POST-TRANSLATIONAL MODIFICATIONS ON THE K2 DOMAIN MODULATE

SOX9 TRANSCRIPTIONAL ACTIVITY IN LUNG CANCER

By

MICHELE PATRIZII

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Cellular and Molecular Pharmacology

Written under the direction of

Sharon R. Pine and Hatem E. Sabaawy

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October 2019

ABSTRACT OF THE DISSERTATION

Post-translational modifications on the K2 domain modulate SOX9

transcriptional ability in lung cancer

by

MICHELE PATRIZII

Dissertation Directors:

Sharon R. Pine and Hatem E. Sabaawy

Lung cancer is the deadliest cancer type, accounting for about 25% of all cancer-related deaths. Despite impressive advancements in the molecular characterization of lung cancer, more than 70% of lung cancer patients are still left with either no known clinically relevant driver mutation or with mutations for which there isn't a targeted therapeutic option available. Thus, there is a great need to understand the underlying biology of this disease and to discover the molecular pathways driving its onset and progression. SOX9 is a transcription factor involved in several processes during embryonic development, and its role in promoting cancer-related features in lung cancer has been established. Here, we characterized the role of post-translational modifications in mediating SOX9-associated functions in the context of lung cancer, with a particular focus on modifications occurring on the K2 domain. We demonstrated that the K2 domain

is able to modulate SOX9 transcriptional ability and that changes in the phosphorylation status of specific residues within K2 affect this mechanism. We were able to show that the K2 domain might represent a region of intrinsic disorder within SOX9, and that post-translational modifications are involved in the maintenance of this disorder, potentially resulting in the modulation of SOX9 ability to interact with binding partners. These results uncover the prominent role of the K2 domain in mediating specific SOX9 functions, thus contributing to dissect the mechanistic details of SOX9 impact on lung cancer onset and progression.

Acknowledgements

My sincerest gratitude goes to my mentors, Dr. Sharon R. Pine and Dr. Hatem E. Sabaawy, for the continuous support and guidance they provided during my Ph.D. training. I could perceive their deep love for research in every discussion we had, and I feel truly fortunate in having had the opportunity to appreciate how their different perspectives have in common the same passion for science. I hope that one day I will be able to train someone with the same dedication, attention and commitment they showed to me during these years.

Additionally, I would like to thank the members of my thesis committee, Dr. Joseph R. Bertino, Dr. Suzie Chen and Dr. Wenwei Hu, for their insightful comments and thoughtful observations, which unquestionably were a significant contribution for the realization of this project. My gratitude is also extended to all the other Professors I had the pleasure to interact with, including Dr. Vikas Nanda and Dr. Hossein Khiabanian, who kindly shared their knowledge with me and were not hesitant in providing help and guidance.

I also want to thank all my fellow lab mates, from both labs, who helped me to navigate through these 5 years and who eased the burden during hard times, even just with a smile or a joke. Particularly, I would like to thank Monica Bartucci, Katie Flaherty, Eric Huselid, Shamila Yusuff and Mohamed Hussein from Sabaawy lab, and Katie Morgan, Irina Teplova, Jamie Shah, Hua Zhong and Gina Castellano from Pine lab. I want you all to know that I truly cherish the moments we spent together, and that I sincerely wish the best, professionally and personally, to every

one of you. Thank you also to all the fellow students, post-docs and staff with whom I didn't physically share a lab but that were a source of discussion and laughs, including Siddharth Marwaha and my friend Luca Tottone.

Moreover, my thank goes to the Rutgers School of Graduate Studies and to all the great people involved in the Ph.D. program, especially to all the Professors who taught me during the program. Their dedication to students was unquestionable, and they were able to transmit their passion for science in an amazing way.

Finally, I would like to thank my parents, I am sure that this achievement means to them as much as it means to me. Thank you to my sister and my gorgeous nieces, and thank you to my extended family in Italy.

At last, a thank coming for the heart of my hearth to the most important people of my life, my beautiful wife Monica and my kids Liam and Emma. You are simply the inspiration of and the reason for everything I do. My hope is to make you proud of me every day of my life.

Table of Contents

Results

List of Illustrations

Introduction

Lung cancer

Cancer represents the second leading cause of death worldwide, preceded only by cardiovascular diseases. Cancer was globally responsible for almost 10 million deaths in 2018 [1], and it is estimated that in 2019 there will be over 600,000 cancer-related deaths in the United States [2]. While prostate and breast cancers are the most prevalent types of neoplasia in men and women, respectively, lung cancer is the deadliest one in both sexes, accounting for about 25% of all cancer-related deaths [2].

Histologically, lung cancer is classified as small cell lung cancer (SCLC) (about 15% of cases) or non-small cell lung cancer (NSCLC) (85% of cases) (*Fig.1*) [3, 4].

Figure 1. Histological classification of lung cancer.
Histological classification of lung cancer.
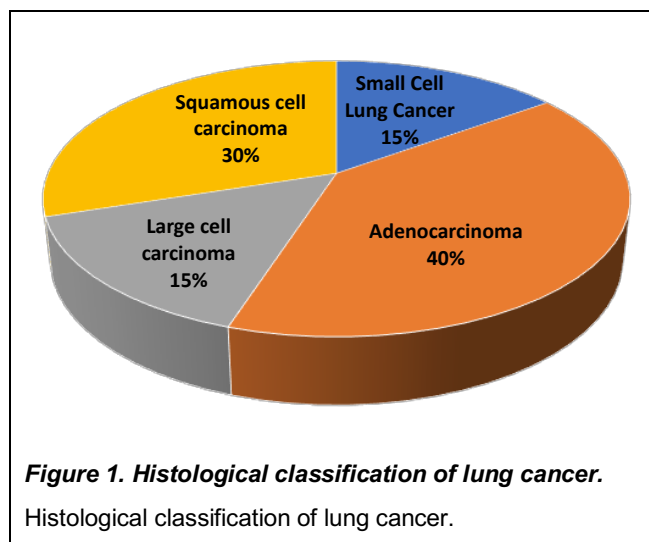
NSCLC is furtherly distinguished into Adenocarcinoma, Squamous cell carcinoma and Large cell carcinoma.

Adenocarcinoma is the most prevalent histologic subtype, accounting for about 40% of all lung cancer cases [5].

Despite advances in the management of non-small cell lung cancer patients, long-term survival remains very low. In the last three decades, the 5-year-relative survival rate has only slightly improved (14.3%, 15.5%, and 18.4%, respectively)

[6, 7]. One of the reasons for these poor survival rates is that frequently the diagnosis occurs in patients with an advanced-stage disease. Approximately two thirds of NSCLC patients present with a locally advanced (Stage III) or metastatic (Stage IV) disease [8], making therapeutic options less effective.

Therapeutic management of NSCLC

There are essentially 5 therapeutic options for the treatment of NSCLC patients [9]:

1. Surgery
2. Chemotherapy
3. Radiation therapy
4. Targeted therapy
5. Immunotherapy

The optimal choice of a treatment regimen depends on many factors, including the stage at which the disease is presented, the overall health status of the patient and the histological subtype.

For Stages I through IIIA, surgery is the treatment of choice, often combined with chemotherapy and/or radiation as neoadjuvant (usually to shrink the tumor before the surgery) or adjuvant therapies (after the surgery, to maximize its effect and prevent relapse of the disease).

For patients with locally advanced or metastatic NSCLC, surgery is usually not a resolutive option. These cases are treated with a multi-modal approach combining chemotherapy with other therapeutic options.

The most common cytotoxic chemotherapy drugs prescribed for NSCLC patients are Taxols (such as Paclitaxel and Docetaxel), Platins (Carboplatin and Cisplatin) and Topoisomerase inhibitors (Etoposide and Irinotecan).

One of the therapeutic approaches for late stage NSCLC is targeted therapy, employing small molecule drugs or monoclonal antibodies to target the genetic aberrations identified as potential drivers of tumor progression.

Currently, the genetic profiling of NSCLC includes more than 15 frequently mutated genes [10] that could be exploited for treatment or used for patients classification (*Fig.2*).

Figure 2. Molecular classification of NSCLC.
Molecular classification of NSCLC. Image from [11] .
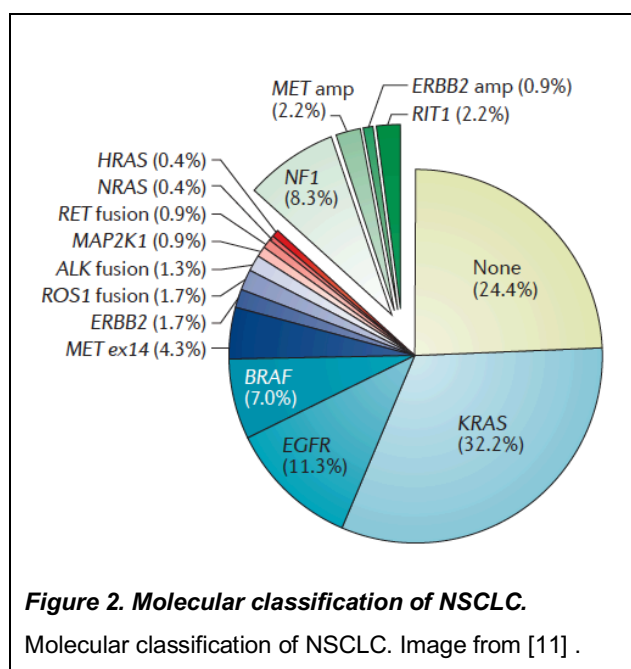
These genetic aberrations encompass EGFR (11% of NSCLC cases), KRAS (32%) and BRAF (7%), among the most common ones. However, the availability of molecular-guided drugs is mainly restricted to EGFR, BRAF and ALK/ROS1 fusions. Thus, more than 70% of NSCLC patients (including the many cases with KRAS mutations) are left with either no known clinically relevant driver mutation or with mutations for which there isn't a targeted therapeutic option available.

In recent years, cancer immunotherapy has surged as an effective option for cancer treatment [12]. Based on the concept of boosting the immune system

against cancer cells, immunotherapy treatments have been shaped in different forms. One of the most commonly used type of immunotherapy is the immune checkpoint inhibitors (ICIs) therapy. The ICIs approach aims at preventing the inhibitory effect of cancer cells on T-cells function, thus allowing T-cells to maintain their tumor-killing action. ICIs-based therapies have been proved efficacious on many different cancer types, including NSCLC. In 2018, the ICI Keytruda (pembrolizumab) has granted full approval from the Food and Drug Administration (FDA) as a first-line treatment (in combination with standard chemotherapy) for patients with metastatic NSCLC lacking EGFR or ALK mutations. However, despite the promising results, immunotherapy does not represent a universal therapeutic option for NSCLC patients. Moreover, its effectiveness varies according to many factors (such as PD-1/PD-L1 expression, the overall molecular profiling of the tumor and the availability of immune cells at the tumor site, to name a few), some of which are not fully understood yet [13].

Hence, while there are different options for treating NSCLC patients, there is still a great need to further understand the underlying biology of lung cancer and to discover the molecular pathways driving its onset and progression.

Embryonic signaling pathways as a target in cancer therapy

During embryonic development, cell fate determination and tissue patterning are orchestrated through the highly regulated expression of specific signaling pathways [14, 15]. These embryonic signaling pathways governing cellular proliferation, migration and differentiation during embryonic development are also frequent drivers of adult malignancy, including cancer [16]. The aberrant activation

of pathways such as Hedgehog, Notch, Wnt, BMP and SOX outside the context of embryogenesis determines the deregulated expression of transcription factors controlling networks of genes, thus concurring to tumor development and progression (Fig.3).

Figure 3. Embryonic pathways in development and tumorigenesis.
Schematic representation of the role of embryonic pathways during development (*left*) and tumorigenesis (*right*).
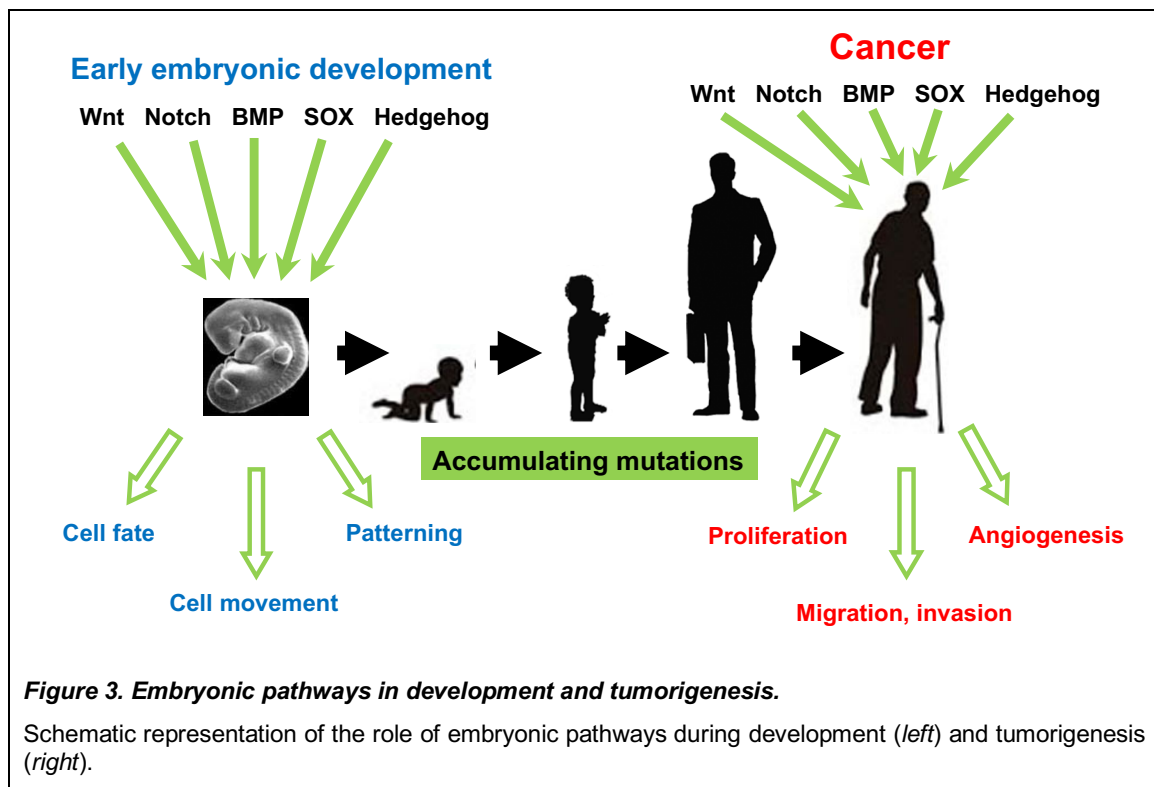
The Hedgehog pathway, for instance, is essential for both embryonic development and organ homeostasis during adult life [17, 18]. Activation of this pathway leads to the expression of target genes involved in proliferation, angiogenesis and stem-cell renewal [19]. However, deregulated and persistent activation of Hedgehog signaling through various mechanisms has been linked to different types of cancer, including prostate, breast, glioma and melanoma [20-23]. Importantly, a conspicuous number of clinical trials are evaluating the impact of targeting Hedgehog signaling in an heterogenous panel of cancers [24].

Interestingly, there are increasing evidence pointing to frequent cross-talks between different embryonic pathways. For instance, the role of Hedgehog pathway in the onset of basal cell carcinoma (BCC) through activation of the glioma-associated oncogene (GLI) requires concomitant activation of the Wnt pathway [25]. Although cross-talks between the two pathways have been observed at different levels, the mechanisms governing it have not been fully elucidated yet [26].

Given the pivotal role of the embryonic pathways, a better understanding of their fine regulation will contribute to deepen our knowledge about the biology of cancer and to advance the generation of additional therapeutic strategies for the treatment of many cancer types, including lung cancer [27, 28].

SOX family of genes

Among the embryonic signaling pathways, of particular interest to our lab is the SOX pathway. SOX is a family of 20 genes encoding for transcription factors, arose from the founding member Sry, the mammalian testis-determining factor [29, 30]. The members of the SOX family are distributed in 9 groups, according to the similarity of their high mobility group (HMG) box domain (*Fig.4*). They function by activating or repressing transcription of target genes, and are implicated in many important biological processes, especially in the differentiation of distinct lineages during embryonic development. Particularly, SOX members are involved in sex determination, neurogenesis, neural crest development, skeletogenesis and hematopoiesis [31].

Figure 4. Structure of human SOX proteins.

Image from [30] indicating the various regions and functional domains of SOX family members.
CC: Coiled coil domain; HMG: High-mobility-group box DNA-binding domain; TAD: Transactivation domain; TRD: Transrepression domain.
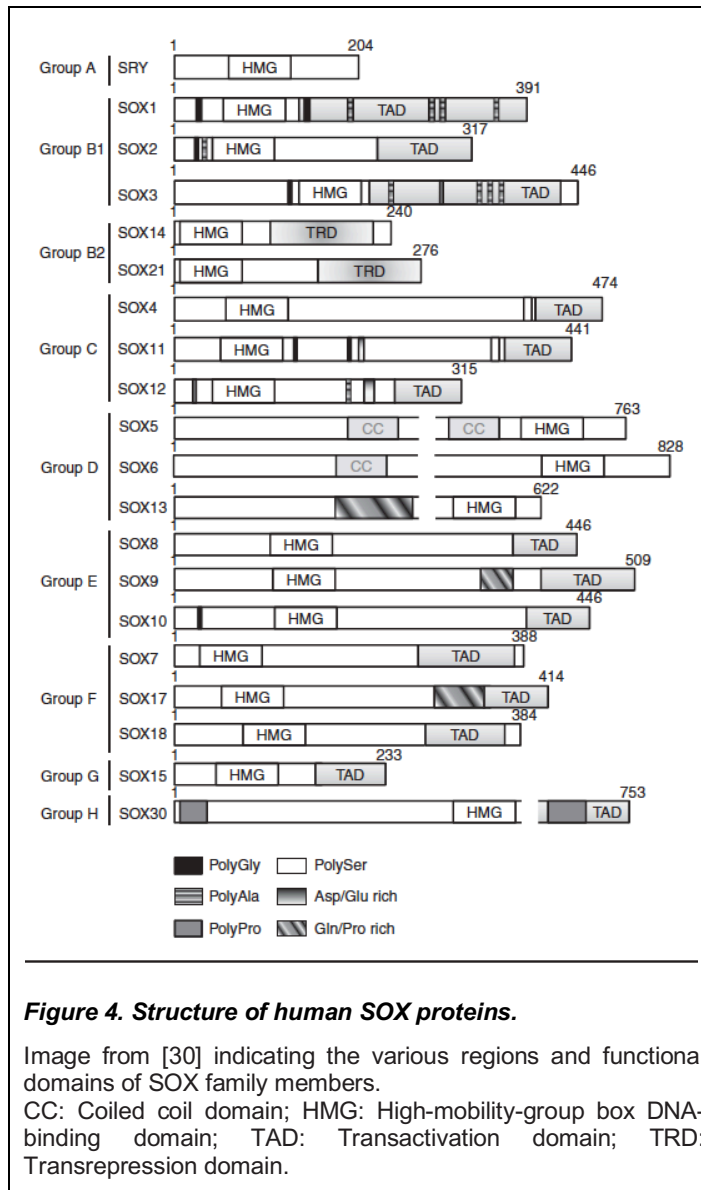
Genetic aberrations or deficiency in the expression of SOX family members during embryonic development is associated to several human diseases, including X-linked and sex-determining syndromes (*SRY, SOX3, SOX9*), neural-related disorders (*SOX2, SOX10*) and congenital skeletal malformation (*SOX9*) [32-36].

Many SOX genes have been proven to play a role in cancer development. The first clue in that sense was the demonstration of the causative role of SOX3 in the oncogenic transformation of chicken embryonic fibroblasts [37]. These finding were confirmed in subsequent studies, where SOX3 was identified as a proto-oncogene in a genome-based analysis of retroviral insertion sites in mouse T-cell lymphomas [38]. Since then, there have been many examples of correlative studies linking SOX genes expression with a broad range of tumors [30, 39].

For instance, SOX2 expression has been associated with tumor-initiating cells (a subpopulation of cancer cells endowed with stemness potential and chemoresistance ability) both in breast cancer [40, 41] and in glioma [42-44]. In colorectal cancer, SOX2 expression has been associated with poor prognosis and metastatic potential [45].

Among the Group C of SOX genes, SOX4 has been found to be upregulated in human acute leukemia [46] and in lung neuroendocrine tumors [47]. Moreover, SOX4 expression participates in colorectal oncogenesis [48, 49], while its knockdown induces apoptosis and growth suppression [50, 51].

SOX11 is a highly specific marker for mantle cell lymphoma (MCL), and it is instrumental in the differential diagnosis of MCL from other B-cell lymphomas [52]. Interestingly, one of the most represented SOX members in the list of genes associated with several different cancer types is SOX9, belonging to the SOX E group [30].

SOX9 in development

SOX9 is a key regulator for the proper embryonic development of ectodermal, mesodermal and endodermal derivatives (*Fig.5*) [53].

During mesodermal development, SOX9 acts, in concert with SRY, as the master regulator of testis formation. The process is started by SRY (localized on the Y chromosome) and is then completed by SOX9 through a dual transactivation mechanism. On one hand, SOX9 promotes the expression of Prostaglandin D2, resulting in the specification of the Sertoli lineage. On the other hand, SOX9

upregulates the production of anti-Mullerian hormone (AMH), thus preventing the development of the female Mullerian ducts [54, 55].

Figure 5. SOX9 in development.

Image from [53]. SOX9 expression in ectodermal, mesodermal and endodermal derivatives, fetal progenitors and adult stem/progenitor cells.
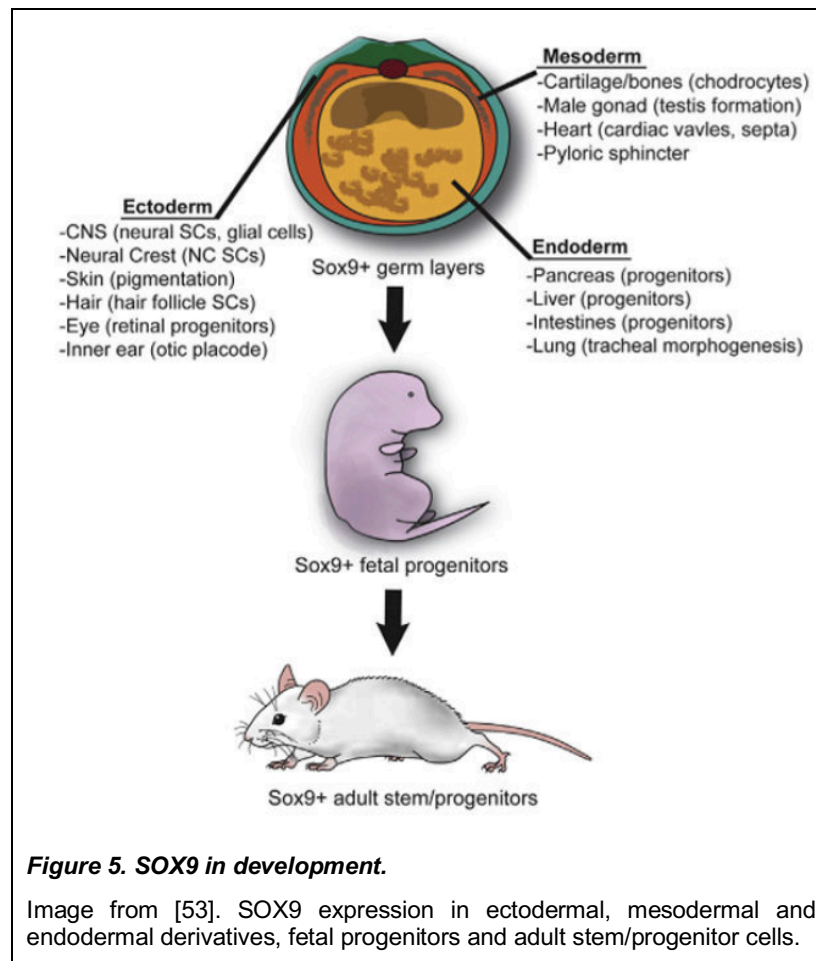
Another mesodermal event regulated by SOX9 is chondrogenesis, a process in which mesenchymal cells condense and differentiate into chondrocytes (one of the elements composing the skeletal system). SOX9 is essential for mesenchymal condensation, and it elicits its activity by differentially regulating many genes involved in chondrocytes proliferation such as COL2A1 and COL10A1, among the others [56, 57].

