-down resulted in the upregulation of SASH1 and KRT13, but not of JUP1, and FLG mRNAs (Fig. 8A-B). The increased SASH1 and KRT13 expression was also confirmed at protein level (Fig. 8C-D). As expected from the literature[88], the down-modulation of miR-9 leads to a strong abrogation of N-Cadherin protein expression, further supporting the involvement of miR-9 in the regulation of a EMT-like phenotype (Fig. 8D). In CAL27 cells, expressing low endogenous miR-9, its overexpression decreased mRNA and protein expression of both SASH1 and KRT13 (Fig. 8E-F).

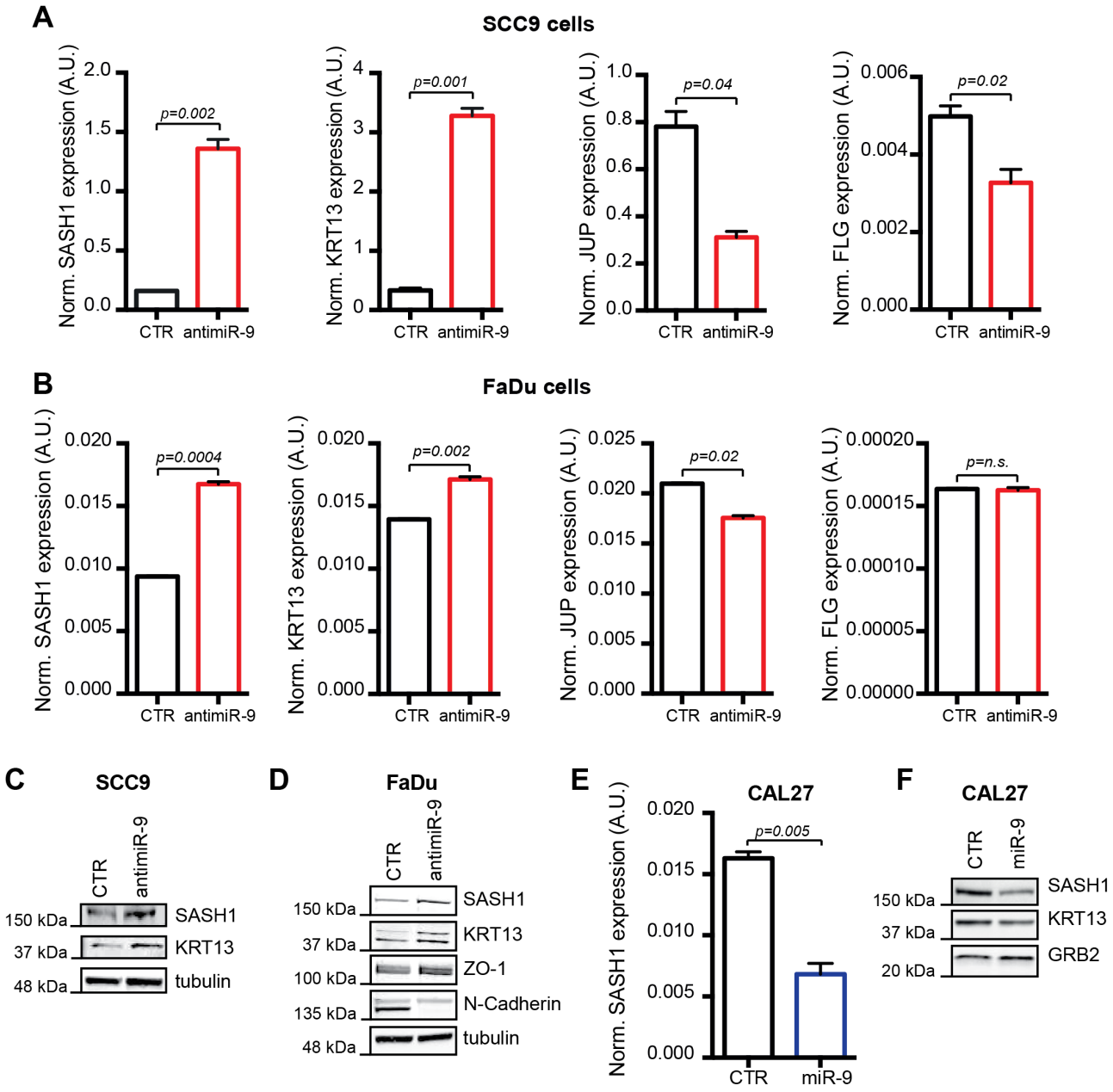

Figure 8. ***In vitro*** **validation of miR-9 putative targets.**
A/B. Graphs showing (from left to right) the normalized expression of SASH1, KRT13, JUP and FLG mRNA of both control and miR-9 knockdown respectively in SCC9 (**A**) and FaDu (**B**) cells. Data report the median value (±SD) of three independent experiments performed in duplicate. Statistical significance was calculated using Student t test. **C.** Western blot analysis of SASH1 and KRT13 protein expressions in control and miR-9 knockdown SCC9 cells. Tubulin was used as loading control. **D.** Western blot analysis of SASH1, KRT13, Zo-1 and N-Cadherin protein expressions in control and miR-9 knockdown FaDu cells. Tubulin was used as loading control. **E.** Graphs showing the normalized expression of SASH1 mRNA in control and miR-9 over-expressing CAL27 cells. Data report the median value (±SD) of three independent experiments performed in duplicate. Statistical significance was calculated using Student t test. **F.** Western blot analysis of SASH1 and KRT13 protein expressions in control and miR-9 over-expressing CAL27 cells. GRB2 was used as loading control.

As a proof of principle, we also tested whether miR-9 could directly regulate SASH1 expression acting on its 3'-UTR, that contains three seed sites for miR-9 (Fig. 9A). Luciferase assay in FaDu cells demonstrated that miRNA-9 knockdown significantly increased the luciferase activity when the first seed site (position 217-223) was tested (Fig. 9B). Overall, these data suggest that SASH1 (and likely KRT13) could represent reliable markers of miR-9 activity in HNSCC cells.

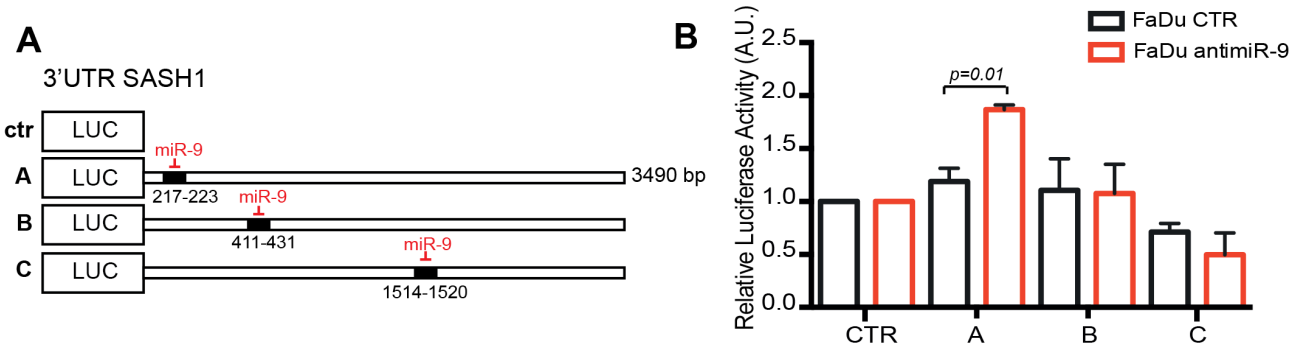

Figure 9. Validation of putative miR-9 seed sites in the 3'-UTR of SASH1 gene.
A. Schematic design of the pGL3-vectors used to test the potential miR-9 seed sites in the SASH1 3'-UTR regulatory region. A, B and C, indicates the 3 regions tested in the luciferase experiments shown in **B**. **B.** Graphs showing the normalized luciferase activity of pGL3-vectors shown in **A** and transfected in control or miR-9 knock-down FaDu cells. Data represent the median values (±S.D.) of three different experiments. Statistical significance is report in the graph and was calculated using Student t test.

We next experimentally validated SP1 as common target of miR-1, miR-133a and miR-150. Overexpression of miR-1, miR-133a or miR-150 alone did not affect SP1 mRNA or protein expression in different HNSCC cell lines. However, the combined overexpression of two and, even better, of the three miRNAs together significantly reduced SP1, at both protein and mRNA levels (Fig. 10A-D).

Expression data paralleled luciferase assay on SP1 3'-UTR, that contains one seed site for miR-1 and miR-133a and two seed sites for miR-150 (Fig. 10E). Each miRNA alone slightly reduced the luciferase activity of each construct (Fig. 10F), further supporting the possibility that the three miRNAs need to work together to reduce SP1 expression.

Members of TGFβ and β-Catenin pathways were also predicted targets of miR-1, miRNA-133a and miRNA-150. Among the predicted targets of miR-1, -133a and -150 (*i.e.* TGFβ-R1, -R2 and -R3, WNT4, WNT5A and CTNN1B), the expression of WNT5A strongly decreased in cells overexpressing miR-133a (Fig.11A) while the expression of TGFβ-R3 decreased following the combined overexpression of the three miRNAs (Fig.11B).

From a functional point of view, we observed that combined inhibition of miR-9 and overexpression of miR-1, -133a and -150 (Fig.11C) up-regulated the epithelial marker E-Cadherin in FaDu cells

(Fig.11D), supporting the hypothesis that balanced expression of miR-9 and miR-1, miR-133a and miR-150 plays a functional role in HNSCC recurrence formation by modulating the EMT process.

