Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
der Fakultät für Biologie der Ludwig-Maximilians-
Universität München

The role of the transcription factor SOX2 in tumorigenesis and development of the stomach

Vorgelegt von

Katharina Antonia Hütz

München

2013

„Wenn Du ein Schiff bauen willst, dann trommle nicht Männer zusammen um Holz zu beschaffen, Aufgaben zu vergeben und die Arbeit einzuteilen, sondern lehre die Männer die Sehnsucht nach dem weiten, endlosen Meer.“

Antoine de Saint-Exupery

Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 (in der Fassung der sechsten Änderungssatzung vom 16. August 2010) von Prof. Thomas Cremer von der Fakultät der Biologie vertreten.

Eidesstattliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, …….

Katharina Hütz

Dissertation eingereicht am 06. Juni 2013

1. Gutachter: Prof. Thomas Cremer
2. Gutachter: Prof. Elisabeth Weiss

Mündliche Prüfung am 21. Oktober 2013

Meinen Eltern und meinen Schwestern

Table of Contents

Table of Contents 9
List of Abbreviations 13
1. Introduction 17
1.1. The stomach 17
1.1.1. Development of the stomach 17
1.1.2. Anatomy and function of the stomach 17
1.1.3. Gastric stem cells 19
1.2. Gastric cancer 24
1.2.1. Epidemiology of gastric cancer 24
1.2.2. Classification of gastric cancer 24
1.2.3. Development of gastric cancer 25
1.2.4. Helicobacter pylori and gastric cancer 30
1.2.5. The Cancer stem cell theory 33
1.3. The transcription factor SOX2 34
1.3.1. General characteristics of SOX (**S**ry b**ox**) proteins 34
1.3.2. Classification and localization of SOX2 35
1.3.3. SOX2 in pluripotency and development 36
1.3.4. SOX2 and cancer 38
1.4. Aims of the study 41
2. Methods 43
2.1. Molecular cloning 43
2.1.1. Cloning strategies 43
2.1.2 Restriction digestion of DNA 44
2.1.3 Ligation 44
2.1.4 Competent E. coli cells 44
2.1.5 Transformation of DNA in bacteria 45
2.1.6 Preparation of plasmid DNA 45
2.1.7 shRNA cloning 45
2.1.8 Agarose gel electrophoresis 46
2.2 Cell culture 46
2.2.1. General cell culture methods 47
2.2.1.1. Cell counting 47
2.2.1.2. Freezing cells 47

2.2.1.3. Thawing and maintaining cells 47
2.2.2. Transfection of cells 48
2.2.2.1. Transient transfection by lipofection 48
2.2.2.2. Transient transfection by electroporation 49
2.2.2.3. Titration of antibiotics (Killing curve) 49
2.2.2.4. Stable transfection of cells 49
2.2.2.5. Inducible cell clones.....50
2.2.3. shRNA mediated RNA interference.....51
2.3. Biological assays 52
2.3.1. Luciferase reporter gene assay 52
2.3.2. Proliferation analysis.....53
2.3.3. Analysis of apoptosis.....53
2.3.4. Cell cycle analysis 54
2.3.5. Wound healing/migration.....55
2.3.6. Analysis of senescence associated β-galactosidase (SA-β-gal) activity 55
2.4. Immunofluorescence 56
2.5. Infection with Helicobacter pylori 56
2.6. Sodium-dodecyl-sulfate- Polyacrylamide Gel Electorphoresis (SDS-PAGE).....57
2.7. Western Blot 58
2.8. RNA Analysis 58
2.8.1. RNA Isolation.....58
2.8.2. Reverse transcription 59
2.8.3. Quantitative real-time polymerase chain reaction (qRT-PCR).....60
2.8.4. RNA Microarray.....61
2.9. In vivo studies.....62
2.9.1. Genotyping transgenic mice 62
2.9.2. Preparation of blastocystes 64
2.9.3. Tamoxifen application.....65
2.9.4. Histology.....65
2.9.4.1. Organ isolation and embedding 65
2.9.4.2. Hematoxylin-Eosin (HE) staining.....66
2.9.4.3. Immunohistochemistry.....66
2.9.5. Analysis of AZ-521 tumorgenicity in vivo.....68
2.10. Materials 68
2.10.1. Buffers and Solutions 68

2.10.2. Chemicals and Kits 76
2.10.3. Antibodies 80
2.10.4. Consumables 81
2.10.5. Instruments 82
3. Results 85
3.1. Role of SOX2 in gastric cancer 85
3.1.1. SOX2 expression in gastric tumors and gastric cancer cell lines 85
3.1.1.1. SOX2 expression in gastric cancer tissue samples 85
3.1.1.2. Cell line screening for SOX2 expression in gastric cancer cells 86
3.1.2. The functional role of SOX2 in gastric cancer cells 88
3.1.2.1. SOX2 knock-down strategies 88
3.1.2.2. The role of SOX2 in cell proliferation 91
3.1.2.3. The role of SOX2 in apoptosis 94
3.1.2.4. The role of SOX2 in cell migration 97
3.1.2.5. The role of SOX2 in cell cycle regulation 98
3.1.2.6. The role of SOX2 in cellular senescence 102
3.1.2.7. The role of SOX2 in tumor growth and metastasis 104
3.1.3. Identification of SOX2 target genes 109
3.1.4. Regulation of SOX2 in gastric cancer 120
3.1.4.1. The regulation of SOX2 by STAT3 120
- The expression of SOX2 and STAT3 in gastric tumors of Gp130 mutant mice 120
- Co-expression and co-localization of SOX2 and STAT3 in GC cells 121
- Regulation of SOX2 by STAT3 in vitro 122
3.1.4.2. The influence of H. pylori infection on SOX2 expression in GC cells .. 124
3.2. The role of SOX2 in stomach development 127
3.2.1. Sox2 in embryonic development 128
3.2.2. The influence of Sox2 on postnatal development 133
3.2.3. Sox2 in the adult murine stomach 135
4. Discussion 137
4.1. Role of SOX2 on gastric carcinogenesis 137
4.2. The regulation of SOX2 in gastric cancer 145
4.3. The role of SOX2 during murine development 147
5. Summary 151
References 153

Appendix 165
1. List of differentially expressed genes (limmatable p 0.01) 165
2. List of DEGs assigned in Cluster 169
3. CoPub gene list..... 183
Acknowledgements.....193
Curriculum Vitae .....**Fehler! Textmarke nicht definiert.**

List of Abbreviations

°C	Degree Celsius
ad	Lat.: fill up until
Amp	Ampicillin
APS	Ammonium persulfate
ATCC	American type culture collection
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
CMV	Cytomegalovirus
DAPI	4’,6-Diamidinio-2-phenylindole
d	Destilled
dd	Double destilled
DMEM	Dulbecco’s Modified Eagles Medium
DMSO	Dimethyl sulfoxide
dn	Dominant negative
DNA	Deoxyribonucleic acid
Dox	Doxycycline
dNTP	Deoxyribonucleoside triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
EtOH	Ethanol
FACS	Fluorescence activated cell sorting
FITC	Fluorescein isothiocyanate
fl	“floxed”
g	Force of gravity
gr	Gram
GFP	Green fluorescent protein
h	Hour

HE	Hematoxylin-Eosine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
H. pylori	*Helicobacter pylori*
IL	Interleukin
JCRB	Japanese Collection of Research Biosources
kDa	Kilo dalton
L	Liter
LB	Lucia broth
Luc	Luciferase
M	Molar
mA	Milliampere
MEM	Minimum essential medium
mg	Milligram
µg	Microgram
min	Minute
ml	Milliliter
µl	Micorliter
mM	Millimolar
mRNA	Messenger RNA
MW	Molecular weight
ng	Nanogram
OD	Optical density
o.N.	Over night
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
qRT-PCR	Quantitative real-time polymerase chain reaction
RLU	Relative light units
RNA	Ribonucleic acid
RNAi	RNA interference
ROI	Region of interest

RPMI	Roswell Park Memorial Institute
RT	Room temperature
SDS	Sodium dodecyl sulfate
sec	Second
s.e.m.	Standard error of means
shRNA	Short hairpin RNA
TAE	Tris-acetate- EDTA
Tam	Tamoxifen
TBS	Tris-bufferd saline
TEMED	Tetramethyldiamine
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
V	Volt
v/v	Volume for volume
WHO	World health organization
w/w	Weight for weight

1. Introduction

1.1. The stomach

1.1.1. Development of the stomach

Early in embryonic development of vertebrates the mono-layered blastula evolves into the so called gastrula. During this process called gastrulation, totipotent cells of the epiblast are partitioned into the three germ layers, i.e. ectoderm, mesoderm and endoderm. The ectoderm gives rise to skin and central nervous system, the mesoderm forms blood, bone and muscle and the endoderm forms glands of thyroid, thymus, pancreas and liver as well as epithelium of lung, esophagus, stomach, intestines, and colon. The latter group is usually referred to as the gut [1].

During early development the gut is arranged into four axes, namely the anterior-posterior (AP) axis, the dorsal-ventral (DV) axis, the left-right (LF) axis and the radial (RAD) axis. Through further differentiation and regional specific morphological development, the AP axis later on forms three different regions: The foregut will develop into pharynx, esophagus and stomach, the midgut will give rise to the small intestine and the hindgut will form the colon [2].

Expression of several transcription factors as well as activity of different signaling pathways during fetal development is influencing the specification of endoderm. Among others *SOX* genes, *HOX* genes, *GATA5* and *FGF10* as well as the canonical WNT/β–Catenin, Hedgehog, Notch, and BMP pathway play a crucial role in foregut and stomach formation. Also, enduring epithelial-mesenchymal crosstalk is mandatory for several steps of stomach development, such as gland formation, regionalization and cell differentiation [3-5]. Disorders in these tightly regulated systems frequently lead to tumorigenesis in the adult stomach.

1.1.2. Anatomy and function of the stomach

The vertebrate gastrointestinal (GI) tract is a specialized vital organ derived from an ordinary tube-like structure and consists of GI tract derivatives, namely thyroid, lung, pancreas, liver and the gut (esophagus, stomach, intestine and colon). The main function

of the gut is to digest and absorb nutrients. After mastication the second phase of digestion proceeds in the stomach where food is enzymatically converted into chyme.

The **stomach** is located between esophagus and small intestine and is connected to both by cardia or pylorus, respectively. It can be divided into three parts, i.e. fundus, body (or corpus) and antrum (Fig.1).

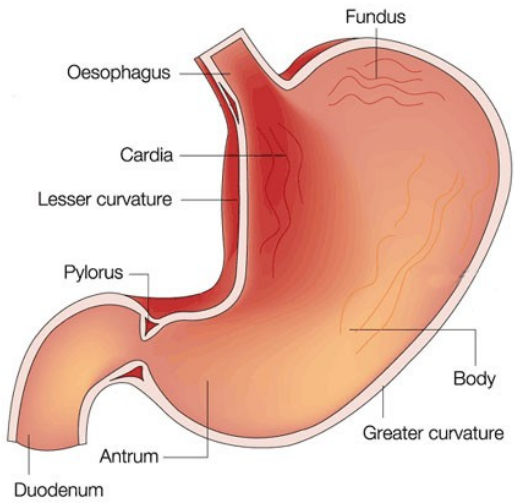


Fig. 1: Anatomy of the stomach (modified from Peek and Blaser [6])

The gastric wall consists of five different layers. The inner lining, namely the mucosa contains the gastric glands. Underneath the mucosa lies the submucosa, followed by a muscle layer and the subserosa. The outer layer is the so called serosa [7].

The gastric mucosa consists of multiple gastric units, comprised of gastric glands. Glands feed into gastric pits that open through a single layer of epithelial cells into the stomach lumen. Depending on anatomical regions of the stomach the gastric units vary in structure and composition [8]. The continuously renewing epithelium of the stomach corpus consists of four different main cell types: Surface mucus faveolar (pit) cells, parietal (oxyntic) cells, zymogenic (chief) cells, and hormone-secreting enteroendocrine cells (Fig.2). In comparison, glands of the stomach antrum have only few parietal and zymogenic cells but a separate population of mucus-secreting cells near the base of the glands [9].

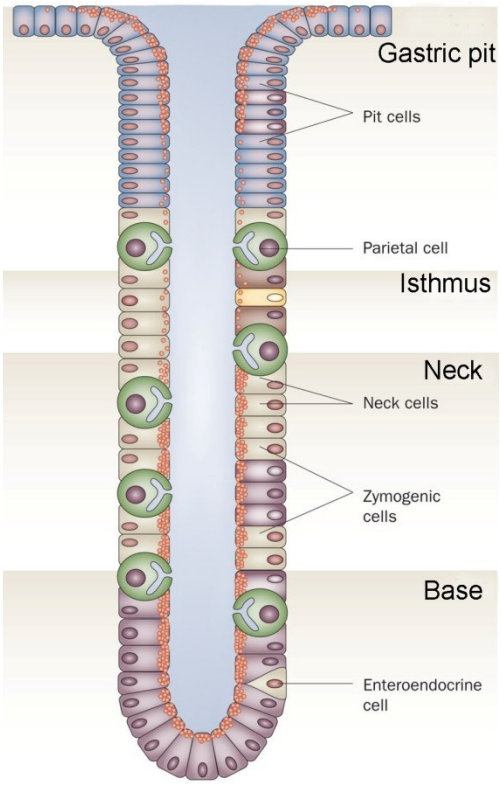


Fig.2: Scheme of a corpus stomach gland [10]. The yellow cell in the isthmus depicts a presumable gastric stem cell.

1.1.3. Gastric stem cells

Almost every tissue of an organism is renewed during lifetime. Differentiated cells are replaced by new mature cells which originate from tissue specific adult stem cells. In several different tissues, e.g. in the intestine, self renewal has been extensively studied. However, only little is known about adult stem cells in the stomach. There is consensus that most adult stomach glands are monoclonal, meaning each gland derives from its own stem cells [11, 12]. According to numerous independent studies using labeled nucleotide incorporation assays and ultrastructural investigations, immature presumable gastric adult stem cells, that maintain and regenerate gastric epithelium, might be located in the isthmus of gastric glands. Characteristics of those cells were high nucleus-to-cytoplasm ratio, lack of granules, open chromatin, undeveloped rough endoplasmatic reticulum, many free ribosomes, and few mitochondria [13-19]. The isthmus lies in the upper third of typical corpus glands and in the lower third of glands of the stomach antrum. Cellular migration takes place bidirectional from this area. Cells

differentiate into gastric surface mucus cells coating the gastric pits and gastric zymogenic and parietal cells at the base of the glands. Zymogenic cells are supposed to be progeny of glandular mucous neck cells. Stem cells are also self-renewing and give rise to enteroendocrine cells (Fig.3) [9, 20, 21].

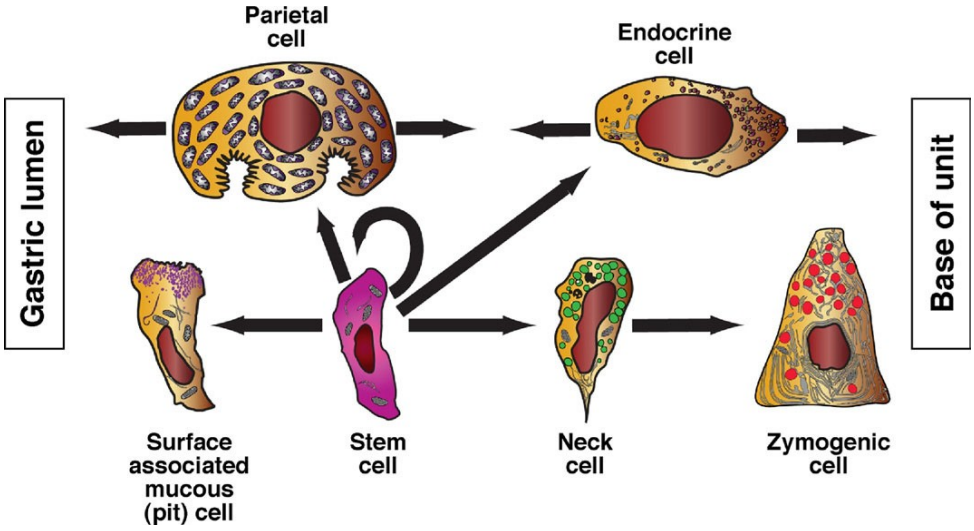


Fig.3: Stem cell theory in corpus epithelial glands [10]. The presumable stem cell is self-renewing and gives rise to all other epithelial lineages.

In contrast to this dogma, studies by Barker *et al.* more recently postulated presumable gastric stem cells being located much lower in the glands, i.e. in the glandular base of stomach antrum and pylorus. Using lineage tracing experiments in mice, they found a *Lgr5* expressing cell population that could give rise to all antral and pyloric unit cells and expressed several proliferation markers. In stomach corpus epithelium, however, *Lgr5+* cells are only detected in neonatal mice and disappear soon after birth (Fig.4) [22]. Therefore, the gastric stem cell maintaining the adult gastric units still need to be identified and specific markers defined.

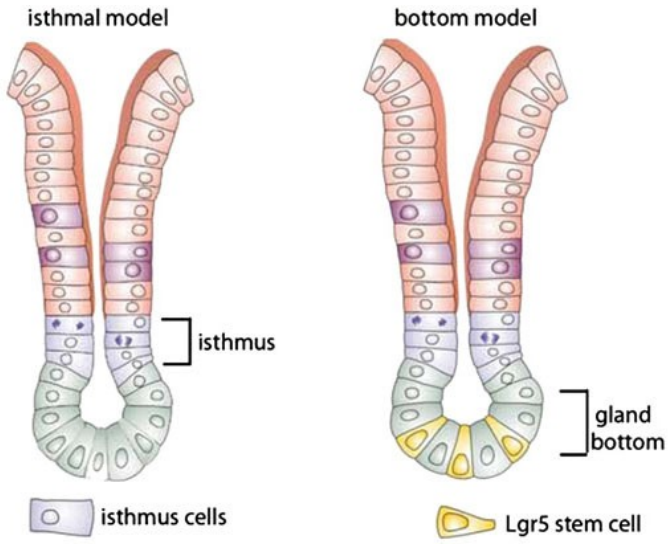


Fig.4: The gastric stem cell; two schools of thought (modified from Vries *et al.* [23]) Proposed localization of adult stem cells in gastric glands.

Furthermore, contrary to current assumptions that adult stem cells remain in a quiescent state, Barker *et al.* reported that *Lgr5+* cells are rapidly dividing under homeostatic conditions. Their findings were based on their earlier lineage tracing studies, in which they identified *Lgr5* to be a putative intestinal stem cell marker [24]. Lineage tracing is, since several years, an excellent tool to test for actual "stemness" of cells. The approach is based on genetic labeling of stem cells. Stem cells are assumed to pass their genetic background to all progeny, thus labeling all daughter cell population which then can be traced *in vivo*. Barker *et al.* used mice harboring a *Lgr5-GFP-IRES-CreER*^{T2} in which *CreER*^{T2} is expressed at all *Lgr5* sites [24]. The CreER^{T2} is an inducible version of Cre-recombinase [25]. The Cre enzyme is inactive in the cytoplasm, being fused to a mutated human estrogen receptor (ER) binding-domain. Upon induction with tamoxifen, a synthetic estrogen antagonist, Cre recombinase can translocate into the nucleus and excise DNA flanked by loxP recognition sites [26]. *Lgr5-CreER*^{T2} mice where crossed into *Rosa26-LacZ Cre*-reporter mice [27]. A loxP flanked stop codon in front of the *LacZ* gene of the *Rosa26-LacZ Cre*-reporter inhibits LacZ expression until tamoxifen is applied to the system, which then subsequently results in irreversible genetic labeling of Cre-expressing *Lgr5+* cells and all their progeny by excision of the stop cassette (Fig.5). The labeling can be visualized by an enzymatic blue staining of the tissue [24, 28]. Observation of blue cell clones over time thus gives insight into growth kinetics, multipotency and longevity of the original marked cell [29].

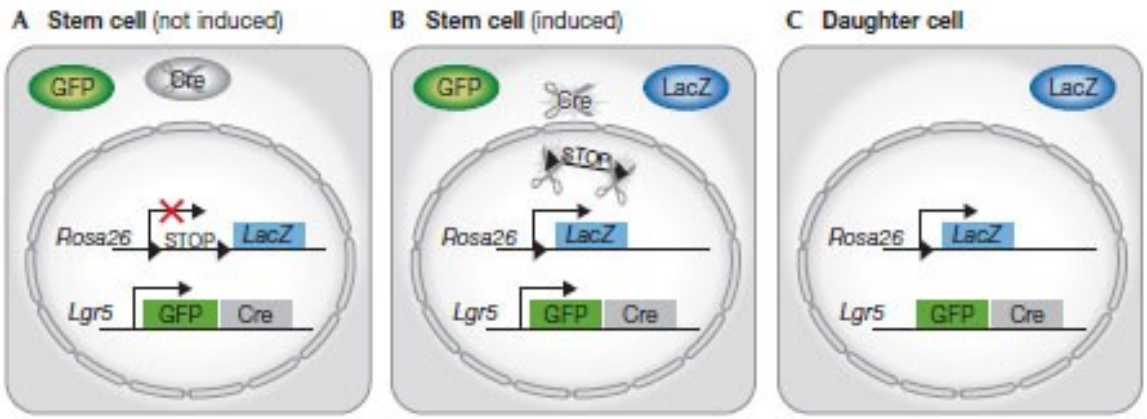


Fig.5: In vivo lineage tracing approach (modified from Snippert et al. [29]) **A)** In not induced stem cells only GFP and an inactive Cre recombinase are expressed **B)** After induction with tamoxifen Cre recombinase can enter the nucleus and excise loxP sites, thus *LacZ* can be transcribed **C)** Daughter cells do no longer express GFP and Cre, however, LacZ expression remains

A further lineage tracing approach in the murine stomach was made by Qiao *et al.* by marking cells expressing the *villin* promoter. Villin is intestine specific and usually not expressed in the stomach. Nevertheless, single *villin* promoter positive cells in varying positions between isthmus and base of antral glands were identified, that gave rise to all cells within an antral gland after stimulation with interferon gamma (IFNy). As these cells seem to replicate only after cytokine stimulation, they are not likely to be real stem cells, but rather a stem-cell like population regulating epithelium renewal, e.g. after injury [21].

In the glands of stomach corpus, Trefoil factor family 2 (*Tff2*) positive cells were found to give rise to a limited spectrum of lineages. TFF2 is a regulator of mucosal repair in the stomach and is also used as an indicator of metaplastic changes [30]. *Tff2* mRNA is expressed in cells in the isthmus of corpus glands. Lineage tracing experiments in mice revealed that parietal cells, as well as mucus neck cells and zymogenic cells are progeny of *Tff2*+ cells. Yet, mucous pit cells and enteroendocrine cells were never labeled and *Tff2*+ cells and their progeny never survived more than 200 days. These findings emphasize that *Tff2*+ cells are not true gastric stem cells [31].

MIST-1 (chief cells of corpus gland base) [32], *CD44* (corpus and antrum gland base) [33], *DCAMKL1* (corpus gland isthmus) [34] and *PROM1* (Prominin-1) (antral gland base) [35] are other candidates for presumable markers for gastric stem cells, but have only marginally been investigated so far and roles need to be clarified.

The transcription factor SOX2 is, in concert with OCT4 and NANOG, an essential factor for pluripotency in ES cells (see **Introduction 3.3**). While analyses of OCT4 and NANOG in reporter animals and knock out studies failed to confirm their contribution to pluripotency in adult stem cells, evidences have accumulated that SOX2 plays an important role in these cells [36]. Lately SOX2 has been reported in various adult progenitors (see **Introduction 3.3**). Therefore, we suspected a role for SOX2 in gastric stem cells. While this thesis work was in progress *Arnold et al.* published their work on adult stem cells. Here *Sox2* was investigated via lineage tracing experiments in several different mouse tissues. In glands of stomach corpus as well as pylorus, *Sox2*+ cells gave rise to all mature glandular cells. *Sox2*+ cells were mostly located in the glandular base, but co-staining experiments revealed that they were rare and not identical with *Lgr5+* cells. *Villin*-expressing cells in antral glands, however, might coincide with *Sox2*+ cells, but further studies need to be done [37]. These findings suggest that there might be several different types of adult stem cells in the stomach, which are, among other factors, dependent on the stomach region. This idea resembles the recently postulated hypothesis for intestinal stem cell markers which also seem to label different cell populations [24, 38-40].

Furthermore, in contrast to intestinal stem cells, gastric stem cells have an essential challenge to overcome: there is a substantial difference in turnover rates of epithelial stomach cells, as surface-associated mucous cells have a life span of 3 – 5 days, whereas the life span of zymogenic cells is several months [13]. Thus, gastric stem cells must generate many more precursor cells of pit cells than of chief cells in every differentiation cycle. How stomach stem cells accomplish to drive their progeny toward the different lineages is not clear yet.

Taken all this together, it is obvious that the characteristics, the markers and the niche of gastric stem cells have not been entirely identifed to date.

1.2. Gastric cancer

1.2.1. Epidemiology of gastric cancer

Gastric cancer (GC) is the fourth most common cancer and the second most common cause of cancer-related death worldwide, as its average 5-year-survival rate is only 25 % [41]. The risk of developing GC is 6 to 10 times higher in undeveloped countries like Latin America and parts of Eastern Europe compared to Northern America and Western Europe [41] and it is more prevalent in African-Americans, Hispanics and Native Americans than in Caucasians [42]. Occurrence of GC shows also a distinct gender specific distribution, as males develop GC about twice as often as females [43]. Highest risk comes with age. 80 % of patients are between 50 and 70 years old [44]. GCs have also strongly been linked to several environmental factors, among others up-take of high-salted food and tobacco smoking, which might nearly double the risk compared to non-smoking population [45, 46]. *Helicobacter pylori* (*H. pylori*) is one of the most important high risk factors for GC [47]. Although early preventions and eradications of *H. pylori* have increased during the last years and therefore incidence of GC has declined [43], every year approximately 500 000 people worldwide newly developed GC [48]. Diagnosis is often made very late, as symptoms (e.g. abdominal pain, blood in stool, weight loss) occur mainly after the tumor has reached an advanced stage and infiltrated the muscularis propria. This results in poor prognosis and low survival rates [49].

1.2.2. Classification of gastric cancer

Different types of gastric tumors can occur in the stomach, e.g. gastrointestinal stromal tumors (GIST), mucus-associated-lymphoid-tissue- (MALT-) lymphomas or squamous epithelial lymphomas, sarcomas, that arise from cells of the muscle layer and carcinoid tumors, arising from neuroendocrine cells [50]. However, the most prevalent gastric tumors are adenocarcinomas (95 % of all gastric tumors) [51] which derive from glandular epithelium of gastric mucosa and are highly malignant.

Many approaches have been made to classify gastric adenocarcinomas, such as the Ming [52] or the World Health Organization (WHO) classification. Another widely used characterization is the DIO system firstly proposed by Laurén in 1965 based on

tumor histology. According to this system gastric adenocarcinomas are classified as diffuse (D), intestinal (I) or other (O). Intestinal-type tumors are highly differentiated and cohesive neoplastic cells show gland-like tubular structures. In contrast, the diffuse-type GC is less differentiated and cells are not cohesive, infiltrating and thickening the stomach wall. Furthermore, in this subtype signet ring-cells are found in the gastric mucosa. The diffuse-type tumor has higher metastatic potential and thus shows poorer prognosis than the intestinal-type tumor. About 10 – 15 % of gastric adenocarcinomas show characteristics of both subtypes and therefore are declared as mixed-type tumors [53, 54].

1.2.3. Development of gastric cancer

Several environmental factors and susceptible genetic variants as well as various genetic and epigenetic alterations in cell cycle regulators, oncogenes, tumor suppressor genes, growth factor systems, DNA repair genes, cell adhesion molecules and telomerase activations are supposed to be involved in the multiple-step development of GC. However, intestinal- and diffuse-type GC seem to differ substantially in decisive combinations of these alterations, suggesting two different carcinogenesis pathways for the two subtypes [55-57].

- *Intestinal-type gastric cancer*

Gastric tumors are not a result of one specific event but develop out of multiple combinations of factors. In 1975 Correa and coworkers for the first time postulated a model which proposed sequential steps leading to changes in histomorphology of the gastric mucosa and eventually to GC. In their model, the so called Correa-pathway, pre-cancerous lesions (gastritis/atrophy) are followed by intestinal metaplasia, dysplasia and finally gastric carcinogenesis (Fig.6) [58]. However, there are two further possible routes leading to intestinal-type gastric carcinomas postulated. Intestinal metaplasia may resume directly in GC and GC may develop *de novo* with no foregoing steps needed [57].

Fig.6: Histology of the Correa-pathway **A)** Normal gastric mucosa **B)** Gastritis **C)** Complete intestinal metaplasia **D)** Dysplasia **E)** Intestinal-type adenocarcinomas (kindly provided by R. Mejias-Luque)

The first step of the Correa-pathway is usually **chronic active non-atrophic gastritis** which is typically triggered by *H. pylori* infection and characterized by infiltration of the mucosa by white blood cells (lymphocytes, plasma cells, macrophages). Additionally, eosinophiles, mast cells and polymorphoculear neutrophils can be detected.

The second stage, **atrophic gastritis**, is typified by changes in the cell cycle which lead to proliferation and apoptosis and results in atrophy (loss of glands) and replacement of parietal cells by mucous secreting cells [59].

This step can be followed by **intestinal metaplasia** (IM). IM shows cells characteristic for the intestine, namely absorptive enterocytes with a brush border, goblet cells containing mucins and paneth cells harboring eosinophilic granules in their cytoplasm. They are typically found at the base of the glands [60]. Accordingly, genes specific for the intestine have been identified in intestinal metaplasia, e.g. *MUC2* and *MUC4* in goblet cells [61], *Villin* in enterocytes or *DEF* (defensines) in paneth cells [62]. Several genes, functions and regulators are known to be deregulated in intestinal metaplasia. Among others these are *CCND2* (Cyclin D2), *COX2*, telomerases, microsatellite instability, reduced expression of *CDKN1B* (p27), *APC* (Adenomatous polyposis coli) loss, mutation/LOH in *TP53* [50, 63]. Furthermore, numerous researchers stated a correlation of deregulation of specific transcription factors and intestinal metaplasia. The intestine specific transcription factor Caudal type homeobox 2 (CDX2) for example is a master regulator in mammalian gut and important for intestinal development and differentiation [64]. Loss of *CDX2* leads to gastric differentiation in the colon [65]. Accordingly, aberrant expression of *CDX2* in the stomach results in intestinalization and is a key event for intestinal metaplasia [66].

Dysplasia is the next step in the Correa-pathway cascade. As it results in pre-invasive neoplastic changes in the gastric glands it is also called intraepithelial neoplasia.

This lesion is characterized by atypical changes in cells and tissue architecture. Nuclei are often enlarged and hyperchromatic, due to a higher proliferation rate in an increased number of cells. Glands are distributed abnormally with irregular lumens. Dysplasia is classified as low- or high-grade depending on the degree of atypia in nuclei and gland deformation [59]. Molecular characteristics of dysplasia are similar to those detected in GC. Typically this is loss of heterozygosity in the *APC* gene, as well as aberrant expression of *TP53* (p53), *CDKN1A* (p21) and *BCL2* [63].

In intestinal-type **adenocarcinomas** numerous genetic and epigenetic alterations can be observed. As mentioned above these are LOH or missense mutations in the *APC* gene as well as reduced expression in tumor suppressor genes, namely *TP53*, *TP73* (p73), *CDKN1B* (p27) and *TFF1* [50]. *TP53* and *TP73* are usually inactivated by LOH whereas loss of *TFF1* expression is believed to arise due to DNA methylation in its promoter region [57]. In addition, TFF1 has been described to be down-regulated by STAT3 in a mouse model bearing a mutation in the GP130 receptor. These mice develop tumors after 4 weeks of age, due to hyperactivated STAT3 signaling, indicating a key role of *STAT3* in gastric oncogenesis [67]. Loss of *APC* expression enhances the expression of *CTNNB1* (β-Catenin), which acts as an oncogene. Other alterations known to be involved in development of intestinal type adenocarcinomas are loss of *RUNX3* expression, LOH on chromosome 7q, amplification of *ERBB2* and *CCNE1* (cyclin E), *CD44* and *CD46* aberrant transcripts, epigenetic alterations in *MLH1*, a mismatch repair gene, *DS191* instability and over expression of several growth factors (*EGF*, *TGFa*, *IGFII*, *bFGF*) (Fig.7) [50].

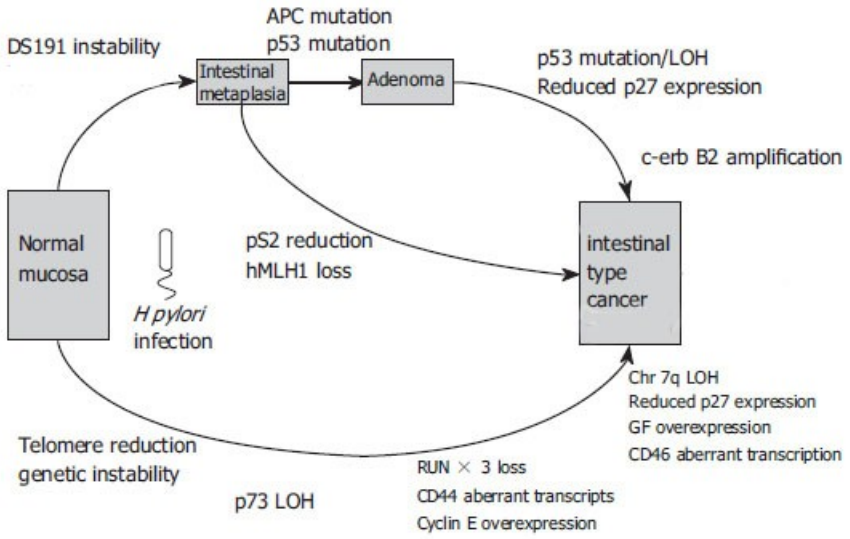


Fig.7: Genetic and epigenetic alterations in intestinal-type GC modified from Smith et al. [50]

- *Diffuse-type gastric cancer*

Although particularly intestinal-type GC has been associated with *H. pylori* infection, there have been some recent studies, showing also a link between diffuse-type GC and *H. pylori [68, 69]*. However, in pathogenesis of diffuse-type GC preceding steps like they are known in intestinal-type cancer are not well understood yet. It is assumed that diffuse-type GC rather arises from single-cell changes in the mucus-neck region of gastric glands. These cells then may proliferate and invade from the crypt in the lamina propria.

Several molecular mechanisms are known to be involved in the generation of diffuse-type GC, including alterations in tumor suppressor genes, activation of proto-oncogenes, aberrant expression of cell cycle regulators and growth factors.

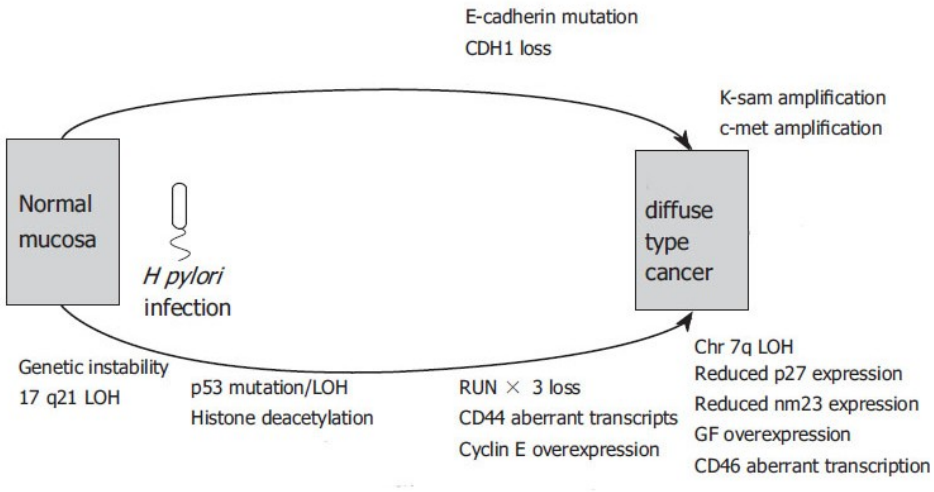


Fig.8: Genetic and epigenetic alterations in diffuse-type GC modified from Smith et al. [50]

A well studied example of alterations in tumor suppressor genes is the autosomal dominantly inherited germ line mutation of *CDH1* (E-cadherin) in patients with hereditary diffuse gastric cancer (HDGC). Carriers of the germ line mutation have a more than 70 % life time risk of developing GC [70]. Furthermore, it is also strongly associated with the risk of women to develop breast cancer [71]. Interestingly, also in 40 - 83 % of sporadically occurring diffuse-type adenocarcinomas of the stomach a somatic mutation of *CDH1* gene has been observed, but was never identified in intestinal-type GC [72]. Additionally, E-cadherin loss can also be found in mixed-type GC, but only in the diffuse section, suggesting it to be a possible genetic variable driving gastric cancer towards diffuse or intestinal differentiation [73]. Another tumor suppressor gene mutated in GC is *TP53*. It is prevalently inactivated by loss of heterozygosity (LOH), missense mutations and framshift deletions, frequently GC-AT transitions in diffuse-type GC [74]. Other tumor suppressor genes that appear to be affected in diffuse-type GC are loss of *RUNX3* and histone H4 deacetylation [50].

Furthermore, several proto-oncogenes are activated in diffuse-type GC. In 39 % of cases this is *MET* (c-met), a tyrosine kinase receptor gene encoding for hepatocyte growth factor receptor. 33 % of diffuse-type GC show over expression of type II *FGFR3* oncogene, a receptor for keratinocyte growth factor.

Other known important factors influencing the development of diffuse GC are *CD44* and *CD46* aberrant transcripts, *CCNE1* (Cycline E) amplification, reduced expression of the CDK inhibitor *TP27*, reduced expression of *NME1* (nm23), LOH on 17q21, including *BRCA1* gene, LOH on 7q, and over expression of growth factors (*TGFb*, *bFGF*, *IGF II*) (Fig.8) [50].