During ectodermal development, SOX9 participates in shaping the central nervous system (CNS) and the neural crest (NC). In CNS formation, SOX9 expression determines the specification of the two main types of glial cells (oligodendrocytes and astrocytes) through the regulation of key metabolic and migratory genes in

astrogliogenesis [58, 59]. Moreover, SOX9 plays a pivotal role in NC development by promoting the epithelial-mesenchymal transition (EMT) of neural crest cells, a process necessary for the migration of these cells to the periphery of the neural crest [60].

SOX9 is also involved in the proper development of endodermal derivates, by being expressed in the progenitor cells of pancreas [61, 62], liver [63] and intestine [64].

Additionally, precise spatial and temporal control of SOX9 expression is essential for proper lung morphogenesis. In mice, SOX9 is expressed at high levels in the proximal tip buds of the branching epithelial tubes at E14.5, and it disappears immediately before birth, when the lung progenitor cells differentiate into alveolar cells [65, 66]. Other studies highlighted the crucial role of SOX9 in tracheal development [67].

After development, SOX9 expression is maintained in adult stem and progenitor cells. For instance, SOX9 is responsible for the maintenance of multipotent neural progenitor cells (NPCs) in the CNS throughout adult life [68]. Additionally, SOX9 contribute to the intestinal epithelium homeostasis. It is expressed in the cells composing the intestinal crypts, such as stem cells, Paneth cells and a subset of transit-amplifying cells [69]. SOX9 inactivation will determine a general dysplasia of the intestine, characterized by the disappearance of Paneth cells and a decrease in the goblet cell lineage [64]. Furthermore, SOX9 promotes repression of the CDX2 and MUC2 genes, normally expressed in the mature villus cells, thus contributing to the Wnt-dependent maintenance of intestinal progenitor cells [70].

SOX9 in developmental and acquired diseases

The main clinical manifestation of SOX9 haploinsufficiency (caused by SOX9 gene disruption) during development is Campomelic Dysplasia (CD), a rare skeletal dysmorphology syndrome characterized by improper growth of long bones, male-to-female sex reversal in XY genotypes and congenital heart defects. Lifespan for patients affected by this disorder is generally restricted to few months, mainly because of the respiratory problems caused by improper tracheal and lung development [71].

Besides developmental disorders, SOX9 is implicated in the pathology of fibrotic diseases, characterized by excessive and/or displaced extracellular matrix (ECM) deposition in response to injury. This is not surprising, given the prominent role of Sox9 in ECM deposition, as evidenced in chondrogenesis. Particularly, aberrant expression of SOX9 is one of the causative agents of liver fibrosis [72]. Upon liver damage, hepatic-stellate cells (HSCs) proliferate into myofibroblasts in a SOX9-dependent mechanism, and then migrate to the parenchymal cells to secrete ECM components for repair. Inappropriate Sox9 expression will determine an abnormal deposition of ECM, thus provoking destruction of the tissue architecture and eventually jeopardizing the organ function.

SOX9 in cancer

The role of SOX9 in cancer has been an active area of research in recent years. SOX9 expression is upregulated in several cancer types (*Fig.6*), and it has been proven to act as an oncogene in many of them.

Particularly, over-expression of SOX9 mRNA has been associated with poor clinical outcome in patients with malignant gliomas [73]. Furthermore, a recent study proved that SOX9 is essential for glioma stem cells (GSCs) self-renewal and chemo-resistance by promoting the expression of pyruvate dehydrogenase kinase 1 (PDK1) [74].

Figure 6. SOX9 is highly expressed in many cancer types.

Percentage of patients with moderate to high SOX9 expression in the different cancer types assessed by immunohistochemistry (IHC). Source: Human Protein Atlas
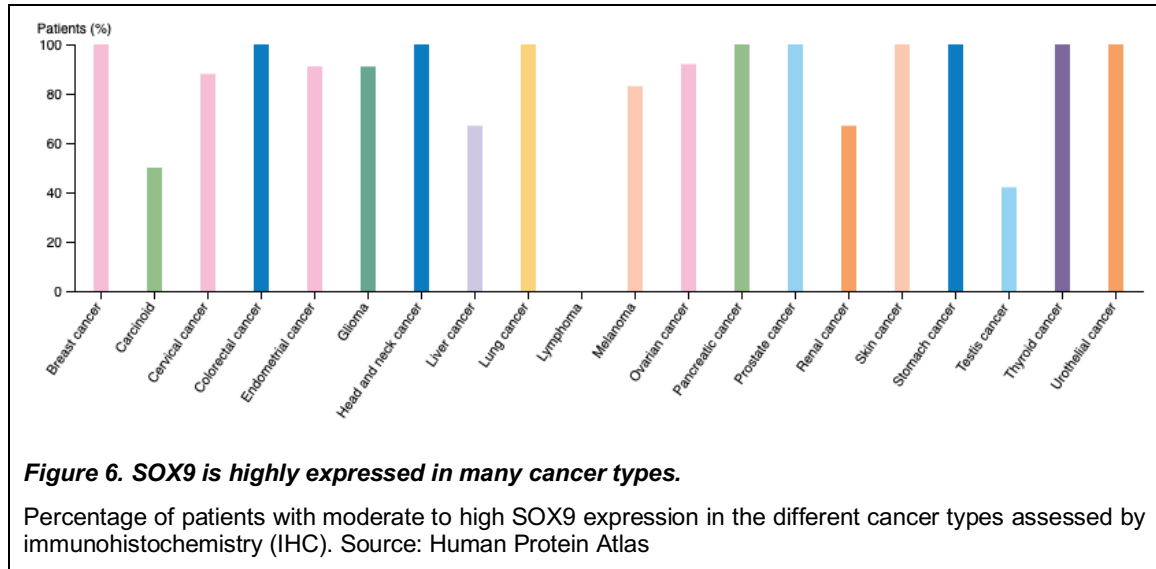
Similarly, the etiology of prostate cancer is influenced by SOX9. In mice with a heterozygous deletion in the PTEN gene (a common mutation found in prostate cancer), over-expression of SOX9 accelerates the onset of high-grade prostate intraepithelial neoplasia (PIN), indicating cooperation between these two players in promoting the disease progression [75]. Another study identified SOX9 as a key determinant of PTEN loss-induced prostate tumorigenesis by eliciting its transcriptional activity on genes such as MIA, involved in cell invasion, and H19, encoding for an RB-targeting microRNA [76]. Additionally, over-expression of SOX9 in mice xenografts from human cancer cell lines has been found to enhance

tumor growth, angiogenesis and tumor invasion, possibly by transactivating the androgen receptor [77].

In pancreatic adenocarcinoma, SOX9 over-expression has been localized to the bottom part of the crypts, suggesting that the neoplasia could arise from dysregulation of the stem cell homeostasis [78]. Additionally, SOX9 accelerates the formation of the KRAS-induced precursor lesions of pancreatic ductal adenocarcinoma [79], and its over-expression has been linked to a NF-kB-mediated epigenetic regulation [80].

In breast cancer, co-expression of SOX9 and the transcription factor SLUG is sufficient to promote tumorigenicity and metastatic propensity in breast cancer cells, and it has been proposed as a poor prognostic marker [81, 82]. A recent paper highlighted the interplay between SOX9 and FXYD3 for the maintenance of ER$^+$ breast cancer stem cells (CSCs) functions [83], and another study identified SOX9 as the main determinant of breast cancer endocrine resistance [84].

SOX9 has been identified as a CSC marker in hepatocellular carcinoma, and has been shown to act in feedback-regulated mechanisms involving the Wnt and NOTCH pathways and their downstream targets OSTEOPONTIN and NUMB [85, 86].

Moreover, a recent meta-analysis of the survival data for more than 3300 solid tumor patients extrapolated from 17 different studies identified SOX9 as a poor prognosis factor across multiple cancer types [87].

While all these scientific findings corroborate the idea that SOX9 functions as an oncogene, there are cancer types in which SOX9 role is still controversial. For

instance, while SOX9 is expressed in over 80% of melanomas, there are studies postulating a tumor-suppressor-like role of SOX9 in this context, based on decrease in cell proliferation (through direct upregulation of the cell cycle arrest gene p21) and restoration of drug sensitivity in both cell lines and xenografts upon SOX9 over-expression [88]. However, a recent study attempted to reconcile the different positions about the role of SOX9 in melanoma by suggesting that SOX9 influence on metastatic propensity is dose-dependent. While sub-optimal levels of SOX9 promote an anti-metastatic phenotype, high levels of SOX9 stimulate metastases in a heterogenous population of melanoma [89].

Colorectal cancer (CRC) is another example of a cancer type in which SOX9 role is still debatable. In fact, many publications attribute an oncogenic role to SOX9 because of its ability to mediate processes such as proliferation, EMT and therapy resistance [90-94] and its correlation with poor prognosis [95, 96]. However, Blache and his collaborators proposed an interesting model that conciliates a potential SOX9 anti-oncogenic role and its high expression levels in CRC. They discovered a different isoform of SOX9, called MiniSOX9, resulting from an alternative splicing event that generates a truncated version of the canonical Sox9, deprived of the carboxy-terminal domains [97]. According to their model, the observed high SOX9 levels in CRC are actually represented by MiniSOX9 that, acting as dominant negative towards the canonical SOX9, inhibits the tumor-suppressor activity the canonical SOX9 is endowed with [98, 99].

SOX9 in NSCLC

The role of SOX9 in NSCLC is devoid of ambiguity, because of the many studies conducted in the recent years which clearly attributed an oncogenic role to SOX9 in this context. The first indications in that sense came from a study showing a significant upregulation of SOX9 in NSCLC, in which SOX9 role in promoting p21 expression and downregulating CDK4 was revealed [100]. A few years later, an analysis of the clinical significance of SOX9 over-expression in NSCLC demonstrated that SOX9 expression correlates with tumor grade and overall survival, thus potentially presenting SOX9 as a prognostic marker in NSCLC [101]. After these initial findings, many other publications contributed to identify the mechanisms through which SOX9 elicits its oncogenic role in NSCLC. Particularly, our group was among the first to demonstrate a functional role for SOX9 in inducing a mesenchymal phenotype in lung adenocarcinoma [102]. We identified SOX9 as a key mediator for NOTCH-induced cell motility and invasion, and we proved that SOX9 negatively regulates E-CADHERIN expression, a key protein lost during EMT. Following our results, other groups confirmed the association between SOX9 expression levels and tumor grade in lung adenocarcinoma, and corroborated the notion that SOX9 is a key inducer of cell proliferation, migration and invasion [103]. Other studies identified an interplay between the tumor-associated macrophages (TAMs) and cancer cells involving SOX9. According to the authors' findings, SOX9 expression in NSCLC cancer cells is stimulated by the secretion of TGF-beta by TAMs, determining the C-jun/SMAD3 pathway activation in cancer cells that is responsible for the high SOX9 levels [104]. SOX9 was also proved to prevent

apoptosis in NSCLC cell lines, acting as a mediator of the Wnt/beta-catenin pathway [105]. Similarly to what have been described for breast cancer [81], association between SOX9 and SLUG plays an important role in NSCLC. Particularly, SLUG controls SOX9 stability by preventing its degradation, thus allowing SOX9 to promote CSC maintenance and metastases formation [106]. Additionally, a gene expression profiling study identified SOX9, together with few other genes, as a potential mediator for resistance to the tyrosine kinase inhibitor (TKI) Erlotinib in EGFR mutated NSCLC cell lines [107]. Taken together, these studies provide compelling evidence about the oncogenic role of SOX9 in NSCLC.

Sox9 structure and functions

SOX9 gene is located at chromosomal position 17q24.3 and it encodes for a protein of 509 aminoacids (aa). SOX9 protein shares structural features with the other two components of the SOXE family, SOX8 and SOX10 (*Fig.7*) [53, 108].

On the amino-terminal side, SOX9 contains a DNA binding domain (HMG), which allows it to bind to its consensus motifs (A/TA/TCAAA/TG) on the minor groove of the DNA determining a bend to the DNA double helix with an angle variable between 30° and 113° [109]. Embedded in the HMG domain are two independent nuclear localization sequences (NLS) and a nuclear export sequence (NES). On the same side, Sox9 also possess a dimerization domain (DIM), since most of its functions are carried out through homodimerization [109]. A central K2 domain and a transactivation domain (TA) on the carboxy-terminus are also shared among the SOXE family members. Unique to SOX9 is the presence of a subdomain rich in proline and glutamine residues (PQA).

Figure 7. Structure of SOXE proteins.

Image modified from [53]. Structural organization of the different SOXE proteins. DIM: dimerization domain; HMG: high mobility group DNA binding domain. For SOX9, aa sequence of the HMG domain is provided, containing the 2 NLS sequences (blue) and the NES (red).
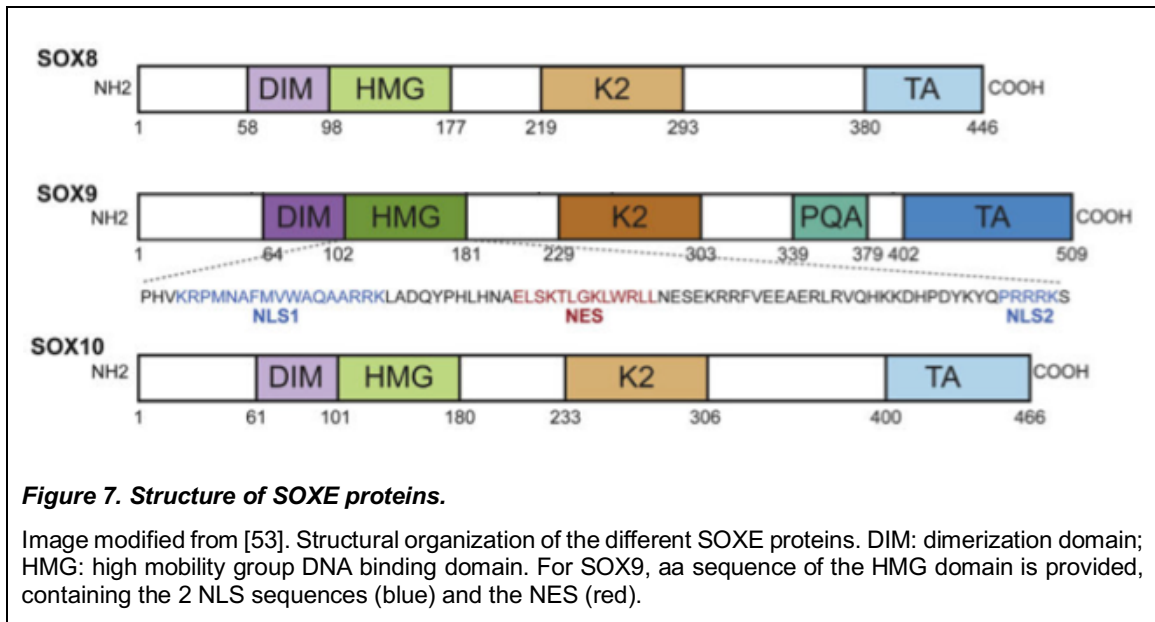
Most the information about SOX9 mechanism of action are derived from developmental studies, particularly in chondrogenesis and CNS development. In this context, SOX9 shares some redundant functions with the other two SOXE members. For instance, separate deletions of either SOX9 or SOX10 will not significantly affect the proper oligodendrocyte development, while that is not the case when both genes are concomitantly deleted [110]. However, the contribution of different SOXE members might differ temporally and in magnitude, in a tissue-specific fashion [111, 112].

SOX9 regulatory functions on target genes are carried out through association between SOX9 homodimers and other binding partners, usually different transcription factors or other SOX proteins, thus forming a complex that would stabilize SOX9 binding to the DNA and that could subsequently recruit additional co-activators or co-repressors [109]. For example, during chondrogenesis, SOX9 binding to its consensus sequence on COL2A1 gene is stabilized by the interaction with the transcription factor SMAD3, bound to a nearby enhancer region. The

formation of the SOX9/SMAD3 complex will determine recruitment of the co-activator CREB-binding protein (CBP/p300), ultimately allowing transcription of the gene [113]. The effect of SOX9 on transcriptional modulation of target genes could be positive (activation) or negative (repression) depending of many factors, including the target site, the binding partners, and the co-activators/co-repressors recruited on the site (*Fig.8A*). For instance, while association with SMAD3 and the co-activator CBP/p300 results in COL2A1 transcription, as just described, the association on the same COL2A1 gene of SOX9 with a different transcription factor, ELF3, determines recruitment of the same CBP/p300 co-activator but results in the opposite effect (repression of gene transcription) [114].

Figure 8. Effect of binding partners on SOX9 activity.

Image modified from [53]. *A.* Scheme of SOX9 requirement for binding partners to elicit its functions. *B.* Negative (up) and positive (down) effects on SOX9 transcriptional modulation of COL10A1 and COL2A1 genes upon interaction with different binding partners.
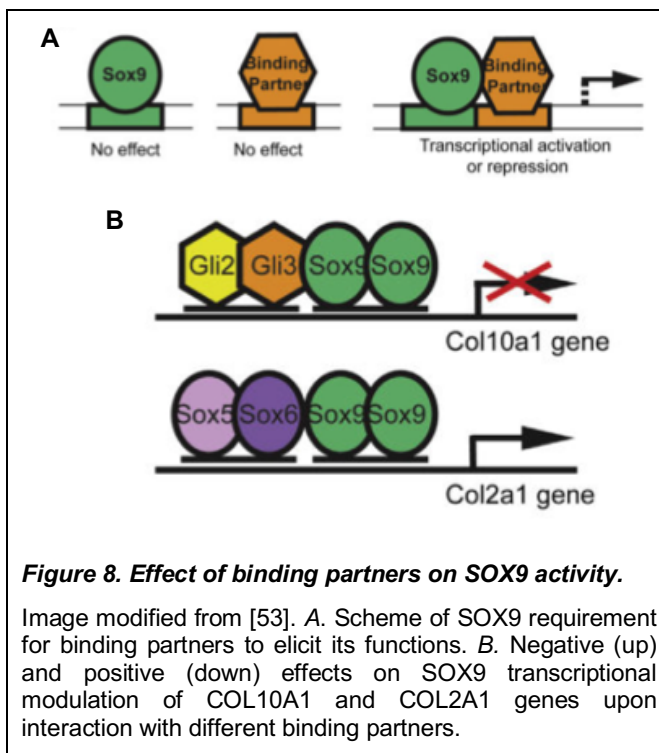
Similar examples are frequent when describing SOX9 activity. For instance, during the same developmental process (chondrocytes maturation), association between the SOX9 homodimer and components of the GLI family will repress transcription of COL10A1 gene. Conversely, the same SOX9 homodimer on a different target site (COL2A1 gene) will activate gene transcription by associating with other SOX members bound to nearby DNA sites (*Fig.8B*) [57].

All these findings clearly indicate that the effects of SOX9 on transcriptional modulation, as well as SOX9 regulation, are strictly context-specific, and most likely this aspect is not restricted to development but can also be applied to SOX9 role in cancer.

Besides its canonical role of transcription factor, SOX9 has been proven to be involved in the modulation of protein stability. Particularly, SOX9 promotes lysosomal degradation of the transcription factor RUNX2 [115] and antagonizes the Wnt/β-catenin pathway with a dual mechanism; through its amino-terminal part, SOX9 directly promotes β-catenin degradation, while through its carboxy-terminal part it inhibits β-catenin transcriptional activity by affecting its nuclear translocation [116].

SOX9 regulation

Similarly to our understanding of SOX9 mechanism of action, most of our knowledge about SOX9 regulation is derived from developmental studies. In this setting, many signaling pathways have been identified as potential upstream regulators of SOX9 expression (through various mechanisms), such as Hedgehog, Wnt/b-catenin, Notch, FGF and TGF-beta [117].

Particularly interesting is the role of post-translational modifications in modulating SOX9 stability, intracellular localization and its overall activity. During testis formation, phosphorylation at the S64 and S181 sites by the cAMP-dependent protein kinase A (PKA) promotes SOX9 nuclear translocation, thus allowing Sertoli cells differentiation [118]. Surprisingly, the same phosphorylation events are also responsible for the enhanced SOX9 ability to transactivate COL2A1 gene during

chondrocytes maturation [119]. Phosphorylation at the S211 site by p38 (through a TGF-beta-mediated mechanism) stabilizes SOX9 protein during chondrogenesis [120]. Post-translational modification by SUMOylation (addition of ubiquitin-related modifiers) has been also linked to modulation of SOX9 activity during chondrogenesis, although there are contrasting reports about the significance of these SUMOylation events on COL2A1 gene modulation [121, 122].

While the role of SOX9 post-translational modifications has been extensively studied for developmental processes, this is not the case for cancer.

Sox9 post-translational modifications in cancer

When considering the role of SOX9 post-translational modifications in the context of cancer, our knowledge is limited to few phosphorylation events (spread throughout SOX9 sequence), registered mainly in breast cancer, by studies looking at global phosphorylation patterns through phosphoproteomic analysis [123-125]. Moreover, these studies didn't provide experimental validation of the role of those phosphorylation events on SOX9 stability, function or localization. A tentative to attribute a role to SOX9 phosphorylation events was made by Schaab and his collaborators. By employing quantitative mass spectrometry analysis, they evaluated changes in the global phosphoproteomic profile of a panel of NSCLC cell lines resistant to the TKI Dasatinib [126]. Through this approach, they identified a number of phosphorylations on several molecules, including phosphorylation on S199 in SOX9, as potential mediators of Dasatinib resistance.

Our group was among the firsts to discover and experimentally validate the role of specific SOX9 post-translational modifications in cancer. We identified a novel

regulatory mechanism of SOX9 stability in response to genotoxic stress that could have implications in cancer therapy resistance [127]. In detail, we found that in response to UV irradiation or genotoxic chemotherapeutics, SOX9 is degraded in the ubiquitin proteasome system. This degradation is mediated by the interaction between SOX9 and FBW7, an E3 ubiquitin ligase that functions in the DNA damage response pathway. The interaction between SOX9 and FBW7 is dependent on a post-translational modification on SOX9, specifically on the phosphorylation of residue T236 on the K2 domain promoted by the kinase GSK3beta. Once the interaction occurred, another post-translational modification (in this case a polyubiquitination) targets SOX9 for degradation in the proteasome. Our findings were confirmed by an independent study, in which the authors validated our results and suggested a role for the SOX9/FBW7 interaction in modulating the SOX9 oncogenic role in medulloblastoma, one of the most common childhood brain tumors [128].

Intriguingly, the post-translational modification at the center of our above-described research is located on the K2 domain, which spans from aa 229 to aa 303 (*Fig.7*). This domain is particularly interesting because its role has not been fully elucidated yet [62]. While for other SOXE family members the K2 domain has been endowed with potential transactivation functions [129, 130], in SOX9 the TA domain (aa 402-509) is considered the main transactivating component [131]. To date, the only role attributed to the K2 domain is the mediation of SOX9 stability through the FBW7-mediated mechanism described above. Interestingly, a recent

study based on the analysis of co-evolutionary patterns of the different SOX9 domains suggested that the K2 domain could potentially cooperate with the canonical transactivation domain as well as with the dimerization domain in mediating SOX9 differential transcriptional ability on target genes [132].