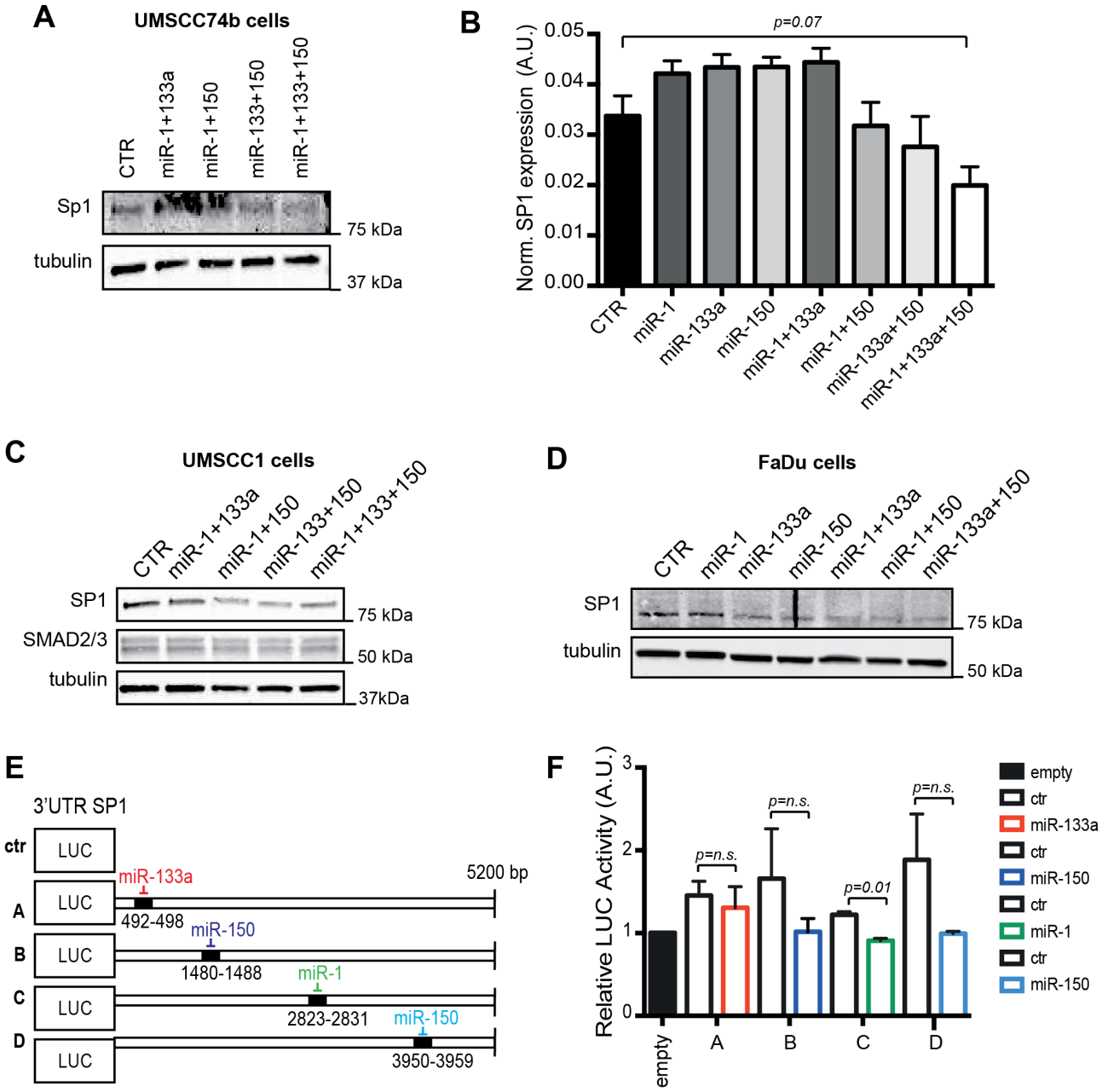

Figure 10. Combined expression of miR-1, -133a and -150 is necessary to down-regulate SP1.
A. Western blot analysis of SP1 protein expression in UMSCC74b cells expressing combination of miR-1, -133a, and -150, as indicated. **B.** Graph showing the normalized expression of SP1 mRNA in UMSCC74b cells transfected as in (**A**). Data report the median value (±SD) of three independent experiments performed in duplicate. Statistical significance was calculated using Student t test. **C.** Western blot analysis of SP1 and SMAD2/3 protein expressions in UMSCC1 cells transfected with miR-1, -133a and -150, as indicated. **D.** Western blot analysis of SP1 protein expression in FaDu cells transfected with miRNA-1, -133a and -150 alone or in combination, as indicated. In **A**, **C** and **D** tubulin was used as loading control. **E.** Schematic design of the pGL3-vectors used to test the potential miR-1, -133a and -150 seed sites in the SP1 3'-UTR regulatory region. A, B, C and D indicates the 4 regions tested in the luciferase experiments shown in **F**. **F.** Graphs showing the normalized luciferase activity of pGL3-vectors shown in **E** transfected in 293T17. Data represent the median value (±S.D.) of three different experiments. Statistical significance is report in the graph and was calculated using Student t test.

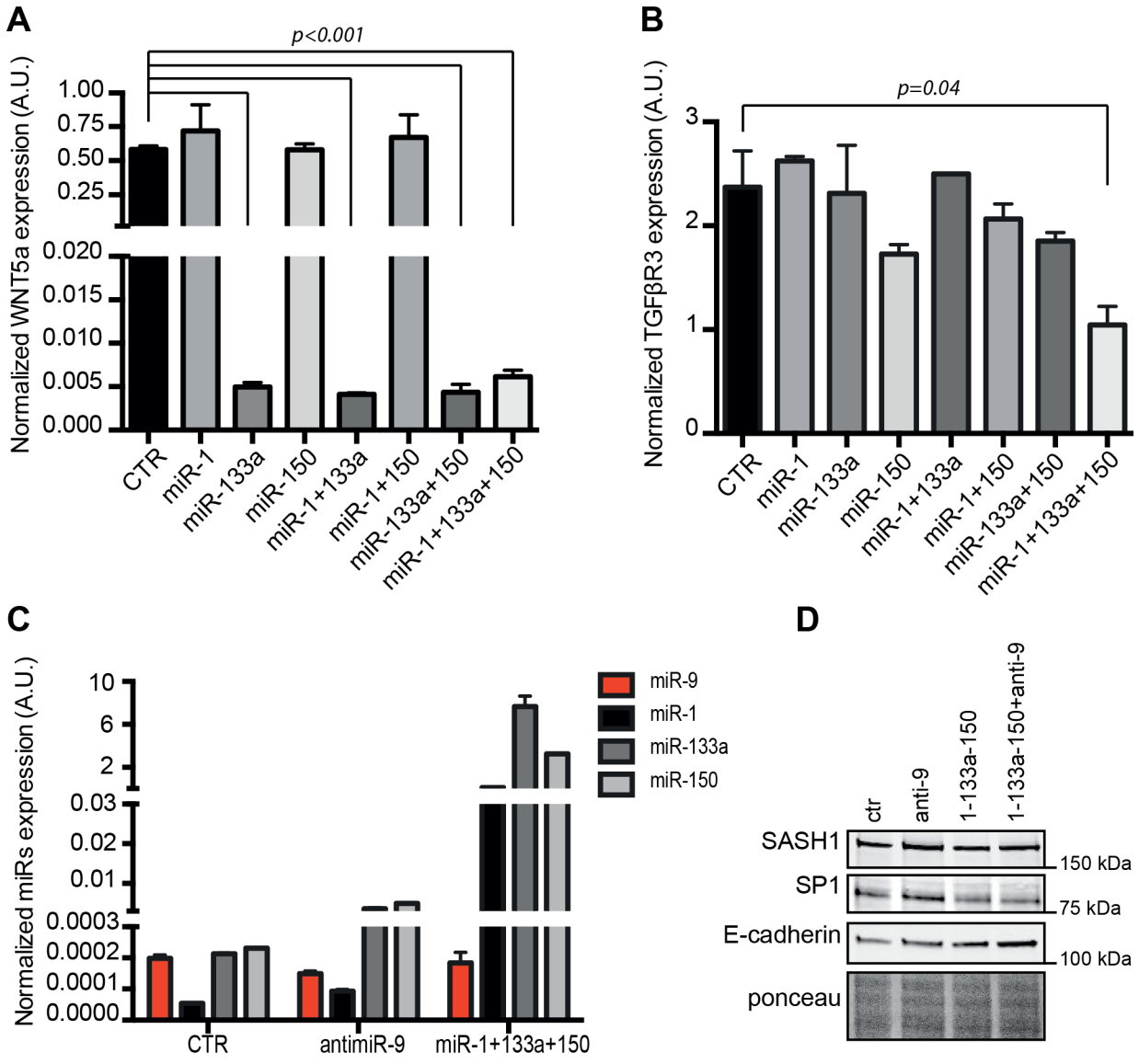

Figure 11. Validation of putative miR-1, -133a and -150 putative targets.
A. Graph showing the normalized expression of WNT5a in CAL27 cells expressing miR-1, - 133a and -150 alone or in combination. **B.** Graph showing the normalized expression of TGFβ-R3 mRNA in CAL27 cells expressing miR-1, -133a and -150 alone or in combination. Data report the median value (±S.D.) of 3 independent experiments performed in duplicate. Statistical significance is reported in each graph and was calculated using Student t-test. **C.** Graph reporting the normalized expression of miR-9, -1, -133a, -150 in FaDu cell line transfected with antimiRNA-9 or with miR-1, -133a and 150 or their combination, as indicated. Data represent the median value (±SEM) of one experiment performed in triplicate. **D.** Western Blot analysis of E-cadherin, SP1 and SASH1 protein expressions in FaDu cell line treated as in **C**. Ponceau staining of the membrane was used as loading control.

4.4 *In vitro* functional implication of miR-9 driven EMT

High expression of miR-9 could modulate an EMT-like phenotype in SCC, both impinging on SASH1 and KRT13 levels, as we observed (Fig. 8), and triggering an up-regulation of N-Cadherin (Fig.8D), suggesting an important involvement of miR-9 in the regulation of tumorigenic and/or survival capabilities.

Moreover, recently it has been demonstrated that in SCC a subpopulation of stem cells display increased miR-9 expression respect to the differentiate tumour cells[88], this is of particular clinical interest since there is a close correlation between the presence of a non-proliferative niche of stem cell and resistance to both chemo or radio-therapy[89,90], dependent on the context, even if the mechanisms involved are not completely understood.

These observations prompted us to investigate whether high expression of miR-9 could be linked to survival advantage of tumour cells, both in basal condition and under the selective pressure of radio- or chemo-therapy.

To verify the role of miR-9 in the promotion of cell survival and tumorigenic potential in HNSCC, we generate stable antimiR-9 FaDu and SCC9 cells. Down-modulation of miR-9 results in a dramatic decrease of clonogenic survival capability in 2D, both in FaDu and SCC9 cells (Fig. 12A). Moreover, when FaDu cells are embedded in 3D-inert matrix, in anchorage-independent assay, antimiR-9 cells display a reduced colony formation ability, both for number and size, respect to the control ones (Fig. 12B). Consistently, when miR-9 is transiently up-regulated in CAL27 cells, it remarkably increased their clonogenic ability in 2D assay (Fig. 12C).

Since it has been demonstrated that both EMT and stem "niche" correlate with radio- and/or chemo-resistance[89], we next investigated the implications of miR-9 knockdown in the standard care of HNSCC focusing our attention in particular on radiotherapy, the elective treatment for early and advanced patients, and on chemotherapy, testing both cisplatin and anti-EGFR targeted agents.

To investigate the role of miR-9 in the response to γ-radiation, we irradiate control or antimiR-9 expressing FaDu and SCC9 cells with ionizing radiation at 5-Gy high dose or 2-Gy, a low dose that more closely resemble the hypo-fractionated regimens in patients[91], and evaluated cell survival fraction. As expected from the previous observation, the abrogation of miR-9, both in FaDu and SCC9 cell lines, resulted in a strong decrease in survival fraction, especially when the highest dose was used (Fig. 14A-B).

Since γ-radiation exerts its major effect inducing DNA damage, we analysed the expression of phosphorylation in Ser139 of H2AX (*i.e.* γH2AX), a very sensitive and well-established marker for DNA double-strand breaks (DSBs)[92]. Following radiation, high levels of DNA damage are often

associated with radio-sensitivity rather than -resistance[93]. Intriguingly, in FaDu miR-9 knock-down cells we observed in increased γH2AX expression, when cells are treated with 2 Gy and even better with 5 Gy dose of radiation (Fig. 14C-D), further supporting that inhibition of miR-9 is sufficient to mediate a γ-radiation sensitivity in HNSCC, eventually impinging on a radio-resistant niche population. It is also interesting to note that when miR-9 expression was inhibited we observed a strong down-modulation of SP1 (Fig. 14D), corroborating the hypothesis that a possible crosstalk exists between miR-9 and EMT pathway targeted by miR-1, 133a and -150.