1.2.4. Helicobacter pylori and gastric cancer

H. pylori is a gram negative, motile, microaerophilic bacterium. It can colonize the human stomach and is responsible for development of atrophic gastritis and peptic ulcer which can later on lead to GC [47]. Therefore, in 1994 the WHO defined *H. pylori* as a type I carcinogen. Prevalence of *H. pylori* infection differs highly among populations [75] and increases with age [76]. Severity of gastric damage and clinical outcome are supposed to be dependent on bacterial virulence factors as well as genetic background of the host.

- Bacterial virulence factors

The best studied virulence factor in *H. pylori* is the cag pathogenicity island (PAI) Type IV secretion system. **CagA**, the product of the terminal gene in the island, is translocated into epithelial cells by this system [77], where it is phosphorylated at certain glutamate-isoleucine-thyrosine-alanine (EPIYA) motifs by Src-kinases. CagA can then activate the Ras-mitogen activated protein kinase (MAPK) pathway and the EGF receptor and thus leads to changes in intracellular signaling and morphological changes in the epithelium. CagA can also bind to numerous host proteins, like SHP-2 and MET thereby inducing subsequent aberrant signaling which leads to apoptosis and increase cell turnover of epithelial cells [78, 79]. Another known pathway influenced by CagA is the E-cadherin/β-catenin pathway, which regulates junction formation, cell adhesion and cell growth [80]. Interaction of CagA with this system can also lead to direct transactivation of CDX1 and thus to mucosal metaplastic changes [81]. Furthermore, CagA can activate NF-κB leading to IL-8 secretion [82].

Another major virulence factor in *H. pylori* is the secreted pore-forming protein vacuolating cytotoxin A (**VacA**). *VacA* gene structure can be divided into a signaling (s), a middle (m) and a recently identified intermediate (i) region. The majority of *H. pylori*

strains express *VacA*, but variations in the subtype result in differences of cytotoxicity. *VacA s1* and *VacA m1* were assumed to be most cytotoxic and thus highly associated with GC, however, recently also *VacA i1* was identified to contribute to malignant invasive tissue formation [83, 84]. VacA induces vacuole formation in epithelial cells, leading to altered antigen presentation [85]. Moreover, disruption of the epithelial barrier as well as pore-formation in mitochondrial membranes which may lead to the activation of apoptosis has been reported [86, 87]. Besides, it interferes with GSK3β pathway which leads to β-Catenin release and modulation of apoptosis and cell cycle regulation [88, 89].

The outer-membrane protein **BabA** (Blood group antigen binding) binds to the Lewis *b* antigen on the surface of gastric epithelial cells and facilitates entering of the host epithelium by the bacteria [90]. *BabA* expressing strains of *H. pylori* promote a more pathogenic phenotype and presence of BabA is significantly associated with higher incidence of gastric adenocarcinomas. *H. pylori* strains expressing *BabA* concomitant with *VacA* and *CagA* are associated with highest risk for developing GC [91].

- Genetic background of the hos*t*

Not only bacterial virulence factors but also host genetic factors are profoundly influencing the outcome of an *H. pylori* infection and progression to GC. It is known that *H. pylori* triggers chronic inflammation in the gastric mucosa through several pro- and anti-inflammatory cytokines, though alterations in genes involved in immune response have been reported to cause an increased risk for the host. IL-1β, a strong inhibitor of acid secretion, is one of the most important pro-inflammatory cytokines produced in response to *H. pylori* infection [92]. Specific single nucleotide gene polymorphisms (SNP) in the *IL1B* locus are associated with a higher risk for developing GC. For example, higher rates of atrophy and gastritis are found in patients with the genotype *IL1B 511 T/T* correlating with an increase in pH of gastric juices [93, 94].

TNF-α, a pro-inflammatory cytokine mainly produced by activated macrophages in immune response is also secreted in response to *H. pylori* infection. It is an endogenous pyrogen and induces inflammation. Like IL-1β, it also has an acid inhibitory effect. Polymorphisms in *TNFa* are reported to correlate with many inflammatory conditions and thus promote the risk of GC [95].

IL-8 is a further important cytokine associated with *H. pylori* pathogenesis. Its primary function is to recruit neutrophiles at the site of inflammation during innate immune response. Promoter polymorphisms at the *-251T* allele influence neutrophil infiltration in the gastric mucosa and increase the risk of atrophic gastritis and GC [96].

IL-10 is an anti-inflammatory cytokine that down-regulates pro-inflammatory cytokines, e.g. IL-1β and TNF-α. Polymorphisms that result in low production of IL-10 elevate the risk of developing GC after *H. pylori* infection.

Furthermore, changes in pattern recognition factors, like *TLR4*, *NOD1*, and *NOD2* as well as aberrant gene expression of proteases, mucins, HLA molecules, xenobiotic metabolism enzymes, cell cycle regulators, and DNA repair enzymes play a major role in *H. pylori* induced disease [81].

- Inflammatory response induced by *H. pylori*

Inflammatory response of *H. pylori* includes the release of pro-inflammatory cytokines such as IL-1β and IL-6. To analyze the etiological role of IL-1β in gastric carcinogenesis a transgenic mouse model over expressing *Il1b* in the stomach was established. *Il1b* transgenic mice developed spontaneous inflammation, metaplasia, dysplasia and carcinoma of the stomach, demonstrating that increased levels of IL-1β can be sufficient to induce neoplasia [97].

Increased levels of IL-6 have been described in the gastric mucosa of *H. pylori* infected subjects [98, 99]. IL-6 signaling is mainly mediated by STAT3. A significant increase in STAT3 activation has been detected in *H. pylori* infected patients as well as in adenocarcinomas [100] and it has been proposed as a prognostic factor for poor survival of GC patients [101]. STAT3 activation involves binding of the cytokine to a membrane bound IL-6 receptor, triggering heterodimerization with GP130 and activation of GP130-associated Janus kinases (JAK). Latent STAT3 monomers are then recruited to the phosphorylated residues of the GP130 receptor and phosphorylated by JAK1 and JAK2, resulting in their dimerization and translocation to the nucleus where the STAT3 DNA-binding domain interacts with STAT binding sites in target gene promoters [102].

1.2.5. The Cancer stem cell theory

The cancer stem cell (CSC) theory is based on the hypothesis that a subpopulation of cells within a tumor is the origin of the tumor, responsible for persistence and spreading. CSCs therefore have characteristics of adult or embryonic stem cells, namely self-renewal, a proliferative ability to drive continuous growth and the generation of differentiated cells [103]. They are supposed to be responsible for tumor initiation, invasion, metastasis and chemoresistance [104]. CSCs may arise from tissue specific adult stem cells or originate through degeneration of somatic progenitors within the tissue (Fig. 9) [105].

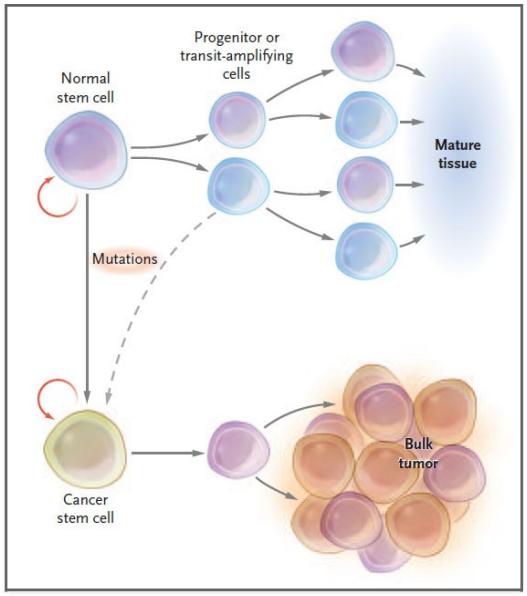


Fig.9: Cancer stem cell theory (modified from Jordan [105]) Normal adult tissue arises from adult stem cells that differentiate in progenitor cells which give rise to mature cells. Throughout mutations in normal stem cells or progenitor cells cancer stem cells can develop, self-renew, grow and differentiate into primary tumors.

CSCs have been characterized in many tumors including breast [106], ovary [107], brain [108], pancreas [109], head and neck [110], colon [111], and cancer of the blood [112, 113]. However, only little is known about cancer stem cells in gastric adenocarcinomas. In 2009 Takaishi *at al.* identified a *CD44*+ self-renewing cell fraction in GCs which possessed the ability to give rise to differentiated daughter cells and was chemo- and radioresistant. Nevertheless, *CD44* is not specific for GC stem cells and further markers need to be identified [33]. Recently Rocco *at al.* disproved *CD44*+ cells to have cancer stem cell properties in primary human gastric tumors [114]. A further

suggestion of the origin of GC stem cells was given by Barker *et al.* They postulated that deletion of *APC* in putative *LGR5*+ gastric cancer stem cells in the stomach leads to rapid development of gastric adenomas, thus implicating that *LGR5*+ gastric stem cells might be the origin of *APC* driven adenomas in GC. However, this is only proven for cells at the base of the pyloric glands in the distal stomach [22].

1.3. The transcription factor SOX2

*1.3.1. General characteristics of SOX (**S**ry b**ox**) proteins*

In 1990 SRY (Sex determing region Y) was the first member of SOX proteins identified [115]. SOX proteins build a group of transcription factors which are all largely conserved in their high mobility group (HMG) domain, which is their DNA binding domain of approximately 79 amino acids. The conserved amino acid motif in the HMG box is RPMNAFMVW. An exception is the first identified SOX protein SRY, which shares only a part (RPMNAF) of the conserved motif [116]. Furthermore, all SOX proteins bind to variants of the DNA sequence 5'- (A/T)(A/T)CAA(A/T)G -3' [117]. The SOX superfamily can be further divided into two groups: The SOX/TCF/MATA family binds specific DNA motifs, whereas members of the HMG/UBF group bind less specific [118]. Members of the SOX/TCF/MATA group usually bind to the minor groove of the DNA double helix. Binding causes a strong DNA bending which might bring distal proteins on gene promoters and enhancer closer together, facilitating functional interaction [119]. SOX proteins are restricted to animal kingdom and can be subdivided according to phylogenetic analysis of their HMG box domain. 10 subgroups (A-J) are identified so far, of which 8 groups (A-H) are present in mice and humans. All SOX proteins share around 50 % homology within their HMG box with SRY, however, group internal amino acid homology can be up to 90 % [116, 120]. Even though classification is solely based on sequence comparison, SOX proteins in the same group have similar biochemical properties and biological functions [121]. SOX transcription factors are expressed in embryos as well as in adults and are known to be involved in many different

developmental and proliferative processes in different tissues. They are also implicated in the etiology of different diseases and certain cancers [116, 122].

1.3.2. Classification and localization of SOX2

The transcription factor SOX2 belongs to the group B of SOX proteins. This group can be further divided into two subgroups, namely group B1 and group B2. SOX2 belongs, among SOX1 and SOX3, to the subgroup B1. Group B2 consists of SOX14 and SOX21 and their protein interaction domain acts as a transcription repressor. In contrast, the protein interaction domain of SOX1, SOX2 and SOX3 acts as an activator of transcription [123].

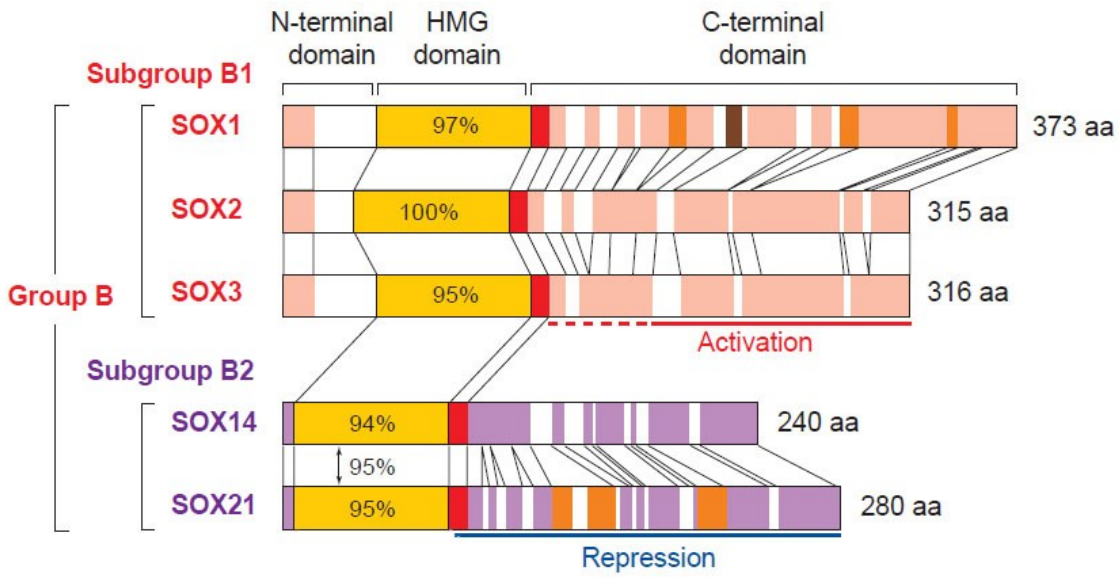


Fig.10: Homology of chicken SOX proteins in group B (modified from Kamachi [124]). The scheme shows similarity of HMG domain amino acids in chicken, relative to SOX2. Color code: Red - group B homology, light red - group B1 homology, purple - group B2 homology, orange - poly alanine sites, brown - PRD repeats

Expression of members of the *SOX* B1 family is prevalent in early embryo, developing testis, and nervous system and plays a crucial role in cell fate determination and cell differentiation during mouse development [125-127]. *SOX2* is the best studied and probably also the most important gene in the *SOX* B1 family. It is located at the short arm of chromosome 3 at position 3q26.3 – q27. Human *SOX2* (*hSOX2*) gene consists of a coding nucleotide sequence of 2511 bp length, is intronless and highly conserved in vertebrates. The *SOX2* gene encodes a 317 aa long SOX2 protein with a molecular mass of about 34 kDa. The chicken homologue *c*Sox2 has 95 % total amino-

acid identity to *h*SOX2. Mouse Sox2 (mSox2) shows even 97 % of similarity to hSOX2 amino acid sequence [128].

1.3.3. SOX2 in pluripotency and development

Embryonic stem (ES) cells emerge from the inner cell mass (ICM) of blastocysts and proliferate while sustaining pluripotency, thus contributing to all three germ layers and the respective tissues of an embryo [129]. Because of their origin they are not only interesting in regard to cellular processes of development but also for therapeutical approaches in respect to malignancies. SOX2 is, among OCT4 (POU5F1) and NANOG, one of the main regulators of pluripotency in ES cells [130-132]. The current model suggests that, on the one hand, SOX2, OCT4 and NANOG bind to promoter regions of genes associated with pluripotency and on the other hand to promoters of genes involved in differentiation. Thereby they operate as activators of transcription of pluripotency genes and repressors of genes for differentiation (Fig. 11) [133].

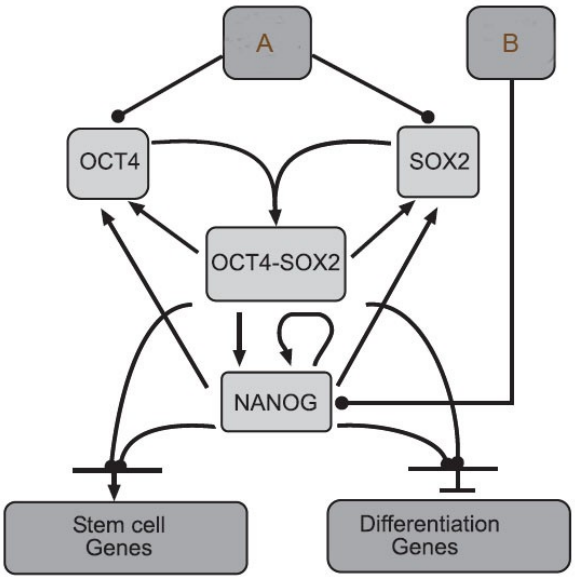


Fig.11: Model of maintenance of pluripotency in ES cells (modified from Chickarmane [134]) OCT4, SOX2 and NANOG build a core unit of a transcriptional network for maintaining pluripotency. They bind to the regulatory region of a number of genes, therefore operating as activators or repressors. They can target genes individually as well as jointly. A and B represent extrinsic and intrinsic factors influencing the system.

But not only in ES cells is SOX2 a marker for pluripotency. A study by Takahashi and coworkers in 2006 for the first time described that ectopic co-expression of four

transcription factors is enough to induce mouse fibroblast to dedifferentiate into pluripotent stem cells, so called induced pluripotent stem cells (iPS). Those four factors were SOX2, OCT4, KLF4 and C-MYC [135]. In the mean time several somatic cells in different species have been reported to dedifferentiate into iPS with similar reprogramming strategies all including SOX2 [136-138]. These studies again elucidate the exceptional role of SOX2 in pluripotency.

Yet, it is known that SOX2 does not influence pluripotency and expression of target genes on its own, but needs binding partners to activate specific transcription of genes. Cooperative binding of SOX-partners depends significantly on the stage of development. In implantation phase of embryos SOX2 binds to OCT4 and activates e.g. FGF4, NANOG and UTF1. Interestingly the SOX2/OCT4 protein complex is also self-activating, thus inducing an auto-regulatory loop [131]. However, in later phases of embryogenesis when OCT4 is no longer expressed, SOX2 can build hetero-dimeric complexes with different partners [139]. One example is the SOX2-PAX6 complex which activates the *CRYBB2* (δ-crystalline) gene during retina development [140]. Moreover, it has been shown lately that the regulation of Sox2 by the transcription factor Stat3 is necessary in early stages of mouse neural development for differentiation of ES cells into neuronal precursor cells (NPC), suggesting a novel signaling pathway [141]. Hence the combination of SOX2 and partner factors provides a code for activation of target genes and specification of cells.

In early phase of development SOX2 expression is limited to cells with stem cell characteristics like germ line cells or trophoblast stem cells. Accordingly, SOX2 is down-regulated in all cells with limited developmental potential [127, 142]. SOX2 is essential during embryogenesis. Homozygous *Sox2* knock out mouse embryos are not able to build the epiblast or the extraembryonic ectoderm and die shortly after implantation. A specific knock-down of *Sox2* in ES cells *in vitro* via dominant negative SOX2 or siRNA results in loss of pluripotency and induces differentiation towards trophoectoderm [127, 143]. Heterozygous loss of *SOX2* provokes several defects in mice and human, one of which is abnormal anterior pituitary development and endocrine deficits [144], which may be linked to the malformation of sensory organs of the inner ear [145]. Mutation of *SOX2* causes also Anophthalmia-Esophageal-Genital- (AEG) Syndrome, which was first

named by Shah et al. in 1997. AEG is an association of anophthalmia/microphthalmia (bilateral absence or malformation of the eye), esophageal atresia with or without tracheo-esophageal fistula (trachea and esophagus fail to separate) and urogenital abnormalities (cryptorchidism, micropenis) [146, 147]. In 2007 Que and coworkers defined a dose depended role for SOX2 in patterning and differentiation of mouse foregut. They demonstrated that heterozygous loss of *Sox2* does not influence foregut development, however, further reductions of *Sox2* using *Sox2* hypomorphic alleles result in various abnormalities according to dose of expression. Complete knock-outs fail to separate trachea and esophagus and die postnatally [148].

Lately SOX2 has been shown to be expressed in some mammalian adult tissue such as progenitors of the brain [149], retina [150], progenitors of pituitary glands [151], trachea [152], tongue epithelium [153], dermal papilla of the hair follicle [154], epithelium of seminiferous and lens, squamous epithelia lining the esophagus, forestomach, anus, and cervix as well as glandular stomach [37]. In most of these tissues *SOX2*+ cells seem to have stem cell properties, thus implicating that SOX2 might be a common marker of pluripotency that activates general pathways essential for self-renewal and differentiation and consequently is crucial for tissue homeostasis.

1.3.4. SOX2 and cancer

The improper expression of *SOX* genes has been implicated in numerous severe clinical disorders including human tumors [122]. In prostate [155] and pancreatic cancer [156] *SOX2* was observed to be amplified and might be involved in later events of carcinogenesis, such as invasion and metastasis. SOX2 was also seen to be over expressed in 43 % of basal-like sporadic breast cancers and in tumors arising from *BRCA1* germ line mutation [157]. *In vitro* studies in breast cancer cells revealed that SOX2 promotes cell cycle progression and tumorigenesis by facilitating G1/S- transition [158]. Accordingly, in tumor-initiating cells (TIC) derived from glioblastoma, the most aggressive cerebral tumor, silencing of *SOX2* via miRNA resulted in stop of proliferation and loss of tumorgenicity [159]. In lung and esophageal squamous cell carcinomas, *SOX2* is highly amplified and acts as a lineage survival oncogene, promoting cell migration and anchorage-independend growth *in vitro* [160, 161]. In a very recent report by Xiang and

co-workers, SOX2 was seen to be over expressed in presumable lung cancer stem cells. Down-regulation of *SOX2* via RNA silencing resulted in inhibition of growth and migration and an increase of apoptosis of cells, suggesting a key role for SOX2 signaling pathway in lung cancer stem cells [162]. Furthermore, a correlation of poorer overall survival and SOX2 expression was found in oral tongue squamous cell carcinomas [163]. In Merkel cell carcinomas nuclear SOX2 expression correlated with tumor thickness. *In vivo* studies in a xenograft model showed that *SOX2* knock down resulted in decrease in mean tumor volume [164]. Furthermore, a recent study identified *SOX2* to contribute to melanoma invasion [165]. Interestingly, SOX2 was also lately detected to be expressed in colorectal cancer (CRC) and was associated with poorer outcome and reduced overall survival due to correlating lymph-node and distant metastases [166, 167]. The majority of CRCs typically develop via an activation of WNT/β–Catenin signaling pathway and accumulation of nuclear β–Catenin is associated with poorer prognosis. Contrary to these results SOX2 was shown to down-regulate β–Catenin/TCF transcriptional activity *in vitro [168]*. However, knock-down of *SOX2* in a colorectal cancer cell line decreased growth rate *in vitro* and in an *in vivo* xenograft model [169]. The results suggest that there is a definite influence of the regulatory feedback mechanism between SOX2 and the β–Catenin/TCF complex in CRC that remains unclear and needs to be clarified. SOX2 was also found to play a role in GC. A recent study showed that SOX2 was regulated by IL-4 through STAT6 signaling in gastric epithelial cells and that SOX2 expression was blocked in *H. pylori*-mediated intestinal metaplasia, triggering Correa's sequence [170]. *In vitro* studies revealed that *SOX2* was down-regulated in some GC cell lines and that exogenous expression led to inhibition of cell growth and apoptosis of cells, suggesting *SOX2* to be a tumor suppressor gene [171]. Tsukamoto *et al.* reported that SOX2 was gradually down-regulated in intestinal metaplasia and intestinal-type gastric adenocarcinomas, inversely correlated with expression of CDX1/CDX2. However, they found SOX2 to be expressed in diffuse- and mixed-type GC being there important for maintenance of the gastric phenotype [172, 173]. Additionally, Matsuoka *et al.* affirmed that SOX2 positive expression might be associated with the invasion of GC [174]. Immunohistochemical experiments conducted in *CEA/SV40* transgenic mice [175], which develop gastric tumors, clearly showed SOX2 expression in tumorigenic tissues (Fig.12).

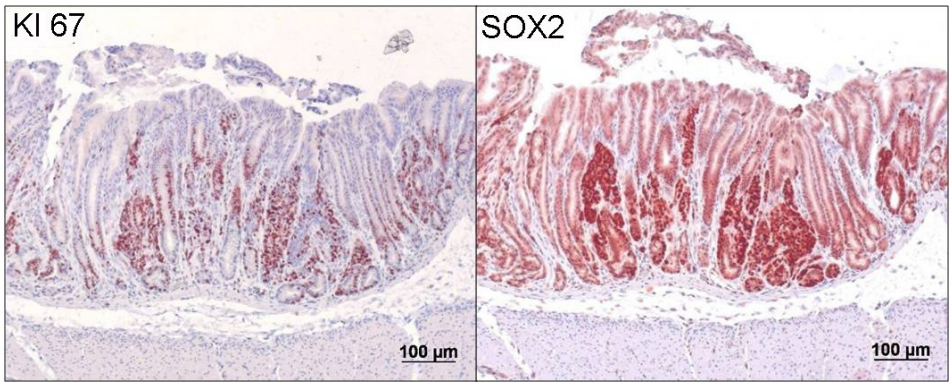


Fig.12: H&E staining of gastric tumors derived from CEA/SV40 transgenic mice. Expression of SOX2 correlates with the expression of the proliferation marker KI67.

In numerous GC cell lines *SOX2* seems to be up-regulated and considering the role for *SOX2* in maintenance of pluripotency in embryonic and adult stem cells and the current cancer stem cell theory it might not be surprising for *SOX2* to have a tumor initiating character in gastric adenocarcinomas.

1.4. Aims of the study

Gastric cancer (GC) arises throughout the influence of different environmental factors and/or the genetic predisposition of the patient. Developing a therapy for the treatment of malignant tumors however, can only be achieved when selective pathways or target genes leading to carcinogenesis are identified.

SOX2 is a very important factor in pluripotency of embryonic stem cells, influencing cell growth and differentiation. Moreover, it is also known to be highly over expressed in different kind of tumors such as breast cancer, glioblastoma and lung cancer, being responsible for their tumorigenicity. Furthermore, SOX2 was also found to be over expressed in some gastric tumors, indicating it might here be also tumorigenic. Thus, the first objective was to analyze the role of SOX2 in tumorigenesis of the stomach.

In order to evaluate the influence of SOX2 in aggressiveness and invasive characteristics of the tumor, GC cells were investigated for changes in proliferation, apoptosis, migration and *in vivo* metastatic behavior after inhibition of SOX2. Additionally, to depict the mechanism underlying possible growth inhibition, changes in cell cycle were also investigated.

To date pathways influencing development of GC are only poorly understood. Furthermore, not many target genes of SOX2 are known up to now. A second objective was to identify genes regulated by SOX2 and related to GC. Therefore, gene expression in GC cells was analyzed at certain time points after SOX2 inhibition using an mRNA microarray approach. Identifying differentially expressed genes might unravel target genes and pathways influenced by SOX2 in gastric carcinogenesis.

To date there are not much data available concerning regulation of SOX2 in gastric tumors. Thus, another objective was to analyze the influence of the transcription factor STAT3 on SOX2 in GC since STAT3 is known to regulate SOX2 expression in neuronal development. Furthermore, it was aimed to identifiy how SOX2 could be regulated by *H. pylori* in GC cells, as *H. pylori* infection is an important key factor in the development of stomach cancer.

SOX2 is not only an important factor concerning tumorigenesis but also important during development of the stomach. Finally, to investigate the role of SOX2 in gastric development, *Sox2* depletion in a conditional knock-out mouse was analyzed. Different knock-out systems as well as different Cre mouse lines were analyzed and compared. Results would give valuable hints how the regulation of SOX2 could influence stomach development. Furthermore, studying the adult stomach would give deeper insight into identification of adult stomach stem cells which still are not fully identified.

2. Methods

2.1. Molecular cloning

2.1.1. Cloning strategies

Insert DNA was amplified by PCR using primer pairs containing enzyme restriction sites to generate segments of inserts with cohesive ends. PCR products were digested with respective restriction enzymes and ligated into the vector system of interest (Tab.1).

Tab.1: List of vectors and primers for cloning strategies

Construct	Vector	Primers	Restriction sites
SOX2 full flag tag	pcDNA4-TO	se 5' GCA**GGTACC**ATGGATTACAAGGATGACGGACGATTAA GATGTACAACATGATGGAGACGGAG 3' as 5' CAG**GGATCC**TCACTTATCGTCGTCATCCTTGTAATCCATGT GTGAGAGGGGCAG 3'	Kpn I BamHI
dnSOX2 1-129 HA tag	pcDNA4-TO	se 5' CCA**GGATCC**ATGTACAACATGATGGAGACGGAG 3' as 5'GCA**GCGGCCGC**TCAAGCGTAATCTGGAACATCGTAT GGGTAACCCCCGCCGGGCAGCGTGTAC 3'	BamH I Not I
dnSOX2 1-158 HA tag	pcDNA4-TO	se 5' CCA**GGATCC**ATGTACAACATGATGGAGACGGAG 3' as 5' GCA***GCGGCCGC***TCAAGCGTAATCTGGAACATCGTAT GGGTAACCACTGTCCATGCGCTGGTTCAC 3'	BamH I Not I
shRNA	**Vector**	**Oligonucleotides**	**Restriction sites**
shSOX2	pLVTHM	se 5' **CGCGT**CCCC**GGTTGATATCGTTGGTAAT**TTCAAGAGA**TT ACCAACGATATCAACC**TTTTTGGAA**AT** 3' as 5' **CGAT**TTCCAAAAA**GGTTGATATCGTTGGTAAT**TCTCTTG AAA**ATTACCAACGATATCAACC**GGGG**A** 3'	Mlu I Cla I
shControl	pLVTHM	se 5' **CGCGT**CCCC**GTACAGCCGCCTCAATTCT**TTCAAGAGA**AG AATTGAGGCGGCTGTAC**TTTTTGGAA**AT** 3'	Mlu I Cla I

		as 5' **CGAT**TTCCAAAAA**GTACAGCCGCCTCAATTCT**TCTCTTGA A**AGAATTGAGGCGGCTGTAC**GGGG**A** 3'	

2.1.2 Restriction digestion of DNA

DNA (0.5 ng - 5 µg) was digested for 1 to 2 h with 1 unit of appropriate restriction enzyme (Promega) per µg DNA in the respective restriction enzyme buffer (Promega) following supplier's instructions. To verify successful digestion, restricted DNA was analyzed by agarose gel electrophoresis. Correct fragments were excised from the gel and purified using *Illustra GFX PCR and Gel Band Purification Kit* (GE Healthcare) according to manufacturer's recommendations.

2.1.3 Ligation

For ligation of cohesive ends a molar ration of 1:8 of digested and purified vector DNA and digested insert were used. All ligations were set up using 1 µl (2 µl) T4 DNA ligase and respective ligation buffer in a total volume of 10 µl (20 µl). Ligation reaction was incubated over night at 16 °C.

2.1.4 Competent E. coli cells

Bacteria which are able to uptake DNA are called "competent". Bacterial competence can be achieved by treating *E. coli* cells with calcium chloride in the early log-phase of growth. Cells can uptake chloride ions but their membrane is not permeable for calcium ions. While positive charged chloride ions enter the cells, they are accompanied by water molecules. This water influx causes the cells to swell which induces uptake of DNA. The exact mechanism of this phenomenon is not yet identified, however, it is known, that a heat shock step (42 °C) is necessary for DNA uptake.

A single bacterial colony of the *E. coli* strain K-12 was transferred into an inoculation tube containing 5 ml TYM medium and cultured over night at 37 °C. The next day culture was transferred into 500 ml TYM medium and growth was monitored every 20 min until OD_{600} of 0.5 -0.6 was reached. Cells were incubated on ice for 30 min and subsequently pelleted by centrifugation (3000 x g, 10 min, 4 °C) and washed with 15 ml

ice cold Tfb-I Buffer. Cells were centrifuged again (3000 x g, 10 min, 4 °C) and resuspended in ice cold 20 ml Tfb-II Buffer. Cell suspension was aliquoted in 100 µl fractions and immediately frozen in liquid nitrogen. Long term storage was done at – 80 °C.

2.1.5 Transformation of DNA in bacteria

For transformation 100 µl competent *E. coli* were thawed on ice for 5 min. DNA (50 – 200 ng) was added and mixture was incubated on ice for 20 min. Subsequently cells were heat-pulsed at 42 °C for 45 sec and immediately cooled down on ice for 1 - 2 min. 1 ml LB medium (37 °C) was added to each tube and cells were incubated for 1 h at 37 °C while shaking continuously. Bacteria were pelleted by centrifugation (5000 x g, 5 min) and plated on prewarmed LB plates (LB medium with 1.5 % Agar) containing the appropriate antibiotics (50 µg/ml ampicillin or kanamycin). Plates were incubated over night at 37 °C.

2.1.6 Preparation of plasmid DNA

Single *E. coli* colonies were transferred from the agar plate into 5 ml LB medium with antibiotics and were grown over night at 37 °C while shaking. For Mini preparations this over night culture was used directly. For Midi preparations an aliquot of this culture was transferred to 100 ml LB medium with appropriate antibiotics and again grown over night at 37 °C while shaking. All plasmid DNA preparations were carried out with either *SV Minipreps Wizard Plus* Kit (Promega) or *PureYield plasmid Midiprep Kit* (Promega) according to manufacturer's recommendations.

2.1.7 shRNA cloning

shRNA oligonucleotides of SOX2 were cloned into the vector pLVTHM according to manufacturer's protocol. In contrast to general cloning procedure, purified oligonucleotides must be annealed prior to ligation. For annealing 1 µl (100 ng) oligos for top and bottom strand were mixed together with 48 µl annealing buffer and boiled for 5

min to remove secondary structure and disrupt hairpin. The mixture was cooled down slowly on ice and stored at -20 °C until ligation.

Before ligation, oligos were phosphorylated. 5 µl (500 ng) of oligos was mixed with 12 µl of dH_2O, 2 µl T4 ligation buffer and 1 µl T4 polynucletide kinase (PNK). Mixture was incubated at 37 °C for 30 min followed by heat inactivation of PNK at 70 °C for 10 min. For ligation, 5 µl (125 ng) of annealed phosphorylated oligos was mixed with 20 – 100 ng of digested and purified pLVTHM vector, as a large excess of oligo will inhibit ligation. Ligation took place at RT for 3 - 4 h. Oligos and digested vector were ligated according to supplier's recommendations and transformed into competent *E. coli* DH5α as described above. After plasmid preparation, the desired recombinant plasmid was identified by screening PCR using the primer H1 (5' - TCG CTA TGT GTT CTG GGA AA – 3'). Insert was verified by sequencing with the same primer. Transduction in the lentiviral system was done as described below.

2.1.8 Agarose gel electrophoresis

Electrophoresis through a horizontal agarose gel of 0.7 % - 2 % (depending on MW of the sample) was used to separate, identify and purify DNA fragments. Gels were prepared by mixing appropriate portions of agarose with 1 x TAE buffer. The mixture was cooked and Roti®-Safe GelStain (Roth) was added in a concentration of 5µl/100ml. Agarose gels were poured with a thickness of 6 - 7 mm in a precast agarose gel chamber and added to a running chamber containing 1 x TAE buffer. DNA was mixed with loading dye and 20 µl of the mix was added to gel pockets. An electric field of 80 - 120 mV was applied for 30 – 45 min to separate DNA fragments. A DNA 1 kb ladder was used for size determination under an UV Transilluminator, Eagle Eye Gel Doc.

2.2 Cell culture

In this work several human tumor cell lines where used. AZ-521 cells were purchased from the Japanese Collection of Research Biosources (JCRB, Osaka, Japan) and cultivated in 75 cm^2 culture flasks in Minimum Essential Medium containing

glutamine (MEM-GlutaMax) supplemented with 10 % tetracycline-free fetal bovine serum and 1 % penicillin/streptomycin. AGS, Kato III, MKN7, MKN45, NUGC4, N87, SNU1, St2957, St3051 and St23132 cells were purchased from the American Type Culture Collection (ATCC). All of these cell lines were cultivated in RPMI-1640 with 10 % FCS and 1 % penicillin/streptomycin. All cells were maintained at 37 °C and 5 % CO_2 in a 95 % air-humidified incubator. Cells were routinely checked for *Mycoplasma* contamination.

2.2.1. General cell culture methods

2.2.1.1. Cell counting

Cell suspension was diluted 1:10 with PBS and 10 µl of this suspension was diluted 1:2 with trypan blue. 10 µl were counted in a Neubauer hematocytometer. Trypan blue is a vital dye that stains broken cell membrane. Thus, stained (dead) cells are not counted. Cells were counted in each of the four corner quarters. One corner quarter represents the area of 1 mm^2 and the height of 0.1 mm, thus holds 0.1 µl cell suspension. The mean of the cell number per quarter was calculated. Mean cell number multiplied by 10^4 (volume of the chamber) and the dilution factor (10 x 2) resulted in the number of cells per ml.

2.2.1.2. Freezing cells

Cell suspensions of approximately 1 x 10^7 cells were centrifuged at 1000 x g for 8 min. Cell pellets were resuspended in 10 ml ice cold FCS containing 5 % DMSO and 1 ml aliquots were prepared in sterile polypropylene tube for cryogenic storage. Tubes were slowly cooled down over night at -80 °C in a polystyrene box. Long term storage took place in liquid nitrogen.

2.2.1.3. Thawing and maintaining cells

Frozen cell suspensions were thawed in a water bath at 37 °C and were added to 50 ml cell growth medium immediately to dilute toxic DMSO. Cells were subsequently centrifuged at 1000 x g for 8 min. Cell pellets were resolved in 12 ml growth medium containing 10 % FCS and seeded into a sterile 75 cm^2 plastic flask. Cells were split every 3

– 5 days, depending on confluence. For splitting old medium was removed and cells were washed once with PBS. Cells were subsequently incubated for 2 – 5 min with 2 ml Trypsin/EDTA at 37 °C. When all cells were detached from the surface, fresh medium was added to the suspension. Cells were transferred to a 15 ml tube and spun down for 5 min at 1000 x g. Supernatant was discarded and cell pellet was resuspended in 5 ml growth medium and seeded at adequate cell density.

2.2.2. Transfection of cells

Cell lines were transfected by lipofection or electroporation. Lipofection is a method to introduce nucleic acids into cells by using vesicles, so called liposomes. Liposomes are positively charged and entrap the negatively charged genetic material. They are made of a phospholipid bilayer, thus they can easily pass the cell membranes as it also consists of lipid bilayer and is negatively charged. Electroporation is a very effective method of transfection. The permeability of the cell plasma membrane is increased by applying an electric field. After disruption of the phospholipid bilayer DNA can penetrate the cells. The membrane spontaneously reassembles and leaves the cell intact. However, many cells die during this method.

2.2.2.1. Transient transfection by lipofection

Cells were seeded in a density of 5 x 10^4 cells per well in a 24-well plate and incubated over night for attachment. Plasmids harboring the gene of interest or a control empty vector were diluted in serum-free OPTIMEM in a final concentration of 20 ng/ml. For reporter assays reporter plasmids containing promoter binding sites and firefly luciferase (200 ng/ml) and transfection control plasmids containing renilla luciferase (20 ng/ml) were added to the mix and cells were transfected with 15 µl Lipofectamine 2000 (Invitrogen) per 24-well plate according to manufacturer's recommendations. For transfection cells were cultured in serum-free OPTIMEM for 24 h. Medium was changed to normal growth medium the day after transfection. Lysates for analysis were taken 48 h to 72 h after transfection, depending on experimental procedures.