Intrinsic disordered proteins

Recent improvements in technology and new findings in the structural biology field led to challenging the scientific dogma that a protein function depends on its three-dimensional (3D) structure. In fact, there are many examples of proteins that, despite being perfectly functional, are unable to fold into a stable 3D structure. This particular macromolecules are called intrinsic disordered proteins (IDPs) [133]. Instead of having a defined 3D structure, they are dynamic entities, fluctuating rapidly through a range of different conformations, going from extended coils to collapsed globules [134]. The propensity of a protein to be classified as an IDP can be assessed by specific features of its aa sequence, such as an overall low complexity (i.e. over-representation of specific residues), a relatively low content of bulky hydrophobic aa and a relatively high proportion of polar and charged aa, instead. While a relative low number of proteins are totally disordered, it is believed that the majority of eukaryotic proteins contain intrinsically disorder regions (IDRs) within an otherwise well-structured globular domains [135]. Intrinsic disorder is functionally relevant because it confers conformational flexibility to the protein, which facilitates the different conformational requirements for binding to other macromolecules. As a result, the same polypeptide can undertake different interactions with different partners on different contexts. As a matter of fact, while

IDPs play an important role in many cellular processes, the regions of intrinsic disorder are particularly enriched in proteins implicated in processes requiring frequent interactions, such as cell signaling, chromatin remodeling and transcription [136].

Rationale

Little is known about the effect of post-translational modifications on SOX9 functions in a cancer context. Particularly, there are very few indications about the impact of post-translational modifications occurring on SOX9 K2 domain, a portion of SOX9 that has been gaining increasing importance in modulating SOX9 function. Given the pivotal role of SOX9 in many different cancer types, and in particular because of its prominent role in promoting cancer-related features (such as proliferation, EMT, self-renewal and chemoresistance) in NSCLC, investigating the role of SOX9 post-translational modifications on K2 domain represents an interesting venue of research, with the potential to undercover details of regulatory mechanisms that could be exploited as therapeutic targets in NSCLC.

We hypothesized that the K2 domain plays a prominent role in modulating SOX9 activities, and that post-translational modifications on K2 are a key component of these regulatory mechanisms. Additionally, we postulated that some of features promoted by SOX9 in NSCLC could be modulated by the post-translational modifications on the K2 domain.

Thus, we aimed at identifying the mechanisms through which post-translational modifications on K2 modulate SOX9 activities and at evaluating the functional impact of those modifications in the context on lung adenocarcinoma.

Materials & Methods

Cell lines

A549, NCI-H1299 (called H1299), NCI-H1975 (called H1975), NCI-H322 (called H322), NCI-H838 (called H838) and HEK-293T (called 293T) cells were obtained from ATCC. KP717 cells were a generous gift from Dr. Guo, Rutgers Cancer Institute of NJ. MEFs SOX9 floxed (MEFs$^{f/f}$) were generated from 6.129S7-Sox9<tm2Crm>/J mice (The Jackson Laboratory, strain # 013106), as previously described [137]. Cells (with the exception of 293T) were cultured in RPMI media (Gibco) supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine (2mM final concentration) and 1% penicillin (100 U/ml final concentration) + streptomycin (100 U/ml final concentration). 293T cells were cultured in DMEM media supplemented as described for RPMI. Mutational profiles of KRAS, P53 and EGFR for the cancer cell lines used are summarized in **Appendix B**. Cells were incubated at 37 °C in a humidified atmosphere of 5% CO_2. When subcultured, cells were washed with 1X Dulbecco's Phosphate Buffered Saline (DPBS), detached by using 0.25% Trypsin, and re-plated in a new culture vessel according to experimental needs.

Reporter systems for DLR assay

Three different reporter systems for the DLR assays were used. Schematic representations of each one of them are in the main text.

4x48-p89 Reporter was a generous gift from Dr. de Crombrugghe, MD Anderson Cancer Center. SOX/SAC Reporters were a generous gift of Dr. Blache, IRCM, France. Gal4 UAS Reporter was obtained from Addgene (64125) [138].

All the reporter systems were co-transfected with pRL-TK (Promega) used to normalize Firefly Luciferase expression.

Plasmids

For the 4x48-p89 reporter system, SOX/SAC reporter system and transient transfection on H1299, H1299 shSOX9 and H322 cells, the plasmids used were all derived from pCMV Tag2-A backbone. pCMV Tag2 SOX9 WT (aa 1-509 of SOX9 CDS) was a generous gift from Dr. Jiang, National Institute of Cancer Research, Taiwan. tr.SOX9 was generated by amplification of aa 1-304 from SOX9 CDS and subcloning into the pCMV Tag2-A backbone.

For the Gal4 reporter system, the second elements (Gal4 DBD fusion proteins) were generated by subcloning Gal4 DBD (aa 1-147) from pCMV Gal4 (Addgene 24345) [139] into pCMV Tag2-A by using NheI and SalI restriction sites. Downstream of the Gal4 DBD, we subcloned SOX9 aa 225-308 for Gal4 DBD + SOX9 K2, aa 335-509 for Gal4 DBD + SOX9 TA and aa 225-509 for Gal4 DBD + SOX9 K2/TA, all inserted by using SalI/PvuI restriction sites.

For re-introduction of exogenous flagged SOX9 WT or SOX9 T240A into H1299, H1975 and A549 after induction of SOX9 KD, the pULTRA plasmid backbone (Addgene 24129) [140] was used. SOX9 WT was generated by amplification of FLAG-SOX9 WT from pCMV Tag2 SOX9 WT and subcloning into the pULTRA backbone by using the XbaI restriction site.

For lentiviral particles generation, psPAX2 (Addgene 12260) and pMD2.G (Addgene 12259) were used as packaging plasmids.

All the mutant plasmids were generated with the QuiKChange XL Site-Directed Mutaganesis Kit (Agilent). Details of the templates and primers used as well as the specific mutations generated are summarized in **Appendix C**.

Sox9 knock-down (KD)

To generate SOX9 KD in H1299, H1975 and A549, we used the TetON SMARTvector inducible shRNA system (Dharmacon), which we customized to carry the PGK promoter and to express tRFP. Three different shRNA sequences targeting SOX9 3' UTR were tested for their ability to induce SOX9 KD:

1. sh_A: V3SH11252-227345754 (gene target sequence AAGGCAACTCGTACCCAAA)

2. sh_B: V3SH11252-228729576 (gene target sequence TGAGGGGATTTATACATAT)

3. sh_C: V3SH11252-229215765 (gene target sequence GATTTAAGGAGGAGCTGCC)

A Doxycycline titration curve was generated for each cell line to find the ideal concentration of the drug for KD induction. We found out that Doxycycline at 1 μg/ml was suitable for the 3 cell lines (data not shown).

Lentiviral particles generation and cell transduction

To generate stable cells for expression of the TetON KD system and re-introduction of flagged SOX9 isoforms, we generated lentiviral particles and performed 2 separate transduction experiments.

Briefly, for the generation of lentiviral particles, we transfected HEK-293T cells with the appropriate expression plasmid (10 μg) together with psPAX2 (8 μg) and pMD2.G (3 μg) by using Polyethyleneimine (PEI) (Sigma), with a DNA:PEI ratio of 1:3 (3 μl of PEI for 1 μl DNA). 48 and 72 hours after transfection, supernatant containing lentiviral particles was collected and stored at 4°C. Lentiviral particles were then centrifuged at 25,000 RPM for 2 hours, and resuspended in 500 μl of OptiMEM. Virus titer was assessed as previously described [141]. Cells were transduced with lentiviral particles (complexed with Polybrene at 8 μg/ml) generated from either the SMARTvector Non-targeting control (NTC) plasmid or the sh_A plasmid, at a multiplicity of infection (MOI) of 5. 72 hrs after transduction, cells were selected with Puromycin at 2 μg/ml for 8 days. NTC cells and sh_A cells (called KD cells from now on) were the transduced a second time with lentiviral particles generated from pULTRA CTR, pULTRA SOX9 WT or pULTRA SOX9 T240A at an approximate MOI of 2.5. 96 hours after transduction, cells were processed at the fluorescence-activated cell sorting (FACS) instrument (FACScalibur, BD Biosciences) and sorted according to the intensity of EGFP expression in EGFP[high] and EGFP[low] cells. Since there is a direct proportionality between EGFP and SOX9 expression levels in the transduced cells (because of the bi-cistronic structure of the pULTRA plasmid, encoding a fusion protein containing SOX9 CDS upstream of EGFP CDS, separated by a P2A sequence), we generated KD SOX9 WT and KD SOX9 T240A populations with low or high expression levels of the exogenous SOX9 isoforms. For each cell line, we selected

for our experimental work the population expressing SOX9 (either WT or T240A) at levels similar to endogenous SOX9. H1299: high; H1975: low; A549: low.

For SOX9 KO in MEFs, LENTI-CRE lentiviral particles (Viral Vector Core, Carver College of Medicine) were transduced in freshly-generated MEFs at an approximate MOI of 2.5. Cells were let recover for 8 days before starting the cell division rate analysis. At the same time protein lysates were prepared for the WB shown in Fig. 24B.

Transient transfections

Cells were plated in 24 multiwells plate in the appropriate culture media at a density that would reach approximately 70-80% confluency at the time of transfection. Transfections were carried out in antibiotic-free media by using the FuGENE HD transfection reagent (Promega) at a DNA:FuGENE HD ratio of 1:3 (3 μl of FuGENE HD every 1 μg of DNA), according to manufacturer's recommendations. Depending on the Reporter system used, for each of the wells 500 ng of Reporter + 10 ng of pRL-TK + 500 ng of expression plasmid A (+ 500 ng or 2500 ng of expression plasmid B, variable according to the plasmid A:plasmid B ratio of 1:1 or 1:5, respectively. The Gal4 Reporter system didn't require the presence of expression plasmid B) were transfected. 16 hours after transfection, transfection media was replaced with the appropriate supplemented media. For DLR assays, 48 hours after transfection cells were lysed with Passive Lysis Buffer (Promega), as recommended for the downstream DLR application.

For the cell proliferation assays, 24 hours after transfections cells were trypsinized and re-plated in a new 24 multiwells plate at a density appropriate for the analysis (variable according to the cell line used), and analyzed as described below.

Dual Luciferase assay (DLR)

The DLR assay was performed by using the Dual-Luciferase Reporter Assay System (Promega), according to manufacturer's recommendations. For each condition tested, a minimum number of 3 biological replicates were analyzed. Briefly, luminescence was read through a luminometer. For each replicate the ratio between the Firefly Luciferase signal (generated from the reporter plasmid) and the Renilla Luciferase signal (generated from the pRL-TK plasmid) was used to generated a Luciferase expression value. Relative luciferase expression for each condition was generated by normalizing the Luciferase expression value of the 3 replicates to the average Luciferase expression value of the control. A bar graph representing the fold change of each condition over the control was generated with Prism GraphPad.

Western Blotting

Protein lysates for western blotting were prepared by adding 1X lysis buffer (Cell Signaling) supplemented with Phosphatase Inhibitor Cocktail 2 and 3 (Sigma) and Protease inhibitor (Roche) directly on the culture vessel. Cells were incubated for 30 min on ice prior to scraping and lysate collection. The samples were then centrifuged at 14,000 RPM for 15 min at 4 °C, and supernatant was collected. Total protein concentration was determined by using the Bradford colorimetric assay

(Protein Assay Dye Reagent Concentrate, Bio-Rad). Samples were then diluted to the desired concentration with Laemmli Sample Buffer (Bio-Rad) and using beta-mercaptoethanol (Sigma) as reducing agent. Samples were boiled for 5 min at 95 °C prior to loading on the WB apparatus. Depending on the sample availability,15 to 25 μg of protein per lane were loaded on a 12% Mini-PROTEAN TGX precast gel for SDS-PAGE and run at 110 Volts for 90 min. Proteins were then transferred to a nitrocellulose membrane (Bio-Rad) at 100 Volts for 90 min and blocked with 5% milk in 1X TBS-T for 30 min. Primary antibodies (in 0.25% milk in TBS-T) were incubated overnight at 4 °C. Secondary HRP-conjugated anti-Mouse or anti-Rabbit antibodies (in 2.5% milk in TBS-T), as appropriate, were incubated for 1 hour at room temperatures. Details of the primary antibodies used for the project are summarized in **Appendix D**. Signals were detected using the Pierce ECL Western Blotting Substrate (ThermoFisher). Membranes were imaged by using the ChemiDoc Western Blotting Imaging system (Bio-Rad).

Immunofluorescence staining

H1299 cells were plated on an 8 well Nunc Lab-Tek chamber slide (ThermoFisher) pre-treated with poly-L-lysine at 0.1 mg/ml. Cell density was adjusted so that 48 hrs after plating (during which cells were exposed to Doxycycline-supplemented media) confluency was about 90%, and at this time cells were fixed with 4% paraformaldehyde (PFA) (Sigma) for 10 min at room temperature. Cells were then permeabilized with 0.3% Triton X-100 (Sigma) in PBS for 10 min at room temperature, and then the blocking solution (10% normal donkey serum, Sigma) was added for 45 min at room temperature. Anti-SOX9 antibody (Millipore AB5535)

was diluted 1:1,000 in the staining solution (PBS 3%, 0.05% Tween, 0.001% NaN_3) and incubated overnight at 4 °C. Secondary antibody (AlexaFluor488, ThermoFisher) at 1:2,000 dilution was incubated for 1 hour at room temperature. DAPI solution (ThermoFisher) at a final concentration of 300 nM was used to counterstain nuclei. Slides were mounted with an anti-fade mounting media (Vectashield) and imaged at the Nikon A1R Si confocal microscope. Images were acquired at 20X magnification in the green (SOX9) and blue (nuclei) channels. Merge represents superimposition of both channels.

Cycloheximide (CHX) pulse-chase and protein half-life quantification

H1299 cells were exposed to Doxycycline for 24 hours before the addition of CHX (Sigma) at a final concentration of 50 μg/ml. After addition of CHX, cells were lysed at the indicated time-points and a WB was performed, as described above. Band intensities were measured by using the ImageJ software as the average of 2 different exposure times of the membranes (30 sec and 2 min) to account for any potential band saturation. Protein half-life was calculated as the time-point at which band intensity was 50% of the intensity at T=0. The line graph showing protein half-life was generated by using Prism GraphPad.

RNA extraction and reverse transcription

Total RNA was isolated from H1299 cells exposed to Doxycycline for 16 hours by using the RNeasy Mini Kit (Qiagen), according to manufacturer's recommendations. Elimination of the genomic DNA during the RNA extraction process was achieved by using the QIAshredder kit (Qiagen). Extracted RNA

quantity and purity was assessed with a NanoDrop UV-Vis Spectrophotometer (Thermo Scientific). cDNA was synthesized using 700 ng of total RNA with the QuantiTect Reverse Transcription Kit (Qiagen) according to the manufacturer's instructions.

RNAseq

For RNAseq, 3 biological replicates per condition were analyzed. Total RNA was extracted as described above and checked for their quality and purity. mRNAseq was performed at Rutgers RUCDR-Infinite Biologics using Illumina Truseq chemistry on the Illumina NextSeq (HO), at 2x150bp paired-end reads, and ~40-50M reads per sample. Quality of the reads resulted from the RNAseq run was assessed using FastQC program. All the reads showed very high quality (all > Q20 and ~94% > Q30), and their length averaged 100-150bp. The STAR program was then used to align the reads to human genome reference GRCh38. RSEM program was used to quantify the gene expression. Finally, we used two tailed t-test to compare differentially expressed genes between different groups in R program. P value 0.05 and fold change 1.5 were selected to get the gene lists.

Quantitative real-time reverse transcription PCR (qRT-PCR)

qRT-PCR was performed by using Power SYBR Green PCR Master Mix (ThermoFisher) on a Stratagene Mx3005p qPCR system (Agilent Technologies). For each target gene, 3 technical replicates were used, and each experiment included a no RT control and a No Template control for each set of primers. The thermal profile used and details of the primers are summarized in **Appendix E**.

Relative expression of the mRNA was estimated using the 2-ΔΔCT method and calculated in Microsoft Excel.

Cell proliferation

Cells were plated on a 24 multiwells plate in the appropriate supplemented-media. For each condition, at least 3 biological replicates were analyzed. For cells transduced with the SMARTvector expression plasmids, Doxycycline was added to the media 8 to 24 hours before plating. For transiently transfected cells, cells were trypsinized and plated 24 hours after transfection, as described above. Cell density was adjusted so that at the start of the imaging process, confluency was about 5-10%. 2 to 6 hours after plating, cells were imaged by using the IncuCyte ZOOM live-cell analysis system (Essen Bioscience) equipped with a 10X objective. Multiple images per well were collected every 2 hours for a period of 4 to 7 days. Images were then analyzed by using the IncuCyte ZOOM's Confluence Processing analysis tool. An image collection and a specific processing definition were generated for each cell line, so to account for any difference in cell morphology. Cell proliferation rate was calculated as the dynamics of cell confluence (express in percentage) over time. Data representing the average of the biological replicates was processed with Microsoft Excel to generate the proliferation rate curves.

Cell migration

Cell migration was assessed by performing the scratch wound healing assay with the Incucyte Zoom system, according to manufacturer's recommendations. Briefly, H1299 cells were plated on specific 96 wells plates designed for the assay

(IncuCyte ImageLock) and exposed to Doxycycline. Cell density was adjusted so that confluency was 95-100% 48 hours after plating. For each condition, at least 3 biological replicates were analyzed. The scratch in each well was made by using the WoundMaker instrument (Essen Bioscience), and the plates were then imaged in the IncuCyte ZOOM system for 24 hours by using default specifications dedicated to this application. Images were then analyzed with the IncuCyte ZOOM's Confluence Processing analysis tool. Cell migration rate was presented as the incremental reduction of wound width over time.

Subcutaneous tumor xenografts

All procedures involving animas were approved by the Rutgers Institutional Animal Care and Use Committee (IACUC) with the protocol I13-051-9. Tumor xenografts were generated by subcutaneously injecting 500,000 H1299 cells resuspended in an equal volume of RPMI media and Matrigel Basement Membrane Matrix (Corning) in the flanks of 6-to-8 weeks old NOD.CB17-Prkdc<scid>/J (NOD/SCID) mice (strain 001303) (The Jackson Laboratory). 8 animals per group were randomly allocated. Animals were kept under a Doxycycline-supplemented diet with Doxycycline-containing food pellets at 650 mg/Kg (Envigo), which was started 7 days prior to human cells injection and was maintained throughout the duration of the experiment. Tumors were measured twice per week with a caliper, and tumor volume was calculated using the formula V = (width^2 x length)/2, in which length is the larger of the two dimensions. At the end of the experiment, animals were sacrificed and tumors were explanted, minced into fragments of about 5 mm^3

and homogenized using a mortar and a pestle. Protein lysates were then made with a protocol similar to that used for monolayer cells, as described above.

Kinases prediction

To predict which kinase is most likely to be responsible to be responsible for T240 phosphorylation, we used the SCANSITE 4.0 software (MIT, https://scansite4.mit.edu/4.0/#home). We searched for the whole SOX9 CDS by setting the stringency of the analysis low. The top 5 hits for the T240 residue were: (score in parenthesis): p38 MAPK (0.426); GSK3 kinase (0.457); CDK5 kinase (0.460); CDK1 motif 2 (0.471); cdc2 (0.491).

Intrinsic disorder prediction and molecular dynamics (MD) simulations

To generate a prediction of the relative intrinsic disorder within SOX9, we used the GlobPlot software, as previously described [142]. Default settings were maintained to analyze SOX9 CDS. The software can be found at http://globplot.embl.de.

The MD simulations were generated by using the software AMBER (http://ambermd.org/index.php). The polypeptide simulated corresponds to SOX9 CDS aa 221-310. Phosphorylation of T240 was including a double deprotonated phosphate group. The runs length were 800 nanoseconds each.

Statistical analysis

Unless otherwise indicated, data is expressed as the mean ± standard deviation (SD). Statistical analysis was performed using Prism GraphPad or Microsoft Excel. To compare 2 conditions for the DLR assays, a paired, two-tailed Student's t-test

was used. Results with a P value less than 0.05 were considered statistically significant.

Results

Identification of specific post-translational modifications on K2 domain

In order to focus our investigational research on post-translational modifications likely to be involved in SOX9 functions, we sought to identify their presence on the K2 domain. During a Mass Spectrometry analysis conducted for the project describing the interaction between SOX9 and FBW7 [127], we were intrigued by the observation that 2 aa on the K2 domain were systematically presented with the presence of a phosphate group, T236 and T240. Both aa belong to the phosphodegron consensus sequence necessary for the interaction with FBW7 (TPPTT, aa 236 to aa 240 on SOX9). However, while pT236 resulted necessary for such interaction, the role of pT240 in this context was less prominent [127, 128]. This interesting observation prompted us to hypothesize a role of pT240 in SOX9 functions modulation, so that was the direction we focused our investigational effort on.

Choice of the main cell model

In selecting the right cell model for our research project, we took into consideration several factors, including the strict context-dependency characterizing SOX9 activity, as previously mentioned, [57] and the fact that the aim of our investigation was to translate our findings into a specific context, which is that of lung adenocarcinoma. For these reasons, we selected as our main cell model H1299, a NSCLC cell line derived from lymph node metastases, characterized by

homozygous loss of P53, which have been extensively used in a plethora of scientific publications pertaining to lung cancer. Moreover, the observation from which this project stemmed (the systematic presence of pT240 on K2) was made on H1299 cells, so it seemed reasonable to start investigating the role of that phosphorylation by using the same cell line it was found on.

In order to evaluate the role of pT240, we employed a 2 steps process to generate our cell model: i) knock-down (KD) of endogenous SOX9 ii) exogenous re-introduction of either flagged SOX9 WT or a flagged form of SOX9 in which the phosphorylation in T240 is inhibited by a Threonine (T) to Alanine (A) site-directed mutation at the residue 240 (SOX9 T240A from now on). The comparison between SOX9 WT (systematically phosphorylated at T240) and SOX9 T240A (in which the same phosphorylation event is inhibited) should shed some light into the role of that specific post-translational modification on the K2 domain. To accomplish the first of the 2 steps composing our experimental design, we used an inducible (TetON) system and screened 3 different shRNA sequences targeting SOX9 3'UTR (*Fig.9A*). We identified the shRNA sequence inducing the highest level of SOX9 KD upon Doxycycline exposure for 48 hours, and we used it to transduce H1299 so to generate cells with stable inducible expression of that sequence. For the second step, these same cells underwent a second transduction process in order to constitutively express flagged SOX9 WT or flagged SOX9 T240A. Particular attention was given to the expression levels of exogenous SOX9, either WT or T240A, to make them as similar as possible to the expression levels of endogenous SOX9 in H1299. In this way, we avoided the situation in which over-

saturation of the cells with exogenous SOX9 generates false results not achievable with physiological levels of Sox9 (*Fig.9B,C*).

Figure 9. Generation of the main cell model.

A. Analysis of three different shRNA sequences for their ability to induce SOX9 KD upon Doxycycline exposure. *B.* WB analysis of SOX9 expression levels among the four different population. *C.* Fluorescent microscope analysis of H1299 KD SOX9 WT cells. 1: Phase; 2: RFP (from activation of the TetON system); 3: GFP (from re-expression of SOX9 WT); 4: merge of the channels.