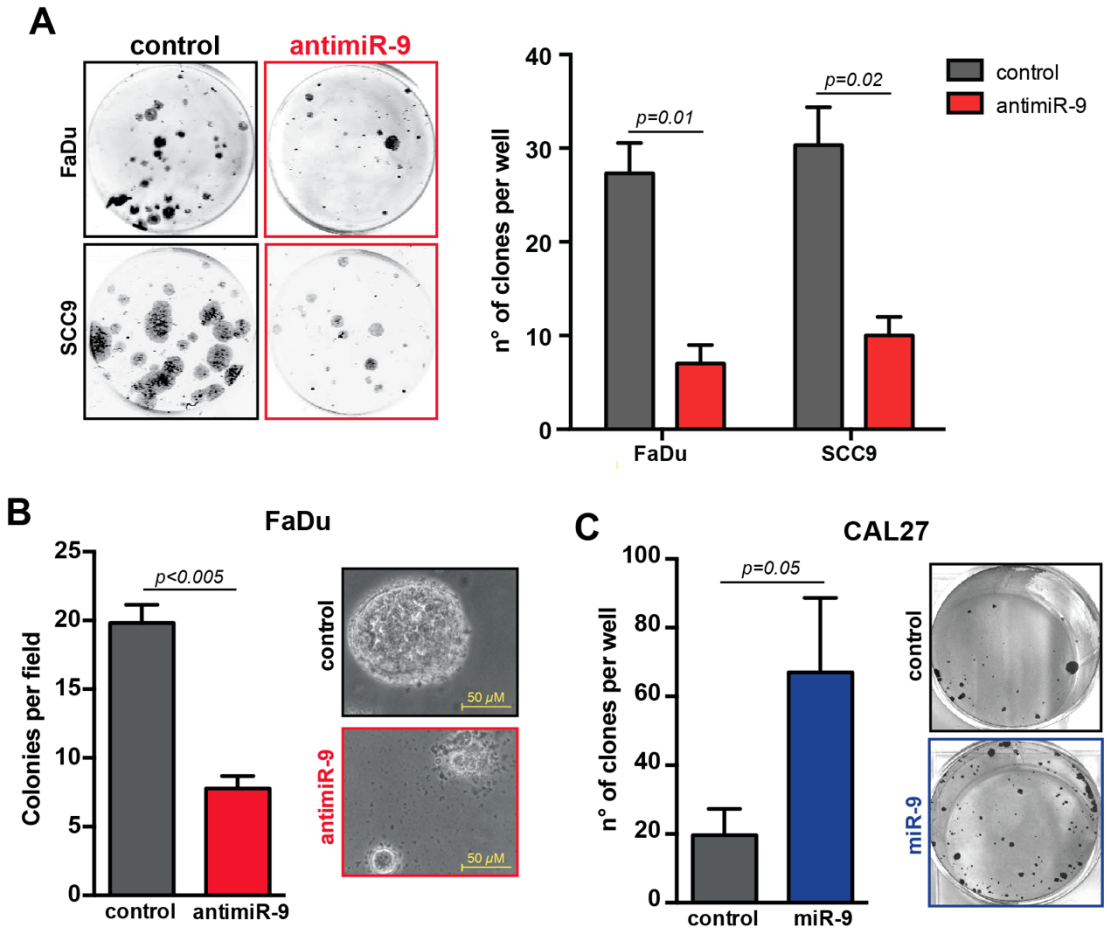

Figure 12. miR-9 promotes cell survival and tumorigenic potential in HNSCC cell lines.
A. Left panel shows representative contrast-phase images of survival clonogenic assay on FaDu and SCC9 cells, expressing or not miR-9. Right panel reports the colony number quantification of the indicated cell lines modified for the expression of miR-9 and allowed to grow for 10 days. **B.** Left panel reports colony quantification of soft agar assay performed with FaDu control or miR-9 knock-down cells allowed to grow for 15 days. Right panel showed representative contrast-phase images. **C.** Left panel reports the colony number quantification of control or miR-9 over-expressing CAL27 cells allowed to grow for 10 days. Right panel shows representative contrast-phase images of survival clonogenic assay on CAL27 cells modified for miR-9 expression. All the experiments were performed three times in triplicate and the graphs report the median values (±SD). Statistical analyses are reported in each graph and were calculated using Student t test.

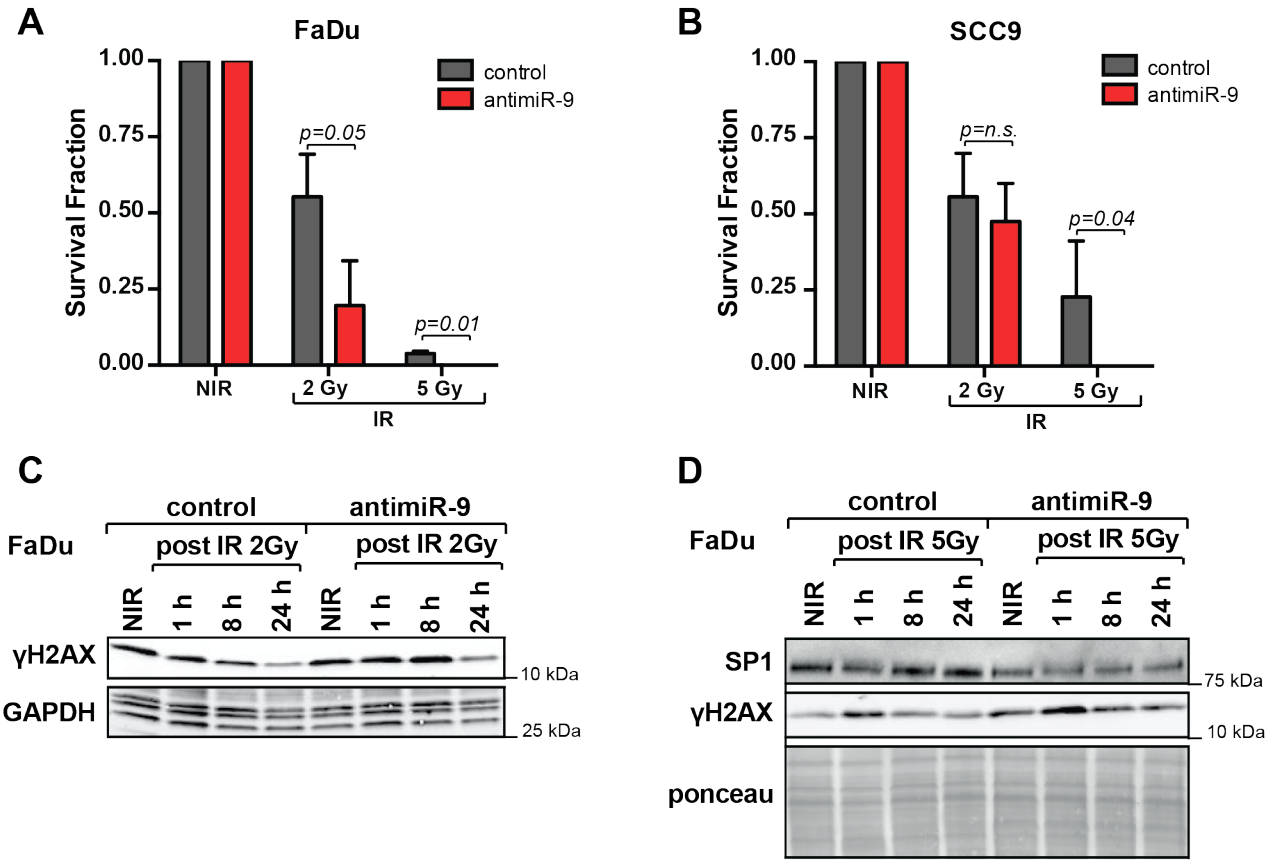

Figure 13. In HNSCC cells miR-9 expression leads to a survival advantage after genotoxic γ-radiation. A/B. Graph reports the survival fraction of control or antimiR-9 FaDu (**A**) and SCC9 (**B**) cells subjected to 2 Gy or 5 Gy γ-radiation and then plated in clonogenic assay. Data are expressed as a mean (±SD) of three independent experiments performed in triplicate. Statistical significance was calculated using Student t test and is reported in each graph. **C.** Western Blot analysis of γH2AX in lysates from control or antimiR-9 FaDu cells non-irradiated (NIR) or irradiated with 2 Gy and harvested after 1, 8 or 24 hours. GAPDH was used as loading control. **D.** Western Blot analysis of γH2AX and SP1 in lysates from control or antimiR-9 FaDu cells non-irradiated (NIR) or irradiated with 5 Gy and harvested after 1, 8 or 24 hours. Ponceau staining of the membrane was used as loading control.

To investigate whether the silencing of miR-9 could modulate also chemo-resistance, we tested the sensitivity of FaDu and SCC9 antimiR-9 cells to incremental dose of cisplatin or EGFR-targeted compounds. We did not notice any differences in cell viability between control and antimiR-9 cells, both in SCC9 and FaDu, when treated with cisplatin (Fig. 14A-B). Whereas, when we tested different anti-EGFR agents, either small inhibitors that exert their activity without the contribution of immune system (*i.e.* Gefitib and Erlotinib) or monoclonal antibody, such as Cetuximab, we observed a significant decrease of IC50 in antimiR-9 FaDu cells respect to the control ones (Fig. 14C-E).

These data strongly suggest that in advanced HNSCC, a relative abundant population of miR-9 expressing cells is intrinsically resistant to the standard radio- and targeted-therapy, thus confirming the potential prognostic value of miR-9 expression and highlighting the urgency of a novel clinical

treatments for this group of patients, which received an unfaithful toxic treatment and that may benefit for a more accurate personalized therapy.

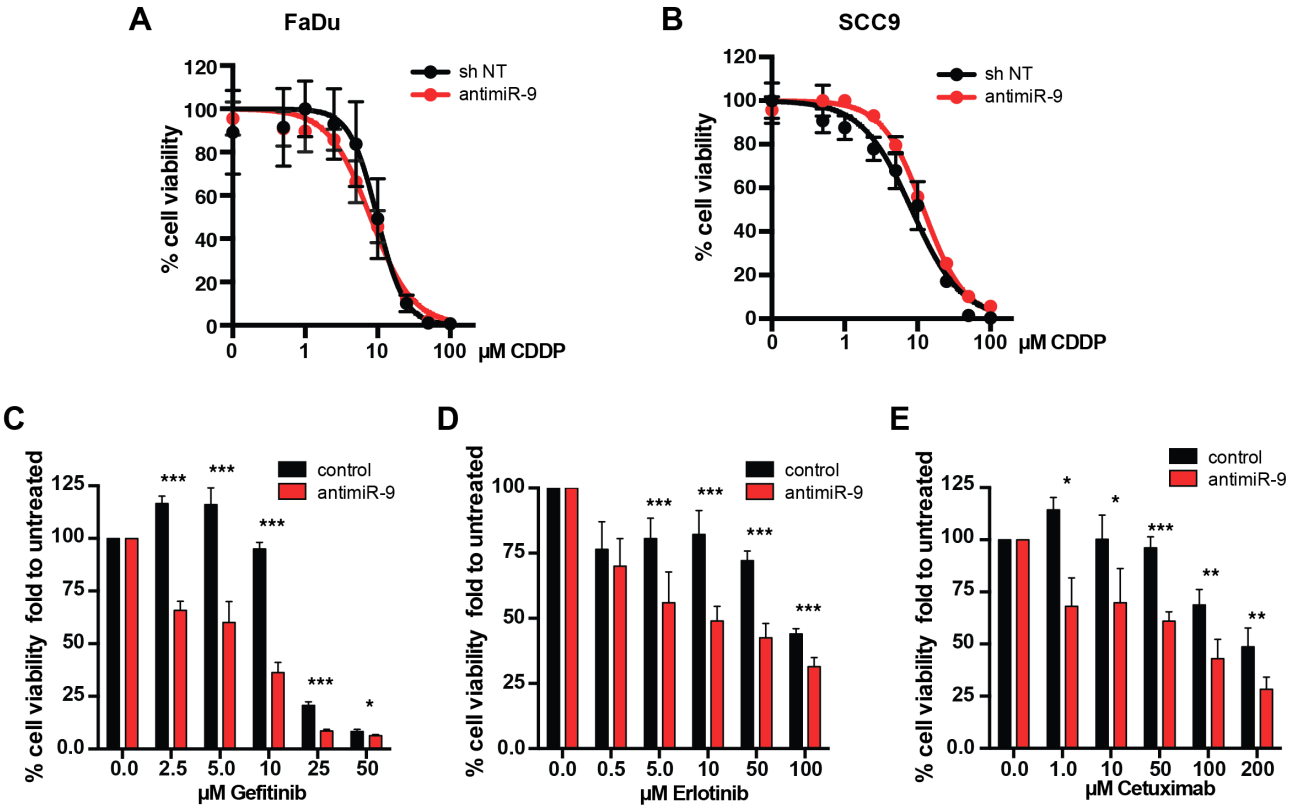

Figure 14. miR-9 silencing sensitizes HNSCC cells to anti-EGFR targeted agents.
A/B. Cisplatin (CDDP) dose-response curve of control or antimiR-9 FaDu (**A**) and SCC9 (**B**) cells expressed as percentage of viable cells with respect to untreated ones. Graphs report the percentage (±SD) of three independent experiments performed in triplicate. **C/D.** Cell viability of control or antimiR-9 FaDu cells treated with incremental dose of Gefitinib (**C**), Erlotinib (**D**) or Cetuximab (**E**) for 72 hours. Graphs report the percentage (±SD) of viable cells folded to untreated cells. Graphs report the statistical significance calculated using Student t test: **p<0.05*, ***p<0.01* and ****p<0.001*.

4.5 *In vivo* validation of identified miRNAs targets involved in the regulation of EMT.

To verify if the predicted targets of the 4 miRNAs were effectively differentially regulated in HNSCC samples, we tested by qRT-PCR the expression SP1, TGFβ-R1, R2 and R3, WNT4 and WNT5A, CTNN1B, SASH1 and KRT13 in the 44 samples described in Table 1-3 and analyzed for the expression of miRNAs (Table 2).

In line with *in vitro* results, we observed that miR-9 targets, SASH1 and KRT13, were both significantly downregulated in patients who experienced recurrence (Fig. 15A) and the expression of SASH1 correlated directly with KRT13 and inversely with miRNA-9 in primary HNSCC (Fig. 15B-C). Among miR-1, -133a and -150 targets, only SP1 and TGFβ-R1 were significantly upregulated in HNSCC samples from recurrent patients (Fig. 15D) and their expression directly correlated (Fig. 15E). No differential expression was observed for TGFβ-R2, and R3, WNT4, WNT5A and CTNN1B between patients with or without recurrence (Fig.16).