2.2.2.2. Transient transfection by electroporation

1 x 10^6 cells in 500 µl OPTIMEM were inserted to a 0.4 cm gap electroporation cuvette together with a plasmid containing the gene to analyze (20 ng/ml), reporter plasmids (200 ng/ml) and transfection control plasmids harboring renilla luciferase (20 ng/ml). Electroporation was carried out in a BioRad electroporator applying 165 V and 1000 µF. 25 µl of cell suspension was subsequently seeded in each well of a 24-well-plate and 500 µl serum-free OPTIMEM was added. Medium was changed to normal growth medium the day after transfection. Lysates for analysis were taken 48 h to 72 h after transfection.

2.2.2.3. Titration of antibiotics (Killing curve)

To determine the optimal concentration for selection of stable cell clones (see below) killing curve experiments were performed. Different cell lines were seeded in 24-well-plates in a density of 5 x 10^4 cells per well. Antibiotics were applied to the cells after attachment. Zeocin was added in concentrations from 100 µg/ml to 1000 µg/ml, blasticidin was added in concentrations from 10 µg/ml to 100 µg/ml. Cell viability was analyzed microscopically after three and seven days. Optimal selection concentration was defined as the lowest concentration of antibiotic in which all cells were dead by seven days at the latest.

2.2.2.4. Stable transfection of cells

Cells were seeded in a 6-well-plate in a concentration of 1 x 10^5 cells per well and incubated over night for attachment. Transfection of cells with a plasmid containing the gene of interest and a eukaryotic antibiotic resistance cassette was achieved by lipofection or electroporation (see above). Each well was transfected with 1 µg of plasmid DNA in OPTIMEM. 24 h after transfection cells were trypsinized, diluted 1:2, 1:5 and 1:10 and seeded into 150 mm cell culture dishes containing 20 ml growth medium and the selective drug. Growth medium and antibiotic was changed every 3 to 7 days. After 2 to 3 weeks, single colonies began to appear and could be harvested. For isolation of colonies, growth medium was removed and cells were rinsed twice with PBS. Cell separating cylinders were dipped into autoclaved vaseline to make the rim sticky. Subsequently, cylinders were put over a cell colony and gentle pressure was applied to

prevent movement. 50 µl of Trypsin/EDTA was added into cylinders and cell culture plates were incubated for 2 min at 37 °C to allow cells to detach. Each trypsin-treated colony was placed into one well of a 24-well-plate containing growth medium and selective antibiotic. Cells were grown until 80 % confluence and split into 24-well-plates and 6-well-plates to check if the gene of interest was present via promoter analysis and protein expression. Positive cell clones were expanded and frozen.

2.2.2.5. Inducible cell clones

The Tet-On system was used to express dnSOX2 at certain time points after induction. The Tet-On system is an aberrant version of the initially established Tet-Off system invented by Bujard and colleges in 1992. In this dual system the regulatory plasmid is a constitutively expressed tetracycline-responsive transcriptional activator (tTA), a fusion protein of tetracycline repressor (TetR) found in *Escherichia coli* and the VP16 domain of *herpes simplex* virus under the control of a CMV promoter. The second plasmid of the system, a responsive plasmid, harbors the gene of interest under the control of a tetracycline responsive element (TRE). In absence of tetracycline, tTA binds to TRE and activates transcription of the gene of interest. When tetracycline is added to the cells it binds to tTA. This complex can no longer bind the TRE and thus transcription is inhibited [176]. The Tet-On system is based on a mutated version of tTA (reverse tTA). It only binds to TRE in presence of tetracycline or a tetracycline derivate, e.g. doxycycline (Dox) and thus induces expression of the gene of interest [177].

A vector containing rtTA was stably transfected into cell lines (see above). Selection took place via transient transfection of clones with a reporter plasmid for rtTA (TO-Luc) and subsequent reporter assays (see below). Successfully transfected stable cell clones were expanded and again stably transfected with a plasmid containing the TRE and dnSOX2 cDNA. Stable clones were selected by reporter assay. Therefore, cells were transiently transfected with a SOX2 reporter plasmid (SOP-Flash), induced with doxycycline 3 h after transfection and analyzed 48 h later.

2.2.3. shRNA mediated RNA interference

To analyze the function of a specific gene product the endogenous mechanism of RNA interference is widely used since the last decade. The knock down of specific mRNAs reduces the expression of the respective protein. RNA interference (RNAi) can be set up by transient transfection of small interfering RNAs (siRNAs) or by stable expression of small hairpin RNAs (shRNAs). The latter are integrated in the genome and processed intracellularly to siRNAs. The stable expression allows a constitutive mRNA knock down.

A lentiviral transfer approach was used to set up stable *shSOX2* cell clones in AZ-521 cells. Lentiviruses are a subclass of retroviruses, which means they integrate their genome into the DNA of their host cells. To ensure safe application these viral vectors never carry genes for their spontaneous self-assembly and thus are unable to self-replicate. In the commonly used lentiviral systems of the so called 2nd and 3rd generation the least necessary number of viral genes therefore is distributed to several plasmids. Producer cell lines, e.g. HEK 293T cells, are transiently transfected with several plasmids (vector, packaging system, envelope plasmid) to initiate assembly of the virus. They secrete the virus into the supernatant where it can be collected and used for transduction.

shSOX2 was cloned into pLVTHM vector as described above. HEK 293T cells were transiently transfected with the recombinant pLVTHM vector, the 2nd generation packaging vector psPAX2, containing a CMV promoter and the 2nd generation envelope vector pMD2.G to produce lentiviral particles. AZ-521 cells were seeded to culture plates and infected with the virus. The production of lentivirus and the infections were done by Dr. Martina Anton (Klinikum rechts der Isar, department for experimental oncology, München, Germany).

2.3. Biological assays

2.3.1. Luciferase reporter gene assay

Luciferase activity was measured using the *Dual-luciferase reporter assay system* (Promega). In this assay two individual reporters are simultaneously expressed within a single system. The first reporter is typically correlated with an effect of experimental conditions, whereas the second co-transfected reporter serves as internal transfection control that gives a baseline signal. Variances caused by differences in transfection efficiency, pipetting volume or cell lysis efficiency can be minimized by normalizing activity of the first reporter to the activity of the internal control. In this system the first reporter was coupled to firefly (*Photinus pyralis*) luciferase. As a control reporter renilla (*Renilla reniformis*) luciferase driven by a CMV reporter was used. Both luciferases have distinct luminescence characteristics which allow discrimination.

Cells were transfected with the plasmids of interest and reporter plasmids as described above. Growth medium was removed from transfected cells and cells were rinsed with PBS. 100 µl of 1 x passive lysis buffer was added to each well. Plates were shock-frozen for 30 min at -20 °C and subsequently incubated 30 min on a shaker at room temperature for complete lysis. 20 µl of cell suspension was transferred in duplicates to a white 96-well-round bottom plate. Firefly luciferase and renilla luciferase activity was quantified in an Orion microplate luminometer using *Dual-luciferase reporter assay system* (Promega). 25 µl of *Luciferase Assay Reagent II* (LARII) was added to each well and relative light units (RLU) of firefly luciferase were measured. The second substrate, *Stop & Glo Reagent*, quenched firefly luminescence and initiated renilla luciferase activity which was quantified immediately. Ratio of firefly luciferase to renilla luciferase was calculated for quantification of relative promoter activity.

The WNT-responsive, TCF-dependent luciferase constructs, pTOPFlash and the control plasmid pFOPFlash, and the SOX2-responsive pSOPFlash and the control pNOPFlash have were kindly provided by Prof. Hans Clevers (NIOB, Utrecht, the Netherlands) and have been described previously [178]. 6x O/S LUC was a kind gift from Dr. Lisa Daily (New York University, New York, USA). pCMV-Renilla was purchased from Promega.

2.3.2. Proliferation analysis

Proliferation of cells was measured using *CellTiter-Glo Luminescent Cell Viability Assay* (Promega) and *Cell proliferation Dye eFluor 670* (eBioscience) according to manufacturer's recommendations.

The *CellTiter-Glo* system is based on the quantification of ATP of metabolic cells via a luminescence signal. ATP-yield is directly proportional to the number of viable cells in culture. Cells were plated in triplicates in a 96-well-plate in a density of 5x 10^3 cells per well. Cells were serum starved for 24 h to synchronize cell cycle. After synchronization cells were treated with 1 µg/ml doxycycline (see above). Measurement was performed in a Berthold Mithras plate reader after 24 h, 48 h, 72 h and 6 days. Cell numbers were normalized to the parental control cell line.

Cell Proliferation Dye eFluor 670 is used to monitor individual cell divisions. The red fluorescent dye binds to primary amines of cellular protein and is distributed equally between daughter cells during cell division, halving the fluorescence intensity. 10^5 cells were seeded into 10 cm dishes and serum starved for synchronization. After 24 h cells were stained with 2.5 µM *eFluor Proliferation Dye* according to recommendations of the supplier and seeded into 10 cm dishes with serum. 1 µg/ml doxycycline was added the same day and FACS analysis using APC channel were done after 24 h, 48 h and 72 h in a CyAn FACS sorter.

2.3.3. Analysis of apoptosis

Apoptosis of cells was measured with *Caspase-Glo 3/7 Assay System* (Promega) and by Annexin V binding. *Caspase-Glo 3/7 Assay* is based on the measurement of Caspase-3 and -7 activities of cells. A luminogenic substrate is cleaved by Caspase 3/7 and releases aminoluciferin which generates a luminescence signal. This signal is directly proportional to the amount of Caspase 3/7 present.

5 x 10^3 cells were seeded in triplicates in each well of a96-well-plate and serum starved over night for synchronization. After synchronization cells were treated with 1 µg/ml doxycycline for 13 h, 24 h and 48 h. Assay was performed according to

manufacturer's protocol and luminescence signal was measured in a Berthold Mithras plate reader. A final concentration of 1 µM Staurosporine was used as a positive control for apoptosis. Staurosporine treatment was done for 8 h.

In viable eukaryotic cells, phosphatidylserine (PS), which is a negatively charged phospholipid, is located at the intracellular site of the plasma membrane. However, when membranes get disrupted during cell death, PS is translocated to the outer membrane. Annexin V has high binding affinity to PS. Therefore, FITC-labeled Annexin V can be used to detect apoptotic cells via FACS. During necrosis cell membranes get disrupted and Annexin V can also bind to PS. To discriminate necrosis from apoptosis cells are co-stained with 7AAD which can only penetrate into necrotic cells. Thus, cells positive for Annexin V and negative for 7AAD are apoptotic cells [179]. For Annexin V staining 2 x 10^5 cells were seeded in a 10 cm dish. After synchronization by serum starvation for 24 h cells were fed with fresh growth medium for another 24 h and afterwards treated with 1 µg/ml doxycycline for 24 h and 48 h. Cells were washed with PBS, trypsinized and centrifuged for 8 min at 300 x g. The cell pellet was resuspended in 500 µl binding buffer containing FITC-Annexin V in a concentration of 10 µg/ml and suspension was incubated for 20 min at RT in the dark. 50 µg/ml 7AAD was added and suspension was incubated for another 10 min at RT in the dark. Subsequently cells were analyzed by flow cytometry in a CyAn ADP Analyzer.

2.3.4. Cell cycle analysis

Cell cycle was analyzed by measurement of propidium iodide (PI) intercalation into cells. PI intercalates into the major groove of double-stranded DNA and is the most commonly used dye for DNA content. When excited with a wavelength of 488 nm, PI fluoresces red, with an emission centered on 600 nm. Since it also can bind to double-stranded RNA it is necessary to treat cells in advance with ribonuclease A (RNaseA) for optimal DNA resolution.

1 x 10^5 cells were seeded into 100 mm cell culture dishes and serum starved for synchronization. After 24 h without serum, growth medium containing FCS and 1 µg/ml doxycycline was added. Subsequent to 24 h or 48 h of treatment with doxycycline, cells were harvested and pelleted by centrifugation of 500 x g for 8 min. Cell pellet was

washed twice with PBS containing 1 % (w/v) BSA. Resulting pellet was resuspended in 500 µl PBS 1 % BSA. For fixation, a single cell suspension was dropped into ice cold 70 % ethanol while vortexing. After 4 h at -20 °C or 24 h at 4 °C fixed cells were centrifuged for 8 min at 500 x g and pellet was washed twice with PBS 1 % BSA. Cells were resuspended in 200 µl PI staining solution containing 2 µg/ml RNase A. Cell cycle profile was analyzed by flow cytometry in a CyAn FACS sorter.

2.3.5. Wound healing/migration

Wound healing assays study directional cell migration *in vivo*. A wound is introduced in a cell monolayer and pictures are taken in regular intervals documenting accomplishment of wound healing and quantifying migration of cells.

Wound healing was analyzed using silicone culture inserts (ibidi) in a 12-well plate. A cell solution of 5 x 10^5 cells was prepared and 70 µl of this solution were applied in both wells of the insert. Cells were incubated over night until 100 % of confluence. The culture insert was removed and left a 500 µM gap between the two defined cell patches. 1 ml growth medium was added to the well. 1 µg/ml doxycycline was added for induction of the dnSOX2 construct. Wound healing was observed microscopically after 0 h, 24 h, 36 h and 48 h. Pictures were taken at a Leica DMB microscope.

2.3.6. Analysis of senescence associated β-galactosidase (SA-β-gal) activity

When diploid cells are no longer able to divide normally theses cells go to so called cellular replicative senescence. Senescent cells express β-galactosidase detectable at a pH 6.0, distinct from the acidic lysosomal β-galactosidase activity all cells express at a pH of 4.0. Using a citric acid/sodium phosphate buffer at pH 6.0, senescence can easily be visualized through the cleavage of X-gal, which yields in an insoluble blue color.

Cells were seeded on a 6 cm cell culture dish and grown until 50 % of confluence. Cell culture medium was removed and cells were washed three times with PBS. After washing, cells were fixed with fixative solution for 5 min at RT. Subsequently fixative solution was removed and cells were washed two times with PBS. 4 ml of staining

solution was added to the dish and cells were incubated at 37 °C without CO_2 over night. Cells were microscopically analyzed and photographed in a Leica DMB Microscope.

2.4. Immunofluorescence

Immunofluorescence is a method to detect proteins in tissue sections or seeded cells. Primary antibodies bind to their corresponding antigen and respective fluorescence-conjugated secondary antibodies build a complex with the primary antibodies which is detectable in fluorescence microscopy.

Cells were grown on cover slips in a 12-well plate until 60 % of confluence. Cover slips were washed with PBS and cells were fixed with a mixture of equal parts methanol and acetone for 15 min at RT. Fixation solution was removed and cover slips were washed three times with PBS. For permeabilization and blocking, cells were treated with blocking buffer for 15 min at RT to prevent unspecific antibody binding and to permeabilize cells for antibodies. Primary antibodies were incubated in antibody buffer over night at 4 °C in a humidified chamber. The following day cover slips were washed three times with antibody buffer. Corresponding secondary antibodies were diluted 1:200 in antibody buffer and incubated for 1 h in the dark in a humidified chamber. Cells were washed once in washing buffer. Plasma membrane was stained using 5 µg/ml *CellMask Deep Red Plasma Membrane Stain* (Invitrogen) according to manufacturer's recommendations. Cover slips were mounted on a microscope slide with *VECTASHIELD Mounting Medium* (Vector Laboratories) and sealed with nail polish. For long term storage slides were kept at -80 °C. Cells were microscopically analyzed and photographed using the confocal microscope, Leica SP5.

2.5. Infection with *Helicobacter pylori*

H. pylori strains (Tab. 2) were cultured on Wilkins–Chalgren plates supplemented with Dent supplement antibiotic mix (Oxoid). The bacteria were harvested, and the OD at 600 nm was measured. OD_{600} of 1 represented 2 x 10^8 bacterial cells. Different tumor cells were infected with *H. pylori* with a MOI (multiplicity of infection) of 50:1, 3 h after

transfection with SOPFlash/NOPFlash (see above). Luciferase activity was analyzed 3 h, 6 h and 24 h after infection.

Tab.2: H. pylori strains used

Strain	Resistance	Reference
G27 wt	-	Baltrus *et al*, Journal of Bacteriology, 2009 [180]
G27 Δ CagA	Kana	Figura *et al*, Gastroenterology & Hepatology, 2004 [181]

2.6. Sodium-dodecyl-sulfate- Polyacrylamide Gel Electorphoresis (SDS-PAGE)

In SDS polyacrylamide gel electorphoresis proteins are separated according to their molecular weight by migration through a gel in an electric field. The anionic detergent SDS denaturates proteins, disrupts hydrogen bonds, blocks hydrophobic interactions and unfolds protein molecules by elimination of their tertiary and structure. The discontinous gel [182] consists of a non-restrictive large-pore stacking gel, which concentrates the protein, on top of a resolving gel. In the latter, samples are separated. With aid of a protein marker the molecular weight of samples of interest can be estimated, as MW equals the logarithm of the distance the protein traveled in the gel.

Proteins were extracted by using 250 – 400 µl SDS lysis buffer depending on confluence of cells. Lysates were sonicated for 10 sec with amplitude of 50 %. An equal volume of each sample (30 µl) was heated for 5 min at 95 °C for denaturation and subsequently applied on the gel. Electrophoresis was performed at 120 V for 1 h.

2.7. Western Blot

For western blot, samples were separated by SDS-PAGE and the gel was blotted in a blotting chamber onto a Nitrocellulose Transfer Membrane. Assembling of the blot was as follows: four sponges, two sheets of whatman paper, nitrocellulose membrane, gel, two sheets of whatman paper and three sponges on top. All layers were soaked in transfer buffer and applied carefully without bubbles. The transfer was accomplished in transfer buffer with a voltage of 230 mA for 90 min. Proteins migrate towards the anode and stick to the membrane due to hydrophobic interactions. After blotting, the membrane was stained with ponceau red (Roth) to check if protein transfer was successful.

The membrane was washed in TBS-T for 5 min, followed by 30 min of blocking in 5 % non-fat milk powder in TBS-T, to prevent unspecific binding of antibodies. Primary antibodies were diluted in 1 % milk solution and incubated over night at 4 °C. Unbound antibody was removed by three wash steps with TBS-T for 5 min. Corresponding secondary antibodies, coupled to horse-radish peroxidase (HRP), were diluted 1:3000 in 1 % non-fat milk in TBS-T and incubated for 1 h at RT, followed by three final 5 min wash steps in TBS-T. All steps were conducted on a shaker. Peroxidase activity of the secondary antibody was detected by chemiluminescence using *SuperSignal West Pico Chemiluminescence Substrate* (Thermo Scientific) according to manufacturer's protocol. Signals were visualized on an X-ray film in a *Curix 60* developing machine.

2.8. RNA Analysis

2.8.1. RNA Isolation

RNA isolation was done by using *GenElute Mammalian Total RNA Miniprep Kit*. Cells were lysed and homogenized in 250 µl of lysis buffer containing guanidine thiocyanate and 2-mercaptoethanol to release RNA and inactivate RNases. Cell debris was removed by spin column. Total RNA was captured by binding to a high capacity silica

column and eluted with 50 µl nuclease free water. RNA was stored at -80 °C. Concentration and quality of RNA was proofed spectrophotometrically by *NanoDrop* at an absorbance of 260 nm.

To get rid of DNA contamination, DNA was digested by using *DNA-free DNase treatment and removal* according to manufacturer's protocol.

2.8.2. Reverse transcription

Reverse transcription is a technique to synthesize complementary DNA (cDNA) from RNA template. 3 µg RNA was incubated with 150 ng random primer for 5 min at 70 °C to melt secondary structure within the template in order to allow primer to bind to the mRNA. To prevent secondary structure to reform, the reaction was immediately cooled on ice for 5 min. The mixture was spun down for 2 min at 10.000 x g and supernatant of each sample was divided into two separate reaction tubes. Reverse Transcription-Mix was added to the samples. RT-Mix without M-MLV reverse transcriptase was used as negative control.

<u>RT-Mix</u>

M-MLV 5x reaction buffer	5 µl
(250 mM Tris/HCl, pH 8.5, 375 mM KCl, 15 mM $MgCl_2$, 50 mM DTT)	
dNTP	
M-MLV or ddH_2O	200 Units (1 µl)
ddH_2O	add 25 µl

The entire mixture was incubated in a thermocycler as follows:

RT (20 °C)	10 min
55 °C	50 min
70 °C	15 min

After reverse transcription 250 µl of nuclease free water was added to each sample to obtain a final volume of 300 µl cDNA with a concentration of about 10 ng/µl. Long-term storage of cDNA took place at -20 °C.

2.8.3. Quantitative real-time polymerase chain reaction (qRT-PCR)

qRT-PCR is a method to quantify mRNA expression levels throughout measurement of respective amount of cDNA via a PCR system, that allows to monitor the increase of DNA amount during amplification. The method is based on the fact that the fluorescence signal of SYBRGreen, intercalating into double-strand DNA is directly proportional to the amount of DNA.

All quantitative real-time PCR gene expression analyses were done in duplicates using *Maxima SYBR Green/ROX qPCR MasterMix* on a *Light Cycler 480 real-time PCR* system. 50 ng of RNA were mixed with 100 ng of forward and reverse primer of gene of interest and 12.5 µl *Maxima SYBR Green/ROX qPCR MasterMix*. dH_2O was added to a final volume of 25 µl to each sample. PCR was running over 40 cycles with 15 sec of 95 °C for denaturation and 1 min of 60 °C annealing temperature. All results were normalized to an endogenous housekeeping control gene Glyceraldehyde-3-phosphate-dehydrogenase (*GAPDH*). Quantification was performed by comparison of the median cycle threshold (C_T) value.

Tab. 3: Human primer used for quantitative real-time PCR

Gene	Primer sequences
C-MYC	fwd - AGCGACTCTGAGGAGGAACA rev - CTCTGACCTTTTGCCAGGAG
Cyclin B1 (*CCNB1*)	fwd - GGAAACATGAGAGCCATCCT rev - TTCTGCATGAACCGATCAAT
Cyclin D1 (*CCND1*)	fwd - CCGCTGGCCATGAACTACCT rev - ACGAAGGTCTGCGCGTGTT
Cyclin E (*CCNE1*)	fwd - TATCCTCCAAAGTTGCACCAGT rev - CAGGGGACTTAAACGCCACTTA
Δ-Catenin (*CTNND1*)	fwd - GATGCTGTCAAGTCCAATGCAG rev - AGTACTGGGATGCCCTTGAGC

DKK4	fwd - GGAGGTGCCAGCGAGATG rev - GGTGCCCAGTTGTTCCTTC
FGF10	fwd - CACATTGTGCCTCAGCCTTCC rev - AGGTGATTGTAGCTCCGCACA
GAPDH	fwd - GAAGGTGAAGGTCGGAGT rev - GAAGATGGTGATGGGATTTC
LEF1	fwd - TCCAGCTCCTGATATCCCTACTTT rev - CTGACCTTGCCAGCCAAGAG
NOTCH1	fwd - GTCAACGCCGTAGATGACC rev - TTGTTAGCCCCGTTCTTCAG
TP21	fwd - CAAAGGCCCGCTCTACATCTT rev - AGGAACCTCTCATTCAACCGC
SOX2	fwd - CCCTGTGGTTACCTTTTCCT rev - AGTGCTGGGACATGTGAAGT
SOX18	fwd - ACGCGTCGACCATGCAGAGATCGCCG rev - GAAGATCTCTAGCCGGAGATGCA
TGFb-RAP1	fwd - GCGGCTGTGTCCTTTCCATA rev - GCGTCTGCTTCTGTTGCTGAT
TP63 TA	fwd - GCATATCTGGGATTTTCTGGAAC rev - GGAGCCCCAGGTTCGTGTA
TP63 ΔN	fwd - GAAAACAATGCCCAGACTCAA rev - TGCGCGTGGTCTGTGTT

2.8.4. RNA Microarray

The principle behind expression microarrays is that complementary nucleic acid sequences specifically hybridize with each other by forming non-covalent hydrogen bonds. Isolated mRNA is transcribed into cDNA, which is subsequently labeled and transferred to a micro chip harboring a large subset of probes. Non-specifically paired strands are removed after washing, leaving only tightly bonded specific sequences for analysis.

AZ-521 dnSOX2 inducible cell clones and their parental cell line were seeded in triplicates in a 6-well plate. Cells were induced with doxycycline for 4 h, 8 h, 12 h, 18 h,

and 24 h and RNA was isolated using TRIzol® Reagent according to manufacturer's protocol. Three independent biological replicates were performed and each was tested for integrity on an agarose gel. Pure RNA was transcribed to cDNA and hybridized to a Gene Chip human gene 1.0 ST array (Affymetrix) according to manufacturer's recommendations. RMA-normalized data were analyzed using LIMMA [183]. Genes were scored as differentially expressed at a false discovery rate ≤ 0.05 and a log2-fold-change ≥ 1. Conduction and evaluation of the RNA microarray experiment was done in collaboration with Dr. Stefan Krebs (Gene center, Großhadern, Munich, Germany). Raw data can be accessed at GEO, accession number GSE42937.

2.9. In vivo studies

All animal experiments were permitted and controlled by a local ethics committee and conducted according to the guidelines of the German law of protection of animal life. Mice were kept in vented cages with cellulose bedding. Autoclaved standard food and water was provided *ad libidum*. The day/night cycle was 12 h.

2.9.1. Genotyping transgenic mice

In order to define genotypes of mice, PCR amplification was used. Primers were designed binding to a site unique to a mutation or a transgenic element and DNA was detected by agarose gel electrophoresis.

Tissues (clipped tail or ear) from mice were digested at 56 °C over night at 750 rpm in 500 µl Laird's Buffer containing 25 µg proteinase K. Suspension was centrifuged at 16.000 x g for 5 min. Supernatant was added to a new reaction tube containing 500 µl ice cold 2-propanol and gently mixed by inverting the tube several times until DNA precipitates. The DNA precipitate was transferred into a new collection tube and dissolved in 200 µl DNase free H_2O for up to 5 h at 37 °C at 750 rpm.

PCR was conducted as follows using *Taq Polymerase PCR Core Kit* (Qiagen):

Amplification Mix

Sample DNA	1.5 µl
Q-Solution	5 µl
Buffer	2.5 µl
dNTP	0.5 µl
Primers (100 ng)	1 µl each
Taq Polymerase	0.125 µl (1 Unit)
dH2O	ad 25 µl

Amplification conditions

Hot Start	95 °C	5 min	
Denaturation	95 °C	1 min	
Annealing	x °C	x min	x 36 cycles
Extension	72 °C	1 min	
Final extension	72 °C	10 min	
End hold	4 °C		

Annealing temperature and duration was dependend on primers used.

Tab. 4: Primers used for gentoyping

Genotype	Primer sequences	Annealing	Product length
Apc flox	fwd - TAGGCACTGGACATAAGGGC rev - GTAACTGTCAAGAATCAATGG	55 °C – 1 min	flox - 380 bp wt - 87 bp
Cea - T	fwd - AATTCTGAAGGAAAGTCCTTGG rev - TAATGGACCTTCTAGGTCTTGA	54 °C – 45 sec	387 bp
Flp-e	fwd - CTAATGTTGTGGGAAATTGGAGC rev - CTCGAGGATAACTTGTTTATTGC	50 °C – 1 min	600 bp
Gp130 k.o. vs. wt	GP130A - CTGAATGAACTGCAGGACGA GP130B - CAAGTGTTCTCAAGGTCCGAGTCCAC GP130C - TGAAGCCACTCGTCTTTAGC	60 °C - 1 min	k.o. – 700 bp wt – 500 bp

Lgr5- CreER^{T2}	fwd - CACTGCATTCTAGTTGTGG rev - CGGTGCCCGCAGCGAG	58 °C – 45 sec	180 bp
Rosa26- CreER^{T2}	fwd - AATCGCCATCTTCCAGCAGG rev - GATCGCTGCCAGGATATACG	58 °C - 45 sec	356 bp
Rosa26-LacZ vs. wt	Rosa1 - AAAGTCGCTCTGAGTTGTTAT Rosa2 - GCGAAGAGTTTGTCCTCAACC Rosa3 - GGAGCGGGAGAAATGGATATG	52 °C – 45 sec	Lac Z – 340 bp wt – 650 bp
Sox2 βgeo	rf 56 - CGTTGGCTACCCGTGATATT rf 24 - AGGCTGAGTCGGGTCAATTA	50 °C - 45 sec	700 bp
Sox2 deleted	rf 33 - GACCTAGCCAGACCCCCTTA rf 24 - AGGCTGAGTCGGGTCAATTA	60 °C - 45 sec	600 bp
Sox2 flox Δ Neo	rf 23 - CAGTCCAAGCTAGGCAGGTT rf 24 - AGGCTGAGTCGGGTCAATTA	56 °C – 45 sec	flox – 500 bp wt – 300 bp
Sox2-RFP-CreER^{T2} k.i. vs. wt	fwd – TACGACGTCAGCGCCCT rev k.i. - GTACAGCATCTCGGTGTTGG rev wt- CCTCCAGATCTATACATGGTC	56 °C – 1 min	k.i. – 882 bp wt – 662 bp
Sox2-RFP-CreER^{T2} ΔNeo	fwd - GATCTGCTCCTCGAAGCCG SOXΔNeo - CGCAGCTGTCGTTTCGCTG	56 °C – 1 min	ΔNeo – 468 bp
Sox17iCre ΔNeo vs. Neo	EP420 - ATTGCATCGCATTGTCTGAGTAG EP401 - CTCAACTGTTCAAGTGGCAG EP564 - GATCTATGGTGCCAAGGATGAC	60 °C – 45 sec	Δ Neo – 446 bp Neo – 374 bp
Sox17iCre ΔNeo vs. wt	EP400 - GTGTATAAGCCCGAGATGG EP401 - CTCAACTGTTCAAGTGGCAG EP564 - GATCTATGGTGCCAAGGATGAC	60 °C – 45 sec	Δ Neo – 446 bp wt - 288 bp

Samples were loaded onto a 1.5 % or 2 % agarose-gel and analyzed under an UV Transilluminator, Eagle Eye Gel Doc (see above).

2.9.2. Preparation of blastocystes

Sox2$^{\beta geo/wt}$ x Sox2$^{\beta geo/wt}$ mice were mated and plugs were checked every day in the morning and in the evening. At day 3 after positive plug check (E 3.5) pregnant mothers were sacrificed and ovaries, oviducts and uteri were prepared and transferred into a petri dish containing M2 medium. Both uteri were flushed with 0.5 µl M2 medium and subsequently streaked out to remove all blastocystes. Blastocystes were carefully

transferred to a new petri dish containing drops of fresh M2 medium covered by paraffin oil and washed several times in M2 medium. Subsequently they were evaluated, counted and photographed.

2.9.3. Tamoxifen application

The CreER^{T2} is an inducible version of Cre-recombinase from bacteriophage P1 [25]. In this system the Cre enzyme is fused to a mutated human estrogen receptor (ER) binding-domain and thus located in the cytoplasm where it is inactive. Upon induction with tamoxifen, a synthetic estrogen antagonist, Cre recombinase translocates into the nucleus. It specifically mediates recombination of two 13 bp long loxP recognition sites. The DNA fragment flanked by these lox p sites ("floxed gene") is excised [26].

20 mg tamoxifen was pestled in a 25 mm mortar for 10 to 15 min. Powder was transferred to a 15 ml tube and bottles and mortar were rinsed with 500 µl absolute EtOH to capture all remains of tamoxifen. 10 ml corn oil was added to reach a final concentration of 2 mg/ml. The suspension was vortexed and subsequently added to an ultrasonic bath for 10 min. This cycle was repeated 5 – 10 times until tamoxifen was completely dissolved. The solution was used immediately or aliquoted in 1.5 ml reaction tubes and stored for up to 3 month at -20 °C. 100 µl of tamoxifen solution (200 µg) was applied to adult mice orally or intraperitoneally. Analyses took place between one day and several weeks after the first application, depending on experimental set up.

2.9.4. Histology

2.9.4.1. Organ isolation and embedding

Mice were killed by cervical dislocation. For paraffin embedding organs of interest were removed and fixed in 4 % paraformaldehyde (PFA) over night at 4 °C. Organs were transferred into embedding cassettes. Dehydration took place in a tissue processor. Organs were subsequently embedded in paraffin. Paraffin blocks were stored at room temperature. Before cutting blocks were cooled on ice or in a -20 °C freezer. Paraffin embedded tissue was cut in 4- 6 µm thick sections using a microtome. Sections were collected on positively charged glass slides and stored at RT or 4 °C until staining.

For cryo-embedded sections organs were fixed for 30 – 60 min in 4 % PFA and subsequently transferred to 15 % sucrose. After 6 – 8 h sucrose was changed to 30 %. Organs were incubated over night at 4 °C. The next day organs were placed in cryo-molds and carefully covered in *Tissue-Tek O.C.T. Compound* without bubbles. Blocks were frozen on dry ice in 80 % ethanol. Frozen organs were stored at -80 °C. Frozen tissue was cut with a cryotome into 5 -10 µm thick sections. Sections were collected on positively charged glass slides, air-dried for 30 – 60 min at RT and stored at -80 °C until staining.

2.9.4.2. Hematoxylin-Eosin (HE) staining

HE staining is a commonly used histological technique to visualize tissue morphology. Hematoxylin has an affinity for negatively charged molecules. It is staining acidic tissue components, e.g. nucleic acids or acidic proteins (purple/blue). Alkaline tissue components such as keratin or collagens are stained by Eosin (pink/red).

To remove paraffin, sections were put in an oven at 60 °C for 20 – 30 min and subsequently washed in AppliClear (3 x 10 min). Rehydration of sections was achieved by putting them in a graded series of EtOH dilutions – EtOH abs. (2 x 10 min) – 90 % EtOH (5 min) – 70 % EtOH (5 min) – 50 % EtOH (5 min) followed by a 2- 3 min washing step in dH_2O. Sections were stained by 20 % hematoxylin solution in PBS for 5 – 15 min, rinsed with tap water until residual staining was removed (3 – 5 min) and put into dH_2O. Subsequently slides were incubated in 80 % EtOH containing 0.15 % HCl for 1 min, followed by washing in dH_2O and a 5 sec incubation in 0.3 % ammonia water. After washing in dH_2O and a 1 min step in 96 % EtOH tissue was stained with 0.5 % Eosin for 30 – 60 sec followed by incubation in EtOH abs. (3 x 5 min) and AppliClear (3 x 5 min). Sections were mounted in DPX mounting medium and pictures were taken using a phase contrast microscope with camera.

2.9.4.3. Immunohistochemistry

Mouse sections were incubated in an oven for 20 – 30 min at 60 °C and subsequently washed with AppliClear (3 x 10 min) to remove paraffin. Slides were hydrated by consecutive incubation in a graded series of EtOH dilutions – EtOH abs. (2 x 10 min) – 90 % EtOH (5 min) – 70 % EtOH (5 min) – 50 % EtOH (5 min). Sections were

washed for 5 min in dH_2O. Unmasking of antigen was performed by boiling for 10 – 15 min in a pressure cooker filled with 0.01 M Sodium citrate pH 6.0 or 1mM EDTA pH 8.0 depending on specific recommendations for the antibody. After washing in dH_2O slides were washed in PBS (2 x 5 min), ice cold acetone (10 min) and PBS (2 x 5 min). For the last wash step slides were transferred to shandon coverplates. After 10 min treatment with peroxidase solution slides were blocked with blocking buffer for 30 – 60 min at RT. Primary antibody was diluted according to manufacturer's recommendations and incubated over night at 4 °C. The next day slides were equilibrated to room temperature for 15 min followed by six washing steps in PBS for 5 min each. Adequate biotinylated secondary antibody was diluted according to recommendations of the manufacturer and incubated for 30 min at RT. If ABC avidin/biotin method was used, *VECTASTAIN Elite ABC Kit* (Vector Laboratories) was prepared according to manufacturer's instructions and incubated for 30 min at RT. Sections were washed five times with PBS for 5 min each. DAB was prepared according to supplier's recommendations and added to slides for 5 min followed by washing in PBS (2 x 1-3 min). Tissue was counterstained with hematoxylin for 30 – 60 sec and washed in tap water until residual staining was removed. Subsequently slides were put in dH_2O and dehydrated by EtOH dilutions – EtOH 50 % (5 min) – EtOH 70 % (5 min) – EtOH 90 % (5 min) – EtOH abs. (2 x 10 min) followed by incubation with AppliClear (3 x 10 min). Coverslips were mounted in DPX mounting medium and pictures were taken using a phase contrast microscope with camera.

Gastric tumor samples (n=31) were obtained from the paraffin-embedded tissue bank of the Institut für Pathologie, Klinikum Bayreuth (Bayreuth, Germany), after approval of the local ethics committee. 4 µm sections were incubated with anti-SOX2 or Ki67 antibodies after antigen retrieval in Epitope Retrieval Solution pH 6. Slides were then incubated with ImmPRESS (peroxidase) *Polymer Detection kit* (Vector Laboratories) for 30 minutes and developed using *AEC Single Solution* (Invitogen) following manufacturer's instructions. Sections were counterstained with hematoxylin. For automated image acquisition Olympus Virtual Slide System VS120 was used.

2.9.5. Analysis of AZ-521 tumorgenicity in vivo

Female 6- 10 week old athymic NMRI nude-mice were used (n = 5 per group). AZ-521 dnSOX2 cells were virally transduced with an eGFP-Luc construct. Cells were sorted for eGFP by fluorescence activated cell sorting and GFP positive cells were grown in cell culture as described above, being kept under antibiotic free conditions prior to injection. To harvest cells they were trypsinized, washed three times with PBS and diluted in ice cooled PBS at a concentration of 10^6 cells per 150 µL. 150 µL of cell suspension was injected subcutaneously or intravenously. Bioluminescence signal was measured 4 h and 24 h after tumor cell injection and then every 3 to 4 days by a CCD camera in an IVIS Lumina system. For measurement mice were anaesthetized by inhalation of isoflurane in oxygen (2.5 % (v/v)) at a flow of 1 L/min. *Bepanthene* was put on the eyes for protection. Subsequently 100 µL luciferin solution (c = 60 mg/mL) was injected intraperitoneally and was allowed to distribute 10 - 15 min prior to bioluminescent measurement. Bioluminescence was measured with a medium or large binning and 1 - 5 min exposure time. Results were evaluated using *Living Image 3.0 software*.