Subcellular localization

We sought to determine if the prevention of T240 phosphorylation was altering any of the SOX9-mediated activities. We started by looking at SOX9 subcellular localization. SOX9 is a transcription factor and, as such, it is mostly localized inside the nuclei of the cells so that it can elicit its transcriptional modulation activity on target genes. When we performed an immunofluorescence staining and looked at

SOX9 representation in the different cellular compartments, we didn't find any significative difference between endogenous SOX9, exogenous SOX9 WT and exogenous SOX9 T240A, they were all mainly nuclear (*Fig.10*). We concluded that T240 phosphorylation does not alter SOX9 subcellular localization.

Figure 10. T240 phosphorylation does not alter subcellular localization.

Immunofluorescence imaging of the 4 cell populations stained for SOX9 (green). Nuclei are identified by DAPI (blue).
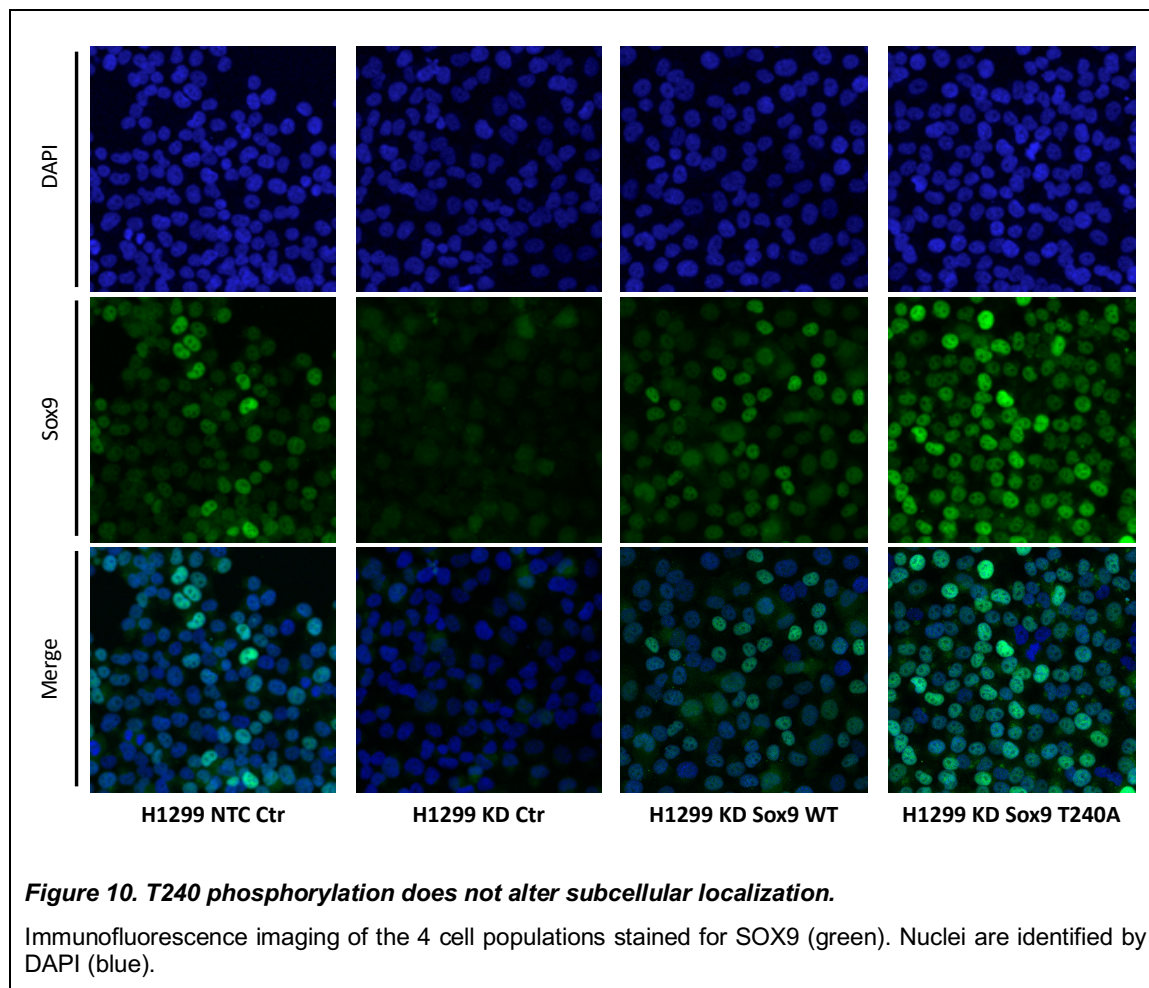
Protein stability

Next, we investigated whether the presence of phosphorylation at T240 was altering SOX9 protein stability. We calculated protein half-life of endogenous SOX9, exogenous SOX9 WT and exogenous SOX9 T240A through a

cycloheximide (CHX) assay. Once again, no difference among the 3 types of SOX9 was found, they all had a half-life of approximately 3.5 hours (*Fig.11*).

Figure 11. T240 phosphorylation does not alter SOX9 stability.

A. WB of SOX9 expression after cycloheximide (CHX) exposure. *B.* Quantification of relative SOX9 expression over time after CHX exposure. SOX9 half-life (dotted line) was calculate as the time-point where protein amount was 50% than at T=0.
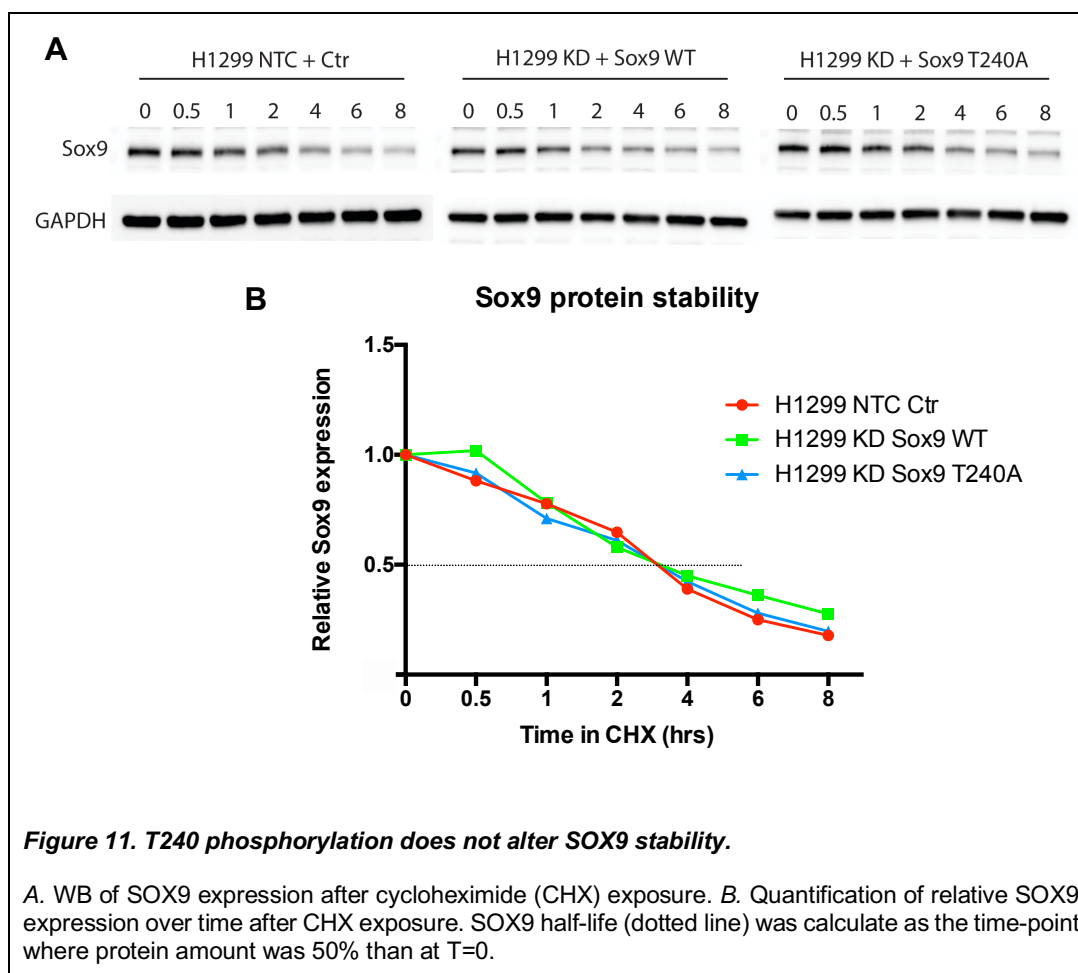
Estimate of SOX9 half-life is consistent to our previous study and to what other studies found [106, 127]. Moreover, in line with the rationale of our project, these results provide evidence that, differently from pT236, pT240 main role is not to modulate SOX9 protein stability (through Fbw7 interaction or different mechanisms).

Transcriptional ability

Then, we sought to evaluate the role of pT240 in modulating SOX9 transcriptional ability. We reasoned that, given the systematic presence of the phosphate group on T240 and the pro-tumorigenic role of SOX9 in the lung adenocarcinoma setting, it is not implausible to predict a reduction of SOX9 transcriptional ability without the phosphorylation in T240. To investigate this, we decided to run a Dual Luciferase Reporter (DLR) assay by using a Luciferase Reporter system previously adopted to evaluate SOX9 transcriptional ability, called 4x48-P89 [143]. Essentially, the Reporter system is composed of a gene encoding Firefly Luciferase under the control of a specific portion (base -89 to base +6) of the promoter of COL2A1 (a known SOX9 target gene during chondrogenesis). Additionally, the Reporter system is influenced by the presence, upstream to the promoter, of 4 repetition of a 48-bp element (located in the intron 1 of murine Col2a1 gene) working as an enhancer (*Fig.12A*).

Figure 12. 4x48-P89 Reporter.

A. Schematic representation of the 4x48-P89 Reporter construct. *B.* Schematic representation of the 3 SOX9 versions transiently transfected together with the 4x48-P89 Reporter.
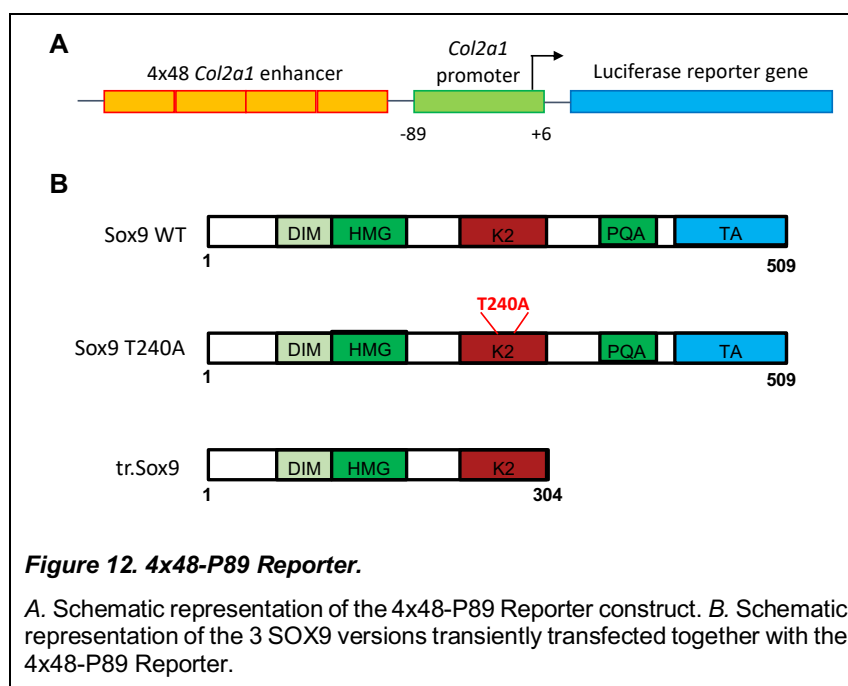
We transiently transfected H1299 cells with the 4x48-p89 Reporter and either flagged SOX9 WT alone (plus an empty control) or flagged SOX9 WT in

combination with flagged SOX9 T240A. We decided to use this experimental design to simulate the physiological conditions in which SOX9 elicits its functions by complexing into a homodimer [109]. When transfected alone, SOX9 WT homodimers modulate transcription of the Firefly gene on the Reporter system. When the combination of 2 plasmids is transfected, at least part of the SOX9 dimers acting on the Reporter is composed of SOX9 WT plus SOX9 T240A, thus allowing us to evaluate the potential dominant negative effect of SOX9 T240A on SOX9 WT. As an additional control, we also included in this assay an experimental group in which we transfected, together with the 4x48-p89 Reporter, SOX9 WT and a truncated version of SOX9, called tr.SOX9 (lacking the carboxy-terminal part of SOX9, from aa 305 to aa 509, including the "canonical" transactivation domain TA), which previous studies have proven to act as dominant negative towards SOX9 WT [144] (*Fig.12B*). When we evaluated the ability of the different experimental conditions to induce Luciferase expression, we were pleased to observe that, when SOX9 WT and SOX9 T240A are co-expressed, there is a significant reduction in Luciferase expression compared to the condition where SOX9 WT alone is transfected. Intriguingly, the extent of this reduction was even greater than the one generated by the concomitant transfection of SOX9 WT and the known dominant negative plasmid, tr.SOX9 (*Fig.13A*).To exclude that an uneven expression of the different plasmids was the reason for the effect we observed, we performed a western blotting (WB) that clearly showed how the different plasmids are all expressed at comparable levels (*Fig.13B*). Thus, we could confirm that the results of this assay suggest that the lack of phosphorylation

in T240 might be responsible for the reduction in SOX9 transcriptional ability and that, consequently, pT240 might play a key role in mediating transcriptional ability for SOX9 WT. Moreover, given the dominant negative effect exerted by SOX9 T240A on SOX9 WT in this assay, there are indications that lack of phosphorylation in T240 in a single monomer of the SOX9 homodimer is sufficient to impact SOX9 activity.

Figure 13. Effects of T240A on the 4x48-p89 Reporter system.

A. Relative Luciferase expression induced by the co-expression of the indicated plasmids. *B.* WB showing that the 3 plasmids have similar expression levels. *C.* Relative Luciferase expression showing a dose-dependent effect of SOX9 T240A.
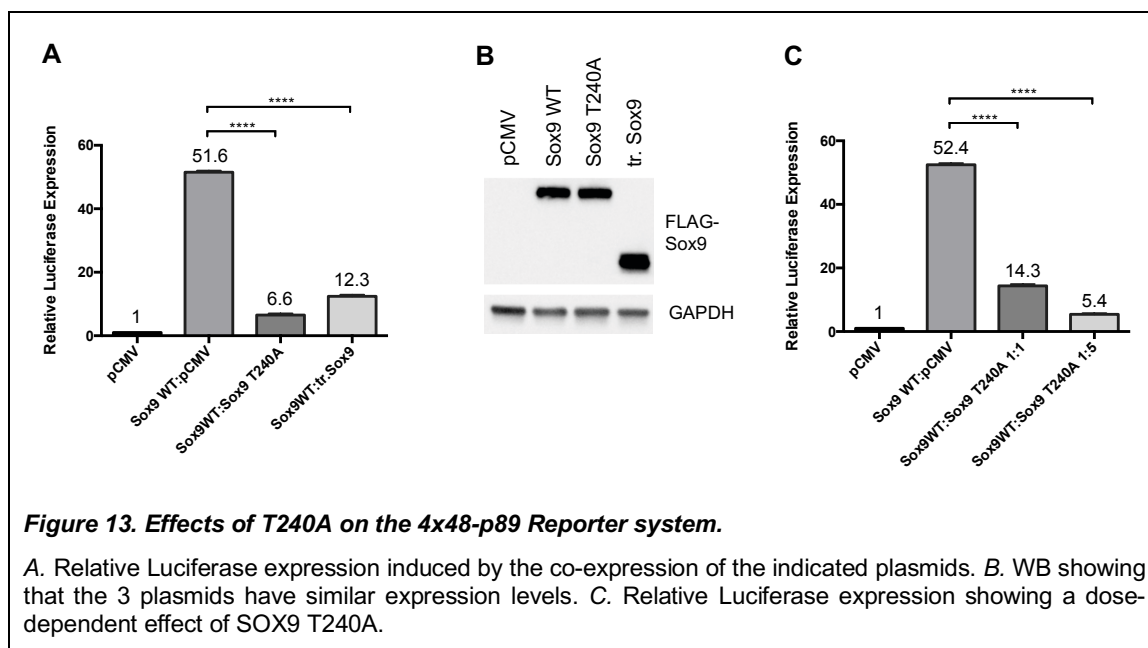
Stimulated by these exciting results, we sought to determine if the effect we observed could be repeated under different experimental conditions. First, we replicated our DLR assay by adding a condition in which the SOX9 WT plasmid and the SOX9 T240A plasmid were co-transfected not only in a 1:1 ratio (as done in the previous experiment) but also in a 1:5 ratio. The purpose of this was to evaluate a potential dose-dependent effect of SOX9 T240A versus SOX9 WT. As shown in *Fig.13C*, the increasing amount of SOX9 T240A relative to SOX9 WT determined a higher reduction of Luciferase expression compared to the condition

where the two plasmids are in the same amount, confirming that the dominant negative effect promoted by the presence of T240A is dose-dependent.

As discussed before, SOX9 activities and regulatory mechanisms have been proved to be strictly context-dependent. Thus, we wondered whether our observations were to be limited to the specific background of H1299 cells or could be extended to other conditions. To evaluate that, we used three additional lung adenocarcinoma cell lines, characterized by a mutational profile different than H1229:

1. A549 – human alveolar epithelial cells from a primary lung tumor. Hypotriploid. Homozygous KRAS G12S mutation.

2. H838 – NSCLC from a lymph node metastasis originating from a smoker. P53 WT, KRAS WT, EGFR WT.

3. H322 – from a cervical node metastasis of a bronchioalveolar lung carcinoma. P53 missense mutation.

Interestingly, our results showed that SOX9 T240A is able to reduce SOX9 WT transcriptional ability in our DLR assay in all the different cellular backgrounds tested (*Fig.14A-C*), although each cell line was characterized by a distinctive magnitude of the effect. Additionally, we were able to observe a reduction in Luciferase expression in the SOX9 WT + SOX9 T240A condition (compared to SOX9 + empty vector) even in a non-human context, by testing our DLR assay in a KRAS mutated and P53 null mouse-derived lung cancer cell line, called KP717 (*Fig.14D*). Taken together, these results provide clear indications that the

phosphorylation in T240 plays an important role in modulating SOX9 transcriptional ability.

Figure 14. 4X48-P89 Reporter on different cell lines.

Relative Luciferase expression induced by the indicated plasmids combinations in A549 cells (A), H838 cells (B), H322 cells (C) and KP717 cells (D).
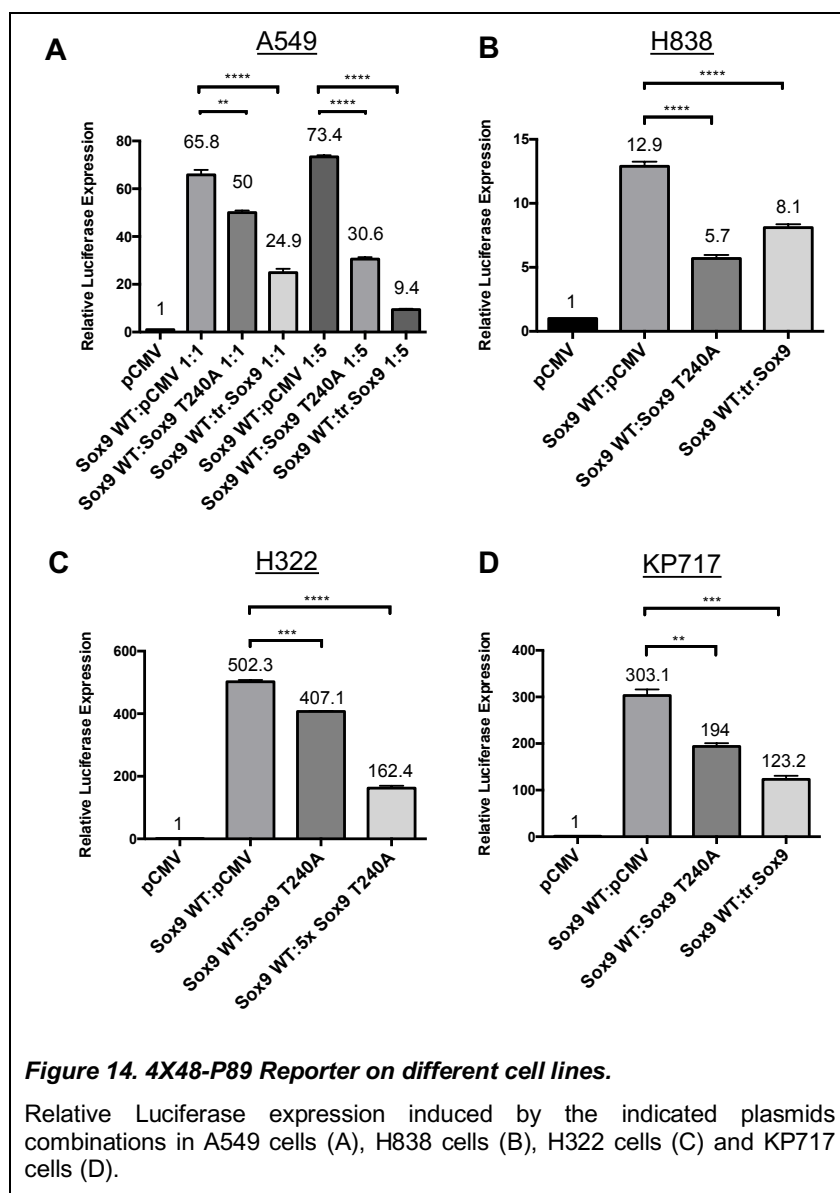
Galvanized by these findings, we wanted to investigate more into this aspect of SOX9 activity. So, we decided to test whether this effect is limited to the specific Reporter system that we have been using (and so specific to the COL2A1 gene) or it could be extended to different target genes. To address this interesting question, we performed our DLR assay by using a different Reporter system, called SOX/SAC, which has been extensively used in the past to evaluate SOX9 transcriptional ability [145, 146]. In brief, in the SOX Reporter system the Firefly Luciferase gene expression is under the control of 7 repetitions of a SOX9

consensus sequence (AACAAAG). In the SAC reporter, functioning as a negative control, the structure is the same but the consensus sequence is mutated to CCGCGGT so to prevent SOX9 binding (*Fig.15A*).

Figure 15. SOX/SAC Reporter system.

A. Schematic representation of the SOX Reporter and SAC (negative control) Reporter constructs. *B.* Relative Luciferase expression induced by the indicated plasmids in the SOX (left) and SAC (right) Reporters.
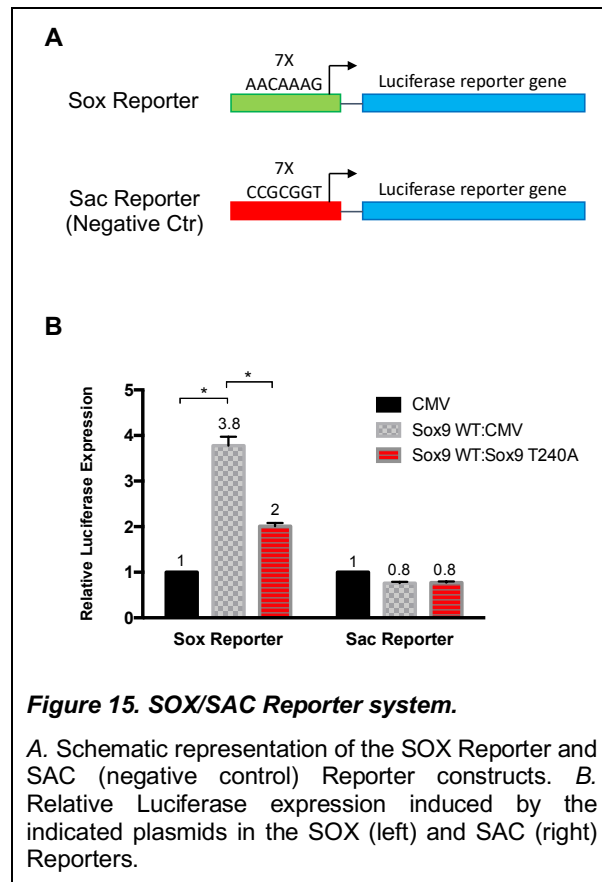
Once again, we were able to observe a reduction in Luciferase expression upon concomitant expression of SOX9 WT and SOX9 T240A (*Fig.15B*), thus confirming that the indications emerged from the DLR assay with 4x40-p89 Reporter were not exclusive of the COL2A1 gene.