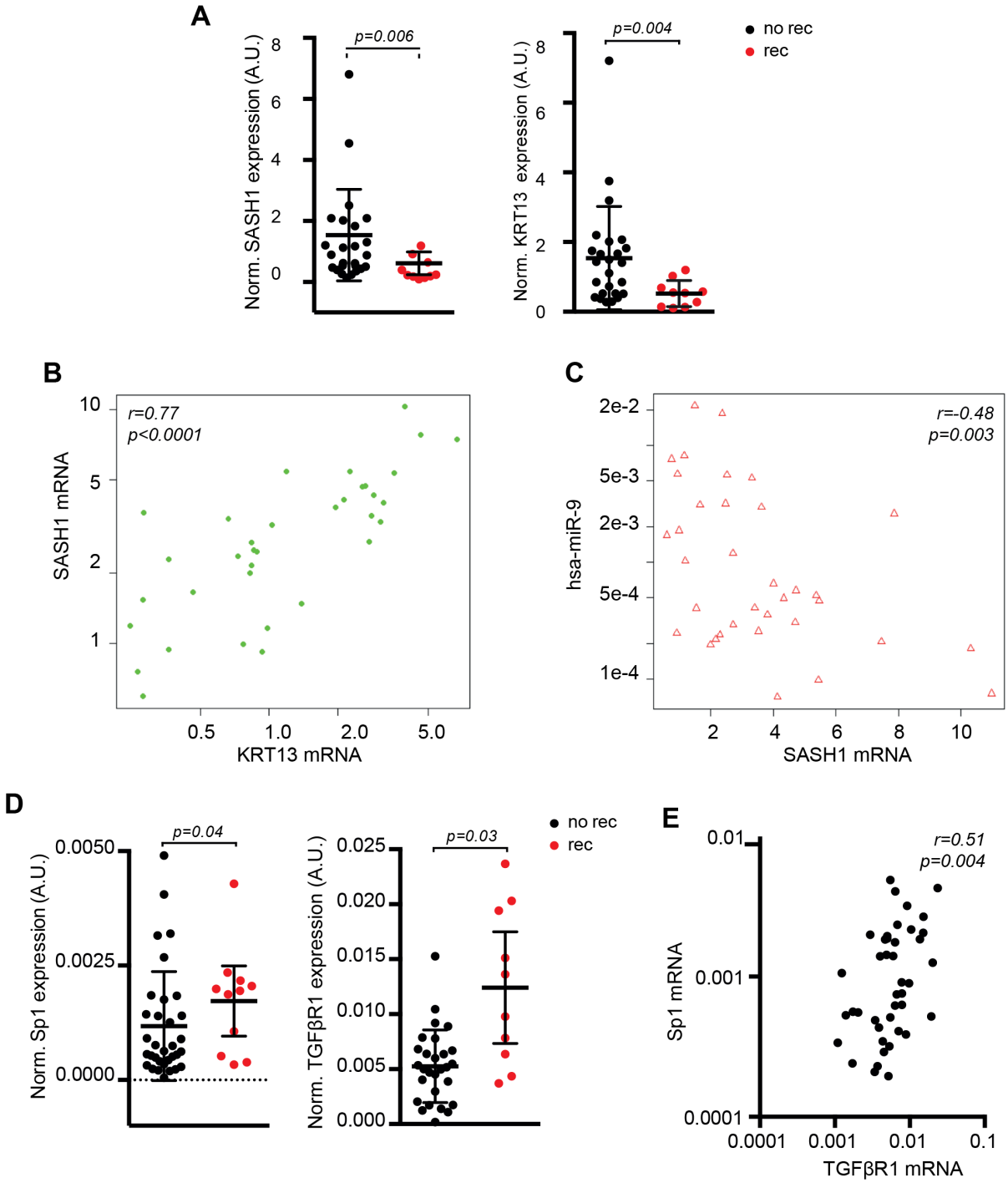

Figure 15. Expression of identified miRNA targets in primary HNSCC from recurrent and non-recurrent patients.

A. Dot plots showing the normalized expression of SASH1 (left) and KRT13 (right) mRNA, in samples from primary HNSCC from recurrent and non-recurrent patients. Bars indicate the mean and the 95% confidence interval. Statistical significance was calculated using the Mann–Whitney test. **B**/**C**. Dot plot showing the correlation of SASH1 and KRT13 (**B**) or SASH1 and miR-9 (**C**) mRNA expression in HNSCC samples described in **A**. Correlation value (r) and statistical significance (P) were calculated using the Spearman correlation test. **D.** Dot plots showing the normalized expression of SP1 (left) and TGFβ-R1 (right) mRNA expressions in samples from primary and recurrent HNSCC samples. Bars indicate the mean and the 95% confidence interval. Statistical significance reported in the graph was calculated using the Mann–Whitney test. **E.** Dot plot showing the correlation of SP1 and TGFβ-R1 mRNA expression in HNSCC samples as in **A**. Correlation value (r) and statistical significance (P) were calculated using the Spearman correlation test.

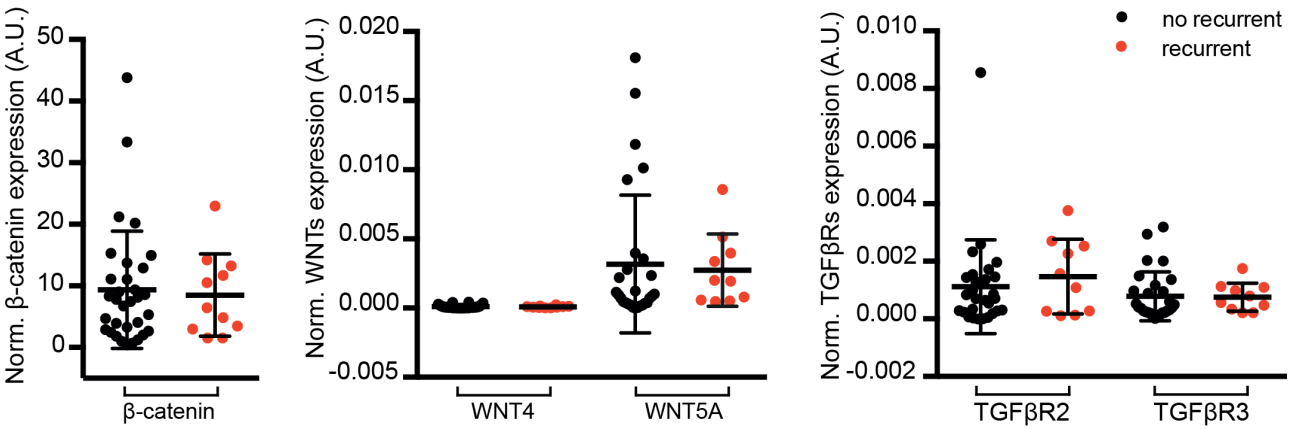

Figure 16. Expression of WNT/β-catenin and TGFβ pathway members in primary HNSCC from recurrent and non-recurrent patients.
From left to right: Dot plots showing the normalized mRNA expression of CTNNB1, WNT4, WNT5, TGFβ-R2 and TGFβ-R3, putative targets of miR-1, 133a and -150, in samples from HNSCC patients who underwent or not loco-regional recurrences, reported in Tables 1 and 3. Bars indicate the mean and the 95% confidence interval. No significant differences have been calculated using the Mann-Whitney test.

To validate our findings in an independent cohort, we collected 78 HNSCC samples in our Institutes (Table 4) and evaluated the expression of the 4 miRNAs, SASH1, KRT13, SP1 and TGFβR1. In accord with previous results, miRNAs expression did not correlate with any biological variable of HNSCC patients, including sex, age, cancer site, T and N stages (Table 4).

The expression of SASH1 inversely correlated with miR-9 and directly correlated with KRT13 (Fig.17A-B). An inverse, although not significant, correlation was noticeable between miR-1 or miR-133a and SP1 expression, as expected from the *in vitro* results (Fig.17C-D).

Characteristics	n	miR-1	miR-9	miR-133a	miR-150
Sex					
Men	40	1.688	11.671	5.809	354.52
Woman	21	1.446	18.657	5.223	609.28
Wilcoxon test		*p=0.7226*	*p=0.5410*	*p=0.9154*	*p=0.5023*
Age					
<60 years	18	1.235	10.037	4.153	497.13
≥60 years	41	1.793	18.657	5.520	418.95
Wilcoxon test		*p=0.4569*	*p=0.3821*	*p=0.3821*	*p=0.5834*
Cancer site					
Oral cavity/Tongue	38	1.688	7.783	5.353	462.50
Oro-/Hypo-/Pharynx	14	1.973	17.535	7.110	236.50
Larynx	6	1.692	35.955	4.595	677.85
Other	3	0.825	20.530	3.921	299.20
Kruskall-Wallis test		*p=0.8006*	*p=0.0561*	*p=0.7513*	*p=0.5434*
cT					
T1-T2	37	1.582	9.675	5.482	466.43
T3-T4	24	1.674	18.125	5.456	317.56
Wilcoxon test		*p=0.8247*	*p=0.6157*	*p=0.9529*	*p=0.3226*
cN					
N0	32	1.916	12.702	5.353	462.50
N1-N2	29	1.292	11.845	5.520	352.91
Wilcoxon test		*p=0.5251*	*p=0.8852*	*p=0.2724*	*p=0.9080*
Adjuvant radiotherapy					
No	34	1.467	11.671	5.809	387.54
Yes	27	1.793	15.720	5.346	458.57
Wilcoxon test		*p=0.5614*	*p=0.8617*	*p=0.6843*	*p=0.6318*

Table 4. Median expression of the identified 4-miRNAs in the Validation Dataset according to the selected characteristics.
Table shows the independent expression of the identified 4-miRNAs from the different covariates considered. Abbreviations: cN, clinical evaluation of node status; cT, clinical evaluation of tumour size.

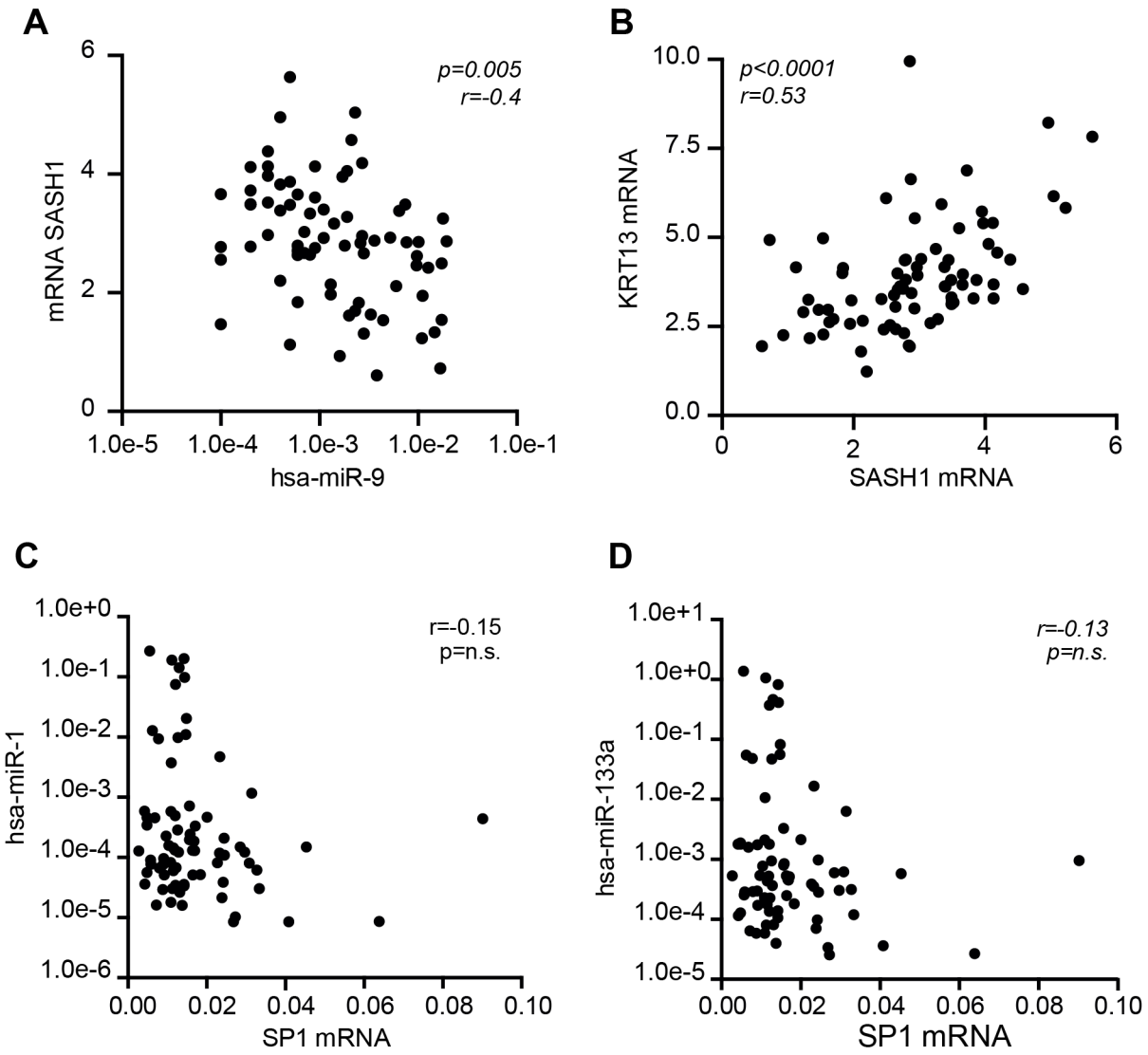

Figure 17. Correlation analyses of miRNA putative targets in primary HNSCC included in the Validation Cohort.