2.10. Materials

2.10.1. Buffers and Solutions

Agarose Gel electrophoresis:

<u>1 x TAE-Buffer</u>

Tris/HCl pH 8.2	40 mM
Acetic acid	20 mM
EDTA pH 7.6	2 mM
In dH_2O	

Apoptosis assay:

<u>Binding buffer</u>

HEPES/NaOH, pH 7.4	10 mM
NaCl	140 mM

$CaCl_2$	5 mM
in dH_2O	

β-Galactosidase staining:

Staining Solution

Potassium Ferricyanide Crystalline	5 mM
Potassium hexacyanoferrate (II) Trihydrate	5 mM
$MgCl_2$	2 mM
X-Gal	1 mg/ml
in PBS	

Staining solution was prepared freshly at the day of experiment. X-Gal was dissolved in dimethylformamide and stored at -20 °C in the dark.

Blastocyste isolation:

M2 medium

NaCl	94.66 mM
KCl	4.78 mM
$CaCl_2 \times 2H_2O$	1.71 mM
KH_2PO_4	1.19 mM
$MgSO_4 \times 7H_2O$	1.19 mM
$NaHCO_3$	4.15 mM
Hepes (pH 7.4)	20.58 mM
Sodium lactate	23.28 mM
Sodium pyruvate	0.33 mM
Glucose	5.56 mM
BSA	4 g/l
Penicillin G, Potassium salt	1000 IU/ml
Streptomycin sulfate	50 µg/ml

Phenol red	0.01 g/l

Osmolarity was measured (should be between 280 and 287 mOsmole) and medium was filtered for sterilization and stored at -20 °C. BSA was added freshly before use.

Cell cycle assay:

PI staining solution

Sodium citrate	3.8 mM
PI	10 µg/ml
in PBS	

Competent *E. coli:*

LB (Luria Broth) medium

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
dH_2O	ad 1 L

Before addition of antibiotics, LB medium was autoclaved and cooled to 50 °C. LB medium was stored at 4 °C.

LB agar (pH 7.4)

NaCl	0.5 %
Yeast Extract	0.5 %
Tryptone	1 %
Agar	1.5 %
in dH_2O	

LB agar was autoclaved and cooled to 50 °C before addition of antibiotics and transfer to plates. LB agar plates were stored at 4 °C.

Tfb-I Buffer

K-acetate	30 mM
$MnCl_2$	50 mM
KCl	100 mM
$CaCl_2$	10 mM
Glycerol	15 % (v/v)
in dH_2O	

$MnCl_2$ and Glycerol were added after autoclaving.

Tfb-II Buffer

HEPES pH 7	10 mM
$CaCl_2$	75 mM
KCl	10 mM
Glycerol	15 % (v/v)
in dH_2O	

TYM medium

Tryptone	2 %
Yeast extract	0.5 %
NaCl	0.1 M
$MgSO_4$	10 mM
in dH_2O	

Histology:

Blocking solution

Normal goat serum	5 %
in PBS	

Peroxidase solution

30 % H_2O_2 (pH 7.5)	200 µl
NaN_3 (10%)	15 µl
PBS	ad 1500 µl

Immunofluorescence:

Antibody buffer

BSA	3 %
Saponin	1 %
Sodium Azide	0.05 %
in PBS	

Blocking solution

BSA	3 %
Triton X-100	0.1 %
Saponin	1 %
Sodium Azide	0.05 %
in PBS	

Washing buffer for Immunofluorescence

Saponin	1 %
in PBS	

Senescence associated –βgalactosidase activity staining:

Citric Acid/Phosphate Buffer, pH 6.0

Citric Acid	0.1 M
Sodium Phosphate Dibasic (Na_2HPO_4)	0.2 M
in H_2O	

Fixative Solution

Formaldehyde	2 % (v/v)
Glutaraldehyde	0.2 % (w/v)
in PBS	

Staining Solution

Potassium Ferricyanide Crystalline	5 mM
Potassium hecacaynoferrate (II) Trihydrate	5 mM
$MgCl_2$	2 mM
$NaCl_2$	150 mM
Citric acid phosphate buffer, pH 6.0	30 mM
X-Gal	1 mg/ml
in dH_2O	

Staining solution was prepared freshly at the day of experiment. X-Gal was dissolved in dimethylformamide and stored at -20 °C in the dark.

SDS-PAGE and Western Blot:

Resolving Buffer

SDS	0.4 % (w/v)
Tris/HCl, pH 6.8	0.5 M
in dH_2O	

Resolving Gel

Resolving Buffer	6 ml
Bis-Acrylamide (30 %)	9.56 ml
APS (10 %)	80 µl
TEMED (10 %)	40 µl
dH_2O	ad 23.92 ml

SDS Electrophoresis buffer

Tris	0.025 M
Glycine	0.2 M
SDS	0.1 % (w/v)
in dH_2O	

SDS lysis buffer (pH 6.8)

Tris/HCL, pH 6.8	62.5 mM
SDS	2 %
Glycerol	10 %
DTT	50 mM
Bromphenol blue	0.01 %
in dH_2O	

Stacking Buffer

SDS	0.4 % (w/v)
Tris/HCL, pH 8.8	1.5 M
APS (10 %)	40 µl
in dH_2O	

Stacking Gel

Stacking Buffer	1.96 ml
Bis-Acrylamide (30%)	1.56 ml
TEMED (10%)	40 µl
dH_2O	ad 7.84 ml

4x SDS sample buffer (pH 6.8)

Tris/HCl	240 mM
Glycerol	40 %
SDS	8 %

β-Mercaptoethanol	5 %
Bromophenol Blue	0.004 %

Ponceau S staining solution

Ponceau S	0.1 %
Acetic acid	5 %

10 x transfer buffer

Tris	25 mM
Glycin	190 mM
SDS	0.01 % (w/v)
in dH_2O	

1 x transfer buffer

10 x transfer buffer	100 ml
Methanol	200 ml
dH_2O	ad 1 L

TBS (pH 7.5)

Tris/HCl, pH7.5	50 mM
NaCl	150 mM
in dH_2O	

TBS-T

TBS	
Tween	0.1 %

shRNA cloning:

Annealing buffer

K-Acetate	100 mM

HEPES pH 7.4	30 mM
Mg-Acetate	2 mM
in dH_2O	

Tissue lysis:

Laird's Buffer

NaCl	100 mM
Tris, pH 8	10 mM
EDTA	10 mM
SDS	0.5 %
in dH_2O	

Buffer was autoclaved after preparation.

2.10.2. Chemicals and Kits

2-Propanol	67523	Roth
30 % Acrylamid/Bis Solution	161-0158	Bio-Rad
4-Hydroxytamoxifen	H-6278	Sigma
Acetic acid	100063	Merck
Aceton	9372	Roth
Agar-Agar	5210	Roth
Agarose	A9539-5009	Sigma-Aldrich
Albumin Fraction V (BSA)	A1391	AppliChem
Ammoniumpersulfat, APS	A2941	AppliChem
Ampicillin	K029.3	Roth
Annexin V-FITC	556419	BD Pharming
AppliClear	A4632	AppliChem
Aqua ad iniectabilia Diaco	2034374	Serag-Wiessner KG
Bench Top 1kb DNA Ladder	G7541	Promega
Bepanthene	7.158	Bayer
Beta-Mercaptoethanol	M3148	Sigma-Aldrich

Blasticidin S	R210-01	Gibco
Bovine Serum Albumin	A9416	Sigma-Aldrich
Bromophenol Blue	8122	Merck
Calcium chloride, $CaCl_2$	2385	Merck
Caspase-Glo 3/7 Assay System	G8090	Promega
CellMask Plasma Membrane Stain	C10046	Invitrogen
Cell Proliferation Dye eFluor 670 Cell Labeling	65-0840	eBioscience
Cell Titer-Glo® Substrate	G755A	Promega
Citric acid	251275	Sigma-Aldrich
Corn Oil	C-8267	Sigma-Aldrich
DAB solution Set	S-99014-103	42 Life Science
Dimethylformamide (DMF)	319937	Sigma-Aldrich
Disodium hydrophosphate, $Na_2HPO_4x2H_2O$	119753	Merck
Dithiothreitol, DTT	A3668	AppliChem
DMSO	A3672	AppliChem
DNA-free DNase treatment and removal	AM1906	Ambion
Doxycycline	44577	Fluka
DPX mounting medium	360294H	VWR
Dual Luciferase Reporter Assay System	E1960	Promega
Dulbecco´s modified Eeagle Medium, DMEM	41965	Gibco
EDTA-free Protease Inhibitor tablets, Complete	4693159001	Roche
Eosin	1024390500	Merck Millipore
Ethanol, absolute	A3678	AppliChem
Ethylenediaminetetraacetic acid, EDTA	A2937	AppliChem
Fetal Calf Serum	F7524-500ML	Sigma-Aldrich
FideliTaq PCR Master Mix (2X)	71182	USB
GenElute Mammalian Total RNA Miniprep Kit	RTN70-1KT	Sigma-Aldrich
Glutaraldehyde	85191	Fluka
Glycerin	A2926	AppliChem
Glycine	3790.3	Sigma-Aldrich
GoTaq Green Master Mix (2X)	M7112	Promega

Helicobacter pylori selective supplement	SE0147	Oxoid
Hematoxylin	1.09253	Merck
Hepes	25245	Serva
Hydrochloric acid, HCl	109063	Merck
Illustra GFX PCR and Gel Band Purification Kit	28-9034-70	GE Healthcare
Isofluran (Forene 100%)	B506	Abbott
Kaliumacetat, KAc	1.04820	Merck
Kaliumchloride, KCl	4936	Merck
LB Agar	75851	USB
LB Broth	75852	USB
Lipofectamine 2000	11668	Invitrogen
Loading Dye, 6X	R0611	Fermentas
Luciferin *VivoGlo*, In Vivo Grade	P1041	Promega
Maxima SYBR Green/ROX qPCR MasterMix (2X)	K0221	Fermentas
MEM-Glutamax	41090	Gibco
Methanol	4627.4	Roth
Milk Powder	T145.2	Roth
Magnesium acetate	105819	Merck
Magnesium sulfated, $MgSO_4$	1.05886	Merck
Manganese Chloride, $MnCl_2$	805930	Merck
M-MLV Reverse Transcriptase, RNase H(-)	M3682	Promega
Monopotassium phosphate, KH_2PO_4	104873	Merck
Normal goat serum	S26-100ML	Millipore
Opti-MEM	51985	Gibco
Orange G	O3756	Sigma-Aldrich
Paraformaldehyde	PZN2652965	Fischar
Penicillin G, Potassium salt	5161	Merck
Penicillin/Streptomycin 100 X Solution	15140	Gibco
Peroxidase 3 %, H_2O_2	-	Apo. Klinikum r.d.I.
peqGOLD Protein-Marker V	27-2210	PeqLab
Phenole red	A7615	AppliChem

Phosphate Buffered Saline, 1X	14190	Gibco
Pierce® ECL Western Blotting substrate	32106	Thermo Scientific
Ponceau Red	5938.1	Roth
Potassium chloride, KCl	P9541	Sigma-Aldrich
Potassium Ferricyanide Crystalline	P8131	Sigma-Aldrich
Potassium hecacaynoferrate (II) Trihydrate	60279	Fluka
Propidium Iodite	P4864	Sigma-Aldrich
Protein A-Agarose	05015979001	Roche
Protein G-Agarose	05015952001	Roche
Proteinase K	P2308	Sigma-Aldrich
PureYield Plasmid Midiprep System	A2492	Promega
Restore Western Blot Stripping Buffer	21059	Thermo Scientific
Ribonuclease A	R6513	Sigma-Aldrich
Roti®-Safe GelStain	3865.1	Roth
RPMI 1640	21875-091	Gibco
Saponine	4185.1	Roth
SDS-Pellets	CN30.2	Roth
SITRAN 1.0 siRNA transfection reagent	TT300002	Origene
Sodium azide, NaN_3	6265	Merck
Sodium chloride, NaCl	3957.1	Roth
Sodium citrate	1.12005	Merck
Sodium hydrogene carbnoate, $NaHCO_3$	HN01.1	Roth
Sodium hydroxide, NaOH	A4422	AppliChem
Sodium lactate	L7022	Sigma-Aldrich
Sodium pyruvate	S8636	Sigma-Aldrich
Staurosporine	S-9300	LC Laboratories
Streptomycine sulfate salt	S9137	Sigma-Aldrich
Sucrose	S0389	Sigma-Aldrich
SuperSignal West Pico Chemiluminescence Subs.	34080	Thermo Scientific
SV Minipreps Wizard Plus	A1460	Promega
T4 polynucletide kinase (PNK)	M0236L	NEB

T4 DNA Ligase	M180	Promega
Taq PCR Core Kit	201225	Qiagen
Tetramethylethylenediamine, TEMED	A1148	AppliChem
Tissue-Tek O.C.T. Compound	SAK 4583	Sakura
Tris	9090.3	Roth
Triton X-100	A1388	AppliChem
TRIzol® Reagent	15596-026	Ambion
Trypan Blue 0.4 %	15250	Gibco
Trypsin/EDTA	41010	PromoCell
Tryptone	A1553	AppliChem
Tween 20	A1389	AppliChem
VECTASHIELD mounting medium	H-1000	Vector Laboratories
VECTASHIELD mounting medium with DAPI	H-1200	Vector Laboratories
VECTASTAIN Elite ABC Kit	PK-6100	Vector Laboratories
Wizard SV Gel and PCR Clean Up System	A9282	Promega
X-Gal (5-Bromo-4-Chloro-3-Indolyl b-DGalacotpyranoside)	B4252	Sigma-Aldrich
Yeast extract	212750	BD Bioscience
Zeocin	R25001	Invitrogen

2.10.3. Antibodies

Tab. 5: Antibodies used in this study

Primary antibodies		
Name	***Implemented concentration***	***Manufacturer***
Anti-β-actin (mouse, monoclonal)	WB 1:5000	Sigma-Aldrich A1978
Anti-HA (mouse, monoclonal)	WB 1:1000 IF 1:200	Sigma-Aldrich H3663
Anti-HA (rabbit, polyclonal)	WB 1:1000 IF 1:200	Sigma-Aldrich H6908
Anti-ΔNp63 (rabbit, polyclonal)	WB 1:200	Santa Cruz SC-8609-R

Anti-SOX2 (rabbit, monoclonal)	IF 1:400 WB 1:1000 IHC 1:100	Cell signaling 3579
Anti- pSTAT3	WB 1:1000 IF 1:100	Cell signaling 9131
p21	WB 1: 1000	Cell signaling 2947
Secondary antibodies		
Name	***Implemented concentration***	***Manufacturer***
Anti-mouse IgG (H&L) HRP conjugate	WB 1:2500	Promega W402B
Anti-rabbit IgG (H&L) HRP conjugate	WB 1:2500	Promega W401B
AlexaFlour488 goat anti-mouse IgG	IF 1:200	Invitrogen A11001
AlexaFlour594 chicken anti-rabbit IgG	IF 1:200	Invitrogen A21442

2.10.4. Consumables

Blotting Paper	GB46	Hartenstein
Cell scrapers, 30 cm	99003	TPP
Cell culture flasks, sterile, 25 cm^2	734-2311	VWR International
Cell culture flasks, sterile, 75 cm^2	734-2313	VWR International
Cell culture flasks, sterile, 175 cm^2	734-2315	VWR International
Cellstar Centrifuge Tubes, 15 ml	188271	Greiner Bio-one
Cellstar Centrifuge Tubes, 50 ml	227261	Greiner Bio-one
Cloning cylinders	TR-1004	Millipore
Cryogenic vial	LW3332	Alpha Laboratories
Cryomolds, Intermediate (15 x 15 x 5)	SAK 4566	Sakura
Cryomolds, Standard (25 x 20 x 5)	SAK 4557	Sakura
Culture inserts	80209	ibidi
DNA Pick-Tips	03DNA65	Süd-Laborbedarf
Electroporation cuvettes, 0.4 cm gap	165-2081	Bio-Rad
Eppendorf Microcentrifuge tubes, 1.5ml	T9661	Sigma-Aldrich
Gel cassettes	NC2010	Invitrogen
Inoculation Spreader	612-1561	VWR International
LightCycler480 Multiwell Plate 96, white	04 729 692 001	Roche

Medical X-ray film (Super RX)	47410 08389	Fujifilm
Microscope Cover Glasses, 18 mm	0101030	Marienfeld
Microscope Slides	ISO8037/1	Thermo Scientific
PCR tubes	6003-SF	Kisker
Plastic embedding device	07-7100	Biooptica
Protan Nitrocellulose Transfer Membrane, 0.45 µm	NC04	Whatman
Round-Bottom Assay Plate, 96 well	3789	Costar
Serological pipettes, 5 ml	606 180/5ml	Greiner Bio-one
Serological pipettes, 10 ml	607 180/10ml	Greiner Bio-one
Serological pipettes, 25 ml	706 180/25ml	Greiner Bio-one
Superfrost plus microscope slides	J1800AMNZ	Thermo Scientific
TipOne Graduated Filter Tips, 10 µl	S1121-3810	Starlab
TipOne Bevelled Filter Tips, 20 µl	S1121-1810	Starlab
TipOne Graduated Filter Tips, 200 µl	S1120-8810	Starlab
TipOne Graduated Filter Tips, 1000 µl	S1122 1830	Starlab
Tips, 0.1 - 10 µl	613-3500	VWR International
Tips, 2 - 200 µl	613-3503	VWR International
Tips, 500 - 1000 µl	613-3508	VWR International
Tissue culture plate, 6 well	353046	BD Bioscience
Tissue culture plate, 12 well	353043	BD Bioscience
Tissue culture plate, 24well	353047	BD Bioscience
Tissue culture plate, 48 well	353078	BD Bioscience
Tissue culture dish, 100 mm	353003	BD Bioscience
Tissue culture dish, 150 mm	353025	BD Bioscience

2.10.5. Instruments

Agate morter (25 x 20 mm) with pistil	neoLab
Agarose Electrophoresis Chambers	Bio-Rad
AxioVert 40 Microscope	Zeiss
Biofuge Pico Centrifuge	Heraeus, Thermo Scientific

Biofuge Primo R Centrifuge	Thermo Scientific
Cell Culture Incubator, HeraCell, 240	Heraeus, Thermo Scientific
Centrifuge 5415D	Eppendorf
Centrifuge 5424	Eppendorf
Centrifuge Micro 200R	Hettich
Confocal Microscope, Leica SP5	Leica Microsystems
Developing Machine Curix60	AGFA
Electrophoresis Power Supply PowerPac300	Bio-Rad
Embedding system TBS88	Medite
Fluorescence Microscope, DMRB	Leica
Freezer -20°C	Liebherr
GenePulserXCell	Bio-Rad
Heatable magnetic stirrer	IKAMAG RET-G
HERAFreeze Basic -80°C Freezer	Thermo Scientific
Hybridization Oven OV3	Biometra
Ice Machine	Ziegra
iCycler	Bio-Rad
Innova 2100 Platform Shaker	New Brunswick Scientific
IVIS Lumina System	Caliper Life Sciences
Laboratory Precision Scale XB120A Precisa	Precisa
Laboratory Scale	Kern
Laminar Airflow Cabinet, HeraSafe	Heraeus, Thermo Scientific
Leica DMRB Microscope	Leica microsystems
LightCycler480	Roche
Mithras LB9440	Berthold
Multifuge 3 S-R	Heraeus, Thermo Scientific
Neubauer hemocytometer cell counting chamber	Marienfeld
Orion Microplate Luminometer	Berthold
Olympus Virtual Slide System VS120	Olympus
pH-Meter	WTW inoLab
Electronic Pipet Filler, Easypet®	Eppendorf

Power Supply PowerPac Basic (for SDS gels)	Bio-Rad
Refrigerator 4°C	Liebherr
Shaker, Titramax 100	Heidolph
Shandon Coverplates	Thermo Scientific
Single Channel Pipettor, 0.5-10 µl	Corning
Single Channel Pipettor, 2-20 µl	Corning
Single Channel Pipettor, 20-200 µl	Corning
Single Channel Pipettor, 200-1000 µl	Corning
Sonoplus HD60	Bandelin
Spectrophotometer NanoDrop 1000	Thermo Scientific
StepOnePlus™ Real-Time PCR System	Applied Biosystems
T10 basix ULTRA-TURRAX Disperser	IKA
T3000 Thermocycler	Biometra
Thermomixer compact	Eppendorf
Tissue processor Shandon Excelsior ES	Thermo Scientific
UV Transilluminator, Eagle Eye Gel Doc	Bio-Rad
Vortex Genie 2	Schultheiss
Waterbath Julabo	Julabo
Waterbath	GFL
XCell IITM Blot Module	Invitrogen
XCell SureLock® Mini-Cell	Invitrogen

3. Results

3.1. Role of SOX2 in gastric cancer

3.1.1. SOX2 expression in gastric tumors and gastric cancer cell lines

3.1.1.1. SOX2 expression in gastric cancer tissue samples

Several GC tissue samples from intestinal and diffuse type tumors were stained and analyzed for SOX2 expression. About 30 % of all tumors were positive for SOX2. 80 % of SOX2 positive tumors were classified as intestinal type, whereas 20 % were diffuse type tumors. Among SOX2 negative tumors 70 % were intestinal type and 30 % diffuse type tumors. Positive and negative type tumors did not significantly differ in histological grade. All SOX2 negative tumors were infected with *H. pylori*. Among SOX2 positive tumors one case was negative for *H. pylori* infection (Tab. 6).

Tab.6: ***Clinico-pathological characteristics of SOX2 positive and negative human gastric tumors (n =31)***

	SOX2 positive (n=11)	SOX2 negative (n=20)
Age	65.1 ± 12.7	73.6 ± 8.6
Sex		
M	5[1] (45.5%)[2]	11 (55.0%)
F	6 (54.5%)	9 (45.0%)
Lauren´s classification		
Intestinal	9 (81.8%)	14 (70.0%)
Diffuse	2 (18.2%)	6 (30.0%)
Histological grade		
Low grade (G1-G2)	5 (45.5%)	11 (55.0%)
High grade (G3-G4)	6 (54.5%)	9 (45.0%)
H. pylori		
Active	8 (72.7%)	15 (75.0%)
H .pylori (eradicated)	2 (18.2%)	5 (25.0%)
Negative	1 (9.1%)	0

[1] Number of cases [2] Percentage of cases

Additionally, KI67 staining was performed, to determine proliferation sites in the tumors. Tumor cells expressing SOX2 showed also KI67 expression (Fig. 13), indicating, that cells positive for SOX2 are highly proliferative and thus potentially tumorigenic.

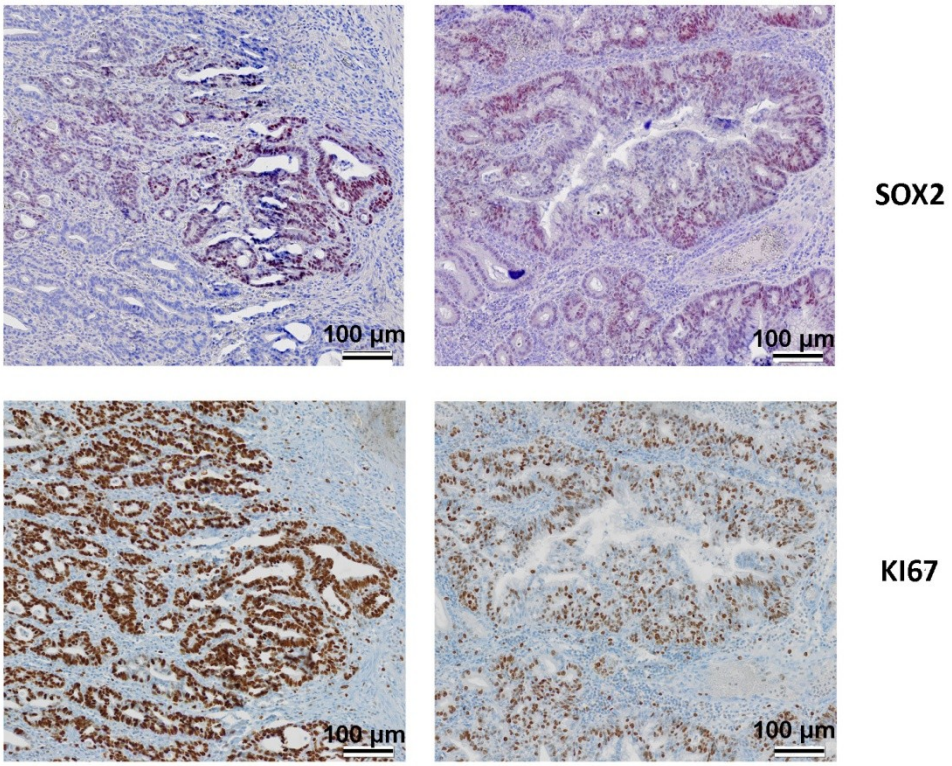


Fig. 13: Immunohistochemical analysis of gastric tumor tiussue samples; SOX2 expression in several gastric tumors correlated with KI67 expression.

3.1.1.2. Cell line screening for SOX2 expression in gastric cancer cells

The expression of the transcription factor SOX2 was analyzed in eleven GC cell lines by quantitative real time PCR and by western blot. The stomach carcinoma cell line AZ-521 showed highest RNA levels of *SOX2*. Significant amounts were also detected in AGS, Kato III and MKN45 cells. In all other cell lines no RNA of SOX2 was detected. Similar results were observed at the level of protein expression. AZ-521, AGS, Kato III and MKN45 cells showed high protein levels. Minor expression levels were detected in MKN7. No SOX2 protein expression could be detected in the other cell lines. Furthermore, SOX2 transcriptional activity was also analyzed in the same GC cell lines. For this purpose a SOX2 reporter plasmid containing a 6-fold SOX2 binding site in front of a luciferase gene was used (see **Material and Methods 2.1.2.1**). The relative level of transcriptional activity was expressed as relative light units (RLU) of luciferase to renilla.

Correlating with the expression levels of SOX2, AZ-521 cells showed very high SOX2 transcriptional activity. Interestingly, some cell lines, such as Kato III or MKN45 cells, exhibiting high *SOX2* mRNA and protein levels, showed only moderate or low transcriptional activity (Fig. 14).

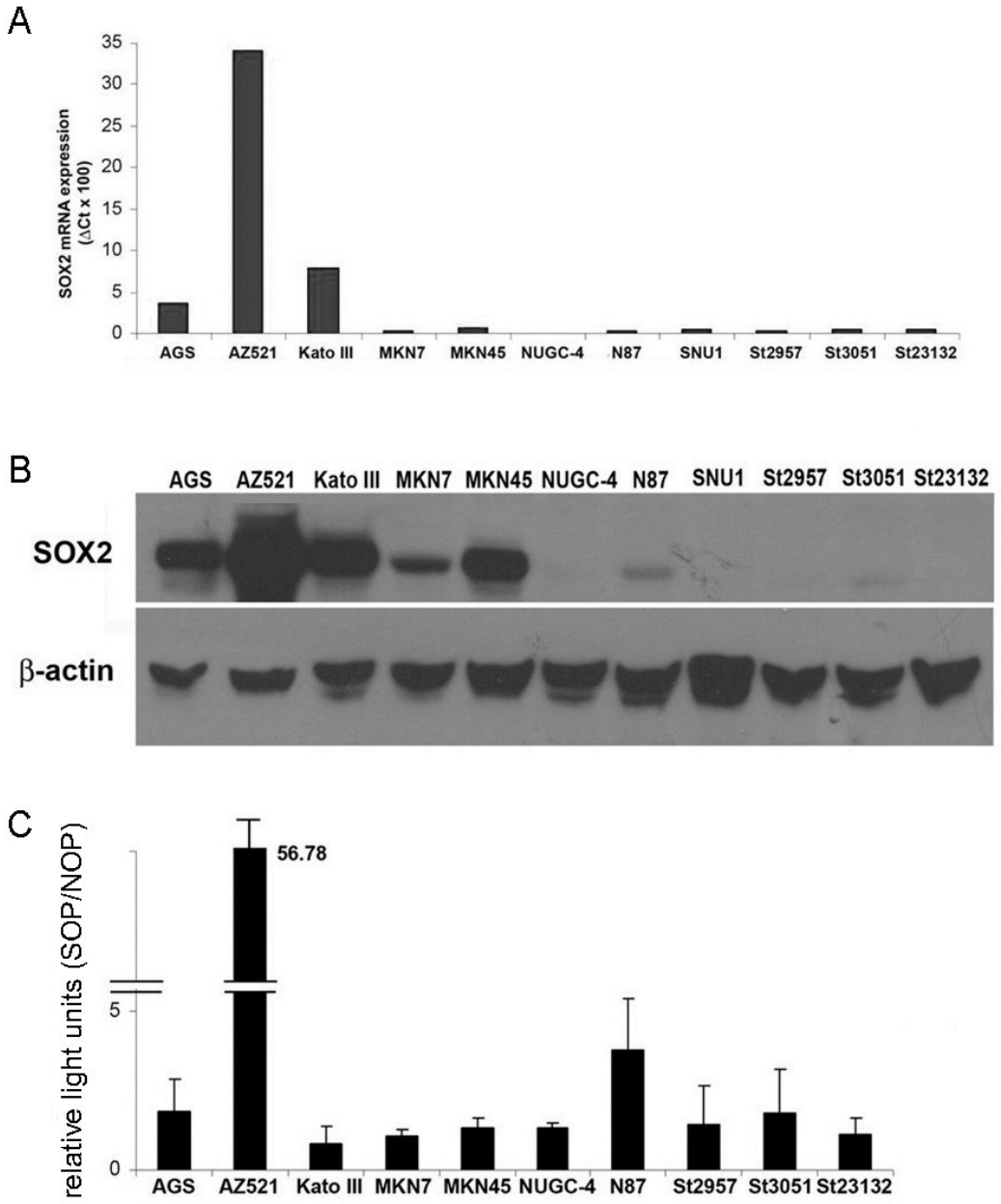


Fig. 14: SOX2 expression and transcriptional activity in GC cell lines; q-RT PCR was performed for analysis of mRNA expression of *SOX2.* N = 3, n = 9 **(A)**. SOX2 protein was detected by Western Blot. N = 3 **(B)** SOX2 transcriptional activity was measured by SOPFlash/NOPFlash assay. N = 3, n = 6 **(C)**

3.1.2. The functional role of SOX2 in gastric cancer cells

3.1.2.1. SOX2 knock-down strategies

To determine the functional role of SOX2 in GC, AZ-521, a cell line highly expressing SOX2, was chosen for further experiments. Two knock-down strategies to inhibit SOX2 were followed. The first strategy was to transiently transfect AZ-521 cells with shRNA of *SOX2* cloned into the expression vector pLVTHM. SOX2 protein expression was analyzed 48 h after transfection. Western blot showed a decrease of SOX2 protein after transfection with *shSOX2* in contrast to control shRNA, proving effectiveness of the *SOX2* shRNA (Fig. 15 A). Moreover, SOX2 transcriptional activity was analyzed 48 h after transfection. SOX2 transcriptional activity was 50 % decreased by *SOX2* shRNA in comparison to the parental cell line and cells transfected with the control shRNA (Fig. 15 B).

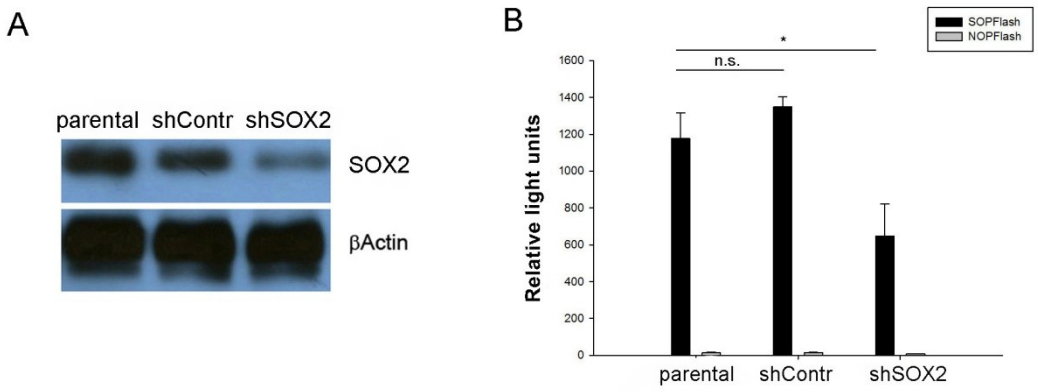


Fig. 15: Analysis of AZ-521 cells transiently transfected with SOX2 shRNA; western blot of AZ-521 cells 48 h after transient transfection with *shSOX2* showed a decrease in SOX2 protein compared to parental cell lines and cells transfected with a control shRNA **(A)**. N = 3; In contrast to the control shRNA, expression of *shSOX2* significantly reduced SOX2 transcriptional activity measured by reporter gene assay 48 h after transient transfection **(B)**. N = 3, n = 6;

In parallel, SOX2 was inhibited by blocking transcriptional activity, using a dominant negative construct for SOX2. Therefore, AZ-521 cells were stably transfected with a regulatory plasmid containing a tetracycline repressor (TR) and a responsive plasmid harboring a C-terminally truncated version of SOX2 under the control of a tetracycline responsive element (TRE). Two different truncated versions of SOX2 were cloned. The longer version consisted of the amino acids 1-158, the shorter version harbored only the amino acids 1-129. Both truncated dominant negative (dn) SOX2 constructs were designed to bind to recognition sites on DNA, as the HMG binding

domain was still functional, but they lacked the activation domain abolishing signaling activity (Fig. 16).

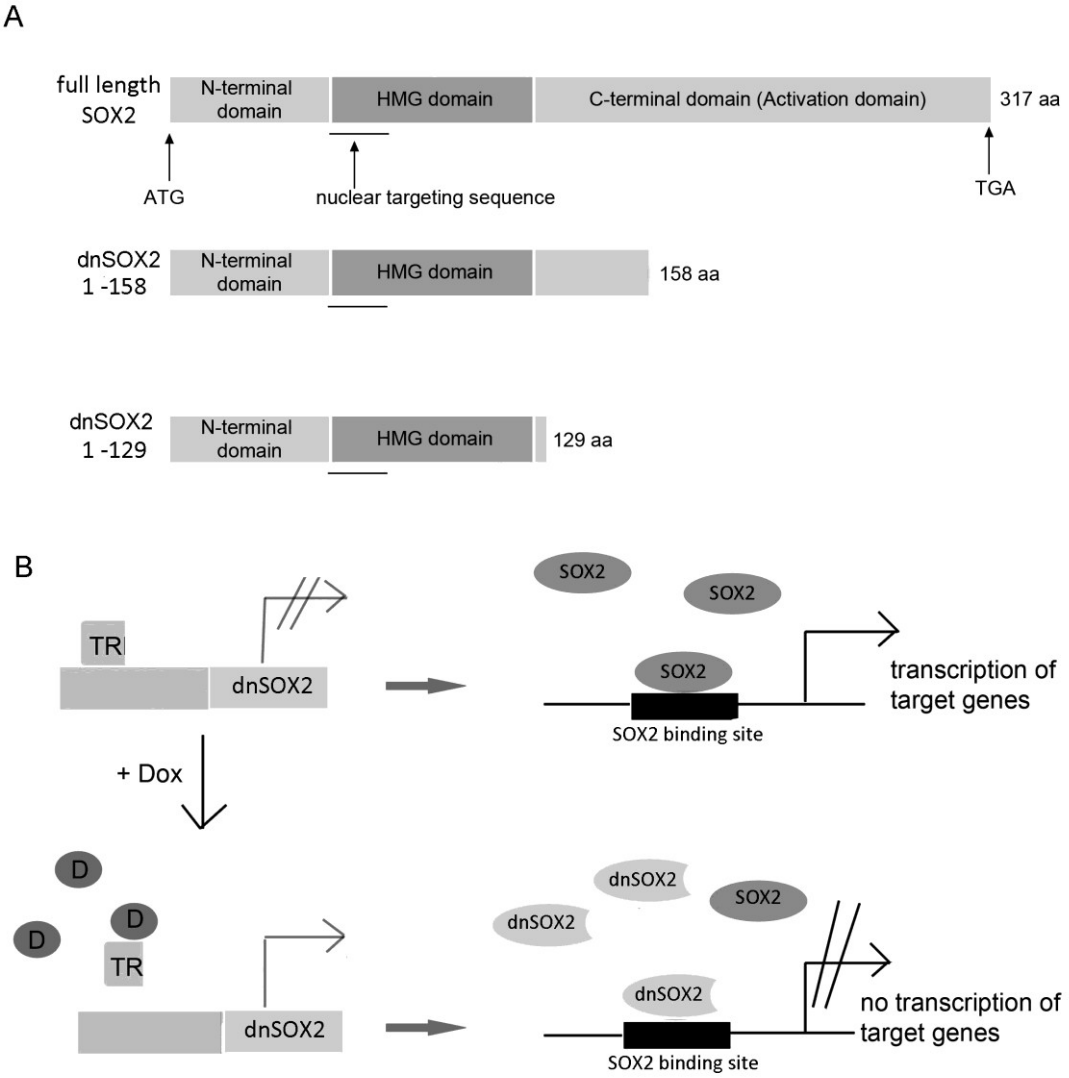


Fig. 16: Scheme of the SOX2 gene, the truncated SOX2 constructs and the Tet-On system; the nuclear targeting sequence lies in the HMG domain of the *SOX2* gene. Both dominant negative SOX2 constructs were C-terminally truncated, leaving the N-terminal domain and the HMG box fully functional **(A)**. In untreated cells dnSOX2 was blocked by a tetracycline repressor (TR) and SOX2 was transcribed normally. After adding doxycycline to the system, dnSOX2 was transcribed, blocking *SOX2* binding sites, therefore inhibiting transcription of target genes **(B)**. Drawings are not to scale.