Subsequently, we sought to test if the effect promoted by the lack of T240 phosphorylation could be reverted by adding that phosphorylation back.

Thus, we generated (by the same site-directed mutagenesis protocol we used for T240A) the two phosphomimetics T240D and T240E, where the Threonine in position 240 was substituted by the negatively charged aa Aspartate (D) or Glutamate (E), to simulate the condition of a constitutive phosphorylation of the T240. We then ran a DLR assay with the 4x48-p89 Reporter co-transfecting the phosphomimetics with SOX9 WT. Surprisingly, none of the phosphomimetics was able to rescue the Luciferase expression to the levels induced by SOX9 WT + empty vector (*Fig.16A*).

Figure 16. Effect of phosphomimetics on the 4x48-p89 Reporter.

A and B. Relative Luciferase expression induced by the indicated combinations of plasmids. In B, all blue bars represent a T to A mutation in a specific aa, while all red bars represent a T to D or T to E mutation in a specific aa. *C*. WB of phosphorylation status of T236 and T240 with the different mutations.
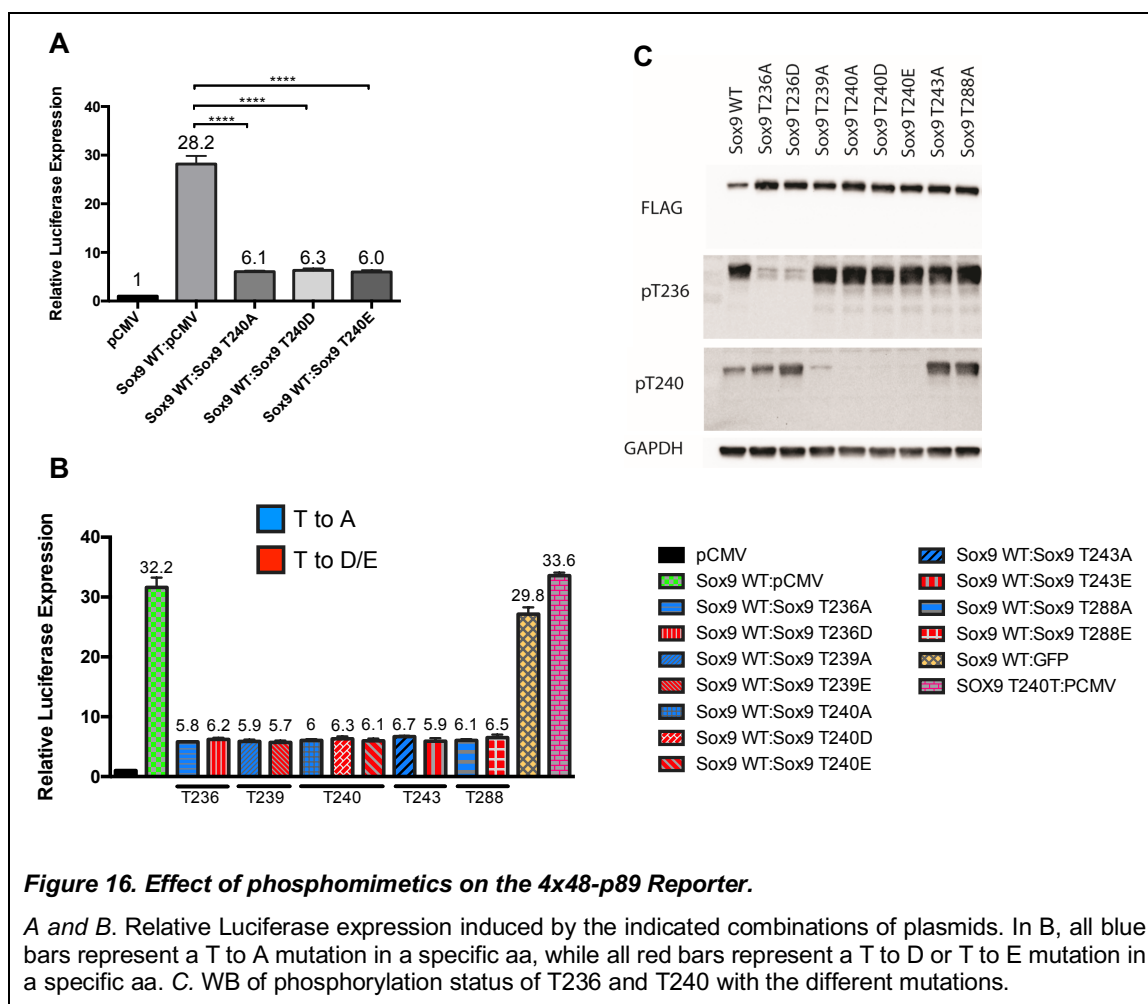
There could be multiple different reasons to explain the lack of phenotype rescue. First, it is not unusual for the phosphomimetic mutants to fail in mimicking the functional effect of a phosphorylated aa [147-149]. Second, the event we observed could be dependent not solely on the phosphorylation status of T240 but also on specific structural rearrangements caused by the T to A mutation. For these reasons, we decided to verify if a T to A mutation in any of the 4 additional Threonine residues present in the K2 (besides T240) could produce the same effect. Thus, through an extensive site-directed mutagenesis work, we generated T236A, T239A, T243A and T288A mutations, as well as a T to D and/or a T to E phosphomimetic for each one of them. When we tested these plasmids with the

4x48-p89 Reporter system, and verified that any T to A mutation was producing the same dominant negative effect on the ability of SOX9 WT to induce Luciferase expression. At the same time, no phosphomimetic was able to rescue this phenotype (*Fig.16B*). Importantly, to exclude that our observations were triggered by technical issues of the assay, we tested 2 controls: i) a T240T plasmid, generated with an A to T mutation by using the SOX9 T240A as a template ii) co-expression of SOX9 WT with a plasmid encoding a totally unrelated gene (EGFP in this case). When tested with the 4x48-p89 Reporter, both controls induced Luciferase expression levels very similar to those induced by SOX9 WT, thus ruling out any technical problem with the site-directed mutagenesis process or the assay design. Moreover, to verify that the phenotype promoted by any single T to A mutation was direct and not determined by an indirect effect of that mutation on the phosphorylation status of an adjacent Threonine, we checked by WB the presence of the phosphate group on T236 and T240 (for which we had custom-made phospho-antibodies available) when the plasmids with the different mutations were expressed. As shown *Fig.16C*, the phosphorylation status of T236 and T240 is affected only by mutations on T236 and T240, respectively. In other words, the effect of any single mutation is direct and not indirect.

Hence, we concluded that the K2 domain plays a pivotal role in modulating SOX9 transcriptional ability and that specific post-translational modifications on K2 are involved in this mechanism.

Excited by these findings, we sought to evaluate the effects of the phosphorylation status of T240 on the global transcriptome of our cell model by performing a

RNAseq, so to identify specific genes or functional pathways whose transcription is affected by the presence of the phosphate group on T240. Our experimental design included filters that would generate two final lists of genes: i) genes activated by SOX9, comprising genes whose expression is decreased in the KD and rescued by re-expression of SOX9 WT or SOX9 T240A ii) genes repressed by SOX9, including genes whose expression is increased upon SOX9 KD and reduced by SOX9 WT or SOX9 T240A re-introduction. For each of these two lists, we had a number of "overlapping" genes, whose expression (compared to KD) was rescued by re-introduction of SOX9 WT as well as SOX9 T240A. However, our focus was concentrated on the "non-overlapping" genes, whose expression was rescued only by the re-introduction of SOX9 WT (but not SOX9 T240A) or, vice versa, only by the re-introduction of SOX9 T240A (but not SOX9 WT) (*Fig.17A and **Appendix F**). These non-overlapping genes represent targets whose expression is potentially modulated by the phosphorylation status of T240, thus making them ideal candidates for our analysis. However, we were surprised to notice how these non-overlapping lists were populated only by few entities, many of which were in reality non-coding genes. Thus, we carefully selected potential targets among the non-overlapping genes (for both, genes activated by SOX9 and repressed by SOX9) and complemented them with previously published SOX9 target genes, and ran a qRT-PCR analysis to validate them. Unfortunately, the results of the qRT-PCR were not encouraging, because most of the analyzed genes resulted in an expression pattern different than what expected from published data or from the RNAseq (*Fig.17B*). At the end, we were able to identify

only 3 genes potentially modulated by the phosphorylation status of T240, COL1A1, IL11 and IL6R, but given the relatively small difference in the expression levels among the 4 populations, we decided not to pursue their characterization any further.

Figure 17. RNAseq and qRT-PCR analysis.

A. Schematic representation of RNAseq results. For genes repressed by SOX9 (left) and activated by SOX9 (right), we compared genes rescued by SOX9 WT and SOX9 T240A. Focus was placed on the non-overlapping genes (indicated by red arrows). *B.* qRT-PCR results showing relative expression of the indicated genes in the 4 populations. Asterisks represent genes potentially modulated by T240 phosphorylation status.

Transactivation ability of SOX9 domains

As mentioned before, SOX9 activity on target genes is promoted by its association with co-activators and co-repressors, thus forming a complex functioning as

transactivator for the expression of the target gene [109]. The TA domain (aa 342-509) has been considered the "canonical" transactivator domain in SOX9, while the K2 domain (aa 229-303) has not been endowed with a strong transactivation potential [62]. However, our experiments clearly showed that the K2 domain (and post-translational modifications on it) have a pivotal role in modulating SOX9 transcriptional ability. Thus, we decided to directly compare transactivation ability of K2 domain and TA domain within SOX9. To achieve that, we used a modified version of the Two-Hybrid System (*Fig.18A*) [150].

Briefly, the system is composed of two elements:

1. A Reporter plasmid encoding the Firefly Luciferase gene under the control of the Gal4 enhancer sequence

2. A plasmid encoding for a fusion protein composed of i) Gal4 DNA binding domain (DBD) ii) the specific domain of your interest.

The Gal4 DBD binds to the Gal4 enhancer sequence on the Reporter plasmid, so that the expression of the Luciferase gene is controlled by the transactivation ability of the domain part of the fusion protein. We generated different fusion proteins by cloning downstream of Gal4 DBD either the K2 domain alone (DBD+SOX9 K2), the TA domain alone (DBD + SOX9 TA) or a combination of the two domains (DBD + SOX9 K2/TA). A plasmid including Gal4 DBD but lacking any other element downstream was used as control. The results of the DLR assay showed that the two domains combined are able to induce higher levels of Luciferase expression than any of the two domains alone. However, it was very interesting to notice that, when considering the single domains alone, the K2 domain has a higher

transactivation activity than the "canonical" TA domain, supporting our hypothesis about its central role in modulating SOX9 activity (*Fig.18B*).

Figure 18. GAL4 Reporter system.

A. Schematic representation of the 2 main elements composing the Gal4 Reporter system. *B.* Relative Luciferase expression induced by K2 alone, TA alone or the combination of the 2(right). WB showing expression levels of the different plasmids (left). Red arrows indicate the bands of interest. *C.* Relative Luciferase expression induced by WT and T240A mutants of K2 and K2/TA (left) and relative WB (right). *D.* Relative Luciferase expression induced by K2/TA and the indicated mutants (up) and relative WB (down).

Excited by these findings, we generated a T240A mutant for the K2 alone (DBD + SOX9 K2 T240A) as well as for the combination of K2 and TA (DBD + SOX9 K2/TA

T240A), and tested their ability to impact the Luciferase expression. Remarkably, while we could not see any difference between the K2 domain alone and its mutant, the T240A mutation was able to reduce the levels of Luciferase expression of the plasmid encoding for both K2 and TA domains (*Fig.18C*). These results imply that the impairment of K2 transactivation ability generated by the lack of phosphorylation in T240 likely requires the presence of both K2 and TA domains, thus suggesting a close cooperation between these two in modulating SOX9 transcriptional ability. Similarly to what we observed with the 4x48-p89 Reporter system, phosphomimetics were unable to rescue the phenotype generated by the T240A mutation. However, reverting back to T the residue 240 (A to T mutation from a T240A template) induces Luciferase expression levels similar to the WT conditions, confirming that these results are technically sound and reliable (*Fig. 18D*).

Overall, these results prove that the K2 domain is endowed with transactivation ability, even more than the TA domain, and that post-translational modifications on K2 play an important role in this process. Moreover, they indicate that a close interaction between the two domains is necessary in order for SOX9 to elicit its full transcriptional modulation potential.

Intrinsic disorder in SOX9

In an effort to decipher the molecular details of the mechanisms through which post-translational modifications on K2 modulate SOX9 activity, we decided to look at the structural characteristic of the K2 domain. Particularly, we noticed that the aa sequence of this domain is characterized by low complexity (i.e. over-

representation of specific residues) and that there is a relatively low content of bulky hydrophobic aa, with a relatively high proportion of polar and charged aa, instead. Intriguingly, as mentioned before, these features are considered to be shared by a particular type of macromolecules, called intrinsically disorder proteins (IDPs) [133]. One of the main features of IDPs and, more frequently, of proteins containing intrinsic disordered regions (IDRs) is the ability to undertake different interactions with different partners on different contexts. As mentioned before, one of the distinctive features of SOX9 is to engage in interactions with different partners in different context, promoting different effects [57]. Moreover, the regions of intrinsic disorder are particularly enriched in proteins implicated in cell signaling, chromatin remodeling and transcription (and SOX9 is, indeed, a transcription factor) [136].

Since all these clues were pointing in the direction of SOX9 being an IDP or, at least, possessing an IDR within its sequence, we decided to investigate whether the K2 domain is indeed a region of intrinsic disorder within SOX9.

We ran a simulation to predict the relative propensity to disorder of the SOX9 sequence by using a software able to evaluate for any given aa the likelihood to be in sheet, helix, or random coil, thus predicting regions of order and disorder. Interestingly, we found that the different domains of SOX9 have a dissimilar propensity to disorder. Particularly, while the DNA binding domain (HMG) has a relative low propensity, the surrounding domains in both directions have a much higher tendency to behave as IDRs, including the K2 domain (*Fig.19A*).

Figure 19. Comparison of disorder propensity in SOX9 and P53.

A. Representative results of a simulation of propensity disorder of Sox9 sequence. The green portion of the bar represents predicted ordered regions (corresponding to HMG), while the blue parts represent predicted disordered regions (corresponding to DIM, K2, PQA and TA). B. Results of experimental validation of P53 intrinsic disorder. Blue: ordered regions; red: disordered regions.

Even more interestingly, when we compared the predicted disorder pattern of SOX9 with that of p53, a master transcription factor for which the intrinsic disorder distribution has been experimentally validated [151], we were pleased to see how similar they were (*Fig.19B*). In fact, p53 has a well-structured region, corresponding to the DNA binding domain, surrounded by much more disordered regions on both sides. Importantly, it has been calculated that over 70% of p53 interactions with other macromolecules occur in the non-ordered domains of the protein [151, 152], confirming the prominent role of these IDRs for binding with other partners.

From these preliminary indications, it is plausible to consider the K2 domain as a potential IDR within SOX9. We already established that K2 plays a pivotal role in modulating SOX9 transcriptional ability, and that specific post-translational modifications on K2, such as phosphorylation in T240, have an impact on this mechanism. As mentioned before, SOX9 transcriptional modulation on target

genes requires interaction with co-activators and/or co-repressors. As demonstrated by the p53 example, IDR are particularly prone to engage in interactions with other macromolecules. Hence, it is possible that the intrinsic disordered K2 domain is modulating SOX9 transcriptional activity by mediating interaction with co-activators or co-repressors, and that the phosphorylation status of T240 has a role in mediating this interaction.

To verify this possibility, we performed molecular dynamics (MD) simulations (measuring the role of factors such as charge interaction, structural stress and random atomic vibrations in predicting protein folding) to infer the structural conformation of the K2 domain in different conditions [142]. Particularly, we evaluated the impact of the phosphorylation status of T240 on K2 structure. Interestingly, we determined that K2 conformation is largely affected by the presence of the phosphate group in T240 (*Fig.20A*). In fact, when T240 is phosphorylated, the K2 domain appears long, relaxed, with lack of well-defined structures. In other words, K2 appears to be relatively disordered under these conditions, and its conformation suggests a propensity to interact with the surroundings. Conversely, when we ran the same simulation removing the phosphate group on T240, we could appreciate a significative change in K2 conformation. The domain structure became compact, much more globular, in a folding-like shape. Overall, we can consider this conformation of K2 less disordered than the previous one, and much less likely to interact with the surroundings (*Fig.20B*).

Figure 20. Impact of T240 phosphorylation status on K2 conformation.

Representative images of molecular dynamics simulation of K2 domain when the T240 residue is phosphorylated (A, orange image) or not phosphorylated (B, blue image). The red and green residues represent T236 and T240, respectively.
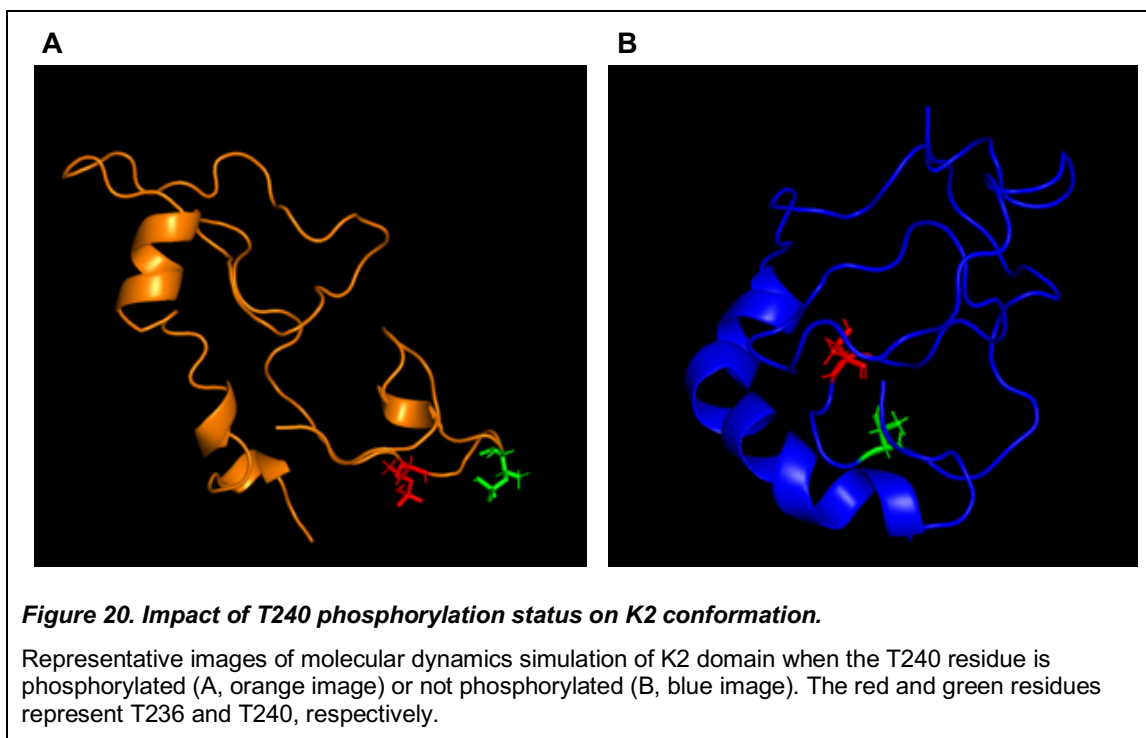
Importantly, we also looked at the relative position of aa T236 and T240 under the two different conditions. When T240 is phosphorylated, both aa are pointing towards the side, exposing the negatively charged phosphate group and making it available for potential ionic bonds. On the contrary, the lack of the phosphate group in T240 displaces both aa towards the inner part of the structure, "burying" them into the globular shape of K2.

Taken together, these experiments suggest that phosphorylation in T240 might contribute to the intrinsic disorder of the K2 domain within SOX9, and that this intrinsic disorder might play a role in the ability of SOX9 to bind co-activators and co-repressors, thus ultimately modulating SOX9 transcriptional ability.

Kinases prediction

Given the central role of the T240 phosphorylation in modulating SOX9 transcriptional ability, we sought to identify kinases that are likely to promote this

phosphorylation. We inquired the SCANSITE 4.0 software, a tool able to predict motifs within proteins that are likely to be phosphorylated by specific protein kinases, with SOX9 aa sequence. Interestingly, we were able to confirm that, as expected, T240 is indeed prone to phosphorylation, as multiple hits for this specific residue were listed in the results. Additionally, the software scored different kinases for their likelihood to promote this phosphorylation, ranking them according to their mechanism of action and to SOX9 structure. The kinases most likely to promote T240 phosphorylation are, in order of probability, p38 mitogen-activated protein kinase (MAPK), GSK3 kinase and CDK5 kinase. Intriguingly, for all these kinases a role in promoting a pro-tumorigenic action in lung cancer has been established [153-155], making them ideal candidates for further analyses.

Significance of post-translational modifications on SOX9-mediated functions in lung adenocarcinoma

After attributing a role to the K2 domain and to the phosphorylation on T240 (and potentially to phosphorylation on other T residues within the K2 domain) in modulating some of the SOX9 activities, the second aim of our project was to assess the functional significance of those post-translational modifications in the context of SOX9-mediated functions in lung cancer.

As previously mentioned, SOX9 has been associated with several cancer-related features in lung cancer, such as proliferation, migration, EMT and in vivo tumorigenesis. Thus, we sought to evaluate the role of pT240 in each one of the SOX9-mediated functions.

We started by looking at cell proliferation in vitro. We transiently transfected H1299 cells with plasmids encoding either SOX9 WT or SOX9 T240A. Additionally, we included a condition in which cells were transfected with the known dominant negative tr.SOX9, to replicate the experimental setting of the DLR assays. After transfection, cells were plated into a live-cell analysis system, which allowed us to capture multiple images for each cells-containing well at defined timepoints, and to calculate in real-time several parameters relative to cell growth and proliferation (cell confluence, cell number, etc.). As a result, we were able to generate a very detailed proliferation curve accurately describing the dynamics of cell proliferation over time. As shown in *Fig. 21A*, transfection of SOX9 WT promoted an increase in the proliferation rate, while tr.SOX9 did not have any effect on cells. Interestingly, transfection of SOX9 T240A did not increasing the cell proliferation rate, but actually determined a reduction in such rate when compared to the CMV control. We repeated this experiment in H1299 cells stably expressing a shRNA targeting SOX9 coding sequence. We confirmed that transfection of SOX9 WT increase cell proliferation rate, while SOX9 T240A-transfected cells reduced their proliferation, similarly to cells transfected by tr.SOX9 (*Fig.21B*). When assessing the same system in a different cellular background, by transfecting H322 cells, we observed a different situation. Both SOX9 WT and SOX9 T240A determined an increase in cell proliferation rate, but the effect promoted by SOX9 T240A was much more pronounced than the one promoted by SOX9 WT (*Fig. 21C*). These results are suggestive of cell-specific effects of the T240A mutation under these conditions, probably influenced by the genetic background of the cell hosts.