A/B. Correlation analyses between SASH1 and miR-9 (**A**) or KRT13 (**B**) expressions in HNSCC samples included in the validation set. **C/D.** Correlation analysis between SP1 and miR-1 (**C**) or miR-133a (**D**). All graphs report correlation value *(r)* and statistical significance *(p)* both calculated with non-parametric two-tailed Spearman correlation test.

4.6 miRNAs' targets predict prognosis in the TCGA HNSCC dataset.

We hypothesized that the expression of SASH1, KRT13, SP1 and members of TGFβ pathway could be used in HNSCC as readout of miRNA activity and tested the expression and the correlation of these genes with miR-9 and miR-1, -133a and -150, in the TCGA dataset.

In accord with the data obtained using our discovery- and validation-cohorts, correlation analyses confirmed that SP1 positively correlated with miR-9 and inversely with miR-133a (Fig. 18A-B). Similar results were observed for SMAD2 and SMAD3 expressions, used as readouts of SP1 and TGFβ pathway activity (Fig. 18C-H).

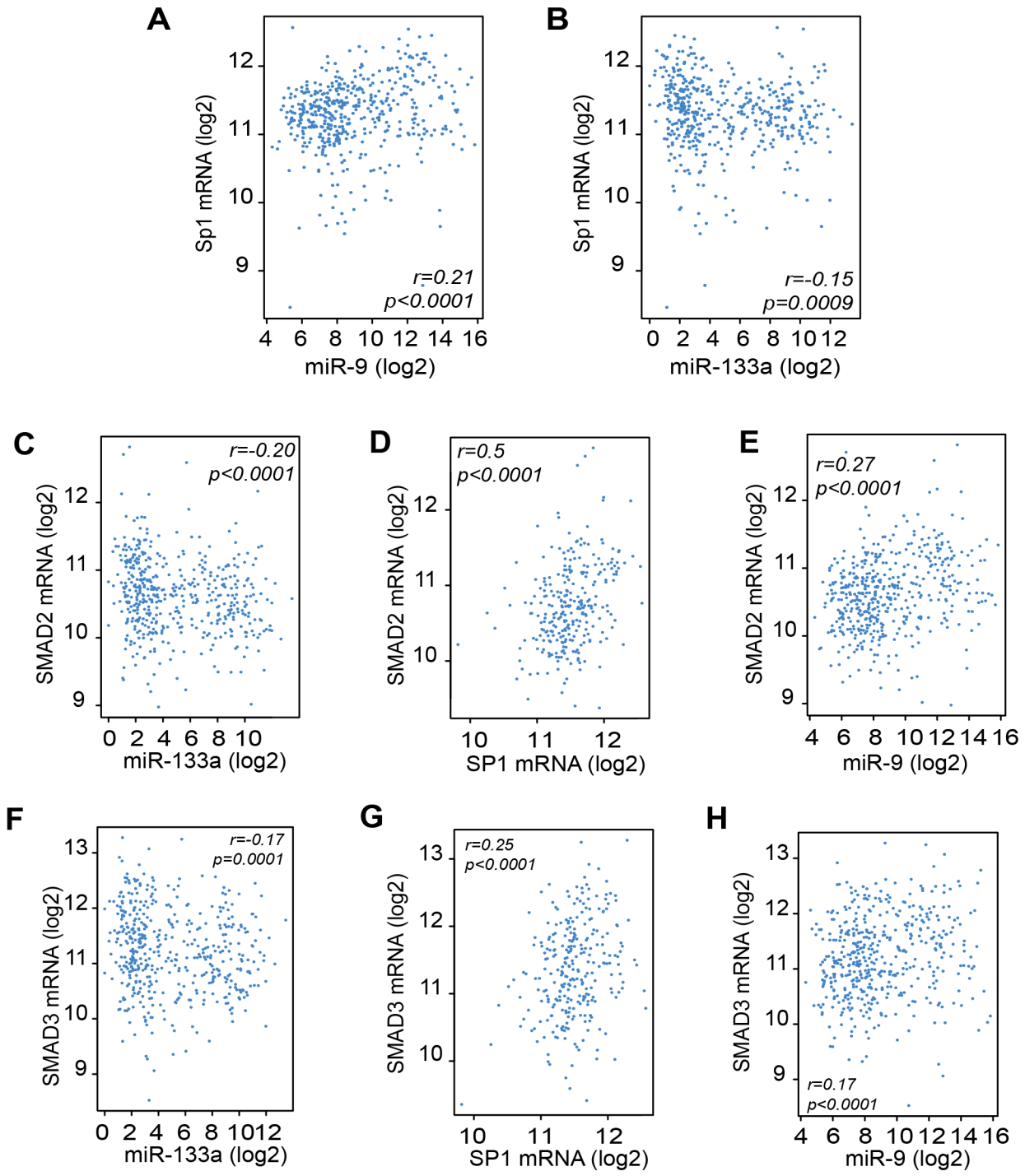

Figure 18. Prognostic significance of miRNA targets in the TCGA HNSCC dataset.
A/B. Dot plots showing the mRNA expression and correlation of SP1 and miR-9 (**A**) or SP1 and miR-133a (**B**) in the HNSCC TCGA dataset. **C/E**. Correlation analyses between SMAD2 and miR-133a (**C**), SP1 (**D**) and miR-9 (**E**) mRNA expression in HNSCC samples included in the TCGA dataset. **F/H**. Correlation analyses between SMAD3 and miR-133a (**F**), SP1 (**G**) and miR-9 (**H**) mRNA expression in HNSCC samples included in the TCGA dataset. Each plot reports the correlation and statistical analyses calculated with nonparametric two-tailed Spearman correlation (*r*) and Statistical significance (*p*).

Moreover, SASH1 expression directly correlated with KRT13 (Fig. 19A) and both were higher in tumour-free cohort (Fig. 19B-C).

Cluster analyses, using the expression of SASH1, KRT13, SP1, TGFβR1, TGFβR2, SMAD2 and SMAD3, divided the HNSCC patients included in the TCGA dataset in two groups: one with low expression of SP1 and members of TGFβ pathway and high expression of KRT13 and SASH1 (Cluster 1) and the other with the opposite gene expression profile (Cluster 2) (Fig. 20A). This clustering had prediction power, since patients included in Cluster 2 displayed worse relapse free survival than patients included in Cluster 1 (HR=2.01; 95% CI 1.1-3.8; p=0.01; Fig. 20B).

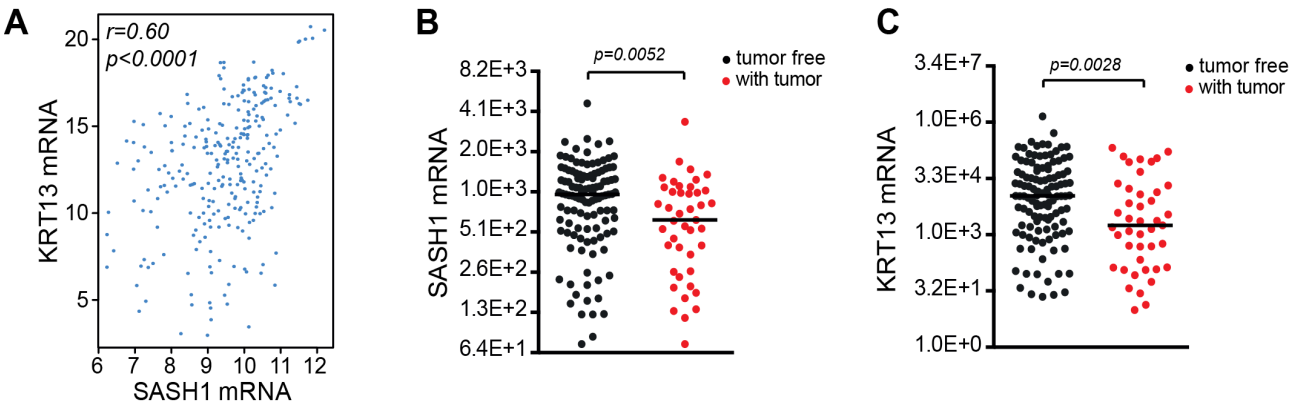

Figure 19. Prognostic significance of miR-9 putative targets in the TCGA HNSCC dataset.
A. Correlation analyses between SASH1 and KRT13 mRNA expressions in HNSCC samples included in the TCGA dataset. Correlation and statistical analyses were calculated with nonparametric two-tailed Spearman correlation (*r*). Statistical significance (*p*) is reported in each plot. **B/C.** Dot plot showing the expression of SASH1 (**B**) and KRT13 (**C**) mRNA in HNSCC samples included in the TCGA dataset and divided based on the presence (red dots) or absence (black dots) of tumor.

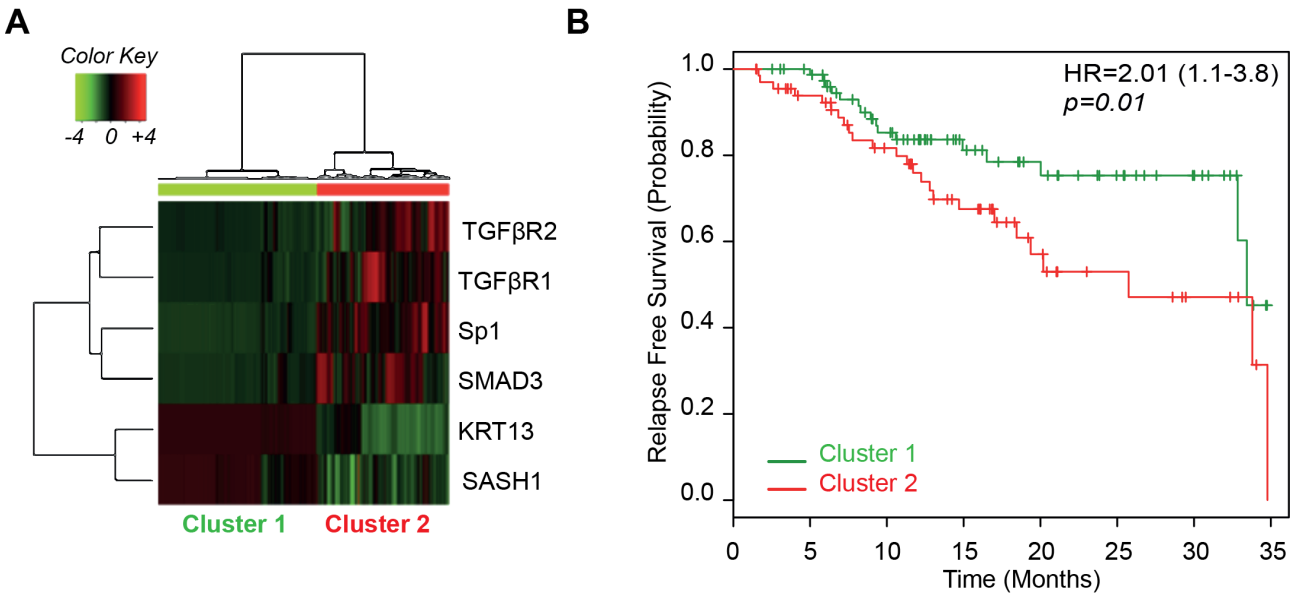

Figure 20. Expression of miRNA targets predict prognosis of HNSCC patients in the TCGA dataset. **A.** Cluster analyses of TGFβ-R1, TGFβ-R2, SP1, SMAD3, KRT13, and SASH1 in HNSCC samples included in the TCGA dataset. **B.** Kaplan–Meier curve evaluating progression free survival of patients with HNSCC divided based on the cluster analysis shown in **A**. HR, confidence interval (between brackets), and significance (P) were evaluated with the log-rank test using the survival package in R.