Both versions were tested in transient transfection experiments for their ability to down-regulate SOX2 transcriptional activity. Significant inhibition of SOPFlash activity could be achieved by both truncated versions after transient transfection (Fig. 17 A). For generating stable cell clones, cells positively selected for the regulatory plasmid containing a tetracycline repressor (TR) by 20 µg/ml blasticidin were transfected with

each construct version of dnSOX2 and selected by applying 400 µg/ml zeocin. 24 clones of each version were picked and further cultured under antibiotic selective conditions. Only a subset of cells was viable after one week of selection and could be extended to test its function. Only cells harboring the shorter version, dnSOX2 1 -129, showed a down-regulation of SOX2 activity after inducing with 1 µg/ml doxycycline (Fig. 17 B). Thus, further experiments were conducted with this shorter dominant negative SOX2 construct in stable cell clones under the regulation of a tetracycline repressor.

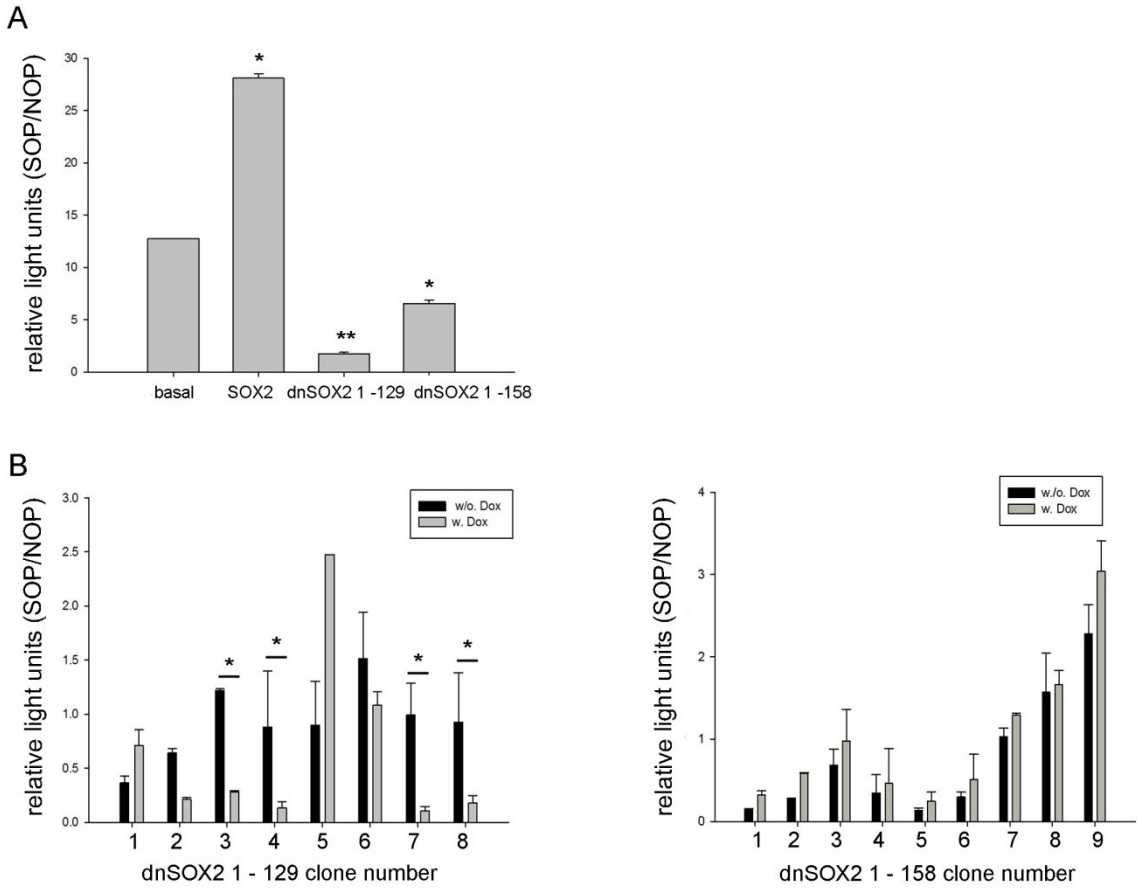


Fig. 17: Analysis of transcriptional activity of SOX2 after transfection with dnSOX2 constructs; SOX2 transcriptional activity was measured after transient transfection by SOPFlash/NOPFlash assay. SOX2 over expression induced SOX2 activity, whereas expression of dnSOX2 1 -129 and dnSOX2 1- 158 both led to a significant decrease of SOX2 activity in AZ-521 cells. **(A)** N = 4, n = 8; Four stable clones of AZ-521 TR dnSOX2 1-129 (no. 3, no. 4, no. 7 and no. 8) led to a decrease of SOX2 activity after induction with doxycycline. In AZ-521 TR dnSOX2 1 - 158 no inducible stable cell clone, down-regulating SOX2 could be identified **(B)**. N = 4, n = 8; Light units were normalized to renilla., * $p < 0.05$, p** < 0.01

Clone no. 4 and clone no. 7 were selected for further experiments as they showed best effects on blocking SOX2 transcriptional activity after induction with doxycycline. First, cell clones were tested for the expression of the HA-tagged dnSOX2 construct. Western blot showed dnSOX2 expression as early as 4 h after induction of

cells with doxycycline. Expression levels were increasing over time (Fig. 18 A). Nuclear expression was analyzed by immunofluorescence and staining was already visible after 8 h. After 12 h approximately 85 % of cells were positive for dnSOX2 staining (Fig. 18 B). Down-regulation of SOX2 transcriptional activity was time dependent and most pronounced 48 h after induction (Fig. 18 C). All following data are shown for the cell clone no. 4.

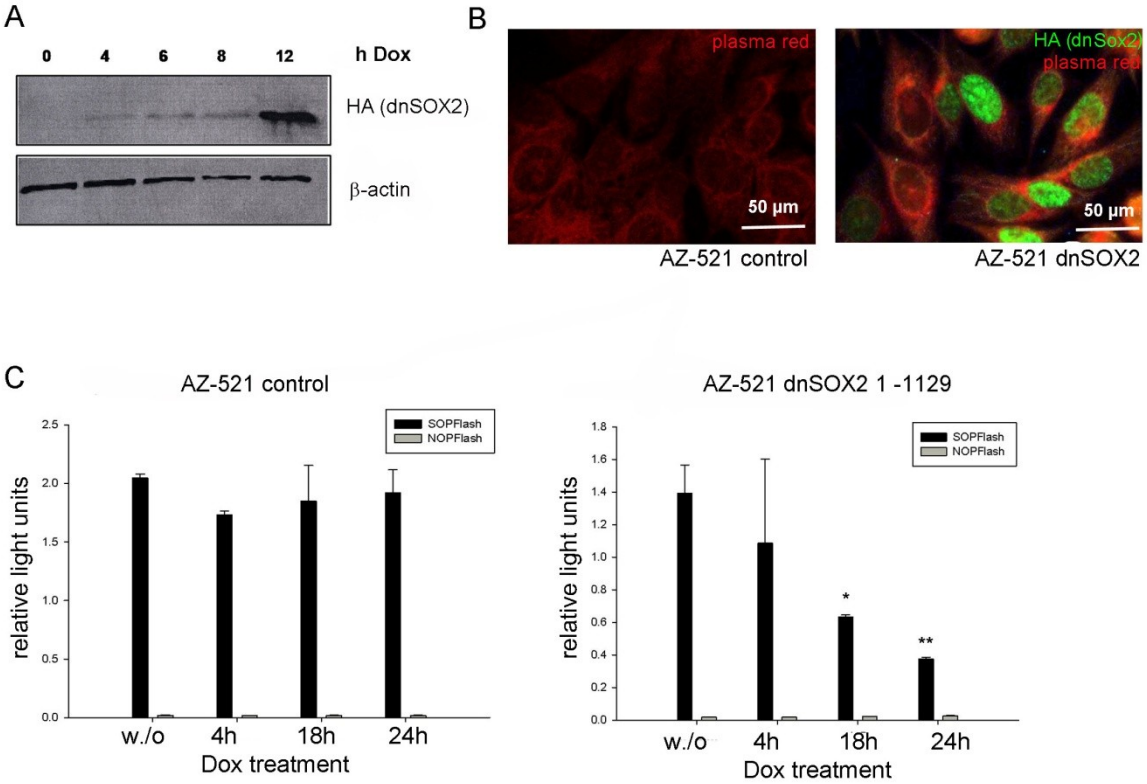


Fig. 18: dnSOX2 expression in doxycycline induced AZ-521 stable cell clones; dnSOX2 is expressed as early as 4 h after induction with doxycycline as shown by western blot against the HA-tag **(A)**. N = 5; Nuclear expression of dnSOX2 12 h after doxycycline treatment detected by immunofluorescence; green = SOX2, red = *deep red plasma membrane stain* **(B)**. N = 5; dnSOX2 expressing cells showed a time dependent decrease in SOX2 activity measured by SOPFlash/NOPFlash assay, whereas control cell line AZ-521 did not show any down-regulation **(C)**. N = 5, n = 10; *$p < 0.05$; ** $p < 0.01$

3.1.2.2. The role of SOX2 in cell proliferation

SOX2 is known to play a crucial role in cell proliferation and cell differentiation in ES and iPS cells. Furthermore, a co-expression of SOX2 and KI67 was seen in gastric tumor samples (see Fig. 13), emphasizing the role of SOX2 in proliferation. Thus, the impact of SOX2 on cell proliferation was analyzed *in vitro*. AZ-521 cells were synchronized by serum starvation and released from cell cycle arrest by addition of 10 %

FCS after 24 h. Proliferation was measured by quantification of ATP after 12 h, 24 h, 48 h, 72 h, and 6 d after transient transfection with *shSOX2* or after induction of dnSOX2 in the stable cell clones. After 12 h (Fig. 19 A) and 24 h (Fig. 19 B) no differences in proliferation of stable AZ-521 dnSOX2 cell clones could be detected compared to parental control cells. Same results were found for cells transient transfected with *SOX2* shRNA. After 48 h a decrease of proliferation was observed in induced stable cell clones (Fig. 19 C) however, this effect could not be seen in cells transfected with *SOX2* shRNA. After 72 h in both, induced stable cell clones (Fig. 19 D) as well as *SOX2* shRNA transfected cells (Fig. 19 F) a significant decrease of proliferation was shown, however, it was more pronounced in the stable cell clones than in *SOX2* shRNA transfected cells. 6 days after induction with doxycycline proliferation levels in cells stably expressing dnSOX2 began to increase again (Fig. 19 E).

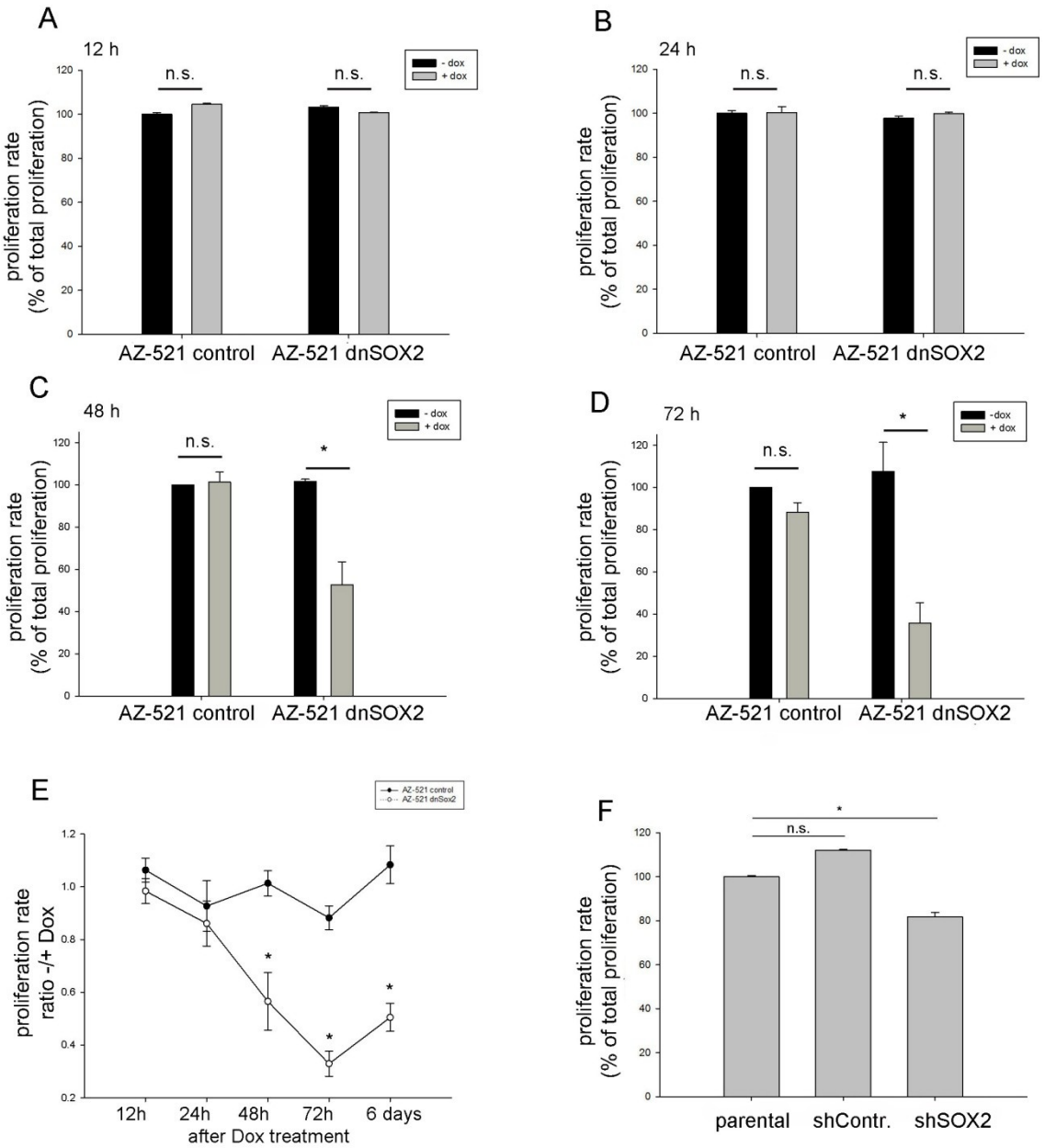


Fig. 19: Proliferation analysis in AZ-521 cells after inhibition of SOX2; 12 h **(A)** and 24 h **(B)** after induction of dnSOX2 no effect of proliferation was observed. After 48 h **(C)** and 72 h **(D)** a significant decrease in cell proliferation could be observed in the stable cell clone compared to the non-induced clone and the parental control cell line. The effect of dnSOX2 in proliferation was most pronounced after 72 h and started to rise again beginning after 6 days **(E)**. N = 4, n = 12; AZ-521 cells transfected with *shSOX2* RNA showed a significant decrease of proliferation only after 72 h **(F)**. N = 3, n = 9; *$p < 0.05$

In addition cell divisions were monitored by staining AZ-521 cells stably expressing dnSOX2 with *Cell Proliferation Dye eFluor 670*. After synchronization, cells were stained and proliferation was measured by FACS analysis 24 h and 48 h after

doxycycline treatment. After 24 h no differences could be seen, comparing stable AZ-521 dnSOX2 cell clones to the parental cell line. However, 48 h (Fig. 20) of dnSOX2 induction resulted in less cell division in the inducible cell clone.

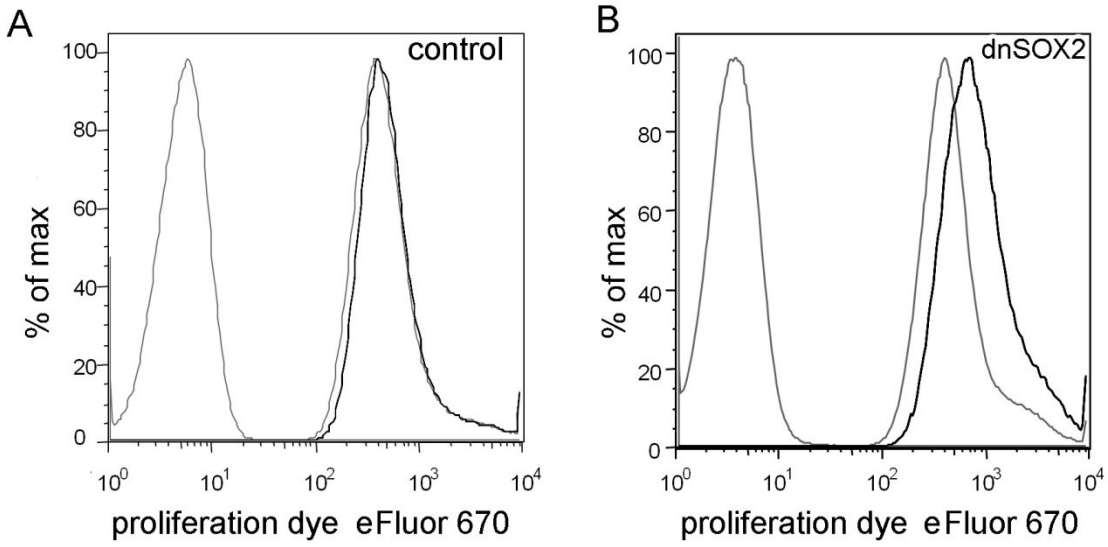


Fig. 20: Cell division in AZ-521 dnSOX2 cells 48 h after doxycycline treatment; AZ-521 control cells do not show differences in proliferation whether treated with doxycycline **(bold line)** or not induced **(A)**. A clear difference could be seen in AZ-521 dnSOX2 cells. Induced cells **(bold line)** proliferate slower compared to the not treated controls **(B)**. The line on the left side of each diagram represents unstained cells. N = 3, n = 6; one representative experiment shown.

3.1.2.3. The role of SOX2 in apoptosis

Dysregulation of apoptosis is an essential mechanism in the development of neoplasms and in cancer progression. Tumor cells are able to avoid apoptosis, leading to an uncontrolled cell replication. To analyze apoptosis of cells after SOX2 blocking, Caspase 3/7 assays were performed, as Caspase 3 is important for the initiation of apoptosis. Prior to experiments cells were serum starved 24 h for synchronization. Stably transfected AZ-521 dnSOX2 cells were analyzed for Caspase3/7 activity after 8 h, 24 h and 48 h of inhibition of SOX2. At 8 h (Fig. 21 A) and 24 h (Fig. 21 B) no significant differences could be observed compared to the non-induced cell clone or the parental cell line. 48 h after doxycycline treatment an increase of apoptosis was observed with 30 % more cells being apoptotic (Fig. 21 C). In cells transfected with *shSOX2*, Caspase 3/7 activity was measured 48 h after transfection. No significant differences could be observed in comparison to control shRNA or the parental cells (Fig. 21 D). This might be due to the exceptional high expression level of SOX2 which could only be marginally

reduced by RNA interference, as also seen in SOX2 transcriptional activity assay and proliferation analysis. As no differences of apoptosis in Caspase 3/7 assay was observed in *shSOX2* transfected cells, Annexin V staining and all following experiments were conducted with cells expressing dnSOX2. Annexin V identifies cells at the early stage of apoptosis, but after induction of Caspase 3. To discriminate apoptotic cells from necrotic cells, they were co-stained with 7AAD. Annexin V staining was performed after 24 h, 48 h and 72 h of induction of dnSOX2 and analyzed by FACS. Concordant with the results of Caspase 3/7 analysis, no differences could be observed after 24 h but an increase of apoptosis was detected after 48 h in the doxycycline treated AZ-521 dnSOX2 cell clone compared to non-treated cells. However, the differences observed were less apparent than seen in the Caspase 3/7 assay (Fig. 21 E). No further increase of apoptosis was seen after 72 h. Differences of results in Caspase 3/7 assay and Annexin V staining might be due to higher sensibility of the Caspase 3/7 assay and more sources of error in Annexin V analysis, as here many factors like cell culture conditions, quality of the binding buffer, time of staining and concentration of 7AAD were highly influencing the outcome of the experiment.

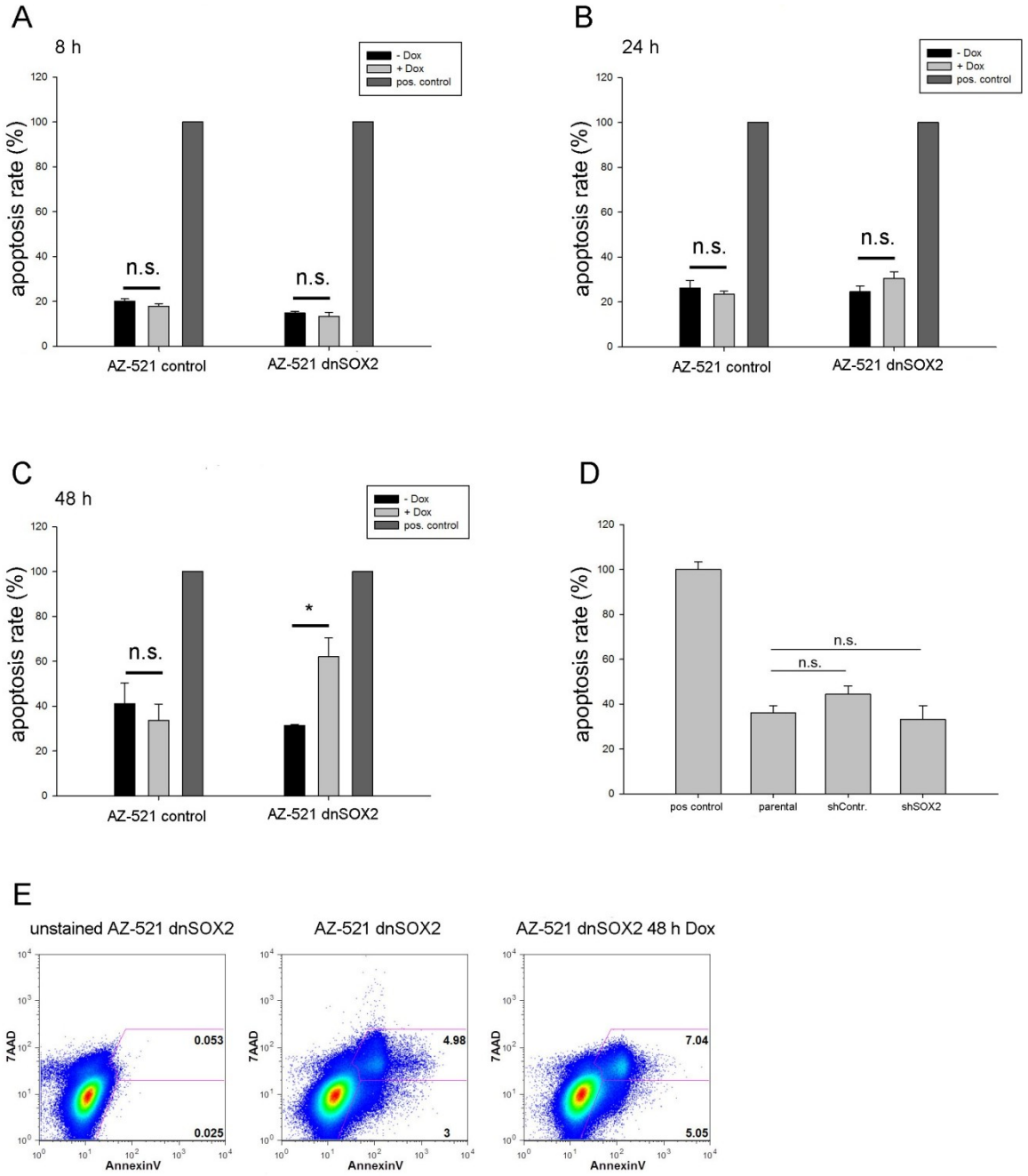


Fig. 21: Analysis of apoptosis via Caspase 3/7 and Annexin V staining after inhibiting SOX2 in AZ-521 cells; Caspase 3/7 activity was analyzed in the inducible stable cell clone. After 8 h **(A)** and 24 h **(B)** no differences could be observed. After 48 h Caspase 3/7 activity was significantly increased in induced AZ-521 dnSOX2 cells compared to non-treated cell clone and the control cell line **(C)**. N = 4, n = 12; Transfection with *shSOX2* did not result in a significant increase of apoptosis **(D)**. N = 4; n = 12; Staurosporine was used as a positive control to induce Caspase 3 activity. Annexin V staining after 48 h showed 5.05 % of cells in the apoptotic phase. In non-treated cell clones only 3 % of cells were apoptotic **(E)**. N = 4; *$p < 0.05$

3.1.2.4. The role of SOX2 in cell migration

Migration is a key process of cancer cells to spread from the initial site of tumor growth into the surrounding tissue. Therefore, the analysis of migratory behavior of tumor cells is essential to define their metastatic potential. With the aim of evaluating differences in migration of AZ-521 tumor cells with or without blocking of SOX2, wound healing assays were performed. A 500 µM wound was introduced into a confluent cell monolayer and AZ-521 parental cells and the cell clone stably expressing dnSOX2 were subsequently treated with doxycycline. Healing of cell wound was observed microscopically 12 h and 24 h after inhibition of SOX2 activity. Compared to non-treated dnSOX2 cells, which achieved 88 % wound closure after 12 h, treated cells showed a wound closure of only 72 %. After 24 h, cells with high SOX2 activity showed a wound closure of 98 % whereas cells with inhibited SOX2 activity only showed 85 % of closure of the wound (Fig. 22). Results indicate that SOX2 plays an essential role in migration and cellular invasion of gastric tumor cells.

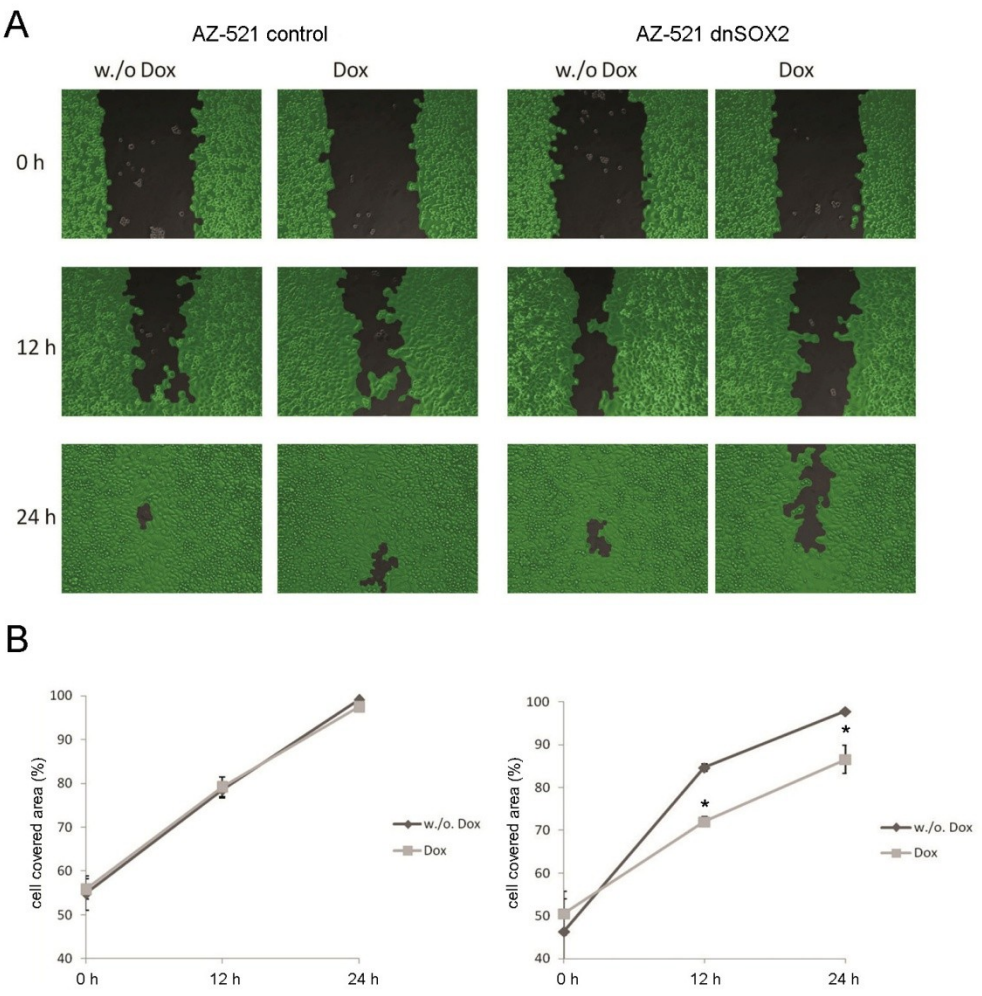


Fig. 22: Migration analysis of AZ-521 cells after inhibition of SOX2 activity; microscopic monitoring of wound healing in AZ-521 control cells and AZ-521 dnSOX2 cell clones after induction with doxycycline **(A)**. Percentage of cell covered area over time in control cell line and the inducible stable AZ-521 dnSOX2 cell clone **(B)**. N = 3, n = 6; *p < 0.05

3.1.2.5. The role of SOX2 in cell cycle regulation

Since a decrease in proliferation as well as slower migration rates were seen after inhibiton of SOX2, which might occure only partly due to apoptosis of the cells, analyzing potential differences in cell cycle was the next step to examine cells for their tumorigenic potential. In order to monitor cells in the different phases of the cell cycle, namely Gap1 (G_1) phase, where cells do not replicate but synthesize new organelles, synthesis (S) phase, in which DNA is synthesized and Gap2/Mitosis (G_2/M) phase, where cells are prepared for mitosis and finally separate into two daughter cells, FACS analysis was performed. Cell cycle profile was analyzed in synchronized cells by flow cytometry using propidium iodide as DNA stain. Analyses were done 24 h and 48 h after

doxycycline induction of the stable dnSOX2 cell clone. No differences of cell cycle profile could be observed after 24 h. This result was concomitant to analysis of proliferation, which also did not reveal any changing of proliferative activity during early time points. A cell cycle arrest could be detected after 48 h in cells where SOX2 activity was blocked through dnSOX2, while distribution of the different cell cycle phases in not induced AZ-521 dnSOX2 stable clones was comparable to the parental cell line. In induced AZ-521 dnSOX2 cell clones, cells accumulated in G_2/M phase of the cell cycle and a concomitant reduction of cell numbers in the S-phase was observed (Fig. 23).

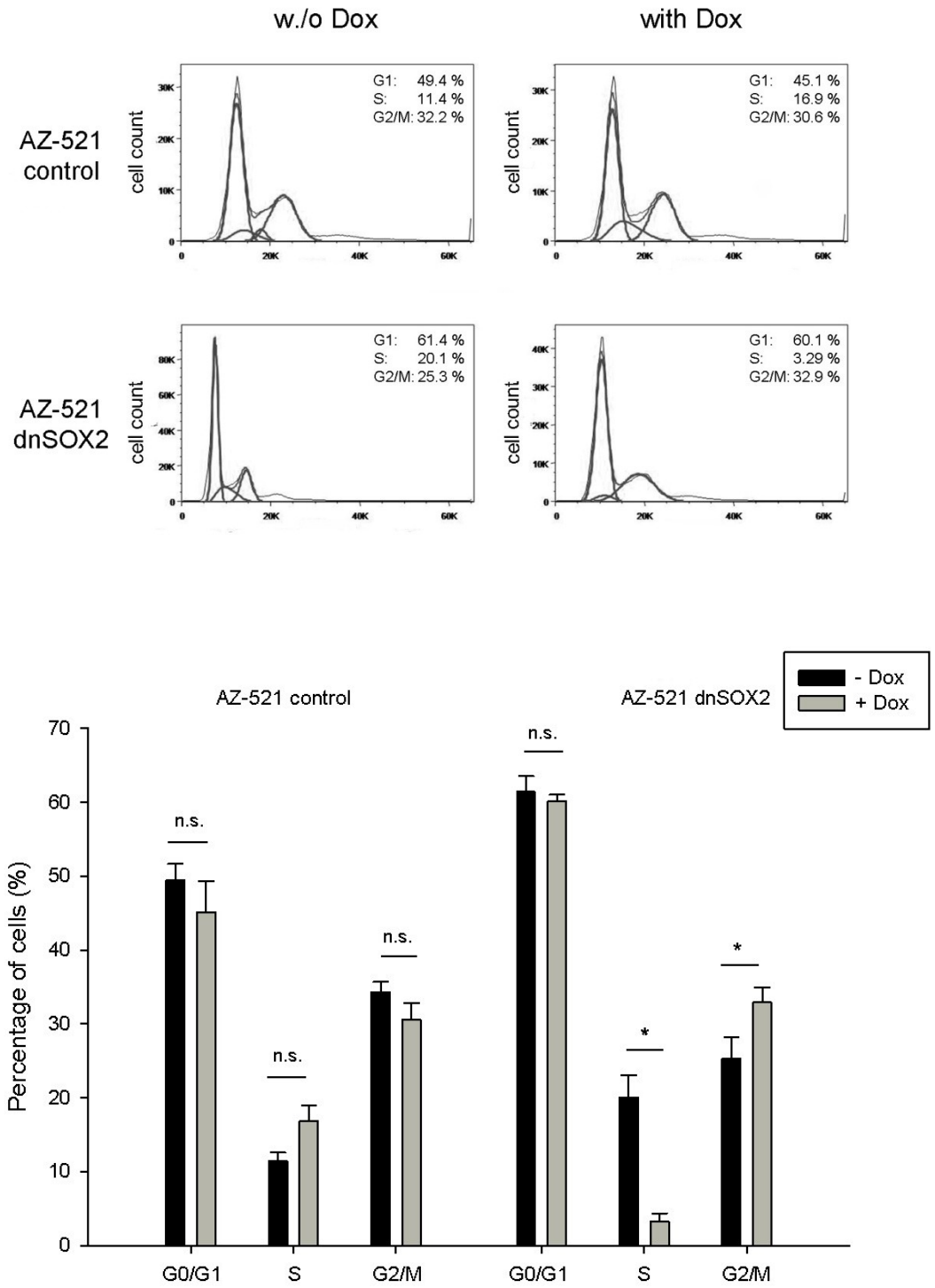


Fig. 23: Distribution of cells in the different cell cycle phases after down-regulating SOX2; after inhibition of SOX2 activity by treatment of stable cell clones with doxycycline for 48 h, fewer cells were detected in S-phase as cell cycle was arrested in G_2/M phase. Upper pictures show one representative experiment; cells were stained with PI and analysed by FACS; N = 3; *$p < 0.05$

Since a cell cycle arrest was observed after induction of dnSOX2, changes in the expression of cell cycle regulators were further analyzed, as cyclins, could be responsible for these changes. First the expression of cyclin B1, cyclin D1 and cyclin D3 was analyzed by quantitative real time PCR. The analysis showed a significant decrease of cyclin B1 during SOX2 inhibition (Fig. 24 A). Furthermore, cyclin D1 mRNA, was significantly up-regulated 12 h upon induction in stable AZ-521 dnSOX2 cell clones with expression returning to basal levels after 48 h (Fig. 24 B), whereas expression levels of cyclin D3 did not show any differences (Fig. 24 C). c-myc is known to be an oncogene and an important factor influencing expression of cyclins, thus, c-myc was also analyzed. However, no significant differences in c-myc mRNA expression were observed (Fig. 24 D).