Figure 21. Proliferation rate in transiently transfected cells.

Effect of transient transfection of empty control (CMV) (*blue*), SOX9 WT (*red*), SOX9 T240A (*green*) and tr.SOX9 (*purple*) on H1299 (*A*), H1299 shSOX9 (*B*) and H322 (*C*).
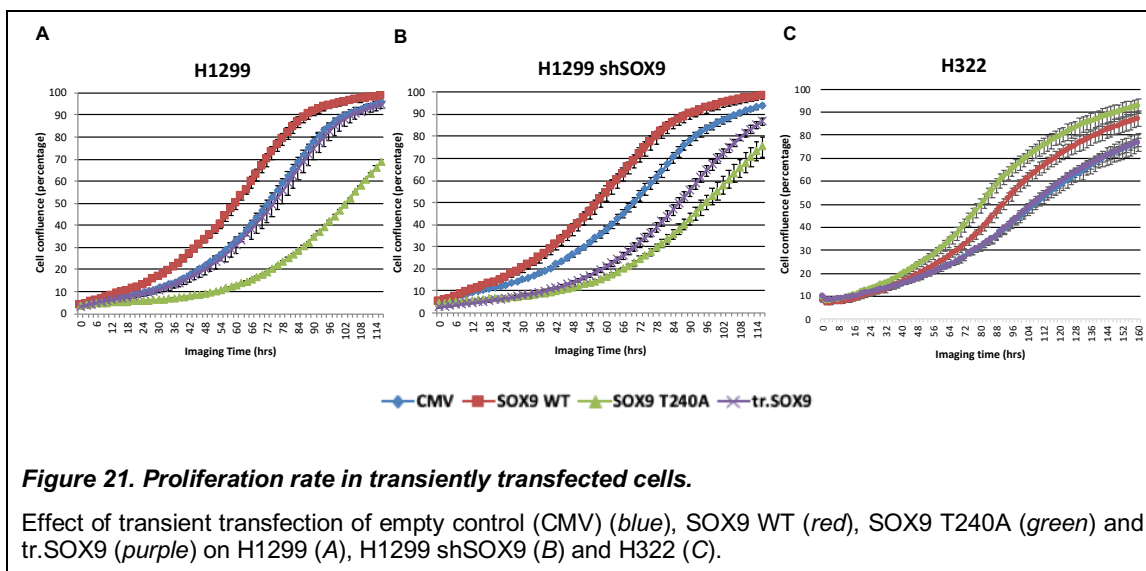
Although interesting, the reliability of these results is limited by the nature of the experimental setting. By transiently transfecting cells, we were over-expressing SOX9 isoforms to levels much higher than those found in a physiological setting, forcing cells to adapt their behavior to these extreme conditions. Thus, to really evaluate the impact of SOX9 (and that of T240 phosphorylation) on lung cancer cells behavior, we focused our investigational effort on our main cell model, H1299 cells with inducible Sox9 KD and exogenous re-expression of SOX9 WT or SOX9 T240A. For each assay, we compared the behavior of 4 different populations:

1. H1299 NTC Ctr – expressing physiological levels of endogenous SOX9
2. H1299 KD Ctr – no endogenous or exogenous SOX9 expressed
3. H1299 KD SOX9 WT – only exogenous SOX9 WT expressed
4. H1299 KD SOX9 T240A – only exogenous SOX9 T240A expressed

Once again, we started by looking at the proliferation rate of these cells. Surprisingly, we were not able to appreciate any significative difference between the 4 populations (*Fig.22A*), indicating that SOX9 (and consequently the T240

phosphorylation) were probably not major players in affecting cell proliferation under these conditions.

Then, we assessed the migration ability of these cells through a wound healing assay, using the same live-cell analysis system above-described. Briefly, in a well with cells at full confluence, a scratch (or ''wound") of specific width was generated. The system collected at defined timepoints multiple images of the cells moving from the edges of the scratch towards its center, until the scratch was filled with cells again. By calculating the time needed for the cells to fill the scratch we could extrapolate their migration propensity.

Figure 22. Assessment of SOX9 functional roles.

A. Cell proliferation rate. *B.* Wound healing assay showing wound closure over time. Left: representative images of the assay at specific time point. Gray: cells; yellow: scratch extension (no cells); purple: front of the cell migration. Right: quantification of the wound extension over time. *C.* WB analysis of the EMT marker Vimentin. *D.* Left: Subcutaneous Tumor Volume curves at the indicated time points. Right: WB of protein lysates from subcutaneous tumors collected at the end of the experiment.
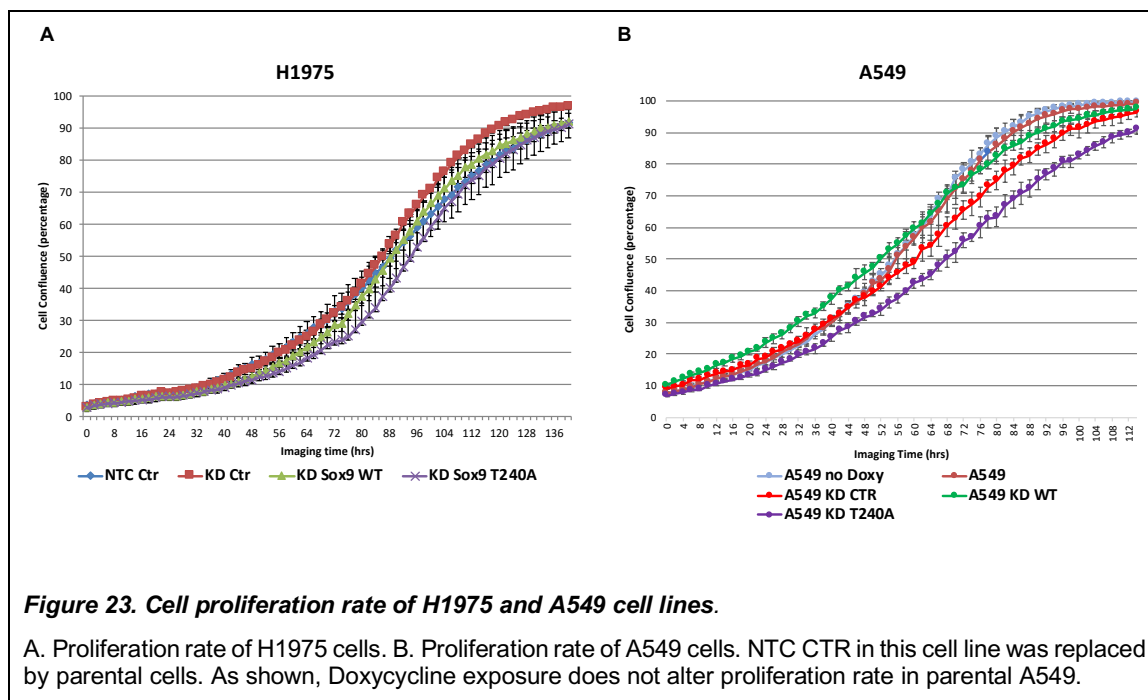
Once again, the slope of the curves describing the migratory dynamics of the cells in the 4 populations were not significantly different, indicating that migration was not affected by SOX9 or pT240 under these experimental conditions (*Fig.22B*).

Next, we evaluated whether another pathological process promoted by SOX9, such as EMT, could be modulated in our experimental conditions. We exposed H1299 cells to Doxycycline for an extended period of time and then we ran a Western Blotting to assess if the expression of any EMT-related marker was changed. Unfortunately, we could not observe any modulation in EMT markers (*Fig.22C*).

We also generated xenografts by injecting H1299 in NOD/SCID mice, in order to test the impact of T240 phosphorylation on tumor growth in vivo. Mice were exposed to Doxycycline-supplemented diet for the duration of the study, and tumors volume were measured at specific timepoints. While SOX9 KD clearly had an impact on the tumorigenic ability of the cells, neither re-expression of SOX9 WT or SOX9 T240A were able to rescue the phenotype (*Fig.22D*).

To evaluate whether the lack of functional phenotype we observed was restricted to H1299 cells, we induced SOX9 KD and re-expression of SOX9 WT or T240A on a different lung adenocarcinoma cell line, H1975, characterized by the presence of 2 common mutation on the EGFR gene (T790M and L858R). However, when we assessed the proliferation rates of the 4 populations, we found that, similarly to H1299, in H1975 cells SOX9 was not a key modulator of cell proliferation under these conditions (*Fig.23A*).

Figure 23. Cell proliferation rate of H1975 and A549 cell lines.

A. Proliferation rate of H1975 cells. B. Proliferation rate of A549 cells. NTC CTR in this cell line was replaced by parental cells. As shown, Doxycycline exposure does not alter proliferation rate in parental A549.

By reviewing the data of a post-doctoral fellow in the lab investigating the interaction between KRAS and SOX9 (data not shown), we started wondering whether the lack of effect seen on H1299 and H1975 cells could be determined by the fact that they are KRAS WT. Thus, we decided to repeat the proliferation assay in a cell lines with mutations in KRAS, such as A549. Interestingly, although we could not appreciate dramatic effects, in this cellular context the absence of SOX9 and the phosphorylation status of T240 resulted in slight variations in the cell proliferation rate (*Fig. 23B*). While not conclusive, these results might justify further investigation in cells characterized by mutations in KRAS.

Taken all together, these experiments provided indications that neither SOX9 nor the phosphorylation status of the T240 residue were able to significantly impact the cancer-associated features we tested in our experimental setting. One potential explanation for the lack of functional phenotypes in a lung cancer context could be the compensation effect generated by other SOX E family members,

which have been shown to share redundant functions with SOX9, especially during development [110]. Thus, we decided to test the relative expression of SOX8 and SOX10 at the RNA levels in our H1299 cells upon SOX9 KD induction. The qRT-PCR results showed that both SOX8 and SOX10 were barely detectable when SOX9 was present in the cells, and their expression was not increased upon SOX9 KD (data not shown).

Although the main focus of this project was to increase our knowledge about the regulatory mechanisms governing SOX9 activity in cancer, we wondered whether our findings could be translated to a non-cancer setting. Given the prominent role of SOX9 in fibrosis [72], we decided to investigate the effects of SOX9 activity modulation on fibroblasts, the principal source of the excessive ECM deposition (one of the main features of fibrosis). Recent studies attributed a role to SOX9 in increasing the proliferation rate of fibroblasts and in protecting them from senescence [156], so we
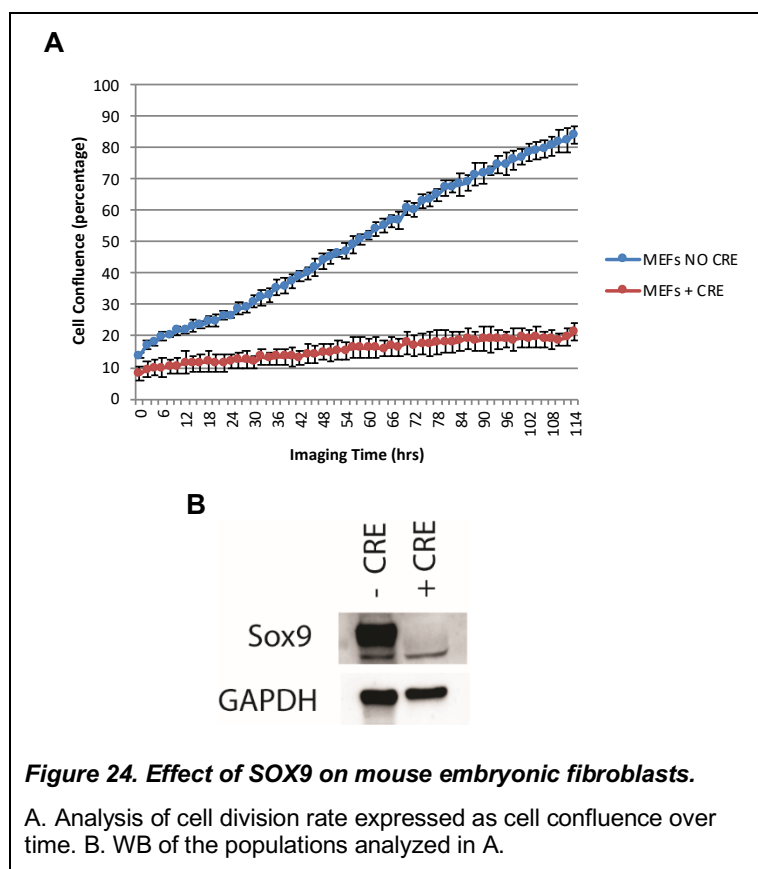
Figure 24. Effect of SOX9 on mouse embryonic fibroblasts.

A. Analysis of cell division rate expressed as cell confluence over time. B. WB of the populations analyzed in A.

thought that it would be interesting to test the role of post-translational modifications on these mechanisms.

We generated mouse embryonic fibroblasts (MEFs) from SOX9 floxed (SOX9$^{f/f}$) mice, engineered so that the SOX9 gene could be removed in a CRE-dependent mechanism. Interestingly, we could confirm that SOX9 knock-out (KO) on these cells has a dramatic effect on cell division rate (*Fig. 24A,B*). These interesting results would be complemented by the analysis of the effects of SOX9 WT and SOX9 T240A re-introduction on SOX9-deficient MEFs, in terms of cell division rate, senescence, and ECM deposition.

Conclusions

The goal of this project was to evaluate the role of the K2 domain in modulating SOX9 functions in the context of lung adenocarcinoma, and to assess how post-translational modifications are involved in this mechanism.

Importantly, we demonstrated that the K2 domain is able to modulate SOX9 transcriptional ability, as shown by the influence of K2 on the induction of Luciferase expression in the different Reporter systems we tested. Moreover, when comparing single domains, the K2 domain showed a higher transactivation capability than the canonical TA domain, as determined by the results of the Gal4 reporter system. Additionally, we proved that specific post-translational modifications, and particularly the phosphorylation in the residue T240, are a key component in mediating the effect we attributed to the K2 domain. In fact, prevention of T240 phosphorylation through a T to A site-directed mutagenesis impaired SOX9 transcriptional ability and also reduced K2 transactivation competence. Phoshomimetics (generated by mutating T to D or E to simulate the presence of the phosphate group on T240) failed in rescuing the phenotype promoted by the T240A mutation, leaving the details of mechanism explaining the T240A-mediated effects an open question. Moreover, we evaluated the influence of K2 on SOX9 transcriptional ability on the global transcriptome of a lung cancer cell line through a RNAseq analysis. We were able to identify a novel set of genes regulated by SOX9 that could open the door to further lines of research.

We verified that the involvement of pT240 was not extended to other SOX9 activities, such as subcellular localization or protein stability. In fact, we

demonstrated that the lack of the phosphate group in T240 does not impact SOX9 nuclear localization and does not determine any change in SOX9 half-life.

Notably, we were also able to show that K2 might represent a region of intrinsic disorder within SOX9, and that the T240 phosphorylation is involved in the maintenance of this disorder. In fact, while K2 lacks ordered structure with the presence of pT240, removal of the phosphorylation induces a drastic change, determining an overall reduction of disorder in favor of a more structured conformation. This is also evident from the relative position of 2 key aa, T236 and T240, which is influenced by the phosphorylation status of T240. When the phosphate group is present, both aa are exposed towards the outside, with their negative charges ready to engage in ionic bonds with the surroundings. However, without the phosphate group, they are brought inside the globular shape of the K2 domain, greatly limiting their ability to interact with the surroundings.

Surprisingly, we were not able to associate the strong findings of our biochemical tests to a clear functional phenotype in the lung cancer setting. We investigated the role of pT240 by using an elegant cell model, in which we could precisely control the timing of SOX9 KD and in which the re-expression levels of SOX9 WT or SOX9 T240A were replicating the physiological expression of endogenous SOX9. However, we could not attribute a prominent role to pT240 in SOX9-mediated features such as proliferation, migratory ability, EMT induction or in vivo tumor growth in the tumor models tested. Nevertheless, these results, while disappointing, do not undermine the validity of our findings.

Discussion

The results of this project contribute to define a prominent role to the K2 domain in mediating some of the SOX9 functions. Particularly, our results show that specific post-translational modifications occurring on this domain, such as phosphorylation of the residue T240, are key mediators of SOX9 transcriptional ability. Moreover, we were able to demonstrate that, when directly compared, the K2 domain has a higher transactivation ability than the canonical TA domain. These findings represent an interesting novelty, since until now the role of the K2 domain in promoting SOX9 transcriptional ability has been considered less prominent than that of the TA domain. Our work sheds some light into the effects of post-translational modifications on SOX9 functions, thus contributing to decipher the mechanisms through which SOX9 elicits its pro-oncogenic role in lung cancer.

While the results from 3 different DLR Reporter systems confirmed that prevention of T240 phosphorylation (through a T to A mutation in that residue) reduces SOX9-mediated transcription, the phosphomimetics we generated failed in rescuing this phenotype. One potential explanation for this failure could be the inability of the phosphomimetics to actually mimic the functional effects of a phosphorylated aa, as previously reported [147-149]. Additionally, the reduction in SOX9 transcriptional ability could be dependent not only on the lack of the negative charge brought by the phosphate (which would be compensate by the negative charges of D and E residues), but also on specific structural rearrangements determined by the T to A mutation. The involvement of structural changes in the pT240-mediated effects were confirmed by molecular dynamics simulations, that

showed the profound changes of K2 conformation promoted by T240 phosphorylation status. In fact, the lack of phosphorylation in T240 determines a drastic reduction of the K2 intrinsic disorder, promoting a spatial re-arrangement of the whole domain (including the key aa T236 and T240) that makes interactions with the surrounding much less likely to occur. These observations are in line with the notion that intrinsic disorder confers to the protein the flexibility needed to assume different conformational requirements for binding to other macromolecules. This differential ability to interact with other macromolecules is particularly interesting, because it could explain the difference in SOX9 transcriptional ability that we observed. In fact, SOX9 elicits its transcriptional modulation on target genes in association with interacting partners. Thus, the conformational changes and reduction in K2 intrinsic disorder induced by the lack of phosphorylation in T240, and the consequent limitation of K2 ability to interact with other macromolecules, could translate in a reduced ability to recruit specific co-activators and co-repressors, thus providing an explanation to the reduction of Sox9 transcriptional ability. Although this effect is mediated by post-translational modifications occurring on the K2 domain, the interaction with co-activators and co-repressors probably involves both the K2 and the TA domains. In fact, in the Gal4 Reporter assay, the combination of the 2 domains resulted in the greatest Luciferase expression and, importantly, T240A mutation caused a reduction in the transactivation ability only when K2 and TA were both present, suggesting that they cooperate in binding partners recruitment.

The confirmation that K2 is an intrinsic disorder within SOX9 would represent an interesting additional novelty element of this project, that we will pursue through other MD simulations and biophysical tests, as described in the next section. Importantly, we verified that the role of pT240 on SOX9 transcriptional ability can't be extended to other SOX9 features, such as subcellular localization or protein stability. Another mechanism potentially impacted by the T240 phosphorylation is SOX9 ability to bind to consensus sequences on target genes. While at this point we haven't tested this possibility yet, we believe that a modulation in DNA binding ability is unlikely, for few reasons. First, there are examples in the literature reporting that mutations in the K2 domain in other SOX E family members do not impact their ability to bind DNA, because of the relative long distance between the HMG (DNA binding domain) and the K2 domain [129]. Second, and more important, we demonstrated that the lack of phosphorylation in T240 has a significative effect on SOX9 transactivation capacity in the Gal4 Reporter system, which in independent of the DNA binding ability.

Given the prominent role of T240 phosphorylation in the effects we observed, we sought to determine what kinases might be promoting for this phosphorylation event. This is an important aspect of the project, because the identification of the kinases phosphorylating T240 could translate into a potential therapeutic opportunity. In fact, the pharmacological target of those kinases would represent a reasonable approach to impair the SOX9-mediated oncogenic functions that are modulated by T240 phosphorylation. Our bioinformatic analysis revealed that the 3 kinases most likely to promote T240 phosphorylation are p38 MAPK, GSK3 and

CDK5. Interestingly, all of them have been already proven to promote tumorigenesis in lung cancer through various mechanisms, and in some cases the effects of their pharmacological inhibition in lung cancer has been already exploited [153-155].

Surprisingly, we could not translate the strong results of our biochemical experiments into a clear functional phenotype in the lung cancer setting. One possible explanation for this is related to the nature of our cell model. In fact, despite the inducible KD system represents a clear improvement and refinement over systems which simply rely on the constitutive overexpression of SOX9 at levels way higher than the physiological ones, it might also introduce an unexpected variable, represented by the antibiotic doxycycline. In fact, there are reports showing that doxycycline (at concentrations normally used for inducible expression) is able to produce confounding effects by altering metabolism and proliferation of human cell lines [157]. Particularly, there are evidence showing that doxycycline can promote changes in gene expression patterns, slow cell proliferation and induce a shift in the cellular metabolism towards a more glycolytic phenotype. Thus, we cannot exclude that the use of doxycycline for KD induction might have contributed to the lack of phenotype in our functional assays. Another factor to keep into consideration when analyzing the results of our functional assays is the mutational profile of the cell lines used. While we clearly showed that the results of our DLR assay with the 4x48-p89 Reporter are independent of the cell host, we could not draw the same conclusion for the functional assays. Many of them were performed in H1299 cells, characterized by loss of P53, among the

other genetic aberrations. While the rationale for the choice of this specific cell line is logic and reasonable (this is the cell line in which we initially observed the constant phosphorylation of T240, which is the foundation of our hypothesis), we cannot exclude that the effects mediated by the T240 phosphorylation status could be masked by mechanisms dependent on the loss of P53. In fact, for at least one of the other cell lines tested in the cell proliferation rate assay (A549, bearing WT P53 but mutant for KRAS), although not dramatic, we were able to see a trend in the results that was consistent with our hypothesis and that might grant further investigation.

Nevertheless, despite our investigational efforts encompassed several SOX9-associated features in cancer such as cell proliferation, cell migration, EMT and in vivo tumorigenesis, we didn't look yet into other well-known SOX9-associated features, such as tumor-formation ability, for instance. In fact, SOX9 has been associated with maintenance of the CSC population in different tumor types, including lung cancer [158]. Thus, an interesting question is whether the phosphorylation status of T240 is influencing this specific subset of cancer cells and their unique features, in lung cancer and possibly also in other cancer types. Additionally, given the preponderant role of SOX9 during chondrogenesis, an interesting venue of research would be to test the functional effects of post-translational modifications on the K2 domain in chondrocytes-related malignancies, such as chondrosarcoma, a form of sarcoma representing about 30% of cancers affecting the skeletal system.

Lastly, although the main focus of this project was lung cancer, we confirmed, as previously reported, that SOX9 plays a prominent role in controlling fibroblasts cell division and senescence, so it would be interesting to test the role of post-translational modifications on these mechanisms.

Future directions

While we are excited to have been able to prove our hypothesis true and to have attributed a key role to post-translational modifications of the K2 domain in modulating SOX9 activity, we believe that there are additional scientific questions stemming from our findings that are definitely worth further investigation. As mentioned before, the confirmation that the K2 domain represents an intrinsic disordered region within SOX9 would be an important finding, potentially explaining the ability of SOX9 to interact with multiple partners and to promote different outcomes in a context-specific manner. To achieve this, our experimental plan includes additional molecular dynamics simulations to infer the conformational changes of the whole SOX9 protein (and not only the K2 domain) upon presence/absence of specific post-translational modifications. To confirm our idea that the inability of phosphomimetic mutants to rescue the phenotype in the DLR assays is determined by their incapacity to mimic the functions promoted by a phosphorylated aa, we will include sequences with phosphomimetics in our future simulations. These informative findings will be validated through biophysical tests (employing techniques such as circular dichroism, hydrogen deuterium exchange and nuclear magnetic resonance (NMR) spectroscopy), to experimentally confirm the conformations predicted by the simulation analysis.