4.7 Peri-surgical treatment with SP1 and TGFβR1 inhibitors prevents local relapse in a xenograft model of HNSCC.

The above data suggested that SP1 and TGFβ pathway together with miR-9 expression could act synergistically in the establishment of recurrence in HNSCC patients. To evaluate this possibility, we first tested the *in vitro* efficacy of Mithramycin A (MTA), a validated SP1 inhibitor, and of two different TGFβ-R1 inhibitors (SB525334, SB52 and SB431542, SB43) on HNSCC cell survival. While MTA was highly active in decreasing HNSCC cell survival, in the nanomolar range, both SB52 and SB43 did not significantly affect cell survival, when used up to 100 μM (Fig. 21A). Yet, in all tested cell lines SB52, used at the ineffective dose of 40 μM, reduced by the half the IC50 of MTA (Fig. 21A), suggesting that they could have synergistic effects. Accordingly, MTA used at the concentration of 20 nM effectively induced PARP1 cleavage (marker of apoptosis) only in combination with SB52 or SB43 (40 μM) in FaDu cells (Fig. 21B).

A

Cell line	IC50 SB43 (μM)	IC50 SB52 (μM)	IC50 MTA (nM)	IC50 MTA +SB52 (nM)	
FaDu	>100	>100	44	23	***
CAL27	>100	>100	5	1.5	***
UMSCC1	>100	>100	6	3	***
UMSCC74b	>100	>100	15	9	***

*** $p<0.005$

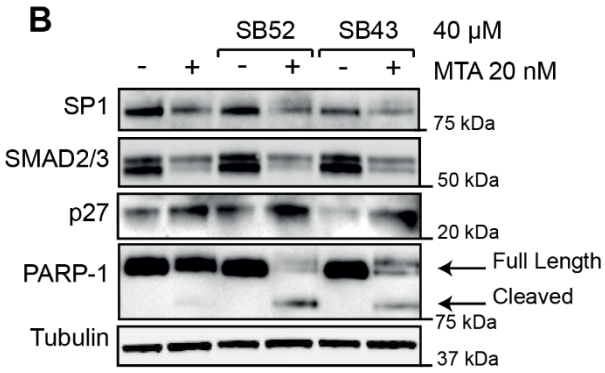

Figure 21. *In vitro* synergistic combination of SP1 and TGFβ pathways blockage.
A. Table reports the IC_{50} values of SB431542 (SB43), SB525334 (SB52), and Mithramycin A (MTA) alone or in combination with SB52 (40 mmol/L) or SB43 (40 mmol/L) in the indicated HNSCC cell lines. Significant difference (***, $p < 0.0001$) between the IC_{50} of MTA, in the presence or not of SB52 or SB43, was calculated with unpaired two-tailed Student t test. **B.** Western blot analysis evaluating the expression of SP1, SMAD2/3 (readouts of MTA activity), p27 (readout of SB52/SB43 activity), and PARP-1 (cleaved form used as marker of apoptosis) in FaDu cells, treated for 24 hours with 20 nmol/L MTA, 40 mmol/L SB52 or SB43, or their combination, as indicated. Tubulin was used as loading control.

Based on these results we tested if combined pharmacological inhibition of SP1 and TGFβ pathways could restrain the formation of local recurrences *in vivo*, using a model of HNSCC recurrences and a peri-surgical treatment schedule (Fig. 22A). Mice were subcutaneously injected with FaDu cells and tumour mass allowed to grow up to ~1cm^3. Mice were then randomly divided in four groups: 1) sham-treated; 2) treated with MTA 1 mg/kg; 3) SB52 15 mg/kg; 4) the combination of MTA and SB52. All groups were treated for 3 consecutive days (day − 1, day 0 and day +1, with respect to surgery) and two more doses (day +3 and +6) only for SB52 treatment. Tumours were surgically removed at day 0 and appearance of local recurrence was monitored over the subsequent 8 weeks of follow-up (Fig. 22A). Pathological analyses of explanted tumours, evaluating the presence and the width of resection margins, demonstrated that no significant differences existed among the four groups in the extent of radical surgery (Fig. 22B)

Explanted tumours (6 from controls and 8/group of treatment, respectively) were analysed for the expression of phosphorylated SMAD2 (pS465-7 SMAD2) (Fig. 22C-D) and SP1 levels (Fig. 22E), as readouts of SB52 and MTA activity, respectively. A single administration of SB52 and/or MTA (administered the day prior to surgery) significantly inhibited SMAD2 phosphorylation and SP1 expression in the tumours (Fig. 22C-E), confirming the *in vivo* efficacy of these drugs.

Control mice developed local recurrence in 50% of injected sites and MTA used alone did not significantly alter the rate of recurrence formation (4/8, 50%) (Fig. 22F). Interestingly, treatment with TGFβ-R1 inhibitor, SB52, as single agent considerably increased local recurrence formation, with recurrence occurring in almost all cases (7/8, 90%) (Fig. 22F). However, the combination of SB52 +

MTA completely prevented recurrence formation (0/8, 0%) (*p=0.005* in log-rank test; HR = 10.01; by comparing control- and MTA+SB treated-mice) (Fig. 22F), suggesting that peri-surgical treatment with MTA and SB52 was sufficient to efficiently suppress HNSCC recurrence *in vivo*.

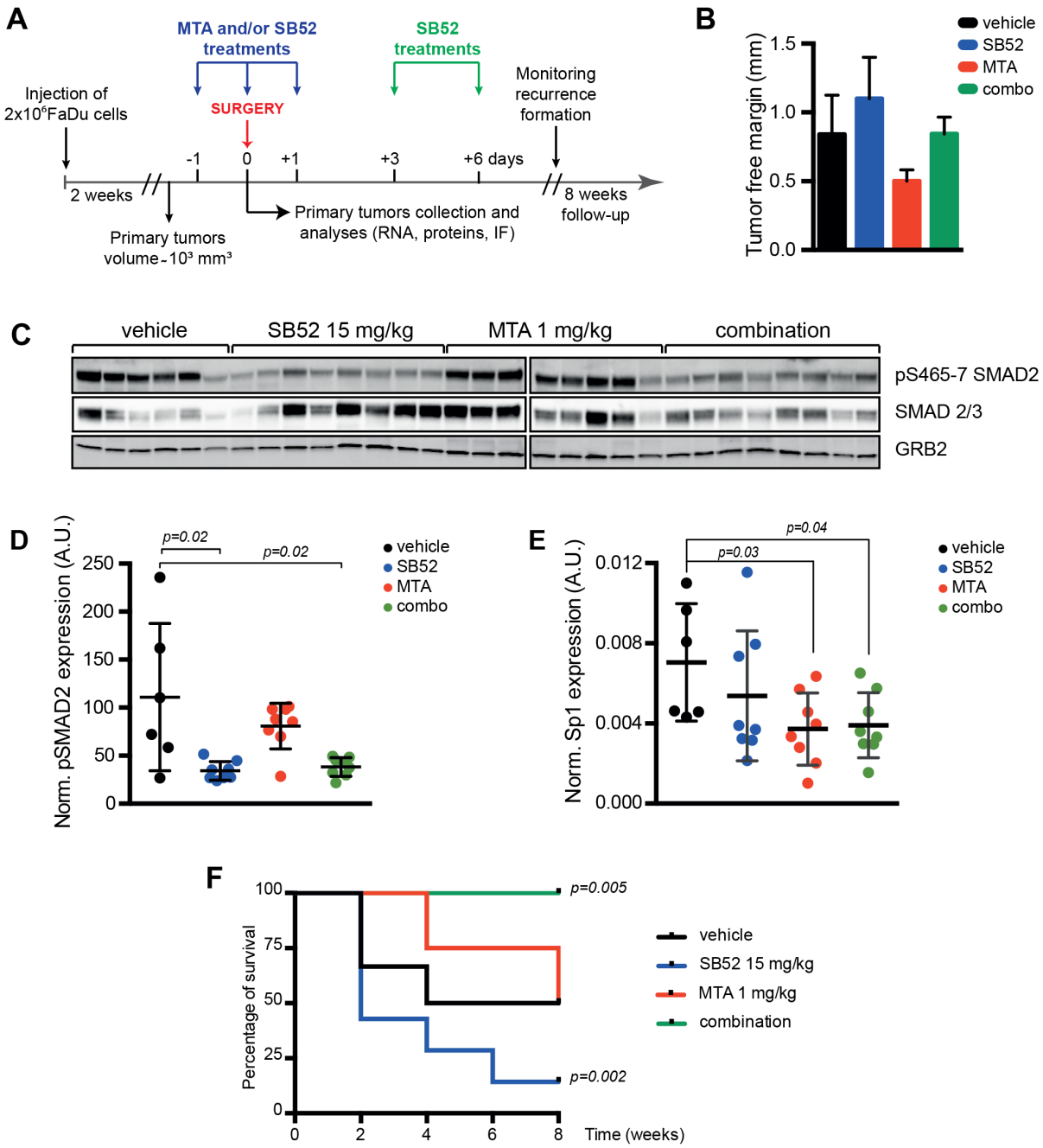

Figure 22. *In vivo* inhibition of SP1 and TGFb pathway effectively reduces recurrence formation in mice. **A.** Schematic representation of the experimental workflow used for the evaluation of recurrence formation in mice xenografted with HNSCC cells. Tumor-bearing mice received treatments at day -1, day 0 and days +1, +3 and +6 with respect to the surgery (day 0), as indicated. Mice have been followed up to 8 weeks to monitor recurrence formation. Control mice were treated with vehicle alone. **B.** Graph reports the evaluation of surgical resection margins in explanted tumors, measured with an optical microscope equipped with an ocular micrometer. Tumor free margins varied from 0.3 to 1.6 mm. No significant differences were observed among the 4 groups of treatment. **C.** Western blot analysis of tumors explanted from mice treated as described in A showing the expression of pSMAD2 (S465/7) and SMAD2/3. GRB2 was used as loading control. **D/E.** Dot plot reporting the normalized expression of pSMAD2 (**D**) or SP1 mRNA (**E**) in tumours explanted at day 0, described in **A**. Each dot represents a primary tumor that received the one dose of vehicle (black) MTA 1 mg/kg (red), SB52 15 mg/kg (blue), or their combination of MTA+SB52 (green). **F.** Disease-free survival of mice subjected to surgery to remove primary tumors (1-1,2 cm^3) and then treated with vehicle (black line), MTA at 1 mg/kg (red line), SB52 15 mg/kg (blue line), or combination of MTA+SB52 (green line). Kaplan-Meier test has been used to calculate the significance of combo-treatment respect to control and to SB52-treated mice.

5. DISCUSSION

Loco-regional and distant relapses are frequent events in HNSCC patients, mainly due to the advanced disease stage at the diagnosis. The appearance of recurrence strongly influences patients' outcome and represents a common and unfavorable prognostic event[1,2,94].

Independently from the etiological risk factors, advanced HNSCC represents a heterogeneous disease, resulting from the accumulation of several genes mutation, leading to the loss of function of important oncosuppressor (*i.e.* TP53) or to the oncogene amplification (*i.e.* EGFR, CCND1), genomic events linked to a poor prognosis[5,16]. Although many studies have been conducted to clarify the molecular mechanisms behind the development of local or distant metastasis, no clear molecular mediators of recurrence formation have been identified so far. In this context, it has been reported that TGFβ pathway alteration, frequently occurring in the advanced stage, plays a pivotal role in the establishment of cell plasticity, eventually leading to a lethal-EMT[43,46,95].