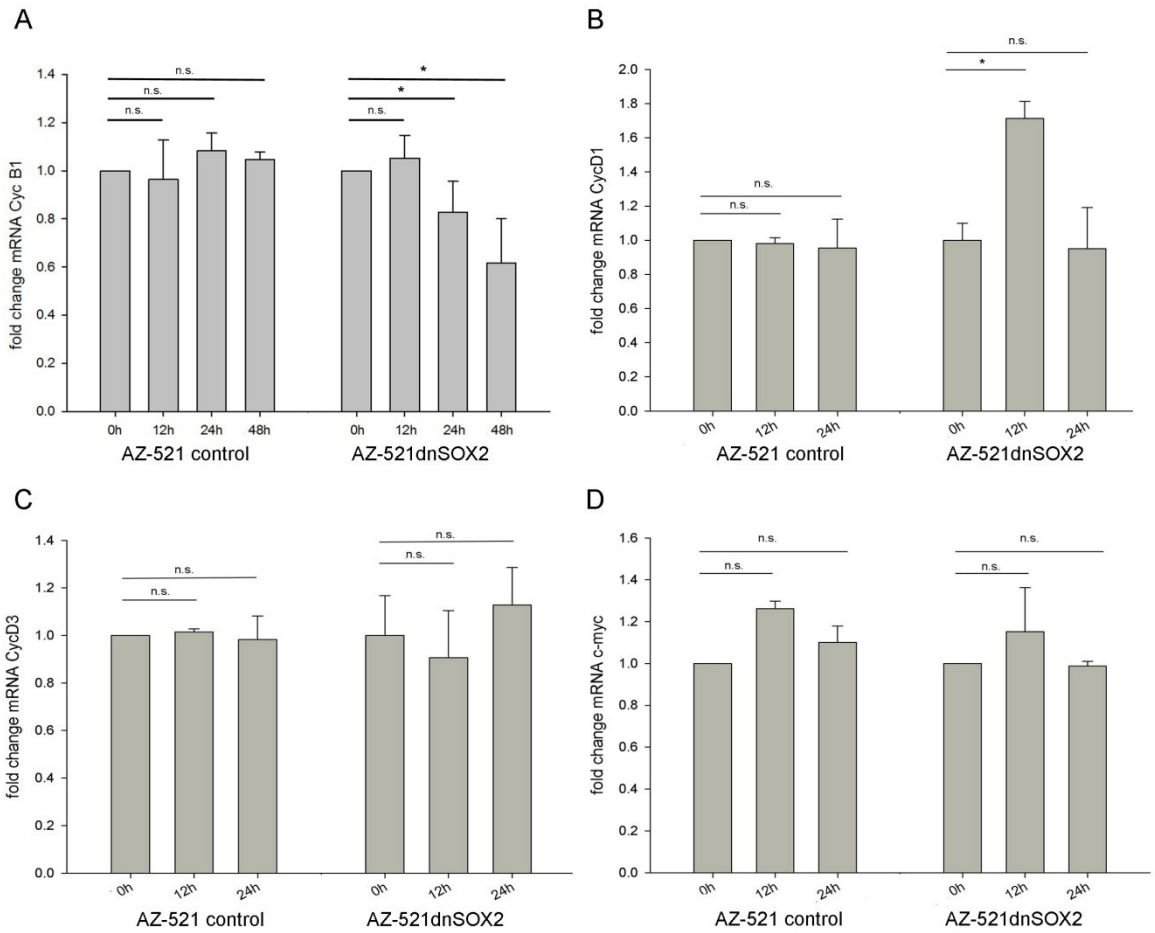
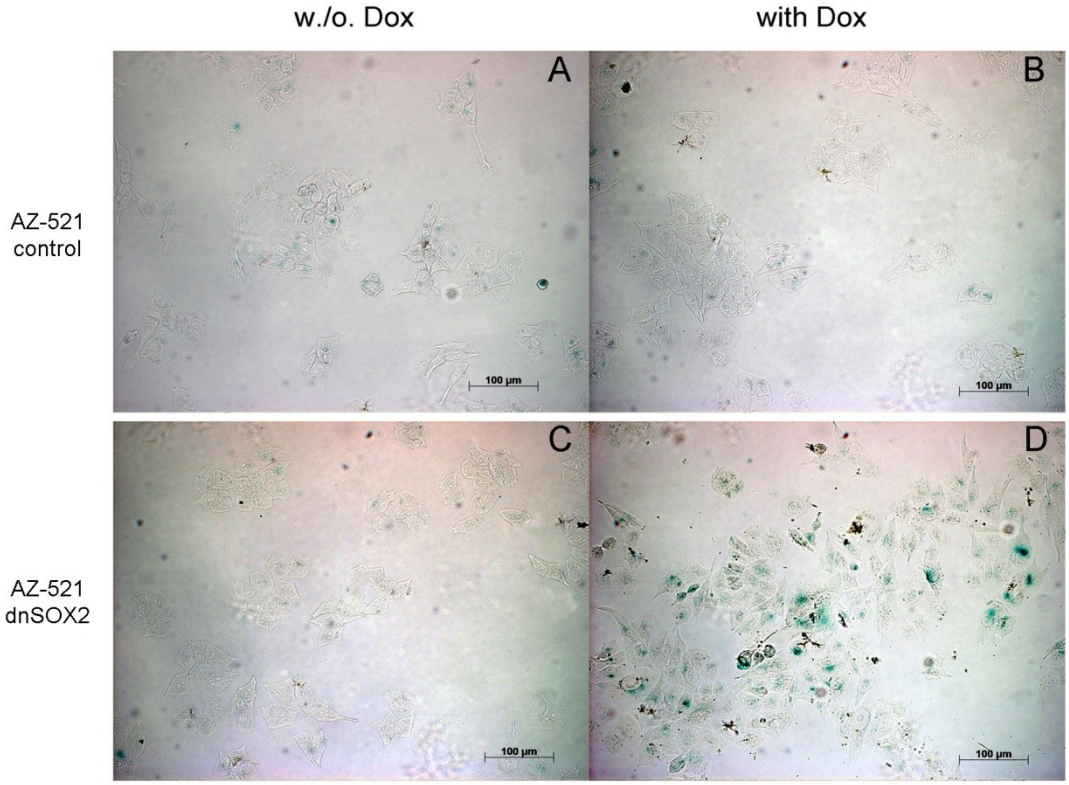


Fig. 24: mRNA expression of different cell cycle regulators; mRNA levels of cyclin B1 decreased over time after SOX2 activity was inhibited **(A)**. Level of cyclin D1 mRNA was increased after 12 h but return to basal levels 24 h after treatment with doxycycline **(B)**. No changes of mRNA levels could be observed in cyclin D3 **(B)** or c-MYC expression **(C)** in the AZ-521 dnSOX2 induced cell clone. N = 4, n = 12; *p < 0.05

3.1.2.6. The role of SOX2 in cellular senescence

Cellular senescence is a molecular mechanism by which cells lose their ability to divide after a limited number of cell divisions *in vitro,* however, remaining metabolic active for a long period of time [184]. Senescence might occur due to changes in gene expression as well as due to DNA damages. Furthermore, it also limits excessive or aberrant cellular proliferation, protecting against the development of cancer. As blocking of SOX2 arrested AZ-521 GC cells, it was sought to analyze if dnSOX2 was also inducing senescence. Senescence associated β-galactosidase (SA- βgal) activity at pH 6.0 and was analyzed in the AZ-521 dnSOX2 stable cell clone and parental control cells after treatment with doxycycline. After 24 h of inhibition of SOX2 activity, approximately 30 % of induced AZ-521 dnSOX2 cells were in a senescent state, in contrast to the non-induced cell clone and the parental cell line, in which the percentage of senescent positive cells did not exceed 5 % (Fig. 25).



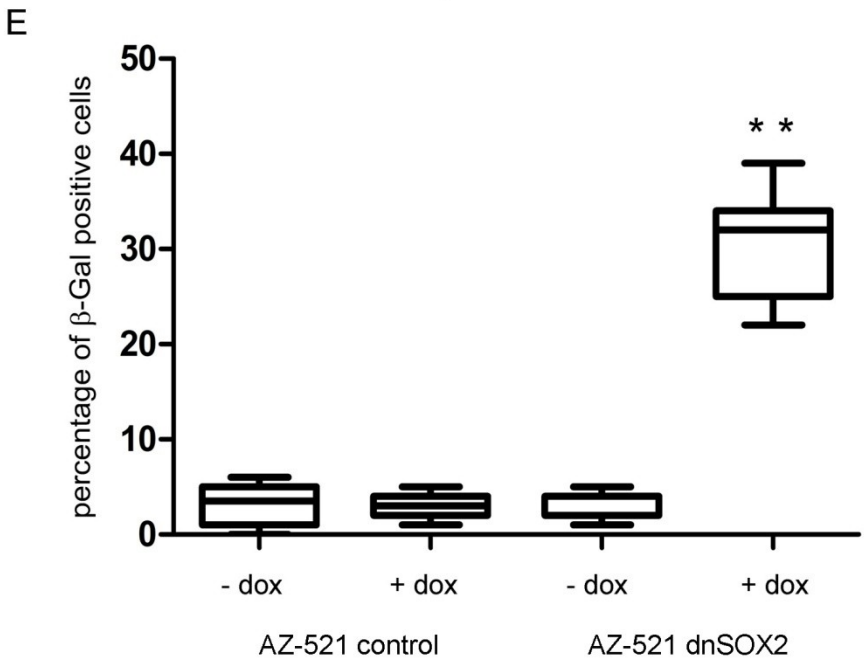


Fig. 25: SA-β-gal activity in AZ-521 cells after SOX2 was blocked; AZ-521 dnSOX2 cell clones were analyzed for cellular senescence. The parental AZ-521 control cell line without **(A)** or with **(B)** doxycycline induction as well as the non-induced cell clone **(C)** did not show SA-β-gal activity, whereas doxycycline treated dnSOX2 cells appeared highly positive for senescence **(D)**. One representative experiment is shown. Cells were analysed in an AxioVert 40 microscope and stained cells were counted per square. Results of seven counted squares are graphically shown as box plot **(E)**. N = 5; **$p < 0.01$

3.1.2.7. The role of SOX2 in tumor growth and metastasis

Having observed a functional role for SOX2 in regulation of proliferation, migration, apoptosis and cell cycle *in vitro*, tumorigenic characteristics of SOX2 expression were subsequently analyzed *in vivo*, regarding tumor growth and migratory behavior of cells. Therefore, AZ-521 dnSOX2 cells were stably transduced with an eGFP-Luc plasmid, allowing monitoring of cell growth via bioluminescence live imaging. NMRI nude mice were injected *s.c.* with the luciferase expressing cell clone of stable AZ-521 dnSOX2 and tumor growth was monitored in an *in vivo* imaging system. Seven days after injection, one group of mice (n = 5) was treated with doxycycline via food uptake. The control group (n = 5) received regular food. Bioluminescence was measured in an *in vivo* imaging system to assess tumor burden every 2 to 4 days. Tumor growth was monitored for 24 days after injection of cells. A significant difference in tumor growth could be observed, comparing treated to untreated mice, beginning 3 days after induction of dnSOX2 expression. This strongly emphasizes the fundamental role of SOX2 in tumor development of GC (Fig. 26).

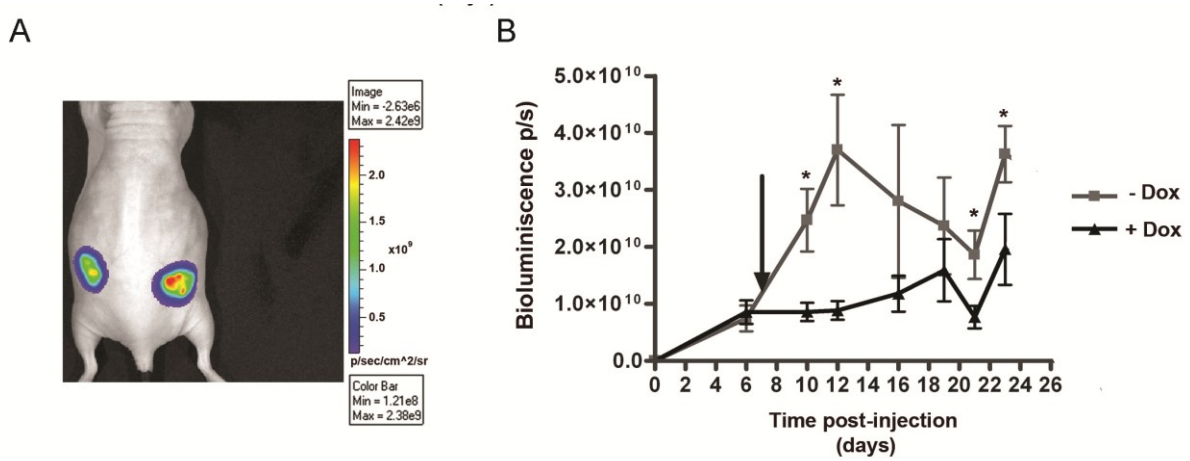


Fig. 26: In vivo analysis of AZ-521 dnSOX2 cells; subcutaneous tumor growth was monitored by bioluminescence **(A)**. Growth was significantly reduced in mice treated with doxycycline **(B)**. Arrow indicates the day of food change. N = 5 per group, *p < 0.01

Tumors were dissected and embedded in paraffin to analyze protein expression of several genes of interest associated to GC by immunohistochemistry. As dnSOX2 induced changes in proliferation, initially the expression of KI67, which is a marker of

proliferating cells, was analyzed. Interestingly no significant differences in KI67 staining could be observed comparing tumors from treated to non-treated mice (Fig. 27). Yet, fewer cells seemed necrotic in tumors from doxycycline treated mice, however, low number of tumor samples did restrict accurate assessment of data (Tab. 7).

Tab. 7: Evaluation of IHC staining of AZ-521 derived s.c. tumors in NMRI nude mice

Block number	KI67	HE
AZ-521 Contr. 1	90 %	Cells pleomorphic, few necrotic
AZ-521 Contr. 2	80 %	Cells pleomorphic, 30 % necrotic
AZ-521 Contr. 3	80 %	Cells pleomorphic, few necrotic
AZ-521 Contr. 4	90 %	Cells pleomorphic, 40 % necrotic
AZ-521 Contr. 5	80 %	Cells pleomorphic, 30 % necrotic
Mean AZ-521 Contr	**84 %**	
AZ-521 Dox. 1	90 %	Cells pleomorphic, few necrotic
AZ-521 Dox. 2	80 %	Cells pleomorphic, 10 % necrotic
AZ-521 Dox. 3	70 %	Cells pleomorphic, 10 % necrotic
AZ-521 Dox. 4	90 %	Cells pleomorphic, 30 % necrotic
AZ-521 Dox. 5	80 %	Cells pleomorphic, few necrotic
mean AZ-521 Dox.	**82 %**	

A panel of genes which were known to be correlated to tumor formation was analyzed in AZ-521 derived tumors in order to find potential targets of SOX2. Staining was done for β-Catenin, Vimentin, P21, P16, CDX2, SNAI2 (slug) and SNAI1 (snail). β-Catenin is an important gene in the WNT signaling pathway. It is accumulating upon WNT activation, leading to the transcription of WNT target genes. Mutation in β-catenin

can cause colorectal cancer. Vimentin is aberrantly expressed in several different cancer tissue, including GCs, where its over expression is associated with poorer outcome and plays an important role in metastasis [185, 186]. P21 is known to prevent cell proliferation and cell growth in various tumors by mediation of the P53 tumor suppressor or by inhibiting the activity of cyclin dependent kinases (CDK) [187]. P16 is, like p21, a tumor suppressor gene inhibiting CDKs [188]. Thus both proteins are adequate markers of neoplastic transformation. CDX2 is a tumor suppressor, which can induce p21 expression and is often down-regulated in colorectal cancer. In intestinal metaplasia expression patterns of CDX2 and SOX2 are inversely related [173]. Aberrant expression of SNAI1 and SNAI2 in human cancers has been correlated with invasive growth potential [189]. IHC staining did not reveal any significant differences in expression of any of those proteins, when comparing tumors from treated to non-treated mice. Nontheless, protein expression was evaluated in both AZ-521 derived subcutaneous tumors. As the tumorigenicity of the cell line is highly dependent on SOX2, a high SOX2 expression was seen in IHC. Furthermore, nuclear β-Catenin expression was observed as well as the expression of vimentin. p21 was only marginally expressed in tumors and seemed to be higher expressed in treated tumors however, differences were not significant. CDX2, P16 as well as SNAI1 and SNAI2 expression was not observed.

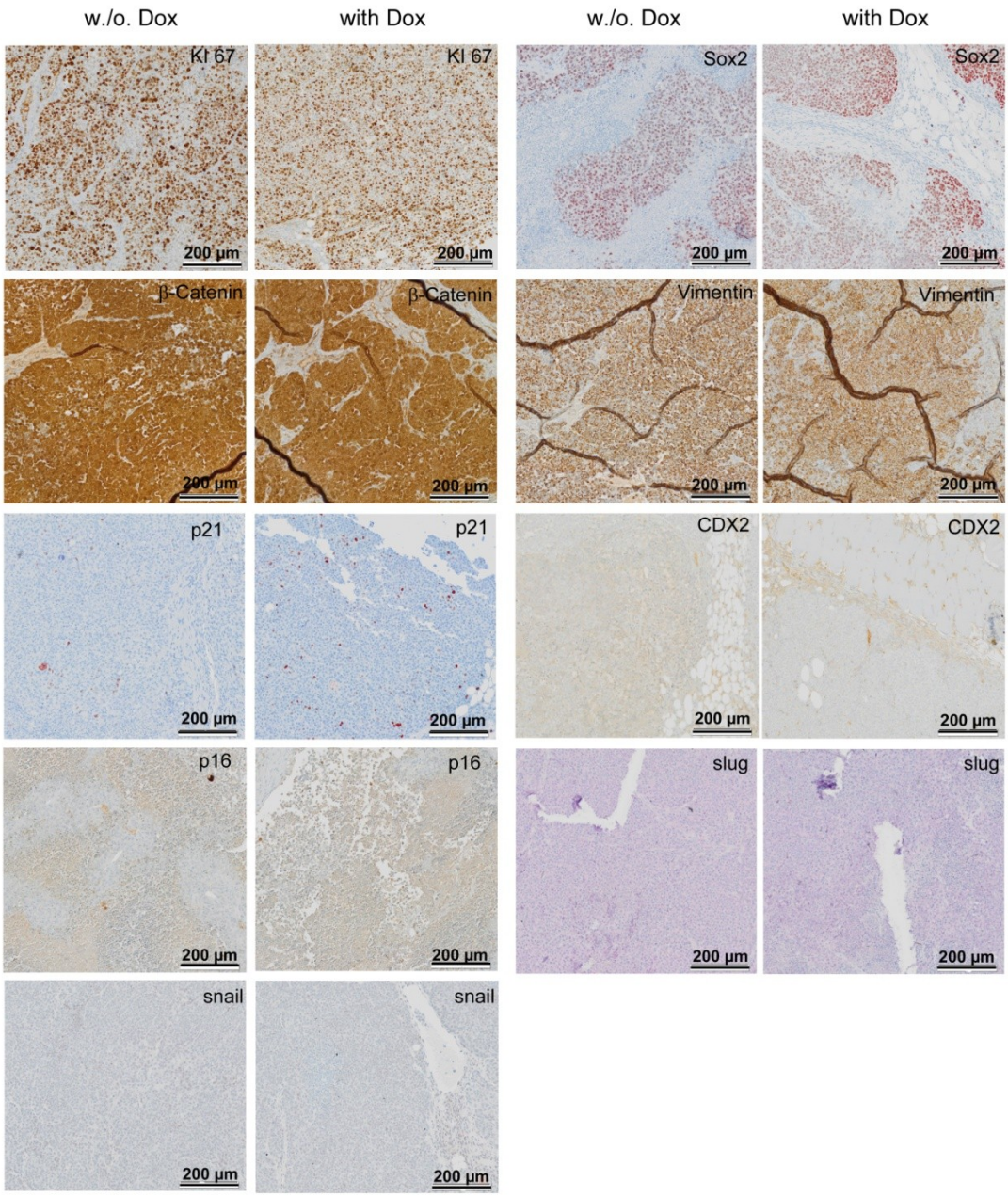


Fig. 27: IHC of AZ-521 derived tumors in NMRI- nude mice; no differences in amount of proliferating cells (KI 67) could be observed when comparing tumors derived from treated to non-treated mice. Nuclear SOX2, β-Catenin and vimentin expression was seen in both tumors. A minor expression of p21 could be observed. No expression of CDX2, P16, slug and snail was seen. N = 2

An assay for experimental pulmonary metastasis *in vivo* was performed in which NMRI nude mice (N = 5) were intravenously injected with 1 x 10^6 AZ-521 cells to monitor metastatic characteristic of the cell line. Cell spreading and tumor growth was monitored after 4 h and 24 h and subsequently twice weekly for one month. A weak signal in the lungs of four mice was observed 4 h after injection, indicating trapping of

the tumor cells in the lung capillaries (Fig. 28 A). The next day no signals could be measured, hence presumably most cells did not survive. One week after injection of cells three mice showed weak signals in the abdomen, but still no signals were observed in the lungs. At day 25 mouse no. 1 and no. 4 showed signals in the lower body part. Mouse no. 3 showed a signal in the snout. No. 2 and no. 5 showed no or background signals (Fig. 28 B). Weight of mice was not changing significantly over time (Fig. 28 C). These data indicate, in agreement with previous observations, that AZ-521 cells are hardly metastatic when injected *i.v.*

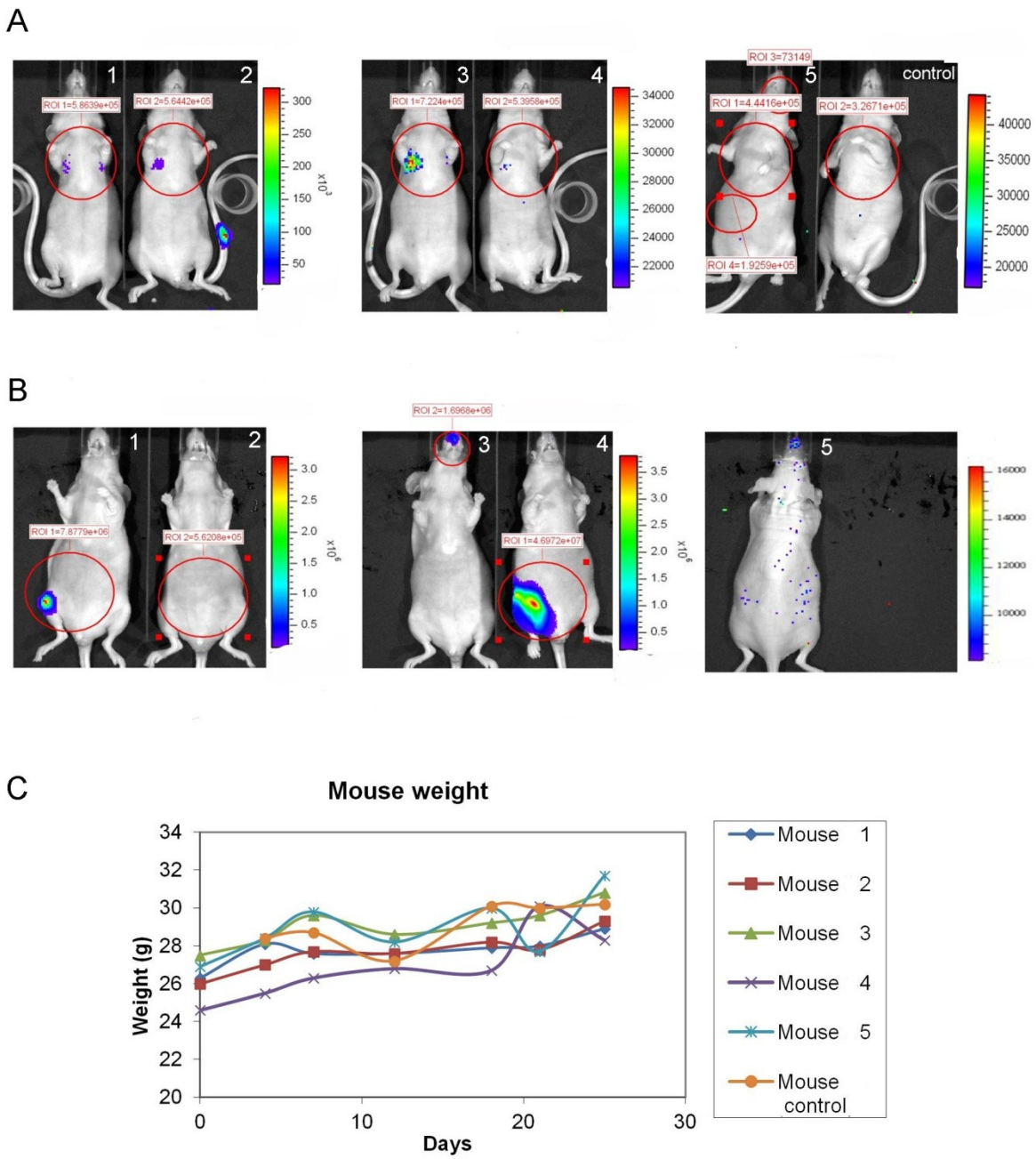


Fig. 28: Monitoring of metastasis of AZ-521 cells in vivo after i.v. injection; 4 h after injection cells could be found in the lungs in 4 of 5 mice **(A)**. After 25 days no lung tumors had grown. Mouse no. 1 and no. 4 showed signals in the abdomen, whereas mouse no. 3 showed tumor cells in the snout region. No signals could be found in mouse no. 2 and no. 5 **(B).** Weight of mice was not significantly changing over time comparing injected mice and control mouse **(C)**. N = 5

3.1.3. Identification of SOX2 target genes

To identify potential target genes and signaling pathways regulated by SOX2 in GC, the gene expression profile in stable dnSOX2 cell clone was analyzed through RNA

microarray analysis and compared to AZ-521 parental cells in collaboration with Dr. Stefan Krebs at the gene center, Munich.

Three independent experiments were performed, in which cells were induced for 8 h, 12 h, 18 h, and 24 h, and RNA expression profile was analyzed together with the corresponding non-induced controls. Since there were no genes differentially expressed (with significance levels set at fdr < 0.05) at the early time points (8 hours and 12 hours) the further analyses were done with the values obtained at later time points.

Quality control of RNA did not show any outliers in their expression data when normalized with robust multi-array average (RMA) algorithm. Hierarchical clustering of the array data showed that induction with dnSOX2 is strongly influencing the expression profile of the cells, but there are only small differences between the two different time points of induction (Fig. 29).

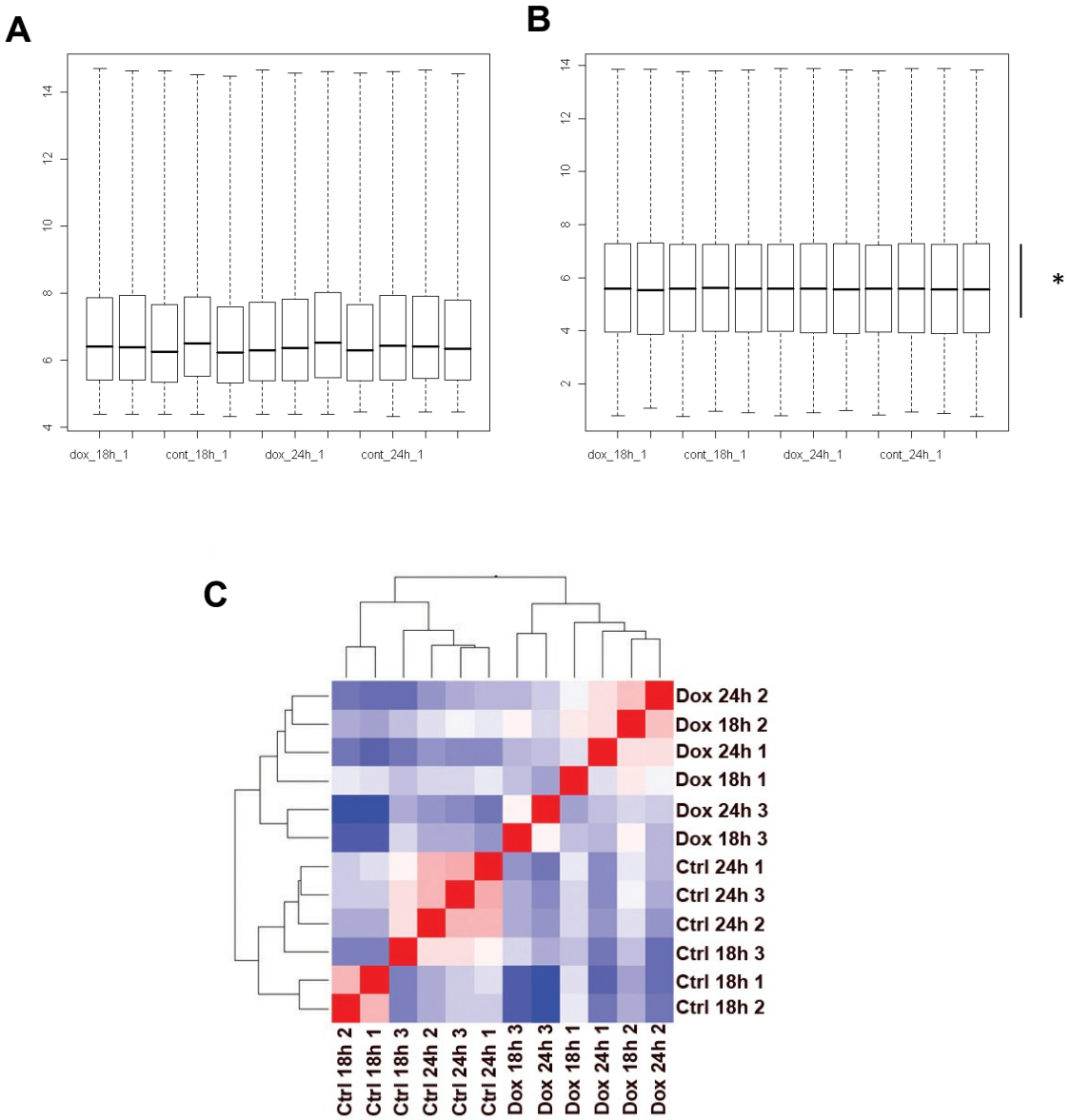


Fig. 29: Quality control of RNA; expression data are shown as box plot before **(A)** and after normalization with robust multi-array average algorithm **(B)**. The dotted line shows the range of signal intensity. 50 % of data are within the frame marked with an asterisk. The bold line indicates the median. Expression profiles were also compared through hierarchic clustering in a heat map **(C)**. Colors of squares correspond to the rank of similarity of expression patterns. They range from dark red (highest similarity) to dark blue (highest differences). The tree structure represents the relation between the expression profiles. Length of branches indicates the grade of difference between the comprised groups. Figures were kindly provided by Dr. Stefan Krebs.

Results were analyzed with LIMMA [190]. Setting the false discovery rate (fdr)-adjusted p-value to 5 %, 578 differentially expressed genes (DEG) could be identified 24 h after induction with doxycycline. However, only a few genes had a fold change of more than two. Cluster analyses were done for the 578 DEGs using the SOTA algorithm as

implemented in the program MeV4.5 (www.tm4.org/mev), which assigned the genes correlating to the progression of their expression at the different time points. The time points of 8 h and 12 h after induction were included in this analysis as it was assumed that genes regulated at later time points are already regulated during these earlier time points. DEGs were grouped into 6 different clusters of up- and down-regulated genes. Cluster 1 showed DEGs which increased their expression continuously during expression of dnSOX2, whereas Cluster 4, 5 and 6 showed all DEGs which decreased during induction with doxycycline. Cluster 2 includes DEGs with a peak at 18 h. DEGs in Cluster 3 also show a peak at 18 h but almost no regulation after 24 h (Fig. 30). A list of all DEGs is included in the appendix.

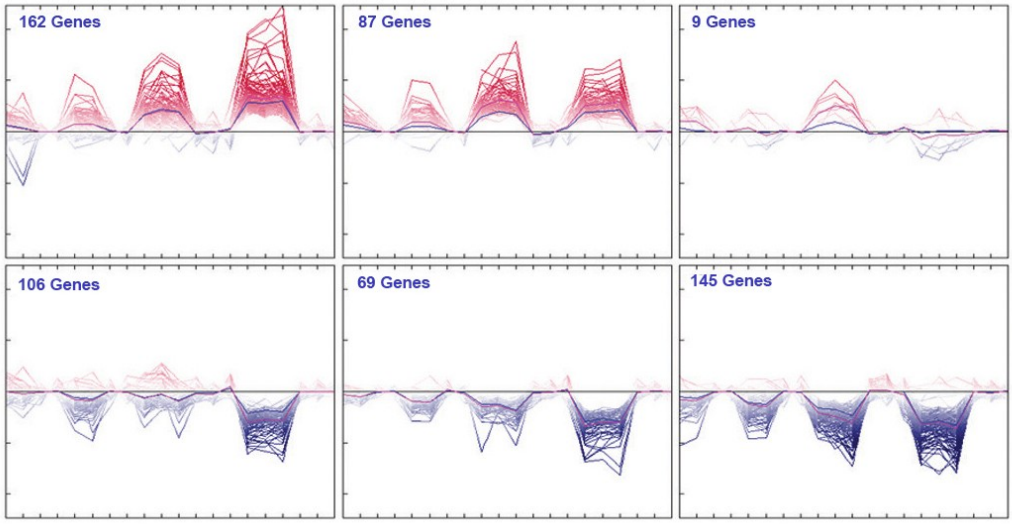


Fig. 30: Cluster analysis of genes; relative alterations of every gene and every sample in comparison to untreated control; Cluster 1 and Cluster 6 show DEG with continuously increasing (Cluster 1) or decreasing (Cluster 6) expression over time. Genes in Cluster 4 and Cluster 5 are also decreasing over time. Genes in Cluster 2 and Cluster 3 show a peak of expression at 18 h with decreasing (Cluster 2) or almost no (Cluster 3) expression of DEG after 24 h. Figure was kindly provided by Dr. Stefan Krebs.

Interestingly, *SOX2* could be found in the list of down-regulated DEGs. It was assumed to find *SOX2* in the list of up-regulated genes, as doxycycline is increasing expression of dnSOX2 which would give a positive signal in mRNA expression analysis. However, probe level plots of *SOX2* after different time points showed a constant reduction of *SOX2* over time, leading to the assumption that the expression of endogenous *SOX2* mRNA was decreased by dnSOX2 expression. This might be due to

disruption of the *SOX2* positive feedback loop. Nevertheless in 3’ region of dnSOX2 the signal of one oligo nucleotide (oligo 7) was continuously increasing. 5’-3’ degradation of dnSOX2 might take place, leaving a residue in 3’ region which is accumulating over time. Accumulation of dnSOX2 might therefore lead to degradation of its own mRNA (negative feedback loop) (Fig. 31).

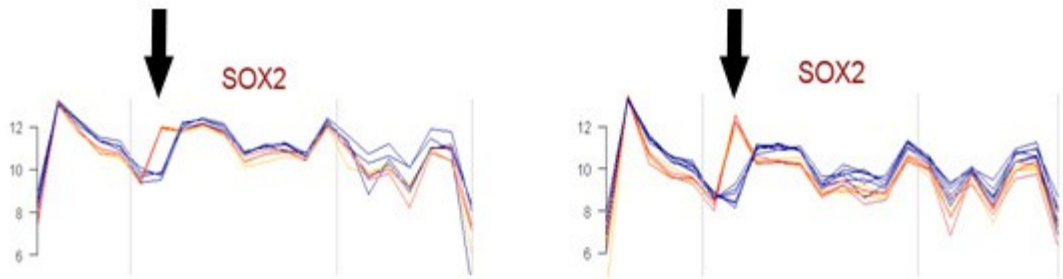


Fig. 31: Probe level plots of SOX2 after different time points of induction; probe level plots showed measured values for every single oligo in the SOX2 transcript. Blue lines represent not induced cells, red and orange lines correspond to values after induction of dnSOX2. The left diagram shows analysis after 8 h and 12 h, the right diagram shows the later time points of 18 h and 24 h. Signals of the SOX2 transcript were continuously decreasing, except one oligo (indicated by arrow). Accumulation of oligo 7 as well as the repression of every other oligo was more pronounced in the later time points. Figure was kindly provided by Dr. Stefan Krebs.

DEGs were analyzed by CoPub [191]. This web-based tool detects keywords which are functionally linked to a set of regulated genes using co-occurrence statistics of genes and keywords in the Medline library. Hints for proceeding differentiation of cells (Cluster 2) and, among others, reduced cell proliferation, cell growth, migration and cell cycle activity (Cluster 6) could be found (Tab. 8). The other clusters did not reveal significant DEGs, functionally linked to any keywords.

Tab. 8: CoPub analysis of DEGs. Biological processes and pathways in Cluster 2 and Cluster 6 which can be linked to the expression of several DEGs; detailed information which gene corresponds to which category can be found in the appendix.

Cluster 2 (up-regulated genes)			
Keyword	**Category**	**p-value**	**Number of genes**
Cell differentiation	biological process	4.09E-03	25
Cluster 6 (down-regulated genes)			
Keyword	**Category**	**p-value**	**Number of genes**
cell growth and-or maintenance, cell growth	biological process	1.15E-03	39
apoptosis	biological process	1.27E-03	49
cell adhesion	biological process	2.01E-03	27
growth	biological process	2.80E-03	61
cell proliferation	biological process	4.38E-03	41
tissue regeneration	biological process	4.38E-03	5
migration	biological process	5.74E-03	23
phosphorylation	biological process	6.87E-03	50
metaplasia	liver pathology	7.76E-03	8
WNT signaling pathway	pathways	2.39E-07	5
motility	biological process	1.1E-07	11
wound healing	biological process	8.45E-03	12

Further analyses focused on DEGs in the LIMMA table with a significance of $p < 0.01$ and with a fold change of 2 or more. 33 genes showed a 2-fold or more change,

among those, 19 genes were down-regulated after inhibition of *SOX2* and 14 genes were found to be up-regulated (Tab. 9).

Tab. 9: DEG with significance of p < 0.01 and a fold change of 2 or more after induction of dnSOX2

Down-regulated genes		Up-regulated genes	
BIRC3	baculoviral IAP repeat-containing 3	*ZNF114*	zinc finger protein 114
XRCC4	X-ray repair complementing defective repair in Chinese hamster cells 4	*APOE*	apolipoprotein E
PCDH18	protocadherin 18	*NELF*	nasal embryonic LHRH factor
FGF10	fibroblast growth factor 10	*SNAPC1*	small nuclear RNA activating complex, polypeptide 1, 43kDa
P2RY5	purinergic receptor P2Y, G-protein coupled, 5	*C18orf19*	chromosome 18 open reading frame 19
COQ3	coenzyme Q3 homolog, methyltransferase (S. cerevisiae)	*SFN*	stratifin
MMP10	matrix metallopeptidase 10 (stromelysin 2)	*CD68*	CD68 molecule
TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)	*GDF15*	growth differentiation factor 15
EGR1	early growth response 1	*LEF1*	lymphoid enhancer-binding factor 1
GOLPH3L	golgi phosphoprotein 3-like	*IFI30*	interferon, gamma-inducible protein 30
ARHGAP24	Rho GTPase activating protein 24	*ELAVL3*	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 3
CYB5R1	cytochrome b5 reductase 1	*DKK4*	dickkopf homolog 4 (Xenopus laevis)
COL3A1	collagen, type III, alpha 1	*IFITM1*	interferon induced transmembrane protein 1 (9-27)
MAMDC2	MAM domain containing 2	*CRABP2*	cellular retinoic acid binding protein 2

COL5A2	collagen, type V, alpha 2		
RBMS3	RNA binding motif, single stranded interacting protein		
CHRDL1	chordin-like 1		
PCOTH	prostate collagen triple helix		
FRMD5	FERM domain containing 5		

Real time qRT-PCR was performed for the most significantly differentially expressed genes listed in table 8, to verify their differential mRNA expression. However, the expression of only very few DEGs could be confirmed by qRT-PCR.

Initially the decrease of *SOX2* expression itself was confirmed by real time mRNA expression analysis (Fig. 32).