Moreover, we will look into the identification of SOX9 binding partners upon changes in the phosphorylation status of T240, which is a central point to demonstrate that the modulation of SOX9 transcriptional ability is dependent on the differential recruitment of co-activators and co-repressors. This type of analysis

will also allow us to pinpoint the specific parts of SOX9 mediating these interactions, thus implementing our knowledge of the specific contribution to the different domains to SOX9 functions. We plan to identify unknown interacting proteins by pulling down Sox9 WT and Sox9 T240A and then by performing a Mass Spectrometry analysis. Any protein differentially present in the two groups represent a potential candidate and its involvement in SOX9 transcriptional ability will be further validate in additional biochemical tests.

While we believe that the DNA binding ability of SOX9 is not modulated by the phosphorylation status of T240, for the above-described reasons, we will validate our ideas on this matter by performing an electrophoretic mobility shift assay (EMSA). In this in vitro binding assay, we will measure the differential ability of SOX9 WT and SOX9 T240A to bind a biotinylated nucleotide probe by quantification of the complex probe + protein in the two different conditions.

As mentioned before, the identification of the kinases promoting phosphorylation in T240 is an important point since it could represent a therapeutic opportunity by targeting the element responsible for the phosphorylation that is causing the SOX9-mediated pro-oncogenic effects. We plan to perform assays such as in vitro kinase assay or kinase inhibitor experiments to verify if any of the predicted kinases is actually promoting T240 phosphorylation.

Additionally, we will broaden our investigational focus into the functional effects of pT240 in SOX9-mediated cancer-related features by testing the effects of T240 mutations in modulating the tumor-formation ability of CSC populations (through anchorage-independent growth assays) in cancer types for which SOX9 has been

associated with CSC maintenance, such as lung cancer, colorectal cancer and breast cancer [83, 158, 159]. As previously mentioned, given the pivotal role of SOX9 in chondrogenesis, we will also evaluate whether the phosphorylation status of T240 causes functional changes in chondrosarcoma cells.

Finally, given the involvement of SOX9 in fibrosis, we plan to follow up on our preliminary findings showing the dramatic effect of SOX9 loss on cell division rate in MEFs by analyzing of the effects of SOX9 WT and SOX9 T240A re-introduction on SOX9-deficient MEFs, in terms of cell division rate, senescence, and ECM deposition.

Appendix A. Abbreviations used.

3D:	three-dimensional
A:	alanine
aa:	amino acid
ALK:	anaplastic lymphoma kinase
BCC:	basal cell carcinoma
BMP:	bone morphogenetic protein
CBP:	CREB-binding protein
CD:	campomelic dysplasia
CDK4:	cyclin dependent kinase 4
CHX:	cycloheximide
CNS:	central nervous system
CRC:	colorectal cancer
CSC:	cancer stem cell
D:	aspartic acid
DAPI:	4',6-diamidino-2-phenylindole
DBD:	DNA binding domain
DIM:	dimerization
DLR:	dual-luciferase assay
DMEM:	Dulbecco's modified Eagle media
DPBS:	Dulbecco's phosphate-buffered saline
E:	glutamic acid
ECM:	extracellular matrix

EGFP: enhanced green fluorescent protein

EGFR: epidermal growth factor receptor

ELF3:   E74-like factor

EMSA: electrophoretic mobility shift assay

EMT: epithelial-mesenchymal transition

ER^+:   estrogen receptor positive

FACS: fluorescence-assisted cell sorting

FBS: fetal bovine serum

FBW7:   F box and WD repeat domain containing 7

FDA: Food and Drug Administration

GFP: green fluorescent protein

GSCs: glioma stem cells

HMG: high motility group

HSCs: hepatic-stellate cells

IACUC: Institutional Animal Care and Use Committee

ICI: immune checkpoint inhibitor

IDP: intrinsic disordered protein

IDR: intrinsic disordered region

KD: knock-down

KO: knock-out

MAPK: mitogen-activated protein kinase

MCL: mantle cell lymphoma

MD: molecular dynamics

MEFs:	mouse embryonic fibroblasts
MIA:	melanoma inhibitor activity
MOI:	multiplicity of infection
mRNA:	messenger RNA
NC:	neural crest
NES:	nuclear export sequence
NF-Kbeta:	nuclear factor kappa-light-chain-enhancer of activated B cells
NLS:	nuclear localization signal
NMR:	nuclear magnetic resonance
NOD/SCID:	non-obese diabetic/severe combined immunodeficiency
NPCs:	neural progenitor cells
NSCLC:	non-small cell lung cancer
PCR:	polymerase chain reaction
PD-1:	programmed cell death 1
PD-L1:	programmed death-ligand 1
PDK1:	pyruvate dehydrogenase kinase 1
PEI:	polyethyleneimine
PFA:	paraformaldehyde
PIN:	prostate intraepithelial neoplasia
PKA:	protein kinase A
PTEN:	phosphatase and tensin homolog
qRT-PCR:	quantitative reverse-transcription and polymerase chain reaction
RB:	retinoblastoma

RFP: red fluorescent protein

RNAseq: RNA sequencing

RT: reverse transcription

SCLC: small cell lung cancer

SCX: scleraxis

SD: standard deviation

SDM: site-directed mutagenesis

shRNA: short hairpin RNA

SUMO: small ubiquitin-like modifier

T: threonine

TA: transactivation

TAMs: tumor-associated macrophages

TBS-T: Tris-buffered saline with Tween

TetON: tetracycline ON

TGF-beta: transforming growth factor beta

TKI: tyrosine kinase inhibitor

UTR: untranslated region

UV: ultraviolet

WB: western blotting

WT: wild type

Appendix B. Mutational profile of cancer cell lines used.

	P53	KRAS	EGFR
H1299	DELETED	WT	WT
A549	WT	G12S	WT
H322	R248L	WT	WT
H1975	R273H	G12D	T790M/L858R
H838	WT	WT	WT
KP717	DELETED	G12D	WT

Appendix C. Site-directed mutagenesis.

Primers were generated using the QuikChange Primer Design software (Agilent).

All the primers are indicated as 5' to 3'.

Mutation	SDM primers	Template	Template codon	Mutated codon
T236A	**F**:GGTGGGTGGGGCCGGTGGGCCCT	SOX9 WT	ACC	GCC
	R:AGGGCCCACCGGCCCCACCCACC			
T236D	**F**:GGTGGTGGGTGGGTCCGGTGGGCCCTGG	SOX9 WT	ACC	GAC
	R:CCAGGGCCCACCGGACCCACCCACCACC			
T239A	**F**:TTTTGGGGGTGGCGGGTGGGGTCGG	SOX9 WT	ACC	GCC
	R:CCGACCCCACCCGCCACCCCCAAAA			
T239E	**F**:GTCGGTTTTGGGGGTCTCGGGTGGGGTCGGTGG	SOX9 WT	ACC	GAG
	R:CCACCGACCCCACCCGAGACCCCCAAAACCGAC			
T240A	**F**:CGGTTTTGGGGGCGGTGGGTGGGGT	SOX9 WT	ACC	GCC
	R:ACCCCACCCACCGCCCCCAAAACCG			
T240D	**F**:CGTCGGTTTTGGGGTCGGTGGGTGGGGTCG	SOX9 WT	ACC	GAC
	R:CGACCCCACCCACCGACCCCAAAACCGACG			
T240E	**F**:CACGTCGGTTTTGGGCTCGGTGGGTGGGGTCGG	SOX9 WT	ACC	GAG
	R:CCGACCCCACCCACCGAGCCCAAAACCGACGTG			
T240T	**F**:CGGTTTTGGGGGTGGTGGGTGGGGT	SOX9 T240A	GCC	ACC
	R:ACCCCACCCACCACCCCCAAAACC			
T243A	**F**:GCACGTCGGCTTTGGGGGTGGTGGGTG	SOX9 WT	ACC	GCC
	R:CACCCACCACCCCCAAAGCCGACGTGC			
T243E	**F**:CCCGGCTGCACGTCCTCTTTGGGGGTGGTGGGT	SOX9 WT	ACC	GAG
	R:ACCCACCACCCCCAAAGAGGACGTGCAGCCGGG			
T288A	**F**:CTCGTTGACATCGAAGGCCTCGATGTTGGAGATGA	SOX9 WT	ACC	GCC
	R:TCATCTCCAACATCGAGGCCTTCGATGTCAACGAG			
T288E	**F**:GTCAAACTCGTTGACATCGAACTCCTCGATGTTGGAGATGACGTC	SOX9 WT	ACC	GAG
	R:GACGTCATCTCCAACATCGAGGAGTTCGATGTCAACGAGTTTGAC			

Appendix D. List of antibodies used.

Antigen	Company	Cat. # (clone)	Concentration primary ab
SOX9	Millipore	AB5535	1:5,000
GAPDH	Millipore	MAB374	1:10,000
FLAG	CST	14793 (D6W5B)	1:1,000
GFP	CST	2955 (4B10)	1:1,000
pT236	Bethyl	N/A	0.5 µg/ml
pT240	Bethyl	N/A	2 µg/ml
GAL4 DBD	Santa Cruz	SC-510 (RK5C1)	1:1,000
Vimentin	CST	5741 (D21H3)	1:1,000

Appendix E. qRT-PCR primers and thermal profile.

Target	Primer F (5' to 3')	Primer R (5' to 3')
COL1A1	TCTGCGACAACGGCAAGGTG	GACGCCGGTGGTTTCTTGGT
IL11	TCTCTCCTGGCGGACACG	AATCCAGGTTGTGGTCCCC
IL6R	CATTGCCATTGTTCTGAGGTTC	GTGCCACCCAGCCAGCTATC
ARHGAP29	GGAATCAGAACGCAAGCAAAATGCG	GGGATGCTGATTCAGCCTCTTGG
SFRP1	TACAAGAAGATGGTGCTGCC	AGATGTTCAATGATGGCCTC
CHD2	GGCATAGTCTATGGAGAAGT	GCTGTTGTCAGAAGTCTCTC
RHOB	AGAACGGCTGCATCAACTG	CTTGTGGGACACGGGTC
ANXA3	CCCATCAGTGGATGCTGAAG	TCACTAGGGCCACCATGAGA
SNAI2	CAGATGAGCCCTCAGATTTGAC	AGGACACATTAGAACTCACACG
SOX8	CCGAGCTCAGCAAGACG	GTGGCTGGTACTTGTAGTCG
SOX10	CTTCATGGTGTGGGCTCAG	GCTTGTCACTTTCGTTCAGC

Thermal profile:

- 10 minutes at 95C

- 40 cycles at:

    - 15 seconds at 95C

    - 1 minute at 60C

- 30 seconds at 55C

- 30 seconds at 95C

Appendix F. RNAseq results.

Results are presented as lists of genes activated or repressed by SOX9. For each category, genes rescued by SOX9 WT, SOX9 T240A and by both isoforms are listed. Genes are presented as Gene symbol and Ensemble Gene ID.

1. **Genes REPRESSED by SOX9**

   a. Genes rescued by both SOX9 WT and SOX9 T240A (n=58)

   AC008695.1|ENSG00000273217
   ADAM19|ENSG00000135074
   ADGRB2|ENSG00000121753
   AKAP2|ENSG00000241978
   AMPD3|ENSG00000133805
   ANKRD2|ENSG00000165887
   ANPEP|ENSG00000166825
   AOX1|ENSG00000138356
   C1QTNF6|ENSG00000133466
   CAMK1D|ENSG00000183049
   CASZ1|ENSG00000130940
   CCDC69|ENSG00000198624
   CDK18|ENSG00000117266
   CES3|ENSG00000172828
   CHST15|ENSG00000182022
   COL1A1|ENSG00000108821
   COL6A1|ENSG00000142156
   COL6A2|ENSG00000142173
   CSGALNACT1|ENSG00000147408
   CYP27C1|ENSG00000186684
   DLX4|ENSG00000108813
   EFEMP1|ENSG00000115380
   EPHB2|ENSG00000133216
   GDPD5|ENSG00000158555
   GPRC5A|ENSG00000013588
   HTRA1|ENSG00000166033
   IGFBP3|ENSG00000146674
   IL11|ENSG00000095752
   IL6R|ENSG00000160712
   ITGB2|ENSG00000160255
   KBTBD11|ENSG00000176595
   KIF21B|ENSG00000116852
   LINC01270|ENSG00000203999
   LRIG1|ENSG00000144749
   MGLL|ENSG00000074416
   MOB3B|ENSG00000120162
   MYPN|ENSG00000138347

NNMT|ENSG00000166741
OAF|ENSG00000184232
PHACTR1|ENSG00000112137
PLPP3|ENSG00000162407
PTGES|ENSG00000148344
RAET1E|ENSG00000164520
SAA1|ENSG00000173432
SEMA3F|ENSG00000001617
SLC19A3|ENSG00000135917
SLC6A17|ENSG00000197106
SLIT3|ENSG00000184347
SPINT2|ENSG00000167642
SPTB|ENSG00000070182
STC1|ENSG00000159167
SYNPO|ENSG00000171992
TGFA|ENSG00000163235
TMEM158|ENSG00000249992
TMEM40|ENSG00000088726
TNFRSF11A|ENSG00000141655
TRIM6|ENSG00000121236
TRPC4|ENSG00000133107

b. <u>Genes rescued only by SOX9 WT (n=29)</u>

ABCC3|ENSG00000108846
AC005839.1|ENSG00000279089
AC092299.1|ENSG00000282416
AC140134.1|ENSG00000179978
ADRA2C|ENSG00000184160
AL121845.3|ENSG00000273154
ALPK2|ENSG00000198796
ARRB1|ENSG00000137486
B3GNT3|ENSG00000179913
C3orf80|ENSG00000180044
COL18A1|ENSG00000182871
DMBT1|ENSG00000187908
EBI3|ENSG00000105246
EPB41L4A|ENSG00000129595
GATA2-AS1|ENSG00000244300
GHDC|ENSG00000167925
HRH1|ENSG00000196639
MAPK13|ENSG00000156711
MISP|ENSG00000099812
NID2|ENSG00000087303
PIANP|ENSG00000139200
PITX1|ENSG00000069011
RRAD|ENSG00000166592
RSPO4|ENSG00000101282
TNFRSF9|ENSG00000049249
TP53TG3|ENSG00000183632
TST|ENSG00000128311

VGF|ENSG00000128564
ZNF628|ENSG00000197483

c. <u>Genes rescued only by SOX9 T240A (n=15)</u>

AC012184.2|ENSG00000260537
AC079328.2|ENSG00000259498
AC079781.5|ENSG00000284707
AL022238.4|ENSG00000284431
AP005212.3|ENSG00000274214
AP005212.5|ENSG00000283294
CLYBL|ENSG00000125246
GALNT16|ENSG00000100626
LINC00707|ENSG00000238266
LINC01106|ENSG00000175772
NGEF|ENSG00000066248
NHSL2|ENSG00000204131
NLRP3|ENSG00000162711
RNASEL|ENSG00000135828
SAMD12|ENSG00000177570

2. **<u>Genes ACTIVATED by SOX9</u>**

a. <u>Genes rescued by both SOX9 WT and SOX9 T240A (n=62)</u>

AC005944.1|ENSG00000267469
AC008764.4|ENSG00000268790
ACOX2|ENSG00000168306
ADAMTS6|ENSG00000049192
ADGRG1|ENSG00000205336
ADRA1B|ENSG00000170214
AK5|ENSG00000154027
AL136164.4|ENSG00000279312
AL139260.3|ENSG00000274944
AL161431.1|ENSG00000275216
ANXA3|ENSG00000138772
ARHGAP29|ENSG00000137962
ARL4C|ENSG00000188042
C10orf10|ENSG00000165507
C15orf52|ENSG00000188549
CALB2|ENSG00000172137
CD22|ENSG00000012124
CHDH|ENSG00000016391
CLIP4|ENSG00000115295
CLMN|ENSG00000165959
EMP1|ENSG00000134531
ENDOD1|ENSG00000149218
FAM26D|ENSG00000164451
FAM95B1|ENSG00000223839

FLI1|ENSG00000151702
FRMD5|ENSG00000171877
GLIS3|ENSG00000107249
GPAT3|ENSG00000138678
IKZF2|ENSG00000030419
ITGA10|ENSG00000143127
LCTL|ENSG00000188501
LINC02454|ENSG00000256268
LRRC15|ENSG00000172061
MELTF|ENSG00000163975
MIA|ENSG0000026185
MPP4|ENSG00000082126
NES|ENSG00000132688
NPAS2|ENSG00000170485
OTUB2|ENSG00000089723
PDP1|ENSG00000164951
PLEKHA7|ENSG00000166689
PMEPA1|ENSG00000124225
PPP2R3A|ENSG00000073711
PRICKLE1|ENSG00000139174
PRR16|ENSG00000184838
RIPK4|ENSG00000183421
RNF128|ENSG00000133135
RP1|ENSG00000104237
S1PR1|ENSG00000170989
SFRP1|ENSG00000104332
SH3TC2|ENSG00000169247
SOX9|ENSG00000125398
SPARC|ENSG00000113140
SRPX|ENSG00000101955
SYTL2|ENSG00000137501
THBS1|ENSG00000137801
TM4SF18|ENSG00000163762
TMSB4X|ENSG00000205542
TP53TG3C|ENSG00000205457
VIPR1|ENSG00000114812

b. <u>Genes rescued only by SOX9 WT (n=19)</u>

AC005538.2|ENSG00000279809
AC010487.3|ENSG00000283088
AC024940.1|ENSG00000177359
AHNAK2|ENSG00000185567
AL158066.1|ENSG00000217576
AL512590.3|ENSG00000255036
COLGALT2|ENSG00000198756
CU633967.1|ENSG00000274333
FAM89A|ENSG00000182118
FOXN3|ENSG0000005325

FP565260.2|ENSG00000276612
LINC00941|ENSG00000235884
MAP2|ENSG00000078018
MYO10|ENSG00000145555
PMP22|ENSG00000109099
RASGRP3|ENSG00000152689
SAV1|ENSG00000151748
SRPX2|ENSG00000102359
TRHDE|ENSG00000072657

c. <u>Genes rescued only by SOX9 T240A (n=8)</u>

AC022167.2|ENSG00000260276
AL049840.2|ENSG00000269910
AL671710.1|ENSG00000273192
ARMCX7P|ENSG00000204072
C9orf43|ENSG00000157653
KCNIP3|ENSG00000115041
RPL17-C18orf32|ENSG00000215472
SLC27A1|ENSG00000130304

Bibliography

1. Ferlay, J., et al., *Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012.* Eur J Cancer, 2013. **49**(6): p. 1374-403.

2. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2019.* CA Cancer J Clin, 2019. **69**(1): p. 7-34.

3. Travis, W.D., et al., *The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification.* J Thorac Oncol, 2015. **10**(9): p. 1243-1260.

4. Maggiore, C., et al., *Histological classification of lung cancer.* Rays, 2004. **29**(4): p. 353-5.

5. Dela Cruz, C.S., L.T. Tanoue, and R.A. Matthay, *Lung cancer: epidemiology, etiology, and prevention.* Clin Chest Med, 2011. **32**(4): p. 605-44.

6. Wang, S., et al., *Survival improvement in patients with non-small cell lung cancer between 1983 and 2012: Analysis of the Surveillance, Epidemiology, and End Results database.* Tumour Biol, 2017. **39**(5): p. 1010428317691677.

7. Molina, J.R., et al., *Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship.* Mayo Clin Proc, 2008. **83**(5): p. 584-94.

8. Morgensztern, D., et al., *Trends in stage distribution for patients with non-small cell lung cancer: a National Cancer Database survey.* J Thorac Oncol, 2010. **5**(1): p. 29-33.

9. Ettinger, D.S., et al., *Non-Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology.* J Natl Compr Canc Netw, 2017. **15**(4): p. 504-535.

10. Vargas, A.J. and C.C. Harris, *Biomarker development in the precision medicine era: lung cancer as a case study.* Nat Rev Cancer, 2016. **16**(8): p. 525-37.

11. Cancer Genome Atlas Research, N., *Comprehensive molecular profiling of lung adenocarcinoma.* Nature, 2014. **511**(7511): p. 543-50.

12. Yousefi, H., et al., *Immunotherapy of cancers comes of age.* Expert Rev Clin Immunol, 2017. **13**(10): p. 1001-1015.

13. Nadal, E., et al., *Immunotherapy with checkpoint inhibitors in non-small cell lung cancer: insights from long-term survivors.* Cancer Immunol Immunother, 2019.

14. Sanz-Ezquerro, J.J., A.E. Munsterberg, and S. Stricker, *Editorial: Signaling Pathways in Embryonic Development.* Front Cell Dev Biol, 2017. **5**: p. 76.

15. Perrimon, N., C. Pitsouli, and B.Z. Shilo, *Signaling mechanisms controlling cell fate and embryonic patterning.* Cold Spring Harb Perspect Biol, 2012. **4**(8): p. a005975.

16. Dreesen, O. and A.H. Brivanlou, *Signaling pathways in cancer and embryonic stem cells.* Stem Cell Rev, 2007. **3**(1): p. 7-17.

17. Zhu, Z.X., et al., *Hedgehog signaling contributes to basic fibroblast growth factor-regulated fibroblast migration.* Exp Cell Res, 2017. **355**(2): p. 83-94.

18. Hao, K., et al., *Hedgehog signaling pathway regulates human pancreatic cancer cell proliferation and metastasis.* Oncol Rep, 2013. **29**(3): p. 1124-32.

19. Liu, J., et al., *Reconstruction of the gene regulatory network involved in the sonic hedgehog pathway with a potential role in early development of the mouse brain.* PLoS Comput Biol, 2014. **10**(10): p. e1003884.

20. Becher, O.J., et al., *Gli activity correlates with tumor grade in platelet-derived growth factor-induced gliomas.* Cancer Res, 2008. **68**(7): p. 2241-9.

21. Kubo, M., et al., *Hedgehog signaling pathway is a new therapeutic target for patients with breast cancer.* Cancer Res, 2004. **64**(17): p. 6071-4.

22. O'Reilly, K.E., et al., *Hedgehog pathway blockade inhibits melanoma cell growth in vitro and in vivo.* Pharmaceuticals (Basel), 2013. **6**(11): p. 1429-50.

23. Sheng, T., et al., *Activation of the hedgehog pathway in advanced prostate cancer.* Mol Cancer, 2004. **3**: p. 29.

24. Salaritabar, A., et al., *Targeting Hedgehog signaling pathway: Paving the road for cancer therapy.* Pharmacol Res, 2019. **141**: p. 466-480.

25. Yang, S.H., et al., *Pathological responses to oncogenic Hedgehog signaling in skin are dependent on canonical Wnt/beta3-catenin signaling.* Nat Genet, 2008. **40**(9): p. 1130-5.

26. Noubissi, F.K., et al., *Cross-Talk between Wnt and Hh Signaling Pathways in the Pathology of Basal Cell Carcinoma.* Int J Environ Res Public Health, 2018. **15**(7).

27. Harris, P.J., G. Speranza, and C. Dansky Ullmann, *Targeting embryonic signaling pathways in cancer therapy.* Expert Opin Ther Targets, 2012. **16**(1): p. 131-45.

28. Borah, A., et al., *Targeting self-renewal pathways in cancer stem cells: clinical implications for cancer therapy.* Oncogenesis, 2015. **4**: p. e177.

29. Berta, P., et al., *Genetic evidence equating SRY and the testis-determining factor.* Nature, 1990. **348**(6300): p. 448-50.