Since, TGFβ has a dual role both oncosuppressive in early tumour stage and oncogenic in the advanced stages, possibly triggering to local and distant relapse, many efforts have been made to translate into clinical practice TGFβ inhibitors for advanced malignance[96]. Unfortunately, TGFβ pathway abrogation has deleterious consequence in term of local relapse, mainly because it inhibits TGFβ-mediated cell apoptosis and strongly increases cell proliferation and invasiveness, through a cell re-program MET[46,97].

Advanced HNSCC, not only displays high percentage of loco-regional or distant relapse, but also the worse response to standard care, even when highly toxic schedule is administered[9,17].

It is not surprising that the two main clinical challenges for this malignance are, one side, the validation of putative biomarker able to classify high-risk patients and, on the other side, the discovery of new potential druggable targets to personalized standard therapy to those subgroup of advanced HNSCC patients.

This PhD project tackled these two unmet clinical needs and aimed to identify a microRNAs signature of loco-regional recurrence in patients with HNSCC.

We started from the profiling of 44 HNSCC primary tumours for miRNAs expression and we showed that four miRNAs can identify patients with HNSCC at high risk of developing recurrence. Intriguingly these 4 miRNAs are deeply involved in the regulation of different "master EMT-genes" involved in the EMT process, both in fibrosis and in solid tumours, in particular in breast cancer.

Using both bioinfarmatic approaches and wet lab techniques, we have linked the balanced activity of these miRNAs to the regulation of EMT, a process considered a critical step in the progression of HNSCC from a pre-invasive to a frankly invasive stage[26,31]. Thus, we identified two principal pathways regulated by the balanced expression of these 4 miRNAs that likely play a role in the regulation of HNSCC recurrence *via* the modulation of EMT, the SP1 and the TGFβ pathways.

On one side, miR-1, miR-133a, and miR-150 synergistically acted to regulate SP1 and, possibly, TGFβ signalling members, two pathways largely involved in cancer progression and EMT. However, miR-9 acted by increasing N-cadherin level, a well-known EMT marker[30], and lowering the expression of SASH1 and KRT13, two known tumour suppressors with potential anti-EMT roles in HNSCC progression[98–100]. Using two independent HNSCC samples cohorts and the TCGA datasets, we confirmed the inverse correlation between the identified miRNAs and their targets' expression, thus validating the six-gene signature as a molecular readout of miRNAs activity.

Since miRNAs are relatively stable molecules, in the clinic-pathological practise their expression can be easily detected directly in the tissue, through *in situ* hybridization (ISH) assay[62,67]. In this context, our finding supports the possibility that expression of miR-1, miR-9, miR-133a, and miR-150 could be estimated directly in primary HNSCC. Alternatively, through immunohistochemistry (IHC) analysis, miRNAs balanced activity could be estimated by the evaluation of their targets' expression (*i.e.* SP1, TGFβ pathway members, SASH1 and KRT13).

In clinic-pathological routine it would be promising the evaluation, as a prognostic marker, of correlation expression between miRNAs and six-gene signature (readout of this balanced activity), more than the sole miRNA expression, to identify patients at high risk of developing recurrence, a subgroup that may benefit for a personalized therapy, avoiding unnecessary toxicity and overcoming possible resistance to standard care.

From a biological point of view, it is interesting to note that we describe here two new targets of miR-9, SASH1 and KRT13, both already linked to the regulation of cell motility and invasion[98–100]. Although the aim of this study was limited to the assessment of their role as potential readouts of high miR-9 activity *in vivo*, it will be worth testing if SASH1 and KRT13 may partially mediate the miR-9-induced local relapse in HNSCC.

Moreover, it has been recently demonstrated that miR-9, targeting E-Cadherin and γ-Catenin respectively in breast cancer and SCC, directly promotes EMT and cancer stem cell phenotype[80,88]. In light of these literature data, we focused on the study of miR-9 role in the establishment of recurrence formation and/or therapy resistance in HNSCC. Our preliminary data clearly indicated that miR-9 expression, even if it did not affect cell proliferation (data not shown), is necessary for the maintenance of survival capabilities and resistance to γ-radiation and EGFR-targeted therapy. Possibly, miR-9 expression could modulate this EMT phenotype also acting on SP1 level, since we demonstrated a strong correlation expression between miR-9 and SP1, both *in vitro* and *in vivo* (Figure 13D and 18A). Further experiments will be performed to investigate the molecular mechanism that orchestrates this phenotype and to dissect the role of miR-9 contribution also *in vivo*. Importantly, our data point to the simultaneous inhibition of SP1 and TGFβ pathway by the concerted

action of three miRNAs, miR-1, miR-133a, and miR-150, as a possible mechanism to prevent disease relapse. At a first glance, our data do not fit in the current literature showing that, in mouse models, the inhibition of TGFβ pathway combined with K-Ras activation is linked to the onset of SCC in the skin and in the oral mucosa[47,88]. However, increasing evidences clearly show that, depending on the stages of tumour progression, TGFβ pathway can exert either pro- or anti-tumorigenic effects[95,96]. For instance, increased TGFβ signalling, in benign tumours or during the course of cancer induction, selects for more aggressive cells and contributes to metastasis formation, in different models of HNSCC[42]. Our data in the mouse model, showing that TGFβ-R1 inhibition could result in different outcomes, depending on the simultaneous inhibition of SP1 by MTA or not (Figure 22F), confirm the hypothesis that TGFβ pathway could act as tumour suppressor or tumour promoter in a context-dependent manner. In strict accord with our data, SP1 is required for TGFβ-induced EMT in pancreatic cancer[101], and SP1 and SMAD2 proteins have been reported to directly interact in different models[102,103], supporting the possibility that SP1 and TGFβ pathway may act together to drive EMT, local invasion and, eventually, recurrence formation in HNSCC. The experimental and *in silico* analyses performed on our and on TCGA samples support this possibility and point to the expression of miR-1, miR-133a, and miR-150 as possible switches of the TGFβ activity.

We are aware that our study has limitations that should be taken into account. The cohort of patients used as a discovery set contains a relatively small number of heterogeneous patients that received different post-surgery treatments. These variables could have an impact on the prognostic value of the miRNA signature and also prevented the possible evaluation of its independent prognostic role in multivariate analyses. Although we experimentally validated the correlation between miRNAs expression and their targets, we could not test the potential prognostic value of these miRNAs in a second independent cohort of samples for the absence of a precise follow-up in this group of patients. Furthermore, it is important to point out that we focused our bioinformatic analyses on the relation between miRNAs expression and EMT regulators, and this represents a possible limitation of our study, since other biological pathways are also significantly altered by the same miRNAs in HNSCC. In perspective, it will be important to verify if these four miRNAs and/or six genes signatures could be prospectively validated to identify patients at high risk of recurrence who may merit to be treated with specific targeted therapies.

The significant synergistic activity of the combined inhibition of SP1 and TGFβ-R1 and the critical role of miR-9 in promoting resistance to both radiotherapy and EGFR-targeted therapy, strongly suggest the urgency to translate our finding to the clinic, this is of potential immediate translational relevance, since both SP1 and TGFβ inhibitors have been already tested in cancer patients.

6. REFERENCES

1. Kamangar, F., Dores, G. M. & Anderson, W. F. Patterns of Cancer Incidence, Mortality, and Prevalence Across Five Continents: Defining Priorities to Reduce Cancer Disparities in Different Geographic Regions of the World. *J. Clin. Oncol.* 24, 2137–2150 (2006).
2. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* 127, 2893–2917 (2010).
3. Cloos, J. *et al.* Genetic susceptibility to head and neck squamous cell carcinoma. *JNCI J. Natl. Cancer Inst.* 88, 530–535 (1996).
4. Ang, K. K. *et al.* Human Papillomavirus and Survival of Patients with Oropharyngeal Cancer. *N. Engl. J. Med.* 363, 24–35 (2010).
5. Poeta, M. L. *et al.* TP53 Mutations and Survival in Squamous-Cell Carcinoma of the Head and Neck. *N. Engl. J. Med.* 357, 2552–2561 (2007).
6. Husain, H. *et al.* Nuclear epidermal growth factor receptor and p16 expression in head and neck squamous cell carcinoma. *The Laryngoscope* 122, 2762–2768 (2012).
7. Smeets, S. J. *et al.* Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus. *Oncogene* 25, 2558 (2005).
8. Wiest, T., Schwarz, E., Enders, C., Flechtenmacher, C. & Bosch, F. X. Involvement of intact HPV16 E6/E7 gene expression in head and neck cancers with unaltered p53 status and perturbed pRb cell cycle control. *Oncogene* 21, 1510 (2002).
9. Leemans, C. R., Braakhuis, B. J. & Brakenhoff, R. H. The molecular biology of head and neck cancer. *Nat. Rev. Cancer* 11, 9–22 (2011).
10. Vergeer, M. R. *et al.* Intensity-Modulated Radiotherapy Reduces Radiation-Induced Morbidity and Improves Health-Related Quality of Life: Results of a Nonrandomized Prospective Study Using a Standardized Follow-Up Program. *Int. J. Radiat. Oncol.* 74, 1–8 (2009).
11. Hutchinson, L. Cetuximab or cisplatin in HNSCC? *Nat. Rev. Clin. Oncol.* 13, 66 (2015).
12. Nör, C. *et al.* Cisplatin Induces Bmi-1 and Enhances the Stem Cell Fraction in Head and Neck Cancer. *Neoplasia* 16, 137-W8 (2014).
13. Blanchard, P. *et al.* Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. *Radiother. Oncol.* 100, 33–40 (2011).
14. Tabor, M. P. *et al.* Persistence of Genetically Altered Fields in Head and Neck Cancer Patients. *Clin. Cancer Res.* 7, 1523 (2001).
15. Califano, J. *et al.* Genetic Progression Model for Head and Neck Cancer: Implications for Field Cancerization. *Cancer Res.* 56, 2488 (1996).

16. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* 333, 1157–1160 (2011).
17. Hammerman, P. S., Hayes, D. N. & Grandis, J. R. Therapeutic insights from genomic studies of head and neck squamous cell carcinomas. *Cancer Discov.* 5, 239–244 (2015).
18. Hama, T. *et al.* Prognostic Significance of Epidermal Growth Factor Receptor Phosphorylation and Mutation in Head and Neck Squamous Cell Carcinoma. *The Oncologist* 14, 900–908 (2009).
19. Khan, H. *et al.* Correlation between expressions of Cyclin-D1, EGFR and p53 with chemoradiation response in patients of locally advanced oral squamous cell carcinoma. *BBA Clin.* 3, 11–17 (2015).
20. Moreira, J., Tobias, A., O'Brien, M. P. & Agulnik, M. Targeted Therapy in Head and Neck Cancer: An Update on Current Clinical Developments in Epidermal Growth Factor Receptor-Targeted Therapy and Immunotherapies. *Drugs* 1–15 (2017).
21. Magrini, S. M. *et al.* Cetuximab and radiotherapy versus cisplatin and radiotherapy for locally advanced head and neck cancer: a randomized phase II trial. *J. Clin. Oncol.* 34, 427–435 (2015).
22. Huang, J., Zhang, J., Shi, C., Liu, L. & Wei, Y. Survival, recurrence and toxicity of HNSCC in comparison of a radiotherapy combination with cisplatin versus cetuximab: a meta-analysis. *BMC Cancer* 16, 689 (2016).
23. Bonner, J. A. *et al.* Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *Lancet Oncol.* 11, 21–28 (2010).
24. Mehra, R., Cohen, R. B. & Burtness, B. A. The Role of Cetuximab for the Treatment of Squamous Cell Carcinoma of the Head and Neck. *Clin. Adv. Hematol. Oncol. HO* 6, 742–750 (2008).
25. Tsai, J. H. & Yang, J. Epithelial–mesenchymal plasticity in carcinoma metastasis. *Genes Dev.* 27, 2192–2206 (2013).
26. Thiery, J. P. Epithelial–mesenchymal transitions in tumour progression. *Nat. Rev. Cancer* 2, 442 (2002).
27. Nieto, M. A. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science* 342, 1234850 (2013).
28. Masuda, M. *et al.* Stat3 orchestrates tumor development and progression: the Achilles' heel of head and neck cancers? *Curr. Cancer Drug Targets* 10, 117–126 (2010).
29. Kojc, N. *et al.* Transcription factors Snail, Slug, Twist, and SIP1 in spindle cell carcinoma of the head and neck. *Virchows Arch.* 454, 549–555 (2009).