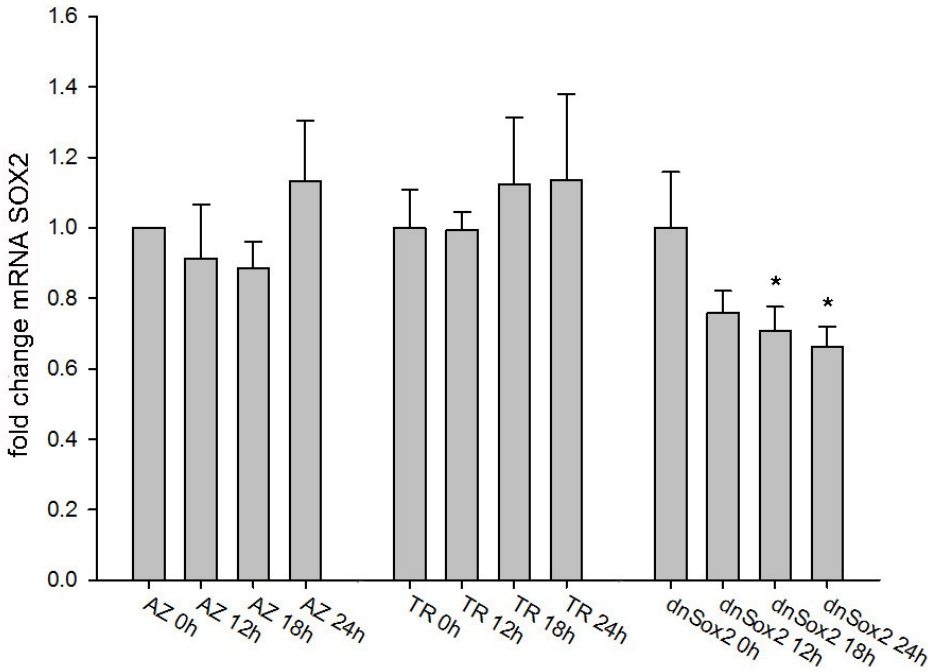


Fig. 32: mRNA levels of SOX2 after inhibition of SOX2 transcriptional activity; AZ-521 parental cell line, an AZ-521 TR control cell clone, only harboring the rtTA plasmid and AZ-521 dnSOX2 inducible cells were analyzed for mRNA expression of *SOX2* via qRT-PCR. When dnSOX2 was expressed and SOX2 transcriptional activity was inhibited, mRNA levels of *SOX2* decreased over time. N = 4, n = 12; *$p < 0.05$

RNA array and real time PCR also showed a significant increase of *LEF1* (lymphoid enhancer-binding factor 1) expression when SOX2 was inhibited in AZ-521 cells (Fig. 33 A). *LEF1* is a well known downstream effector of the canonical WNT signaling pathway. WNT signaling leads to activation of LEF1/TCF-mediated transcription and SOX2 is known to antagonize WNT signaling. To support these results, TOPFlash assays in AZ-521 dnSOX2 cells treated with doxycycline were performed, to detect TCF4 transcriptional activity. A significant up-regulation of TCF4 transcriptional activity was seen, when SOX2 was down-regulated (Fig. 33 B). Another gene found to be up-regulated in the array, *DKK4* was confirmed to be up-regulated in by qRT-PCR. The DKK4 protein is also known to be a TCF/LEF target gene (Fig. 33 C). Another interesting candidate of genes was *FGF10*, since it is also involved in WNT signaling. However, during development of stomach and in several other tissues *FGF10* is known to antagonize *SOX2* and down-regulate its expression. Nonetheless, a down-regulation of *FGF10* expression was verified after inhibition of SOX2 in AZ-521 cells in qRT-PCR analysis (Fig. 33 D).

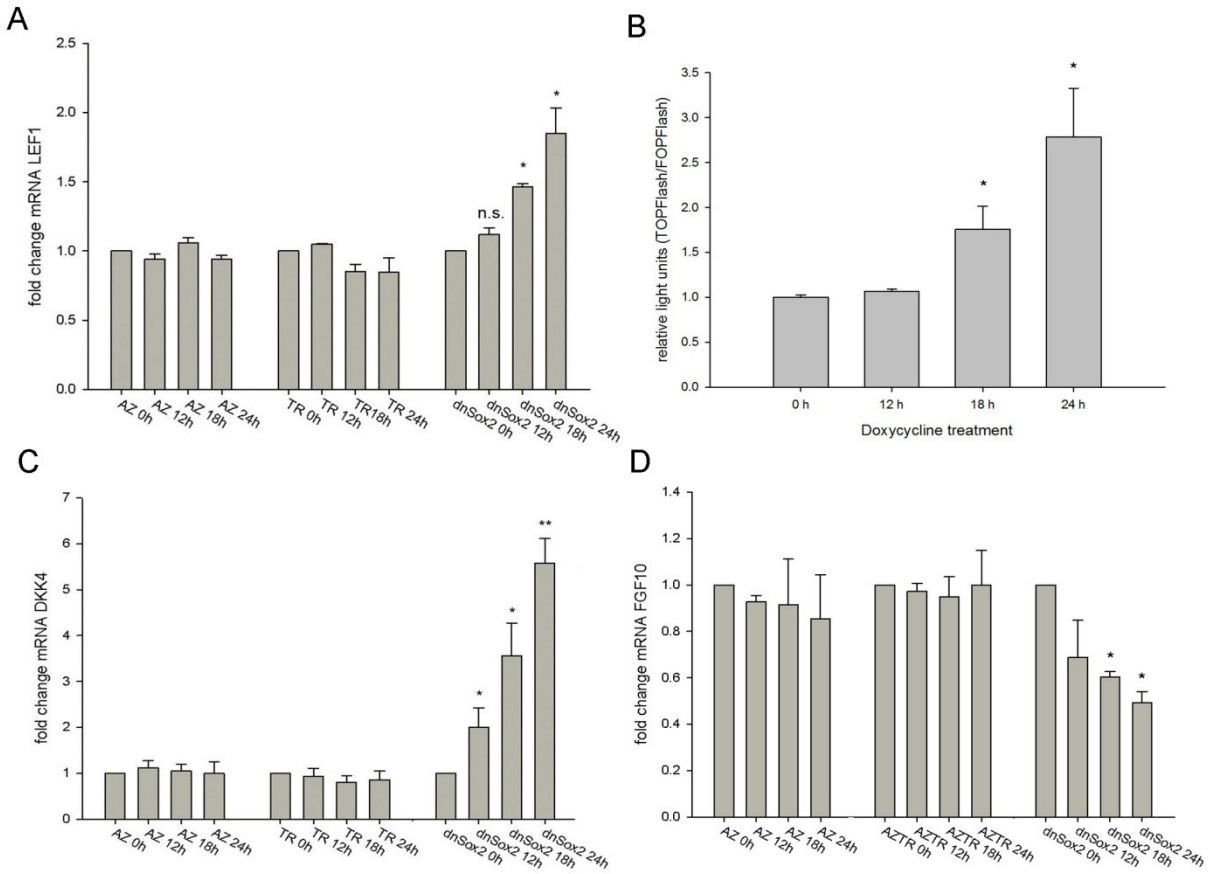


Fig. 33: Analysis of mRNA of WNT target genes and TCF4 transcriptional activity in cells with decreased SOX2 transcriptional activity; mRNA levels of *LEF1* **(A)** and *DKK4* **(C)** were increasing over time, as found in

the array. Concordant to this data, TCF4 transcriptional activity was increasing over time when SOX2 was inhibited in AZ-521 dnSOX2 cell clones **(B)**. *FGF10* could be proven to be down-regulated **(D)**; N = 5, n = 15; *p < 0.05; ** p < 0.01

Furthermore, some genes which did not show a fold change of 2 or more were also analyzed as they were assumed to be important in SOX2 regulatory pathways. From the genes that were shown to be down-regulated by SOX2, TGFβ receptor associated protein 1 (*TGFb RAP1*), *TP63*, and *CTNND1* (*δ-Catenin)* were analyzed. TGFβ RAP1 binds to the TGFß receptor and plays a role in TGFβ signaling. δ-Catenin over expression was recently associated with tumor invasion in small-cell lung cancer and glioma [192, 193]. *TGFb RAP1* (Fig. 34 A) and *CTNND1* (Fig. 34 B) did not show significantly down-regulation of mRNA, but down-regulation of the ΔN variant of *P63* (Fig. 34 C) was confirmed. P63 down-regulation after repression of SOX2 activity could also be observed by western blot (Fig. 34 D). Tumor protein P63 belongs to the *P53* gene family and encodes for two main isoforms, namely TAP63 and ΔNP63. The latter is known to be involved in adult stem cell and progenitor regulation. A differential expression of the *TAP63* isoform was not observed in qRT-PCR.

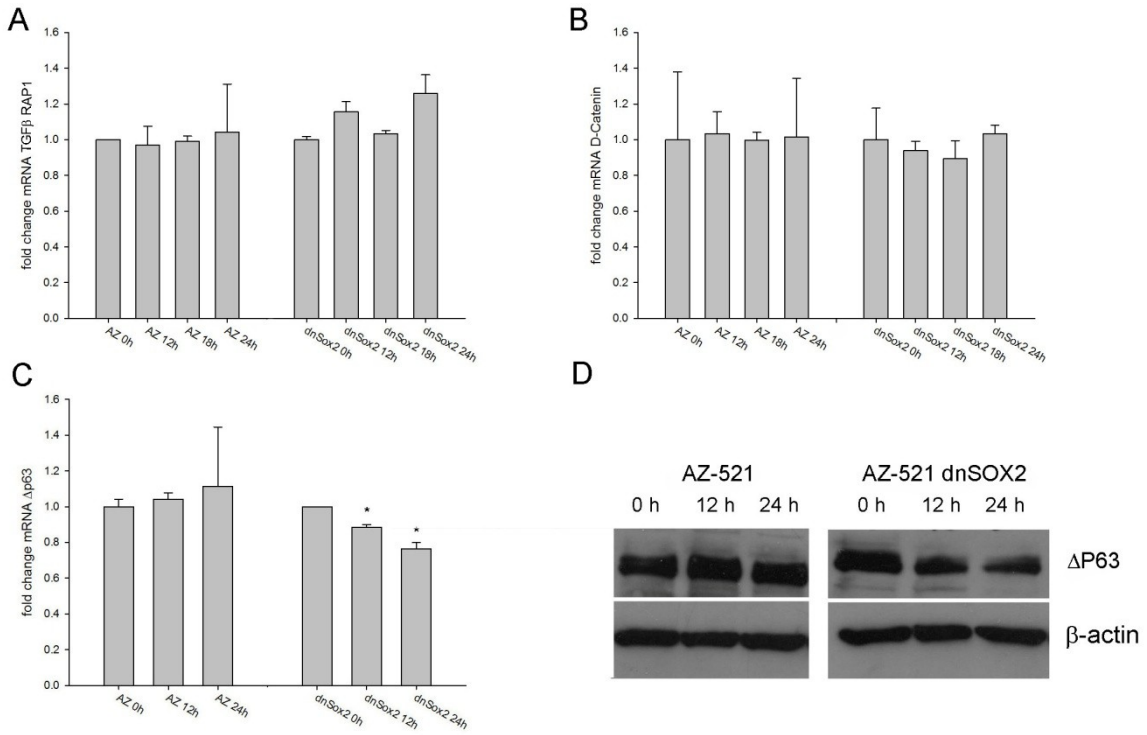


Fig. 34: mRNA expression of genes down-regulated after of SOX2 inhibition; *TGFb RAP1* **(A)** as well as *CTNND1* (δ-Catenin) **(B)** did not show any significant changes in mRNA. The transcription variant *ΔNP63* of

P63 was significantly down-regulated, as seen in qRT-PCR **(C)** and Western blot analysis **(D).** N = 3, n = 9; *p < 0.05

From the genes up-regulated after SOX2 inhibition, *SOX18*, *NOTCH1* and cyclin dependent kinase inhibitor 1 (*p21, CIP1*) were tested in qRT-PCR. SOX18 is, like SOX2, a transcription factor, involved in the regulation of embryonic development, especially in vascular epithelium formation and the generation of hair follicles [194]. NOTCH1 is a transmembrane receptor which functions as a tumor suppressor in skin through suppression of the WNT pathway [195] and is also known to have growth suppressive function in hematopoetic cells, pancreatic epithelium and hepatocytes [196]. *SOX18* (Fig. 35 A) as well as *NOTCH1* (Fig. 35 B) did not show significant differential expression of mRNA. Nevertheless, a significant up-regulation of *p21* (*CIP1*) was observed by qRT-PCR (Fig. 35 C) as well as western blot (Fig. 35 D). p21 is commonly regulated by P53 but can also be expressed without being induced by P53, as a regulation of *P53* tumor suppressor gene was not observed. p21 prevents cell proliferation and reduction of cell growth promoted by p21 expression can lead to cell differentiation.

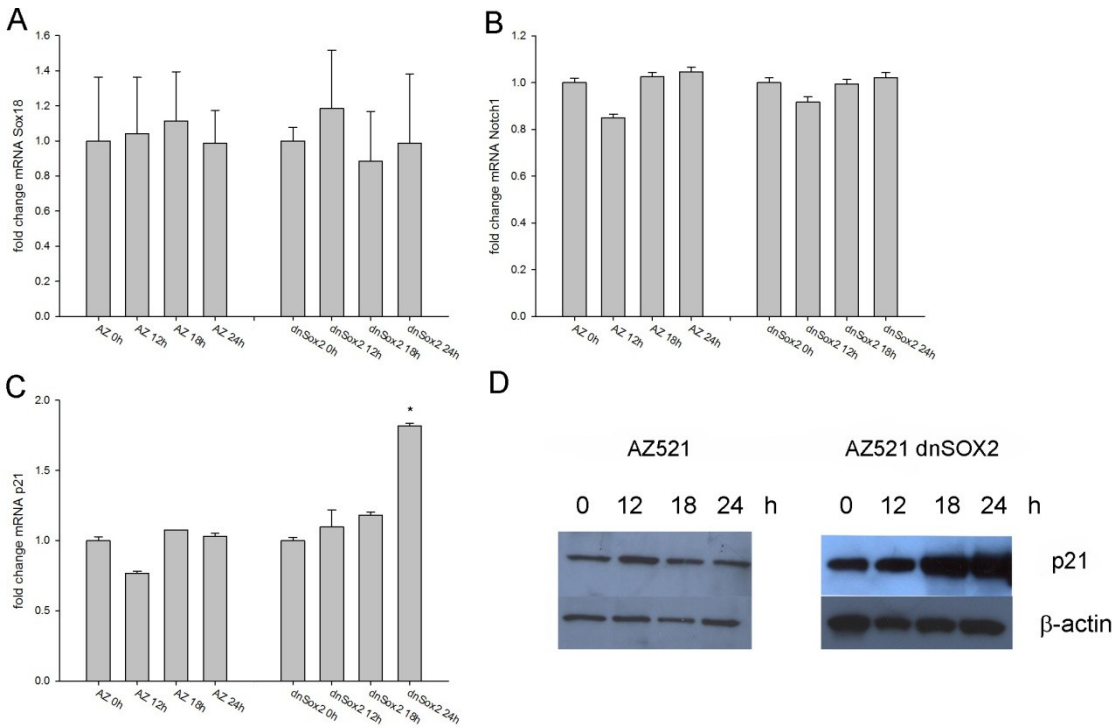


Fig. 35: mRNA expression of up-regulated genes after down-regulation of SOX2; *SOX18* **(A)** and *NOTCH1* **(B)** did not show up-regulation of mRNA in qRT-PCR; *p21 (CIP1)* mRNA was up-regulated 24 h after doxycycline induction of the AZ-521 dnSOX2 cell clone **(C)**. N = 3, n = 9; an up-regulation of P21 protein was also observed by western blot **(D)**. N = 3; *p < 0.05

3.1.4. Regulation of SOX2 in gastric cancer

Moreover, it was not only important to identify target genes of SOX2 but also to elucidate how SOX2 expression can be modulated in GC.

3.1.4.1. The regulation of SOX2 by STAT3

During neuronal development SOX2 has been described to be regulated by STAT3 [141]. Since *STAT3* is also a key gene in promoting oncogenesis, specifically in gastric neoplasia, and is involved in all aspects of carcinogenesis, including cell proliferation, apoptosis, angiogenesis, invasion, migration, and disruption of immune surveillance it was interesting to determine if STAT3 regulated SOX2.

- *The expression of SOX2 and STAT3 in gastric tumors of Gp130 mutant mice*

First, possible co-expression of SOX2 and STAT3 was analyzed in a *Gp130* mutant mouse model, developing spontaneous gastric tumors at 4 weeks of age due to STAT3 hyper activation as a consequence of a knock-in mutation in the GP130 receptor subunit. *Gp130* heterozygous mice were sacrificed at the age of 80 days and expression of SOX2, p-STAT3 and KI67 was detected in paraffin embedded stomachs by IHC. High levels of SOX2 were observed in the tumors that correlated with p-STAT3 expression and proliferation sites (Fig. 36).

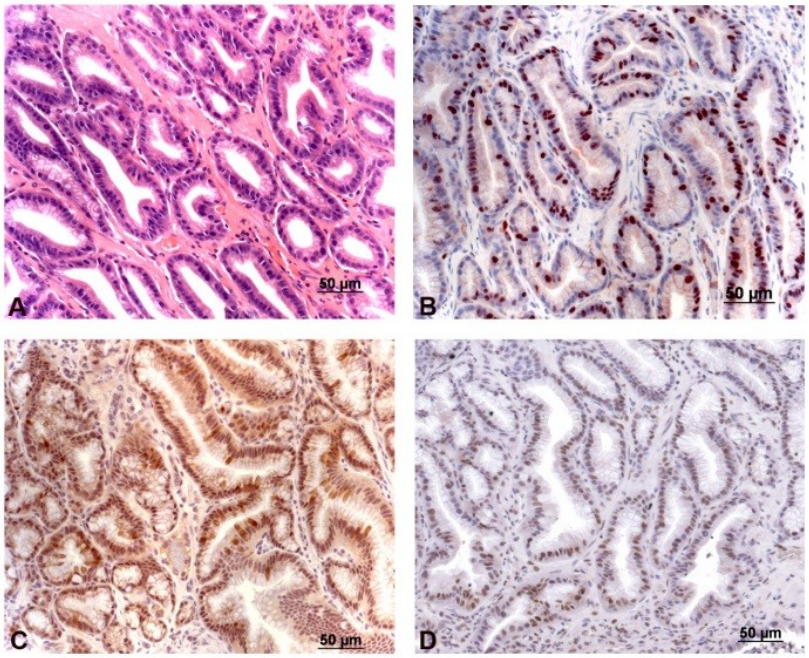


Fig. 36: Tumor sections of Gp130 mutant mice; HE staining (**A**) and expression of Ki67 (**B**), SOX2 (**C**) and p-STAT3 (**D**) in gastric tumors from *Gp130* mutant mice. Co-expression of SOX2 and p-STAT3 was observed, which also correlated with proliferation sites (Ki67).

- *Co-expression and co-localization of SOX2 and STAT3 in GC cells*

SOX2 and p-STAT3 protein levels were detected by western blot in GC cell lines (Fig. 37). SOX2 and p-STAT3 expression correlated in all cell lines, except in AZ-521 cells, which presented high SOX2 levels but not activated STAT3.

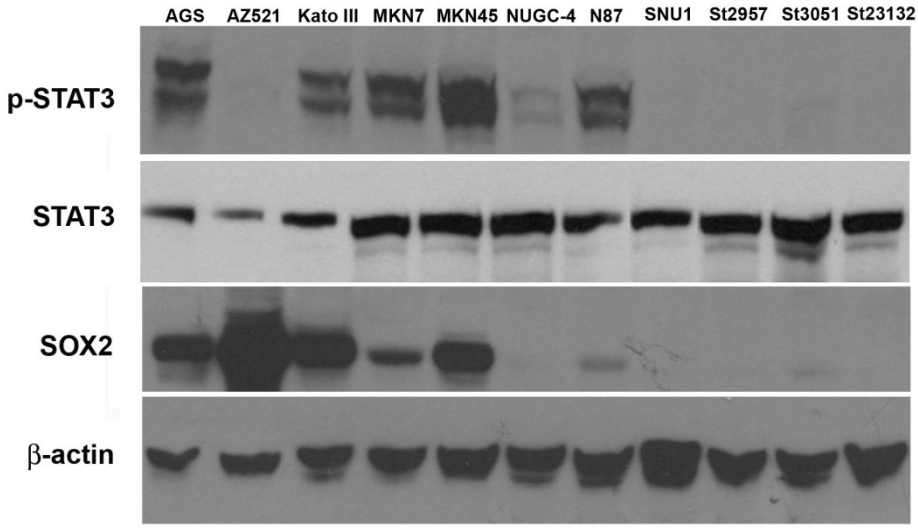


Fig. 37: p-STAT3, STAT3 (79-86 kDa) and SOX2 (35 kDa) protein expression in GC cell lines; β-actin (45 kDa) was used as a loading control. N = 3

As AGS and MKN45 cell lines showed highest expression levels of SOX2 and p-STAT3 they were used for further experiments. First, co-localization analysis of both proteins was performed by immunofluorescence. In both cell lines SOX2 and p-STAT3 co-localized (Fig. 38).

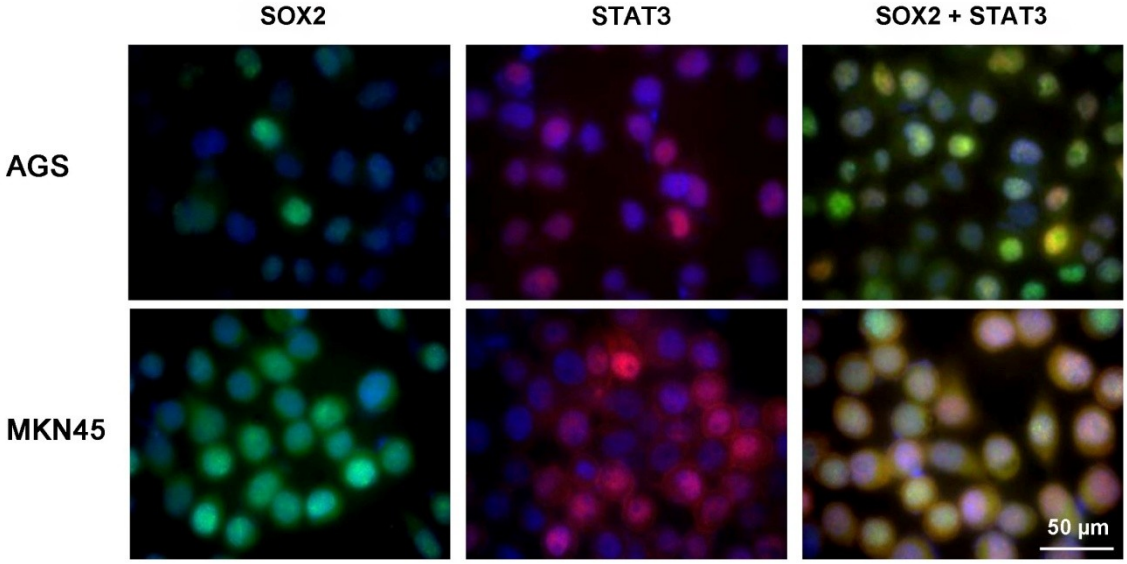


Fig. 38: Co-localization of SOX2 and STAT3 in AGS and MKN45 GC cells; AGS and MKN45 cells expressed high levels of SOX2 and STAT3. Expression of both transcription factors was observed in the nucleus of the cells. Nuclei were stained with DAPI. Co-localization was seen in overlay as yellow colour. N = 3

- *Regulation of SOX2 by STAT3 in vitro*

To determine whether STAT3 can modulate the transcriptional activity of SOX2, AGS and MKN45 cells were co-transfected with 100 ng of SOX2 reporter (SOPFlash) and different concentrations of three different STAT3 constructs: STAT3 wild type, STAT3 constitutively active (cSTAT3) and the dominant negative splice variant βSTAT3. A significant increase in SOX2 transcriptional activity was observed in both cell lines transfected with 250 ng and 500 ng STAT3 wild type, and was enhanced when transfecting the cells with the constitutively active STAT3 plasmid. In AGS cells even 100 ng of cSTAT3 led to a significant increase of SOX2 transcriptional activity. In contrast, transfection with βSTAT3 did not induce significant changes (Fig. 39).

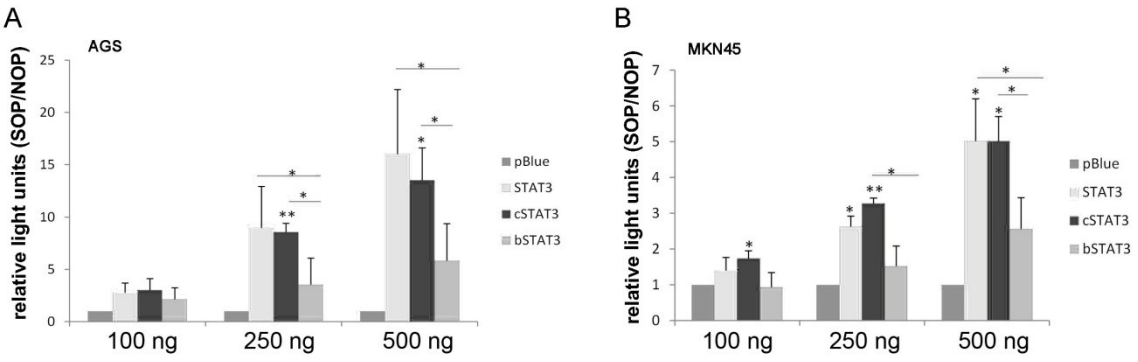


Fig. 39: SOX2 levels after STAT3 transfection; STAT3 induced SOX2 transcriptional activity in AGS **(A)** and MKN45 **(B)** cells in a dose dependent manner. This effect was higher when using the constitutive active STAT3 construct. No significant changes were observed after transfection with βSTAT3. N = 3, n = 6; * $p < 0.05$, ** $p < 0.01$.

To analyze if the IL-6/STAT3 pathway regulates SOX2 expression *in vitro*, GC cells were treated with IL-6 to activate the STAT3 pathway. Subsequently *SOX2* mRNA levels were analyzed. After treating the cells with 5 ng/ml IL-6 for 24 h an increase in *SOX2* mRNA was observed in AGS, NUGC-4, MKN7 and N87 cells (Fig. 40).

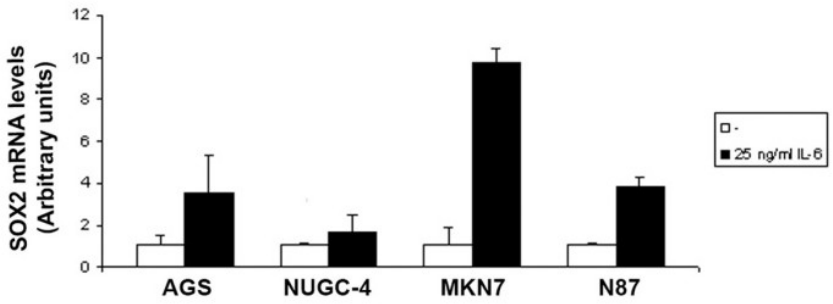


Fig. 40: SOX2 mRNA amounts after IL-6 treatment of GC cells; AGS, NUGC-4, MKN-7 and N87 cells were treated for 24 h with 25 ng/ml IL-6. All cell lines showed a significant increase of *SOX2* mRNA. Results from one representative experiment are shown.

These results show that there is a co-expression of SOX2 and STAT3 *in vivo* and *in vitro* and that STAT3 can regulate the expression of SOX2 via IL-6 in GC cells. However, further investigation should be done to elucidate the mechanisms by which STAT3 regulates SOX2 in GC.

3.1.4.2. The influence of H. pylori infection on SOX2 expression in GC cells

Since another possible mechanism regulating SOX2 in GC is the inflammatory response associated to *H. pylori*, the direct effect of *H. pylori* infection on SOX2 expression was analyzed in GC cells. AZ-521 cells, presenting high SOX2 activity, AGS cells, showing intermediate activity and St2957 and St3051 cells with no SOX2 activity were co-incubated with the *H. pylori* strain G27 or its isogenic CagA deficient mutant at a MOI 50. 24 h after infection, SOX2 transcriptional activity was measured by SOPFlash/NOPFlash assay. AZ-521 cancer cells showed a decrease of SOX2 activity when treated with *H. pylori*. However, this decrease was independent from CagA. AGS cells showed a significant down-regulation of SOX2 activity after *H. pylori* wt infection, whereas no changes were induced with the CagA deficient strain. In St2957 and St3051 *H. pylori* infection did not have any effect on SOX2 transcriptional activity as both cell lines do not present basal SOX2 expression. To induce SOX2 transcriptional activity, these cells were stimulated with IL-4. A strong increase of SOX2 activity was observed after IL-4 treatment, which was blocked by *H. pylori* infection. In the absence of CagA, the inhibition of SOX2 transcriptional activity after IL-4 stimulation was significantly lower, suggesting that CagA might be an important factor in the regulation of SOX2 in gastric tumor cells (Fig. 41).

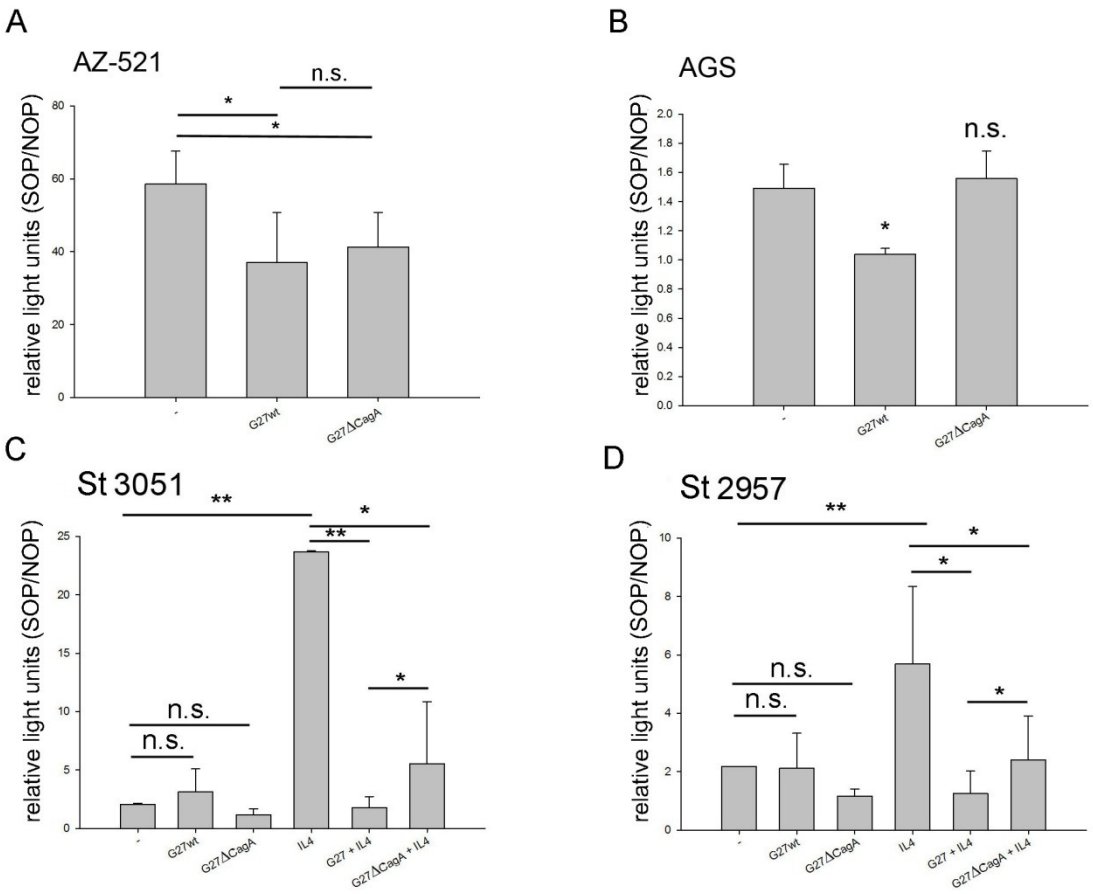


Fig. 41: The influence of H. pylori on SOX2 transcriptional activity in GC cell; AZ-521 cells showed a significant decrease in SOX2 transcriptional activity when infected with *H. pylori*. However, this effect was independent of CagA **(A)**. *H. pylori* down-regulated SOX2 transcriptional activity in AGS cells, and this effect was dependent on CagA expression **(B)**. St3051 **(C)** and St2957 **(D)**, which did not show basal SOX2 transcriptional activity, were stimulated with IL-4. IL-4 significantly up-regulated SOX2 activity and this effect was reversed by *H. pylori.* SOX2 activity was measured as SOP/NOP ratio. N = 3, n = 6; *$p < 0.05$; ** $p < 0.01$

Since reporter assays showed less SOX2 activity after infection with *H. pylori,* changes in SOX2 protein expression levels were analyzed. As SOX2 was seen to be regulated by STAT3, additionally STAT3 activation was also analyzed in parallel. AGS cells showed a reduction of SOX2 expression, correlating with SOPFlash assay results. Concomitant to SOX2 decrease, a decrease in p-STAT3 protein levels was observed, beginning 6 h after *H. pylori* infection and being most pronounced after 24 h (Fig. 42 A). Similar results were found in MKN45 cells 24 h and 48 h after treatment with *H. pylori* (Fig. 42 B).

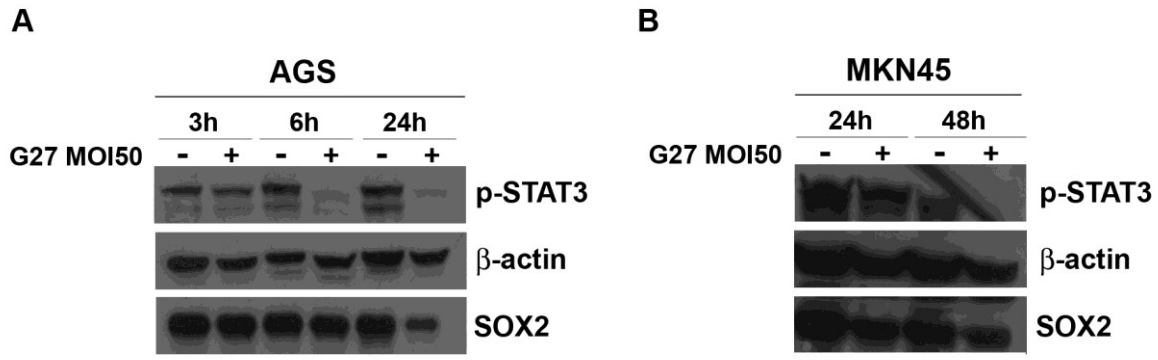


Fig. 42: Protein levels of SOX2 and p-STAT3 after H. pylori infection; AGS cells showed a decrease of SOX2 levels concomitant to pSTAT3 expression **(A)**. Same results were observed in MKN45 cells after 24 h and 48 h **(B)**. N = 3

As reporter assays indicated an involvement of CagA in the regulation of SOX2 activity, AGS cells were additionally screened for SOX2 protein expression levels after *H. pylori* infection with the wild type strain and the CagA deficient mutant. 6 h after infection with MOI 10 and 50 of the wild type strain a decrease of SOX2 protein levels could be observed, but no decrease of SOX2 protein did occur after infection with the CagA deficient mutant. The effect was even more pronounced 24 h after infection (Fig. 43). Those results correlated with the findings on SOX2 transcriptional activity.

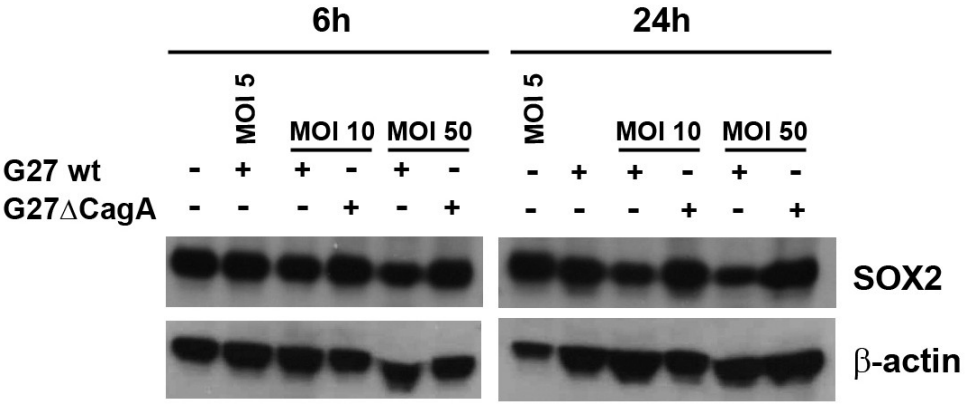


Fig.43: Expression levels of SOX2 after infection of AGS cells with H. pylori in dependency of CagA; SOX2 protein levels were significantly decreased after infection with H. pylori G27 wt, whereas the knock out mutant H. pylori G27 ΔCagA could not down-regulate SOX2 expression. N = 3

These results indicate that *H.pylori* infection influences SOX2 expression in a subset of GC cells and that these effects are correlating with expression of p-STAT3.

Furthermore, reduction of SOX2 expression might be dependent on the presence of CagA.

3.2. The role of SOX2 in stomach development

SOX2 is known to be highly expressed in endoderm derived tissue of vertebrates. As it is an important factor for pluripotency of stem cells, it may thus play a crucial role in stomach development and maintenance of adult gastric stem cells. Therefore, *Sox2* knock out during development and in the adult stomach could give fundamental information about the role of *Sox2* and possible involved pathways.

Initially the expression of SOX2 in various embryonic stages, from E 10.5 to E 18.5, was assessed in order to find out when SOX2 is expressed in the stomach, and is shown representatively for embryonic stage E 14.5 (Fig. 44). Examining expression pattern was also important for the planning of temporal knock-out experiments using inducible $CreER^{T2}$, as well as to foresee possible off target effects due to deletion of *Sox2* in other tissues. High SOX2 expression could be observed as early as E 10.5 in the central nervous system (CNS), in the retina of the eye, in hair follicles, as well as in stomach, esophagus and lungs.