30. Castillo, S.D. and M. Sanchez-Cespedes, *The SOX family of genes in cancer development: biological relevance and opportunities for therapy.* Expert Opin Ther Targets, 2012. **16**(9): p. 903-19.

31. Kamachi, Y. and H. Kondoh, *Sox proteins: regulators of cell fate specification and differentiation.* Development, 2013. **140**(20): p. 4129-44.

32. Ragge, N.K., et al., *SOX2 anophthalmia syndrome.* Am J Med Genet A, 2005. **135**(1): p. 1-7; discussion 8.

33. Mollaaghababa, R. and W.J. Pavan, *The importance of having your SOX on: role of SOX10 in the development of neural crest-derived melanocytes and glia.* Oncogene, 2003. **22**(20): p. 3024-34.

34. Laumonnier, F., et al., *Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency.* Am J Hum Genet, 2002. **71**(6): p. 1450-5.

35. Wagner, T., et al., *Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9.* Cell, 1994. **79**(6): p. 1111-20.

36. Sinclair, A.H., et al., *A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif.* Nature, 1990. **346**(6281): p. 240-4.

37. Xia, Y., et al., *The oncogenic potential of the high mobility group box protein Sox3.* Cancer Res, 2000. **60**(22): p. 6303-6.

38. Kim, R., et al., *Genome-based identification of cancer genes by proviral tagging in mouse retrovirus-induced T-cell lymphomas.* J Virol, 2003. **77**(3): p. 2056-62.

39.	Dong, C., D. Wilhelm, and P. Koopman, *Sox genes and cancer.* Cytogenet Genome Res, 2004. **105**(2-4): p. 442-7.

40.	Rodriguez-Pinilla, S.M., et al., *Sox2: a possible driver of the basal-like phenotype in sporadic breast cancer.* Mod Pathol, 2007. **20**(4): p. 474-81.

41.	Lengerke, C., et al., *Expression of the embryonic stem cell marker SOX2 in early-stage breast carcinoma.* BMC Cancer, 2011. **11**: p. 42.

42.	Ben-Porath, I., et al., *An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.* Nat Genet, 2008. **40**(5): p. 499-507.

43.	Gangemi, R.M., et al., *SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity.* Stem Cells, 2009. **27**(1): p. 40-8.

44.	Jeon, H.M., et al., *ID4 imparts chemoresistance and cancer stemness to glioma cells by derepressing miR-9*-mediated suppression of SOX2.* Cancer Res, 2011. **71**(9): p. 3410-21.

45.	Saigusa, S., et al., *Correlation of CD133, OCT4, and SOX2 in rectal cancer and their association with distant recurrence after chemoradiotherapy.* Ann Surg Oncol, 2009. **16**(12): p. 3488-98.

46.	Aue, G., et al., *Sox4 cooperates with PU.1 haploinsufficiency in murine myeloid leukemia.* Blood, 2011. **118**(17): p. 4674-81.

47.	Castillo, S.D., et al., *Novel transcriptional targets of the SRY-HMG box transcription factor SOX4 link its expression to the development of small cell lung cancer.* Cancer Res, 2012. **72**(1): p. 176-86.

48.	Andersen, C.L., et al., *Dysregulation of the transcription factors SOX4, CBFB and SMARCC1 correlates with outcome of colorectal cancer.* Br J Cancer, 2009. **100**(3): p. 511-23.

49.	Sinner, D., et al., *Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells.* Mol Cell Biol, 2007. **27**(22): p. 7802-15.

50.	Liu, P., et al., *Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells.* Cancer Res, 2006. **66**(8): p. 4011-9.

51.	Pramoonjago, P., A.S. Baras, and C.A. Moskaluk, *Knockdown of Sox4 expression by RNAi induces apoptosis in ACC3 cells.* Oncogene, 2006. **25**(41): p. 5626-39.

52. Ek, S., et al., *Nuclear expression of the non B-cell lineage Sox11 transcription factor identifies mantle cell lymphoma.* Blood, 2008. **111**(2): p. 800-5.

53. Jo, A., et al., *The versatile functions of Sox9 in development, stem cells, and human diseases.* Genes Dis, 2014. **1**(2): p. 149-161.

54. Wilhelm, D., et al., *SOX9 regulates prostaglandin D synthase gene transcription in vivo to ensure testis development.* J Biol Chem, 2007. **282**(14): p. 10553-60.

55. Lasala, C., et al., *SOX9 and SF1 are involved in cyclic AMP-mediated upregulation of anti-Mullerian gene expression in the testicular prepubertal Sertoli cell line SMAT1.* Am J Physiol Endocrinol Metab, 2011. **301**(3): p. E539-47.

56. Leung, V.Y., et al., *SOX9 governs differentiation stage-specific gene expression in growth plate chondrocytes via direct concomitant transactivation and repression.* PLoS Genet, 2011. **7**(11): p. e1002356.

57. Ikeda, T., et al., *The combination of SOX5, SOX6, and SOX9 (the SOX trio) provides signals sufficient for induction of permanent cartilage.* Arthritis Rheum, 2004. **50**(11): p. 3561-73.

58. Kang, P., et al., *Sox9 and NFIA coordinate a transcriptional regulatory cascade during the initiation of gliogenesis.* Neuron, 2012. **74**(1): p. 79-94.

59. Stolt, C.C., et al., *The Sox9 transcription factor determines glial fate choice in the developing spinal cord.* Genes Dev, 2003. **17**(13): p. 1677-89.

60. Sakai, D. and Y. Wakamatsu, *Regulatory mechanisms for neural crest formation.* Cells Tissues Organs, 2005. **179**(1-2): p. 24-35.

61. Seymour, P.A., et al., *SOX9 is required for maintenance of the pancreatic progenitor cell pool.* Proc Natl Acad Sci U S A, 2007. **104**(6): p. 1865-70.

62. Belo, J., et al., *The role of SOX9 transcription factor in pancreatic and duodenal development.* Stem Cells Dev, 2013. **22**(22): p. 2935-43.

63. Carpentier, R., et al., *Embryonic ductal plate cells give rise to cholangiocytes, periportal hepatocytes, and adult liver progenitor cells.* Gastroenterology, 2011. **141**(4): p. 1432-8, 1438 e1-4.

64. Bastide, P., et al., *Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium.* J Cell Biol, 2007. **178**(4): p. 635-48.

65. Okubo, T., et al., *Nmyc plays an essential role during lung development as a dosage-sensitive regulator of progenitor cell proliferation and differentiation.* Development, 2005. **132**(6): p. 1363-74.

66. Lu, Y., et al., *Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells.* Dev Biol, 2007. **310**(2): p. 442-53.

67. Turcatel, G., et al., *Lung mesenchymal expression of Sox9 plays a critical role in tracheal development.* BMC Biol, 2013. **11**: p. 117.

68. Scott, C.E., et al., *SOX9 induces and maintains neural stem cells.* Nat Neurosci, 2010. **13**(10): p. 1181-9.

69. Mori-Akiyama, Y., et al., *SOX9 is required for the differentiation of paneth cells in the intestinal epithelium.* Gastroenterology, 2007. **133**(2): p. 539-46.

70. Blache, P., et al., *SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes.* J Cell Biol, 2004. **166**(1): p. 37-47.

71. Mansour, S., et al., *A clinical and genetic study of campomelic dysplasia.* J Med Genet, 1995. **32**(6): p. 415-20.

72. Hanley, K.P., et al., *Ectopic SOX9 mediates extracellular matrix deposition characteristic of organ fibrosis.* J Biol Chem, 2008. **283**(20): p. 14063-71.

73. Wang, L., et al., *Oncogenic role of SOX9 expression in human malignant glioma.* Med Oncol, 2012. **29**(5): p. 3484-90.

74. Wang, Z., et al., *SOX9-PDK1 axis is essential for glioma stem cell self-renewal and temozolomide resistance.* Oncotarget, 2018. **9**(1): p. 192-204.

75. Thomsen, M.K., et al., *SOX9 elevation in the prostate promotes proliferation and cooperates with PTEN loss to drive tumor formation.* Cancer Res, 2010. **70**(3): p. 979-87.

76. Wang, G., et al., *Zbtb7a suppresses prostate cancer through repression of a Sox9-dependent pathway for cellular senescence bypass and tumor invasion.* Nat Genet, 2013. **45**(7): p. 739-746.

77. Wang, H., et al., *SOX9 is expressed in normal prostate basal cells and regulates androgen receptor expression in prostate cancer cells.* Cancer Res, 2007. **67**(2): p. 528-36.

78. Sakamoto, H., et al., *SOX9 Is Highly Expressed in Nonampullary Duodenal Adenoma and Adenocarcinoma in Humans.* Gut Liver, 2013. **7**(5): p. 513-8.

79. Kopp, J.L., et al., *Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma.* Cancer Cell, 2012. **22**(6): p. 737-50.

80. Sun, L., et al., *Epigenetic regulation of SOX9 by the NF-kappaB signaling pathway in pancreatic cancer stem cells.* Stem Cells, 2013. **31**(8): p. 1454-66.

81. Guo, W., et al., *Slug and Sox9 cooperatively determine the mammary stem cell state.* Cell, 2012. **148**(5): p. 1015-28.

82. Fazilaty, H., et al., *SLUG and SOX9 Cooperatively Regulate Tumor Initiating Niche Factors in Breast Cancer.* Cancer Microenviron, 2016. **9**(1): p. 71-4.

83. Xue, Y., et al., *SOX9/FXYD3/Src Axis Is Critical for ER(+) Breast Cancer Stem Cell Function.* Mol Cancer Res, 2019. **17**(1): p. 238-249.

84. Jeselsohn, R., et al., *Embryonic transcription factor SOX9 drives breast cancer endocrine resistance.* Proc Natl Acad Sci U S A, 2017. **114**(22): p. E4482-E4491.

85. Kawai, T., et al., *SOX9 is a novel cancer stem cell marker surrogated by osteopontin in human hepatocellular carcinoma.* Sci Rep, 2016. **6**: p. 30489.

86. Liu, C., et al., *Sox9 regulates self-renewal and tumorigenicity by promoting symmetrical cell division of cancer stem cells in hepatocellular carcinoma.* Hepatology, 2016. **64**(1): p. 117-29.

87. Ruan, H., et al., *Upregulated SOX9 expression indicates worse prognosis in solid tumors: a systematic review and meta-analysis.* Oncotarget, 2017. **8**(68): p. 113163-113173.

88. Passeron, T., et al., *Upregulation of SOX9 inhibits the growth of human and mouse melanomas and restores their sensitivity to retinoic acid.* J Clin Invest, 2009. **119**(4): p. 954-63.

89. Yang, X., et al., *SOX9 is a dose-dependent metastatic fate determinant in melanoma.* J Exp Clin Cancer Res, 2019. **38**(1): p. 17.

90. Choi, B.J., et al., *Hypoxia induces epithelial-mesenchymal transition in colorectal cancer cells through ubiquitin-specific protease 47-mediated*

stabilization of Snail: A potential role of Sox9. Sci Rep, 2017. **7**(1): p. 15918.

91.   Panza, A., et al., *Interplay between SOX9, beta-catenin and PPARgamma activation in colorectal cancer.* Biochim Biophys Acta, 2013. **1833**(8): p. 1853-65.

92.   Qian, Y., S. Xia, and Z. Feng, *Sox9 mediated transcriptional activation of FOXK2 is critical for colorectal cancer cells proliferation.* Biochem Biophys Res Commun, 2017. **483**(1): p. 475-481.

93.   Shen, Z., et al., *Identification of the interplay between SOX9 and S100P in the metastasis and invasion of colon carcinoma.* Oncotarget, 2015. **6**(24): p. 20672-84.

94.   Matheu, A., et al., *Oncogenicity of the developmental transcription factor Sox9.* Cancer Res, 2012. **72**(5): p. 1301-15.

95.   Marcker Espersen, M.L., et al., *SOX9 expression predicts relapse of stage II colon cancer patients.* Hum Pathol, 2016. **52**: p. 38-46.

96.   Bruun, J., et al., *Prognostic Significance of beta-Catenin, E-Cadherin, and SOX9 in Colorectal Cancer: Results from a Large Population-Representative Series.* Front Oncol, 2014. **4**: p. 118.

97.   Abdel-Samad, R., et al., *MiniSOX9, a dominant-negative variant in colon cancer cells.* Oncogene, 2011. **30**(22): p. 2493-503.

98.   Prevostel, C. and P. Blache, *The dose-dependent effect of SOX9 and its incidence in colorectal cancer.* Eur J Cancer, 2017. **86**: p. 150-157.

99.   Prevostel, C., et al., *SOX9 is an atypical intestinal tumor suppressor controlling the oncogenic Wnt/ss-catenin signaling.* Oncotarget, 2016. **7**(50): p. 82228-82243.

100.   Jiang, S.S., et al., *Upregulation of SOX9 in lung adenocarcinoma and its involvement in the regulation of cell growth and tumorigenicity.* Clin Cancer Res, 2010. **16**(17): p. 4363-73.

101.   Zhou, C.H., et al., *Clinical significance of SOX9 in human non-small cell lung cancer progression and overall patient survival.* J Exp Clin Cancer Res, 2012. **31**: p. 18.

102.   Capaccione, K.M., et al., *Sox9 mediates Notch1-induced mesenchymal features in lung adenocarcinoma.* Oncotarget, 2014. **5**(11): p. 3636-50.

103. Wang, X., et al., *Upregulation of SOX9 promotes cell proliferation, migration and invasion in lung adenocarcinoma.* Oncol Lett, 2015. **10**(2): p. 990-994.

104. Zhang, S., et al., *Tumor-associated macrophages promote tumor metastasis via the TGF-beta/SOX9 axis in non-small cell lung cancer.* Oncotarget, 2017. **8**(59): p. 99801-99815.

105. Guo, Y.Z., et al., *SOX9 regulated proliferation and apoptosis of human lung carcinoma cells by the Wnt/beta-catenin signaling pathway.* Eur Rev Med Pharmacol Sci, 2018. **22**(15): p. 4898-4907.

106. Luanpitpong, S., et al., *SLUG is required for SOX9 stabilization and functions to promote cancer stem cells and metastasis in human lung carcinoma.* Oncogene, 2016. **35**(22): p. 2824-33.

107. Yu, Y., et al., *Exploring the mechanism of non-small-cell lung cancer cell lines resistant to epidermal growth factor receptor tyrosine kinase inhibitor.* J Cancer Res Ther, 2016. **12**(1): p. 121-5.

108. McDowall, S., et al., *Functional and structural studies of wild type SOX9 and mutations causing campomelic dysplasia.* J Biol Chem, 1999. **274**(34): p. 24023-30.

109. Kamachi, Y., M. Uchikawa, and H. Kondoh, *Pairing SOX off: with partners in the regulation of embryonic development.* Trends Genet, 2000. **16**(4): p. 182-7.

110. Finzsch, M., et al., *Sox9 and Sox10 influence survival and migration of oligodendrocyte precursors in the spinal cord by regulating PDGF receptor alpha expression.* Development, 2008. **135**(4): p. 637-46.

111. Stolt, C.C., et al., *Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy.* Development, 2004. **131**(10): p. 2349-58.

112. Kellerer, S., et al., *Replacement of the Sox10 transcription factor by Sox8 reveals incomplete functional equivalence.* Development, 2006. **133**(15): p. 2875-86.

113. Furumatsu, T., et al., *Smad3 induces chondrogenesis through the activation of SOX9 via CREB-binding protein/p300 recruitment.* J Biol Chem, 2005. **280**(9): p. 8343-50.

114. Otero, M., et al., *ELF3 modulates type II collagen gene (COL2A1) transcription in chondrocytes by inhibiting SOX9-CBP/p300-driven histone acetyltransferase activity.* Connect Tissue Res, 2017. **58**(1): p. 15-26.

115. Cheng, A. and P.G. Genever, *SOX9 determines RUNX2 transactivity by directing intracellular degradation.* J Bone Miner Res, 2010. **25**(12): p. 2680-9.

116. Topol, L., et al., *Sox9 inhibits Wnt signaling by promoting beta-catenin phosphorylation in the nucleus.* J Biol Chem, 2009. **284**(5): p. 3323-33.

117. Pritchett, J., et al., *Understanding the role of SOX9 in acquired diseases: lessons from development.* Trends Mol Med, 2011. **17**(3): p. 166-74.

118. Malki, S., et al., *Prostaglandin D2 induces nuclear import of the sex-determining factor SOX9 via its cAMP-PKA phosphorylation.* EMBO J, 2005. **24**(10): p. 1798-809.

119. Huang, W., et al., *Phosphorylation of SOX9 by cyclic AMP-dependent protein kinase A enhances SOX9's ability to transactivate a Col2a1 chondrocyte-specific enhancer.* Mol Cell Biol, 2000. **20**(11): p. 4149-58.

120. Coricor, G. and R. Serra, *TGF-beta regulates phosphorylation and stabilization of Sox9 protein in chondrocytes through p38 and Smad dependent mechanisms.* Sci Rep, 2016. **6**: p. 38616.

121. Oh, H.J., T. Kido, and Y.F. Lau, *PIAS1 interacts with and represses SOX9 transactivation activity.* Mol Reprod Dev, 2007. **74**(11): p. 1446-55.

122. Hattori, T., et al., *Interactions between PIAS proteins and SOX9 result in an increase in the cellular concentrations of SOX9.* J Biol Chem, 2006. **281**(20): p. 14417-28.

123. Yi, T., et al., *Quantitative phosphoproteomic analysis reveals system-wide signaling pathways downstream of SDF-1/CXCR4 in breast cancer stem cells.* Proc Natl Acad Sci U S A, 2014. **111**(21): p. E2182-90.

124. Mertins, P., et al., *Proteogenomics connects somatic mutations to signalling in breast cancer.* Nature, 2016. **534**(7605): p. 55-62.

125. Mertins, P., et al., *Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels.* Mol Cell Proteomics, 2014. **13**(7): p. 1690-704.

126. Klammer, M., et al., *Phosphosignature predicts dasatinib response in non-small cell lung cancer.* Mol Cell Proteomics, 2012. **11**(9): p. 651-68.

127. Hong, X., et al., *SOX9 is targeted for proteasomal degradation by the E3 ligase FBW7 in response to DNA damage.* Nucleic Acids Res, 2016. **44**(18): p. 8855-8869.

128. Suryo Rahmanto, A., et al., *FBW7 suppression leads to SOX9 stabilization and increased malignancy in medulloblastoma.* EMBO J, 2016. **35**(20): p. 2192-2212.

129. Schreiner, S., et al., *Hypomorphic Sox10 alleles reveal novel protein functions and unravel developmental differences in glial lineages.* Development, 2007. **134**(18): p. 3271-81.

130. Cossais, F., et al., *SOX10 structure-function analysis in the chicken neural tube reveals important insights into its role in human neurocristopathies.* Hum Mol Genet, 2010. **19**(12): p. 2409-20.

131. Wegner, M., *From head to toes: the multiple facets of Sox proteins.* Nucleic Acids Res, 1999. **27**(6): p. 1409-20.

132. Geraldo, M.T., et al., *Dimerization and Transactivation Domains as Candidates for Functional Modulation and Diversity of Sox9.* PLoS One, 2016. **11**(5): p. e0156199.

133. Wright, P.E. and H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation.* Nat Rev Mol Cell Biol, 2015. **16**(1): p. 18-29.

134. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.

135. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins.* Chem Rev, 2014. **114**(13): p. 6589-631.

136. Shammas, S.L., *Mechanistic roles of protein disorder within transcription.* Curr Opin Struct Biol, 2017. **42**: p. 155-161.

137. Durkin, M.E., et al., *Isolation of Mouse Embryo Fibroblasts.* Bio Protoc, 2013. **3**(18).

138. Nihongaki, Y., et al., *CRISPR-Cas9-based photoactivatable transcription system.* Chem Biol, 2015. **22**(2): p. 169-74.

139. Potter, C.J., et al., *The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis.* Cell, 2010. **141**(3): p. 536-48.

140. Lou, E., et al., *Tunneling Nanotubes: A new paradigm for studying intercellular communication and therapeutics in cancer.* Commun Integr Biol, 2012. **5**(4): p. 399-403.

141. Sastry, L., et al., *Titering lentiviral vectors: comparison of DNA, RNA and marker expression methods.* Gene Ther, 2002. **9**(17): p. 1155-62.

142. Linding, R., et al., *GlobPlot: Exploring protein sequences for globularity and disorder.* Nucleic Acids Res, 2003. **31**(13): p. 3701-8.

143. Murakami, S., V. Lefebvre, and B. de Crombrugghe, *Potent inhibition of the master chondrogenic factor Sox9 gene by interleukin-1 and tumor necrosis factor-alpha.* J Biol Chem, 2000. **275**(5): p. 3687-92.

144. Lefebvre, V., et al., *SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene.* Mol Cell Biol, 1997. **17**(4): p. 2336-46.

145. Girardot, M., et al., *SOX9 has distinct regulatory roles in alternative splicing and transcription.* Nucleic Acids Res, 2018. **46**(17): p. 9106-9118.

146. Jay, P., P. Berta, and P. Blache, *Expression of the carcinoembryonic antigen gene is inhibited by SOX9 in human colon carcinoma cells.* Cancer Res, 2005. **65**(6): p. 2193-8.

147. Capasso, H., et al., *Phosphorylation activates Chk1 and is required for checkpoint-mediated cell cycle arrest.* J Cell Sci, 2002. **115**(Pt 23): p. 4555-64.

148. Paleologou, K.E., et al., *Phosphorylation at Ser-129 but not the phosphomimics S129E/D inhibits the fibrillation of alpha-synuclein.* J Biol Chem, 2008. **283**(24): p. 16895-905.

149. Anthis, N.J., et al., *Beta integrin tyrosine phosphorylation is a conserved mechanism for regulating talin-induced integrin activation.* J Biol Chem, 2009. **284**(52): p. 36700-10.

150. Bruckner, A., et al., *Yeast two-hybrid, a powerful tool for systems biology.* Int J Mol Sci, 2009. **10**(6): p. 2763-88.

151. Oldfield, C.J., et al., *Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.* BMC Genomics, 2008. **9 Suppl 1**: p. S1.

152. Uversky, V.N., *p53 Proteoforms and Intrinsic Disorder: An Illustration of the Protein Structure-Function Continuum Concept.* Int J Mol Sci, 2016. **17**(11).

153. Koul, H.K., M. Pal, and S. Koul, *Role of p38 MAP Kinase Signal Transduction in Solid Tumors.* Genes Cancer, 2013. **4**(9-10): p. 342-59.

154. Yoshino, Y. and C. Ishioka, *Inhibition of glycogen synthase kinase-3 beta induces apoptosis and mitotic catastrophe by disrupting centrosome regulation in cancer cells.* Sci Rep, 2015. **5**: p. 13249.

155. Zeng, J., et al., *CDK5 Functions as a Tumor Promoter in Human Lung Cancer.* J Cancer, 2018. **9**(21): p. 3950-3961.

156. Zhu, Z., et al., *Sox9 Protects against Human Lung Fibroblast Cell Apoptosis Induced by LPS through Activation of the AKT/GSK3beta Pathway.* Biochemistry (Mosc), 2017. **82**(5): p. 606-612.

157. Ahler, E., et al., *Doxycycline alters metabolism and proliferation of human cell lines.* PLoS One, 2013. **8**(5): p. e64561.

158. Voronkova, M.A., et al., *SOX9 Regulates Cancer Stem-Like Properties and Metastatic Potential of Single-Walled Carbon Nanotube-Exposed Cells.* Sci Rep, 2017. **7**(1): p. 11653.

159. Carrasco-Garcia, E., et al., *SOX9-regulated cell plasticity in colorectal metastasis is attenuated by rapamycin.* Sci Rep, 2016. **6**: p. 32350.