30. Nguyen, P. T. *et al.* N-cadherin expression is correlated with metastasis of spindle cell carcinoma of head and neck region. *J. Oral Pathol. Med.* 40, 77–82 (2011).
31. Thiery, J. P. & Sleeman, J. P. Complex networks orchestrate epithelial–mesenchymal transitions. *Nat. Rev. Mol. Cell Biol.* 7, 131–142 (2006).
32. Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat. Med.* 19, 1438 (2013).
33. Haslehurst, A. M. *et al.* EMT transcription factors snail and slug directly contribute to cisplatin resistance in ovarian cancer. *BMC Cancer* 12, 91 (2012).
34. Hsu, D. S.-S. *et al.* Regulation of excision repair cross-complementation group 1 by Snail contributes to cisplatin resistance in head and neck cancer. *Clin. Cancer Res.* 16, 4561–4571 (2010).
35. Ota, I. *et al.* Snail-induced EMT promotes cancer stem cell-like properties in head and neck cancer cells. *Oncol. Rep.* 35, 261–266 (2016).
36. Kemler, R. *et al.* Stabilization of β-catenin in the mouse zygote leads to premature epithelial-mesenchymal transition in the epiblast. *Development* 131, 5817–5824 (2004).
37. Oloumi, A., McPhee, T. & Dedhar, S. Regulation of E-cadherin expression and β-catenin/Tcf transcriptional activity by the integrin-linked kinase. *Biochim. Biophys. Acta BBA-Mol. Cell Res.* 1691, 1–15 (2004).
38. Islam, S., Carey, T. E., Wolf, G. T., Wheelock, M. J. & Johnson, K. R. Expression of N-cadherin by human squamous carcinoma cells induces a scattered fibroblastic phenotype with disrupted cell-cell adhesion. *J. Cell Biol.* 135, 1643–1654 (1996).
39. Jolly, M. K. *et al.* Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front. Oncol.* 5, (2015).
40. Chen, C., Zimmermann, M., Tinhofer, I., Kaufmann, A. M. & Albers, A. E. Epithelial-to-mesenchymal transition and cancer stem (-like) cells in head and neck squamous cell carcinoma. *Cancer Lett.* 338, 47–56 (2013).
41. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* 331, 1559–1564 (2011).
42. Ikushima, H. & Miyazono, K. TGFβ signalling: a complex web in cancer progression. *Nat. Rev. Cancer* 10, 415 (2010).
43. Inman, G. J. Switching TGFβ from a tumor suppressor to a tumor promoter. *Genet. Cell. Mech. Oncog.* 21, 93–99 (2011).
44. David, C. J. *et al.* TGF-β Tumor Suppression through a Lethal EMT. *Cell* 164, 1015–1030 (2016).

45. Bian, Y. *et al.* Progressive tumor formation in mice with conditional deletion of TGF-β signaling in head and neck epithelia is associated with activation of the PI3K/Akt pathway. *Cancer Res.* 69, 5918–5926 (2009).
46. Honjo, Y. *et al.* TGFβ Receptor I Conditional Knockout Mice Develop Spontaneous Squamous Cell Carcinoma. *Cell Cycle* 6, 1360–1366 (2007).
47. Bornstein, S. *et al.* Smad4 loss in mice causes spontaneous head and neck cancer with increased genomic instability and inflammation. *J. Clin. Invest.* 119, 3408–3419 (2009).
48. Izeradjene, K. *et al.* KrasG12D and Smad4/Dpc4 haploinsufficiency cooperate to induce mucinous cystic neoplasms and invasive adenocarcinoma of the pancreas. *Cancer Cell* 11, 229–243 (2007).
49. Truty, M. J. & Urrutia, R. Basics of TGF-ß and Pancreatic Cancer. *Pancreatology* 7, 423–435 (2007).
50. Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* 11, 228–234 (2009).
51. Lee, Y., Jeon, K., Lee, J., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* 21, 4663 (2002).
52. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23, 4051–4060 (2004).
53. Han, J. *et al.* Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell* 125, 887–901 (2006).
54. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* 17, 3011–3016 (2003).
55. Hutvágner, G. *et al.* A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. *Science* 293, 834 (2001).
56. Kawamata, T., Seitz, H. & Tomari, Y. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat. Struct. Mol. Biol.* 16, 953–960 (2009).
57. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235–240 (2004).
58. Mulrane, L., McGee, S. F., Gallagher, W. M. & O'Connor, D. P. miRNA dysregulation in breast cancer. *Cancer Res.* 73, 6554–6562 (2013).
59. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105 (2009).
60. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* 11, 597–610 (2010).

61. Witwer, K. W. Circulating microRNA biomarker studies: pitfalls and potential solutions. *Clin. Chem.* 61, 56–63 (2015).
62. Iorio, M. V. & Croce, C. M. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO Mol. Med.* 4, 143–159 (2012).
63. Zavadil, J., Narasimhan, M., Blumenberg, M. & Schneider, R. J. Transforming growth factor-β and microRNA: mRNA regulatory networks in epithelial plasticity. *Cells Tissues Organs* 185, 157–161 (2007).
64. Villanueva, A. *et al.* Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. *Oncogene* 31, 2062 (2012).
65. Avissar, M., Christensen, B. C., Kelsey, K. T. & Marsit, C. J. MicroRNA Expression Ratio Is Predictive of Head and Neck Squamous Cell Carcinoma. *Clin. Cancer Res.* 15, 2850 (2009).
66. Babu, J. M., Prathibha, R., Jijith, V. S., Hariharan, R. & Pillai, M. R. A miR-centric view of head and neck cancers. *Biochim. Biophys. Acta BBA - Rev. Cancer* 1816, 67–72 (2011).
67. Jamali, Z. *et al.* MicroRNAs as prognostic molecular signatures in human head and neck squamous cell carcinoma: A systematic review and meta-analysis. *Oral Oncol.* 51, 321–331 (2015).
68. Corcoran, C. D., Senchaudhuri, P., Mehta, C. R. & Patel, N. R. Exact inference for categorical data. *Encycl. Biostat.* (2005).
69. Hall, M. *et al.* The WEKA data mining software: an update. *ACM SIGKDD Explor. Newsl.* 11, 10–18 (2009).
70. Klement, W., Wilk, S., Michalowski, W. & Matwin, S. Classifying Severely Imbalanced Data. in *Advances in Artificial Intelligence: 24th Canadian Conference on Artificial Intelligence, Canadian AI 2011, St. John's, Canada, May 25-27, 2011. Proceedings* (eds. Butz, C. & Lingras, P.) 258–264 (Springer Berlin Heidelberg, 2011). doi:10.1007/978-3-642-21043-3_31
71. Shirdel, E. A., Xie, W., Mak, T. W. & Jurisica, I. NAViGaTing the Micronome – Using Multiple MicroRNA Prediction Databases to Identify Signalling Pathway-Associated MicroRNAs. *PLOS ONE* 6, e17429 (2011).
72. Pavón, M. A. *et al.* Gene expression signatures and molecular markers associated with clinical outcome in locally advanced head and neck carcinoma. *Carcinogenesis* 33, 1707–1716 (2012).
73. Brown, K. R. *et al.* NAViGaTOR: network analysis, visualization and graphing Toronto. *Bioinformatics* 25, 3327–3329 (2009).
74. Brown, K. R. & Jurisica, I. Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.* 8, R95 (2007).

75. Rahmati, S., Abovsky, M., Pastrello, C. & Jurisica, I. pathDIP: an annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis. *Nucleic Acids Res.* 45, D419–D426 (2017).
76. The Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 517, 576 (2015).
77. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* 6, pl1 (2013).
78. Aguirre-Gamboa, R. & Trevino, V. SurvMicro: assessment of miRNA-based prognostic signatures for cancer clinical outcomes by multivariate survival analysis. *Bioinformatics* 30, 1630–1632 (2014).
79. Chan, W.-C. *et al.* MetaMirClust: discovery of miRNA cluster patterns using a data-mining approach. *Genomics* 100, 141–148 (2012).
80. Ma, L. *et al.* miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat. Cell Biol.* 12, 247 (2010).
81. Liu, Y.-N. *et al.* MiR-1 and miR-200 inhibit EMT via Slug-dependent and tumorigenesis via Slug-independent mechanisms. *Oncogene* 32, 296 (2012).
82. Muraoka, N. *et al.* MiR-133 promotes cardiac reprogramming by directly repressing Snai1 and silencing fibroblast signatures. *EMBO J.* 33, 1565–1581 (2014).
83. Yokobori, T. *et al.* MiR-150 is associated with poor prognosis in esophageal squamous cell carcinoma via targeting the EMT inducer ZEB1. *Cancer Sci.* 104, 48–54 (2013).
84. Haddad, R. I. & Shin, D. M. Recent advances in head and neck cancer. *N. Engl. J. Med.* 359, 1143–1154 (2008).
85. Pastrello, C. *et al.* Visual Data Mining of Biological Networks: One Size Does Not Fit All. *PLOS Comput. Biol.* 9, e1002833 (2013).
86. Reis, P. P. *et al.* A gene signature in histologically normal surgical margins is predictive of oral carcinoma recurrence. *BMC Cancer* 11, 437 (2011).
87. Citron, F. *et al.* An Integrated Approach Identifies Mediators of Local Recurrence in Head and Neck Squamous Carcinoma. *Clin. Cancer Res.* 23, 3769 (2017).
88. White, R. A. *et al.* Epithelial stem cell mutations that promote squamous cell carcinoma metastasis. *J. Clin. Invest.* 123, 4390–4404 (2013).
89. Krause, M., Dubrovska, A., Linge, A. & Baumann, M. Cancer stem cells: Radioresistance, prediction of radiotherapy outcome and specific targets for combined treatments. *Radiother. Cancer Present Future* 109, 63–73 (2017).

90. Abdullah, L. N. & Chow, E. K.-H. Mechanisms of chemoresistance in cancer stem cells. *Clin. Transl. Med.* 2, 3–3 (2013).
91. Marcu, L. G. Altered fractionation in radiotherapy: From radiobiological rationale to therapeutic gain. *Cancer Treat. Rev.* 36, 606–614
92. Sak, A. & Stuschke, M. Use of γH2AX and Other Biomarkers of Double-Strand Breaks During Radiotherapy. *Harnessing DNA Repair Improve Radiother. Outcome* 20, 223–231 (2010).
93. Olive, P. L. Retention of γH2AX foci as an indication of lethal DNA damage. *Radiother. Oncol.* 101, 18–23
94. Michiels, S. *et al.* Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data. *Lancet Oncol.* 10, 341–350 (2009).
95. Glick, A. B. The Role of TGF Signaling in Squamous Cell Cancer: Lessons from Mouse Models. *J. Skin Cancer* 2012, (2012).
96. Bierie, B. & Moses, H. L. Tumour microenvironment: TGFβ: the molecular Jekyll and Hyde of cancer. *Nat. Rev. Cancer* 6, 506–520 (2006).
97. Qiu, W., Schönleben, F., Li, X. & Su, G. H. Disruption of Transforming Growth Factor β-Smad Signaling Pathway in Head and Neck Squamous Cell Carcinoma as Evidenced by Mutations of SMAD2 and SMAD4(1). *Cancer Lett.* 245, 163–170 (2007).
98. Zeller, C. *et al.* SASH1: a candidate tumor suppressor gene on chromosome 6q24. 3 is downregulated in breast cancer. *Oncogene* 22, 2972–2983 (2003).
99. Sun, D. *et al.* SASH1 inhibits proliferation and invasion of thyroid cancer cells through PI3K/Akt signaling pathway. *Int. J. Clin. Exp. Pathol.* 8, 12276–12283 (2015).
100. Yanagawa, T. *et al.* Loss of cytokeratin 13 expression in squamous cell carcinoma of the tongue is a possible sign for local recurrence. *J. Exp. Clin. Cancer Res. CR* 26, 215–220 (2007).
101. Jungert, K. *et al.* Sp1 Is Required for Transforming Growth Factor-β–Induced Mesenchymal Transition and Migration in Pancreatic Cancer Cells. *Cancer Res.* 67, 1563 (2007).
102. Docagne, F. *et al.* Sp1 and Smad transcription factors co-operate to mediate TGF-β-dependent activation of amyloid-β precursor protein gene transcription. *Biochem. J.* 383, 393–399 (2004).
103. Feng, X.-H., Lin, X. & Derynck, R. Smad2, Smad3 and Smad4 cooperate with Sp1 to induce p15(Ink4B) transcription in response to TGF-β. *EMBO J.* 19, 5178–5193 (2000).