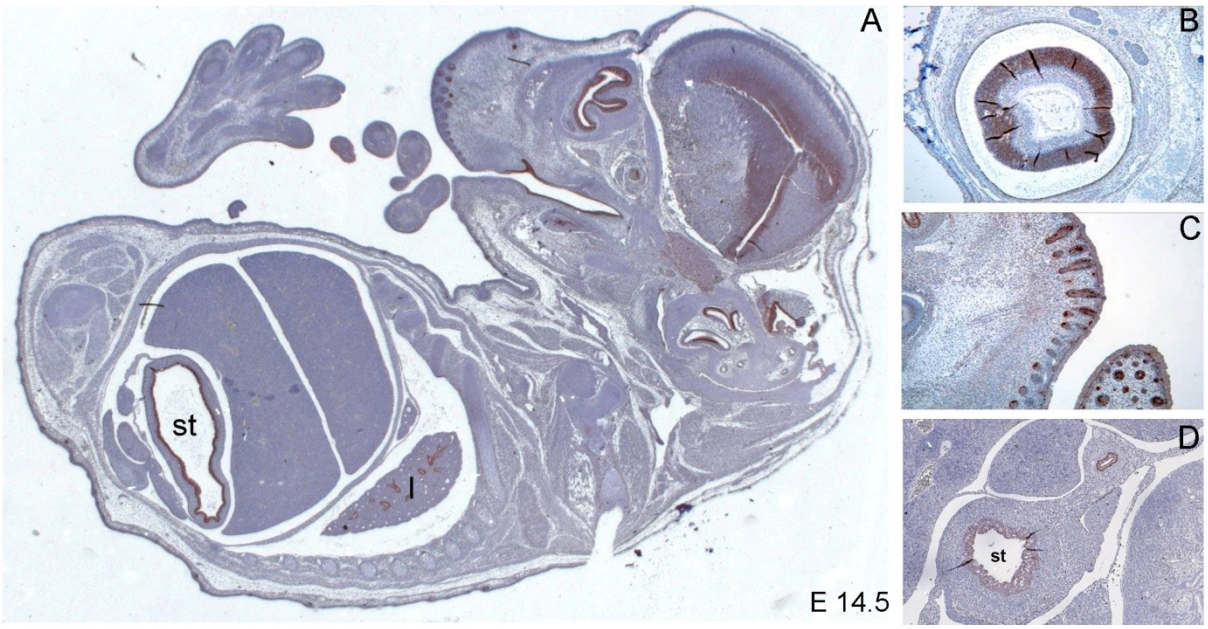


Fig. 44: Expression of Sox2 in a mouse embryo at E 14.5 analysed by IHC; Sox2 is highly expressed in several tissues and in the central nervous system (CNS) at day E 14.5 **(A)**. Expression could be observed in the retina **(B)**, in hair follicles **(C)** and in anterior endoderm of the stomach and in the esophagus **(D)**. (st = stomach, l = lungs)

A mouse model was employed in which the one exon gene *Sox2* was flanked by two lox p sites ("floxed" = fl). Crossing these *Sox2*^{fl} mice into Cre-mice holding the ubiquitous *Rosa26-CreER*^{T2} [197] and activating the Cre-system by injection of tamoxifen, the floxed *Sox2* allele was excised, leading to a conditional *Sox2* knock out. As excision is strongly dependent on the dose of tamoxifen reaching the cells, this knock out is usually not 100 % complete, leaving some cells still with remaining *Sox2* expression. Therefore, an additional mouse line was used, in which the *Sox2* allele was replaced by a β-galactosidase gene ("β-geo" knock in) resulting in a complete *Sox2* knock out in one allele [127]. This reduces the bias in results which might occur due to application variations of tamoxifen.

3.2.1. Sox2 in embryonic development

For embryonic analysis *Sox2*$^{fl/fl}$ x *Rosa26-CreER*^{T2} mice were mated with *Sox2*$^{fl/fl}$ or *Sox2*$^{fl/wt}$ mice and pregnant mothers were *s.c.* injected with 200 µg tamoxifen at embryonic day 7.5 (E 7.5) and E 18.5. Embryos were dissected at E 19.5. *Sox2*$^{fl/fl}$ Cre-positive tamoxifen induced pups (*Sox2* deleted) showed striking differences in phenotype. Ears and eyes were degenerated when compared to control embryos.

Embryos also were smaller. The stomach was smaller in size and showed less lumen when analyzed microscopically (Fig. 45 A and B). Additionally, stomach crypts showed an abnormal branched like structure (Fig. 45 C). The phenotype of the *Sox2* deleted embryos was highly depending on dose and time point of tamoxifen injected as it was observed in a further experiment with more induction time points. Here, four subsequent induction time points during pregnancy at E 7.5, E 9.5, E 14.5 and E 18.5 seemed to result in stagnancy of embryonic development in the last third of pregnancy. At E 19.5 a lower number of embryos was found in the uterus than usual, with about 1/2 to 1/4 of them being much smaller and not properly developed (Fig. 45 D). Genotyping revealed all fully developed intact embryos being either heterozygous for *Sox2*^{fl} or negative for CreER^{T2} or both, compared to degenerated embryos, which all showed *Sox2* deletion. Furthermore, a more severe phenotype was observed when pregnant mothers were positive for *Sox2*$^{fl/fl}$ and *Rosa26-CreER*^{T2}compared to mothers who did not harbor homozygosity for *Sox2*^{fl} and a *Rosa26-CreER*^{T2}. Similar results could be achieved with *Sox2*$^{fl/\beta geo}$ x *Rosa26-CreER*^{T2} breedings.

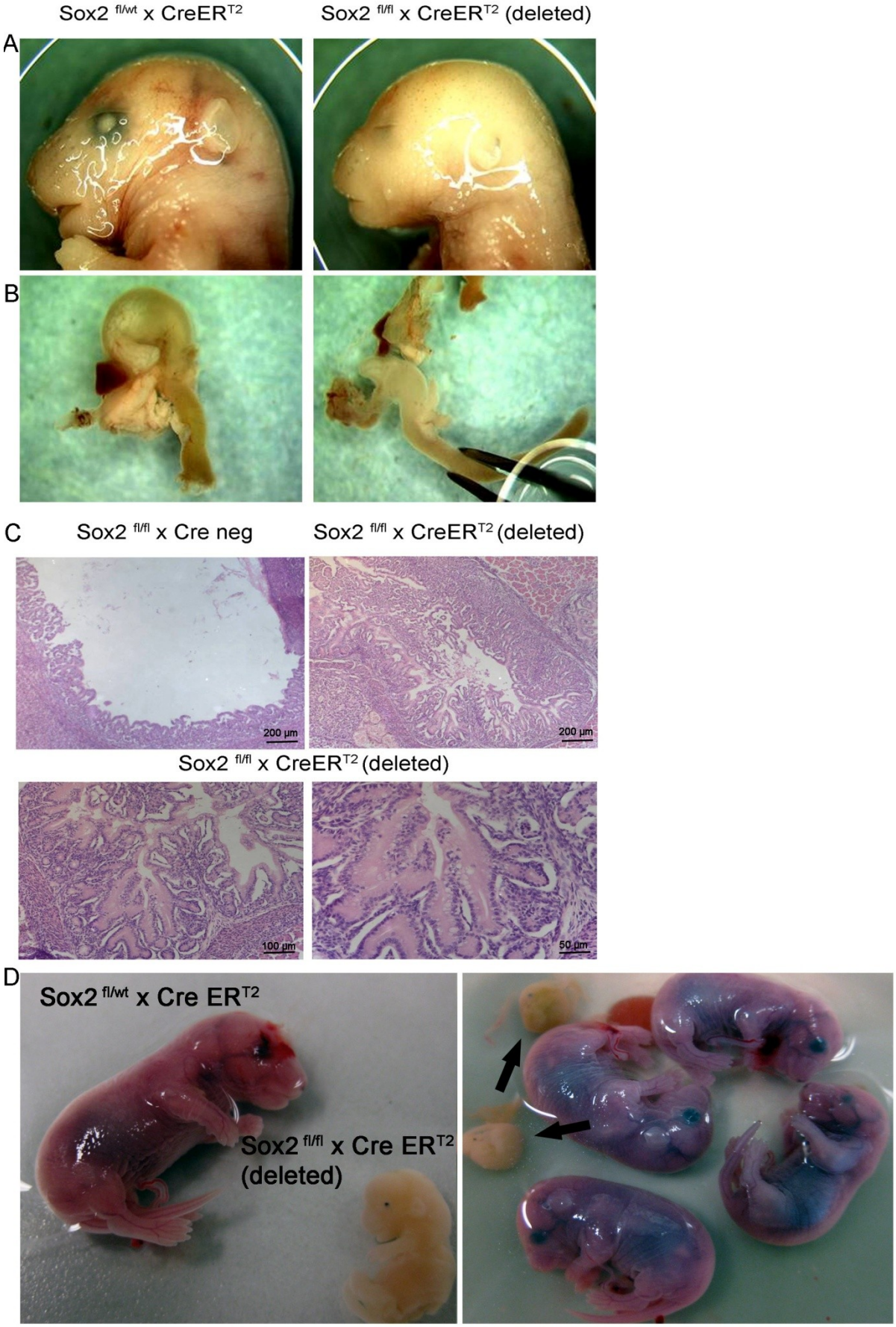


Fig. 45: Embryonic analysis of Sox2 deletion in mice; *Sox2* deleted mice showed significant differences in

phenotype. They developed degenerated eyes and aplastic ears as well as a much smaller stomach **(A)**. Furthermore, the stomach lumen appeared smaller **(B)** and stomach crypts had an abnormal branched like structure **(C)**. More induction time points during pregnancy of a *Rosa26CreER*^{T2} positive mother led to a more severe phenotype with embryos holding a homozygous *Sox2* deletion being absorbed or development being stagnated in the last third of pregnancy. Arrows indicating *Sox2*$^{fl/fl}$ *x CreER*^{T2} (deleted) embryos **(D)**. N = 4 (pregnant mothers), n = 26 (embryos)

As Sox2 is a crucial factor in early embryonic development, influencing gastrulation, homozygous *Sox2*$^{\beta\text{-}geo}$ mice, having a complete *Sox2* knock out are not viable. Analyses of blastulation at E 3.5 showed no significant differences of numbers of blastocystes, late morulae or unfertilized oocytes in comparison to wt-mice (Tab. 10), however, at day E 7.5 about 1/3 to 1/4 of the embryos of a *Sox2*$^{\beta\text{-}geo/wt}$ inmate breeding were degraded and genotyping revealed that the remaining intact embryos were heterozygous for the β-geo allele. These results lead to the assumption that *Sox2*$^{\beta\text{-}geo}$ homozygous embryos fail to complete gastrulation due to the lack of *Sox2*.

Tab. 10: Analysis of E 3.5 in a Sox2$^{\beta\text{-}geo/wt}$ x Sox2$^{\beta\text{-}geo/wt}$ mating

	Wild type (wt)			**Total wt**	***Sox2$^{\beta\text{-}geo/wt}$ x Sox2$^{\beta\text{-}geo/wt}$* (k.i.)**			**Total k.i.**
Mating no.	**No. 1**	**No. 2**	**No. 3**		**No. 4**	**No. 5**	**No. 6**	
Blastocyste	**11**	**10**	**14**	**35**	**12**	**10**	**12**	**34**
Late morula	0	3	1	4	1	2	2	5
Unfertilized oozyte	2	1	0	3	1	0	1	2

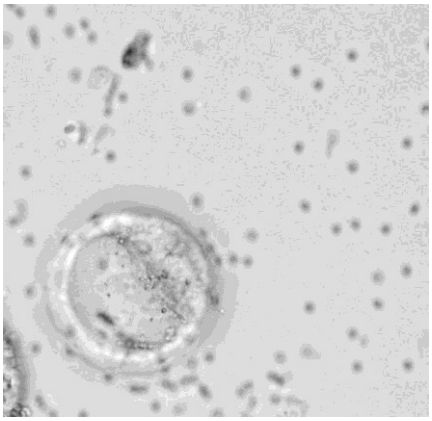


Fig. 46: Blastocyste (E 3.5) in Sox2$^{\beta\text{-}geo}$ homozygous embryo; pregnancies for each condition: N = 3

A limitation for the experimental set up was that the Rosa26-CreER^{T2} is a ubiquitous Cre and *βgeo* k.i also influences the whole mouse organism and not only the stomach. Thus, to have a look at the influence of *Sox2* solely on the stomach, it was planned to work with another mouse model holding a different Cre-system. Unfortunately, a stomach specific Cre-line has not been developed until now. Therefore, a *Sox17iCre* mouse line was used, which was kindly provided by Dr. Heiko Lickert, Institute of Stem Cell Research, Helmholtz center, Neuherberg. A constitutive active iCre is here under the control of the *Sox17* promoter. *Sox17* is expressed in the definitive endoderm which gives rise to lung, liver pancreas, stomach and the gastrointestinal tract. *Sox17* expression initiates around E 6.0 in the extraembryonic visceral endoderm [198]. Since *Sox2* is also expressed in early endoderm, presumably at *Sox17* positive sites, it was expected to see a stronger effect on the phenotype of the mice. In 10 breeding experiments, viable *Sox17iCre x Sox2$^{fl/fl}$* pubs or embryos could not be obtained (Tab. 11) leading to the assumption that Sox2 expression in the definitive endoderm is pivotal for the development and viability of embryos.

Tab. 11: Expected and observed genotypes in ten [Sox17iCre x Sox2$^{fl/wt}$] x [Sox2$^{fl/fl}$] breedings (n = 45 pubs)

Breeding [*Sox17iCre x SOX2$^{fl/wt}$*] x [*Sox2$^{fl/fl}$*]		
Expected	***Sox2$^{fl/fl}$***	***Sox2$^{fl/wt}$***
Sox17iCre +	25 %	25 %
Sox17iCre -	25 %	25 %

Observed	***Sox2$^{fl/fl}$***	***Sox2$^{fl/wt}$***
Sox17iCre +	0 %	11 %
Sox17iCre -	57 %	32 %

Therefore, selective deletion of *Sox2* in the stomach to further analyze the observed phenotype is not yet possible due to the lack of stomach specific Cre-lines.

3.2.2. The influence of Sox2 on postnatal development

To investigate how *Sox2* knock down could interfere in early postnatal development of pups, nursing mothers were induced with tamoxifen 2 times a week after birth. Since tamoxifen is highly lipophilic, pups receive tamoxifen via the maternal breast milk. *Sox2*$^{fl/fl}$ *x Rosa26-CreER*^{T2} pups seemed retarded in development, concerning weight (Tab. 12), size and overall fitness compared to their control siblings.

Tab. 12: Comparison of body weight of wt vs. k.o. pups; data of one representative experiment are shown.

	Body weight (g) at age (d)			**genotype**		**sex**
number	10	12	14	*Sox2*	*Rosa-Cre*	
1	5.9	6.7	7.3	fl/fl	neg	m
2	6.2	6.7	7.3	fl/fl	neg	m
3	6.2	6.8	7.4	fl/fl	neg	m
4	5.8	6.5	7.0	fl/fl	neg	m
5	5.9	6.4	6.9	fl/fl	neg	w
6	5.9	6.5	6.9	fl/fl	+/-	m
7	6.1	6.5	6.8	fl/fl	+/-	m
8	5.6	6.2	6.8	fl/fl	+/-	w
9	5.7	5.9	6.5	fl/fl	+/-	w

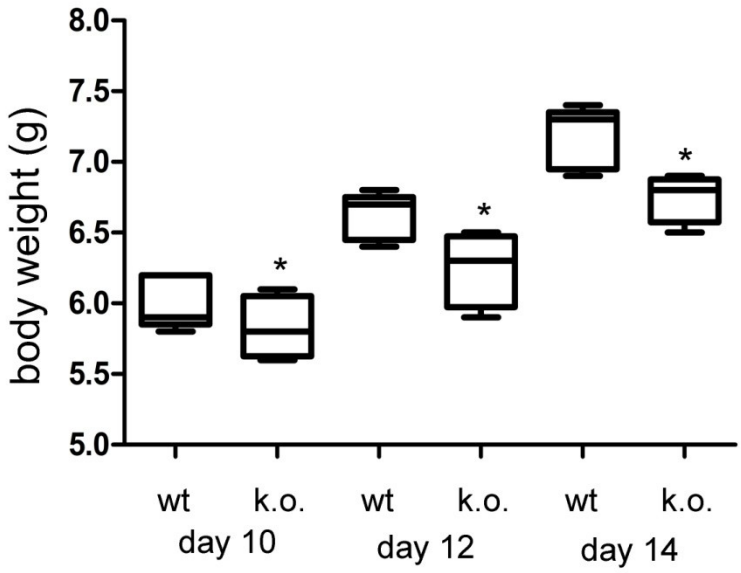


Fig.47: Body weight of wt pups vs. Sox2 k.o. pups; in comparison to wt pups, k.o. pups showed significant lower body weight, when receiving tamoxifen via breast milk. *p < 0.05

Furthermore, mice with *Sox2* deletion did not grow proper fur (Fig. 48). The stomach phenotype of *Sox2* deleted pups resembled the one in *Sox2* deleted embryos, yet the phenotype was not as pronounced. Knock-out pups died at an age of about two to three weeks. It is important to mention, that tamoxifen treated mothers in this experiment were never positive for both genotypes (*Sox2*$^{fl/fl}$ *and Rosa26-CreER*^{T2}), as in 80 % of all cases pups of these mice died one day after birth with no milk in the stomach. It was assumed that *Sox2* knock-out might influence breast milk production in mothers. However, no further investigations were done to strengthen this hypothesis.

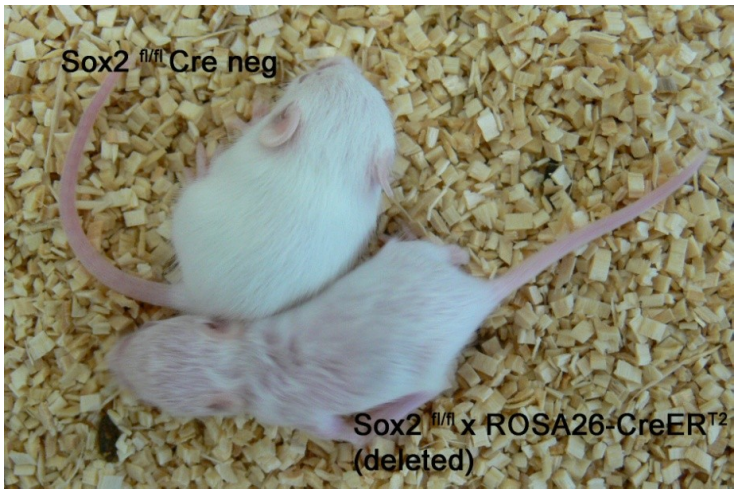


Fig. 48: Two week old pups after treatment with tamoxifen via breast milk; less fur growth is observed in *Sox2* deleted pups compared to control siblings without *Sox2* deletion.

These results demonstrate that *Sox2* contributes to various stages of development and is specifically important in stomach formation. Moreover, the embryonic phenotype highly depends on dosage and time point of the knock-out, implying that *Sox2* expression needs to be accurately balanced during stomach formation.

3.2.3. Sox2 in the adult murine stomach

Adult *Sox2*$^{fl/fl}$ *x Rosa26-CreER*^{T2} mice were injected s.c. with tamoxifen for 3 weeks every other day. Stomach was analyzed by IHC for SOX2 expression. In wt mice, SOX2 expression was adjacent to proliferating cells, marked by Ki67 expression (Fig. 49 A). After *Sox2* knock out the proliferation zone was reduced as assessed by BRDU staining (Fig. 49 B). Same results were obtained with *Sox2*$^{fl/\beta geo}$ *x Rosa26-CreER*^{T2} mice.

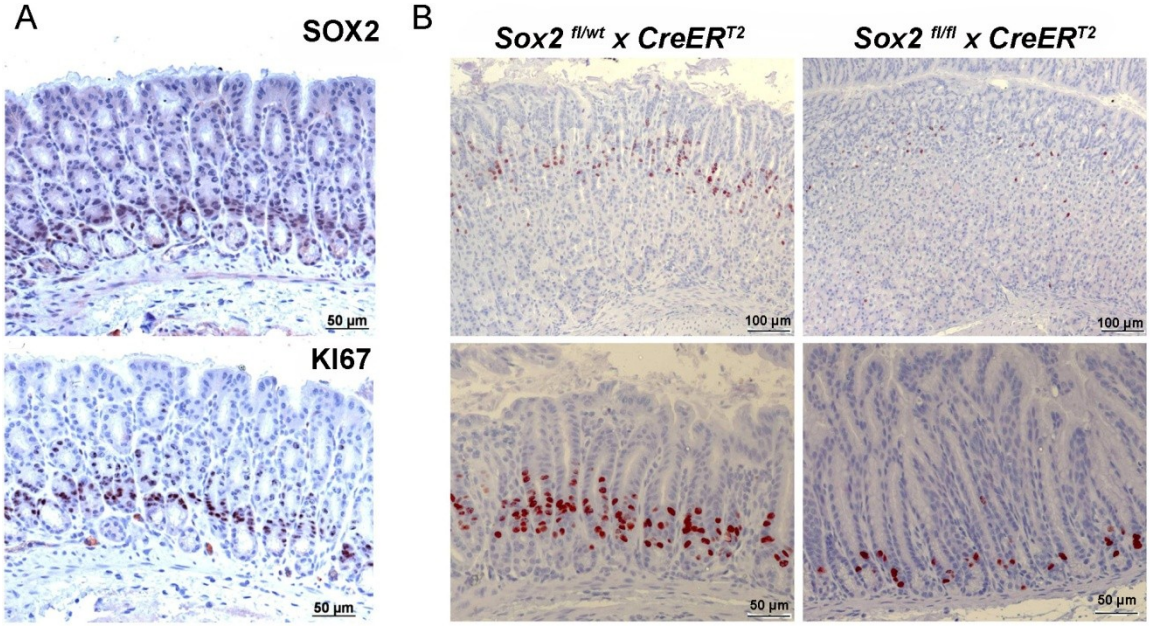


Fig.49: Sox2 expression and proliferation in adult murine stomach; in adult gastric murine glands SOX2 expression was adjacent to the proliferation zone **(A)**. *Sox2* knock down in adult mouse stomach led to a decrease of proliferating cells **(B)**. N = 5

These results indicate that in adult mouse stomach Sox2 is essential to sustain proliferation in gastric crypts, suggesting *Sox2* to be a potential factor of maintenance of adult gastric stem cells.

4. Discussion

4.1. Role of SOX2 on gastric carcinogenesis

Gastric cancer (GC) is currently the fourth most common cancer in the world and still one of the leading causes of cancer-related death worldwide [41]. Although *H. pylori* infections, one of the most important risk factors for GC, are declining and preventive measures are becoming more commonplace, prognosis of advanced GC and hence survival rate still remains poor [49]. Few specific treatments are available to date, making it important to develop novel therapeutic approaches. Thus, identifying essential genes involved in gastric carcinogenesis is mandatory to develop specific therapies.

Numerous different molecular alterations involving various pathways have been implicated in the development of GC. For example the WNT signaling pathway contributes to gastric carcinogenesis by stimulating the migration and invasion of GC cells [199]. Moreover, *CTNNB1* (β-catenin) is known to be frequently mutated in stomach cancer [200]. In several types of GC a full epithelial-mesenchymal transition (EMT) characterized by *SNAI1* (Snail) induction and *CDH1* (E-cadherin) reorganization is induced [201]. Furthermore, over-expression of the proto-oncogene *MET* was found in a subset of GC [50].

The transcription factor SOX2 plays a major role in self-renewal and differentiation of cells [122, 130] and has been shown to be overexpressed in GC as well. It is well known that SOX2 is highly expressed in diffuse-type GC [172]. In contrast to those studies, in the present work SOX2 was also found to be over expressed in intestinal and mixed-type gastric tumors. In some *in vitro* settings [168], as well as in differentiation of adult stem cells [202] and osteoblast differentiation [203], SOX2 was reported to antagonize TCF/β–Catenin activity and therefore repress the WNT pathway by promoting the transcription of its negative regulators. However, contrary to these findings, in retinal development of *Xenopus,* it was described that inhibition of WNT signaling is accompanied by a decrease of SOX2 expression [204]. Moreover, Neumann

et al. recently discovered that SOX2 was expressed in colorectal cancer (CRC) and was associated with poorer outcome [167]. They showed that high expression levels of SOX2 or nuclear β-Catenin were associated with metastasis in these tumors. Nearly all CRCs exhibit a chronic activation of the WNT/β–Catenin signaling pathway [205], implying that the up-regulation of SOX2, which was shown in CRCs, might be involved in the regulation of this pathway or at least is not inhibiting WNT signaling. Thus, intestinal type GC, showing high structural and biochemical similarities to CRC, might also be dependent on dysregulation of the canonical WNT pathway and is, for that reason, likely to be influenced by SOX2 expression. These facts can explain the findings of this study: that SOX2 is expressed also in some intestinal and mixed-type GC. Investigations also revealed that cells positive for SOX2 showed strong expression of KI67, a marker for proliferation. Since increased cell proliferation is a key aspect in development of neoplasms, driving cells towards uncontrolled growth and hence accumulation in the tissue, these data suggest that SOX2 expression in GC cells is an important factor for tumorigenesis.

To investigate the role of SOX2 in gastric tumors, several different GC cell lines were screened for SOX2. A heterogeneous expression of *SOX2* mRNA and protein was found in the different cell lines, with some cells highly expressing *SOX2* and others lacking *SOX2* expression. Those results also reflected the heterogeneous SOX2 expression patterns which were found in gastric tumor tissue samples. Unexpectedly, in some of the cell lines, expression levels of SOX2 did not correlate with its transcriptional activity. SOX2 can undergo different post translational modifications, such as sumoylation [206], phosphorylation [207] and acetylation [208]. The latter is involved in nuclear translocation, which could explain the lack of correlation between protein levels and transcriptional activity, since all such modifications affect SOX2 binding to DNA. Moreover, some cell lines seem to be independent from SOX2 expression. This indicates that the development of GC does not occur due to one specific event but arises from different combinations of factors and is therefore not necessarily dependent on dysregulation of the same gene or pathway [55, 56]. Furthermore, it also explains the heterogeneous levels of SOX2 expression in the gastric tumor samples.

To further pinpoint the role of SOX2 in GC cells, SOX2 was knocked down via RNA interference and via a dominant negative SOX2 construct. The advantage of a dominant negative approach was, that in contrast to RNA interference, the mutated transcription factor directly blocks the wild-type transcription factor from binding to the DNA binding site in a competitive manner, leading to a reduced level of gene activation and therefore more effectively inhibing SOX2 signaling. Although a significant down-regulation of SOX2 protein expression and transcriptional activity was also achieved by shRNA, concomitant with down-regulation of cell proliferation, it was not possible to obtain sufficient inhibition of SOX2 to conduct more extensive experiments. It was assumed that the extraordinarily high SOX2 levels in AZ-521 cells could not be overcome with this approach. Another reason for the strategy of inhibiting SOX2 transcriptional activity lay in the post-transcriptional modifications explained above, since the capacity of SOX2 to bind DNA, and not merely its expression, is essential to regulate target genes that might, in turn, be involved in gastric tumorigenesis. For that reason further experiments were performed with the dnSOX2 approach. An additional advantage was that using a tetracycline repressor provided the opportunity to control the onset of expression of dnSOX2, making it easy to conduct time dependent experiments. The system was validated by expression analysis of the HA-tagged dnSOX2 protein by western blot and confirmation of nuclear localization of the truncated protein by immunofluorescence. Furthermore, decrease of SOX2 transcriptional activity was observed to occur in a time dependent manner. However, a time shift was noted between protein expression and effects on transcriptional activity. This time shift between protein expression and actual activity of protein can be explained by the fact that the expressed protein needs to first accumulate and then localize to the SOX2 binding sites of target genes.

The function of SOX2 in GC was further analyzed by investigating its role in different aspects related to tumorigenesis. Cancer cells are characterized by a high proliferative activity, leading to accumulation of cells at the tumor site and uncontrolled growth. Since SOX2 expression was observed at the proliferation sites within the tumor tissue, the influence of SOX2 in proliferation of GC cells on SOX2 was subsequently verified *in vitro*. It was proven that the proliferative activity of cells decreased after inhibiting SOX2 in a time dependent manner. As this effect was reversible after some

days, it was supposed that this was a true effect on proliferation and was not due to cell death. As soon as dnSOX2 was degraded and the cells could regain their normal activity, the proliferation rate rose almost to basal levels. This led to the conclusion that the proliferation of AZ-521 GC cells is controlled by SOX2 activity. These results resemble findings in glioma cells as well as in breast cancer cells in which SOX2 was shown to be necessary for continuous proliferation [158, 159].

To explore whether the inhibition in cell proliferation observed was also related to an induction of apoptosis, Caspase 3/7 assays were performed. Dysregulation of apoptotic pathways can lead to uncontrolled cell growth, since bypassing apoptosis promotes cancer progression. An up-regulation of Caspase 3/7 activity was observed after blocking SOX2. However, this effect could not be assigned to a particular apoptotic pathway influenced by SOX2. These findings are consistent with recent studies showing that down-regulation of SOX2 inhibits proliferation and induces apoptosis in human lung cancer cells *in vitro* as well as *in vivo* [160, 162].

Recently, *SOX2* has been described to correlate with lymph-node metastases and distant spread in right-sided colon cancer [167]. To determine if SOX2 could also be involved in GC cell migration, wound healing assays were performed. The significant slower wound healing after blocking SOX2 indicated that SOX2 is contributing to the migration of cancer cells. A role for SOX2 in migration and invasion has also been shown in malignant gliomas [209, 210], strengthening the findings that SOX2 is essential for migration in GC cells. Certainly there is a strong correlation between proliferation rate and migration *in vitro* as proliferation of cells strongly influences migration. However, it is assumed that in the present setting a valid effect of SOX2 down-regulation on migration was observed, independent from a decreased rate of proliferation, since the effect on migration occurred earlier after SOX2 inhibition than the time point at which a decrease of proliferation was evident. Hence, observations of proliferation, apoptosis and migration imply that SOX2 expression is a major effector involved in different aspects related to the tumorigenic potential of AZ-521 stomach cancer cells. Furthermore, interruption of SOX2 activity leads to a less tumorigenic phenotype of cells.

To depict if decrease of proliferation was related to senescence, SA-β-gal activity was examined. Analysis of AZ-521 cells revealed that almost 30 % of cells undergo senescence after inhibition of SOX2 activity. In the senescent phase, cells are no longer dividing, nor are they dying, but in a state of inactivity. Senescence is a condition of irreversible cell growth arrest, protecting cells against aberrant proliferation, restricting tumor progression and can be assumed to be a plausible tumor suppressor mechanism [211]. Previous studies suggested that activation of senescence in tumors might contribute to the success of chemotherapy [212, 213]. In this study inhibition of SOX2 led to an increase of senescent cells, hence it might contribute to the protection against tumor formation. Studies done by *Yoon et al.* suggested that replicative senescence in mesenchymal stem cells may be dominantly regulated by SOX2 [214, 215], supporting the observation in the present work. Moreover, the onset of senescence in human neural stem cells was also associated with decreased expression of SOX2 [215].

Furthermore, the inhibition of cell growth was found to be due to changes in the cell cycle, since inhibition of SOX2 led to cell cycle arrest in G2/M phase, and translated into a lower cell migration rate. A recent study by *Lin et al.* showed that SOX2 is known to regulate cell cycle proteins in prostate cancer, and that inhibition of SOX2 induces cell cycle arrest [216]. Moreover, attenuated S-phase entry was observed in human glioma cells upon inhibition of SOX2 [209]. These data support findings presented here indicating that SOX2 is a major factor driving GC cells towards tumorigenesis. Subsequently it was analyzed if SOX2 influences the expression of certain cell cycle regulatory proteins. C-MYC is an oncogene, altered in human cancers and known to be deregulated in about 50 % of all tumors. Among other properties it influences the cell cycle by up-regulation of cyclins [217]. However, an altered expression of C-MYC after inhibition of SOX2 could not be observed. Nevertheless, lower *CCNB1* (*Cyclin B1*) levels were detected in cells with abolished SOX2 activity. It has recently been shown for colorectal cancer cells that down-regulation of Cyclin B1 induces G2/M cell cycle arrest and consequently inhibits proliferation of cells [218]. Moreover, *Xu et al.* reported that during autophagy mRNA expression of Cyclin B is decreasing concomitant with a decrease of OCT4, SOX2 and NANOG [219]. In contrast to findings of the present work, SOX2 was shown to inhibit the transition from G_1 to S phase in prostate cancer cell lines

by targeting Cyclin E (CCNE1) and P27 [216]. This could be due to the different cellular context, or a differential dependency of the cell lines employed on cell cycle regulators. On the other hand, the data repoted here are in contrast to the observations by Otsubo et al. [171], describing inhibition of proliferation and apoptosis upon over expression of SOX2 in GC cell lines. The use of different cell lines with distinct tumor origin could explain these discrepancies. Several of the cell lines employed in the work by Otsubo are known to depend on WNT signaling. The results presented here and the work of other groups [168, 202, 203] indicate that SOX2 negatively regulates the WNT pathway by inhibiting TCF4 transcriptional regulation. The fact that Otsubo *et al.* observed a SOX2-mediated decrease of Cyclin D1 (CCND1) and an increase of P27 (KIP1) levels, both of which are regulated by TCF4, clearly indicates a loss of TCF4 activity upon SOX2 over expression.

In vitro results indicated an essential role of SOX2 in gastric tumorigenesis, which were validated in a xenograft mouse model *in vivo,* where blocking of SOX2 resulted in reduced subcutaneous tumor growth in mice. Dissected tumors where stained for KI67 in order to evaluate proliferation rate of cells. However, no significant differences in proliferating cells were found, when comparing tumors of treated mice with blocked SOX2 to non-treated mice. Nevertheless, a tendency of non-treated tumors to exhibit more necrotic cells was detected. Since tumors are growing faster in those mice, there might be a higher tendency of tissue to undergo necrosis. However, this has to be investigated in more detail with a larger group of mice to assign data more accurately. Furthermore, tumors were stained for several genes known to be involved in gastric tumorigenesis. However, no significant differences in protein expression were seen. Both groups of tumors were negative for P16, CDX2, slug and snail and showed low expression of P21. Moreover, both exhibited high expression of β-Catenin and vimentin. β-Catenin is involved in the WNT signaling pathway and is highly expressed in most types of colorectal cancer. A co-expression of SOX2 and nuclear β-Catenin was seen in one third of all SOX2 positive CRCs and was associated with nodal status and distant metastasis [167]. Interestingly, SOX2 was found to be inhibiting WNT signaling in AZ-521 cells *in vitro*. It was therefore speculated that inhibition of WNT signaling might occur through blocking of β-Catenin by SOX2. It has been shown by Mansukhani *et al.* that

SOX2 binds to nuclear β-Catenin, thus preventing its binding to TCF/LEF during osteoblast differentiation [203]. The same mechanism might be true for GC but has to be further investigated. Over expression of Vimentin in GC was correlated to higher metastasis [185, 186]. Since high expression of Vimentin was seen in both tumor groups, as well as nuclear β-Catenin the metastasis rate of AZ-521 cells was investigated by injecting cells *i.v.* into athymic nude mice. However, AZ-521 cells were not metastatic and were cleared from the lungs one day after *i.v.* injection of cells. AZ-521 cells have been previously shown to have very low metastatic potential [220]. It can be assumed that missing discrepancies in tumor sections occur due to the exceptional high level of SOX2 in AZ-521 cells which is, however blocked by dnSOX2 and therefore regulated on a sub-cellular level, are still not low enough to detect differences by IHC.

To date not many SOX2 target genes in tumors have been found. To identify potential SOX2 targets, comparative RNA gene expression microarrays were performed at different time points after inhibition of SOX2 and differentially expressed genes were analyzed. A report by Tani *et al.* identified Pepsinogen A (PGA5) as potential target gene of SOX2 in a subset of GC cell lines [221], yet, in this study, differential expression of *PGA5* in AZ-521 cancer cells was not seen after blocking SOX2. This might be due to a different cellular background. However, regulations of several other genes could be verfied. Up-regulation of *CDKN1A* (p21*)* mRNA as well as protein was seen after down-regulating SOX2. A deregulation of p21 is known from several human cancers [187]. In GC tissue samples it has been shown that negative expression of p21 is correlated with poor survival and advanced stage and lymph node metastasis [222]. Similar results were found for laryngeal and oral carcinomas [223]. Furthermore, p21 knock-out mice developed spontaneous tumors [224]. Thus, it is suggested that the tumor suppressive role for p21 in GC might be a direct link to SOX2 regulating the expression of p21. Moreover, a down-regulation of ΔNP63, a splice variant of P63, was observed. In squamous cell carcinomas of the GI tract and of the lung SOX2 is co-expressed with P63 [225, 226]. Yet, the role of P63 and its different isoforms, TAP63 and ΔNP63 in tumorigenesis is still not clearly understood. A work by Barbieri *et al.* suggested, that the alpha splice variant of ΔNP63 has an anti-apoptotic effect [227]. Tissue culture experiments as well as tumor sample analysis revealed ΔNP63 to be over expressed in

several cancers and to contribute to proliferative characteristics, tumor growth and the inhibition of senescence [228-230]. Furthermore, p63 knock-out mice showed reduced proliferative capacity in keratinocytes which might be due to an increase of p21 since it is directly transcriptionally repressed by ΔNp63 [231]. Present results suggest that SOX2 might be inducing the expression of ΔNP63, which in turn represses p21 expression, contributing this way to proliferation and apoptosis of GC cells. These data for the first time show a potential correlation between SOX2, p21 and ΔNP63 in GC.

A further interesting observation concerning differentially expressed genes was the regulation of several WNT pathway related genes. Numerous *in vitro* experiments [168] as well as observations of differentiation of adult stem cells [202] and osteoblast differentiation [203] report SOX2 to antagonize TCF/β–Catenin activity and therefore to repress the WNT pathway by promoting the transcription of its negative regulators. An up-regulation of *LEF1* mRNA was confirmed in AZ-521 cells after blocking SOX2. *LEF1* is a WNT effector gene activated upon WNT signaling. Concomitant to this up-regulation an increase in TCF/LEF transcriptional activity was observed. Moreover, an up-regulation of *DKK4*, a direct target gene and upstream antagonist of the WNT pathway was detected. DKK4 acts as a WNT inhibitor, though. Its over expression in hepatocellular carcinoma cells was found to inhibit cell proliferation, reduce colony formation and delay cell migration [232]. As a WNT target gene it is up-regulated upon WNT signaling acting as a negative feed back regulator to attenuate WNT signaling after its activation. An induction of DKK4 by the canonical WNT pathway has also been reported by *Bazzi et al.* [233]. In the present setting DKK4 up-regulation may follow SOX2-mediated WNT activation in order to regulate aberrant WNT signaling. Additionally, a down-regulation of *FGF10* was verified, which is also known to be a WNT target gene. It has been shown earlier that WNT signaling suppresses the activity of FGF10 in branching morphogenesis of lung and the lacrimal glands and leads to a decrease in proliferation [234]. Furthermore, during osteogenesis, FGF10 signals counteract WNT signaling [235]. In human breast cancer and GC over expression of FGFR2, the binding receptor for FGF10, is linked to tumorigenesis and leads to more malignant phenotypes [236, 237]. To date there is no link known between SOX2, the interaction of FGF10 and the canonical WNT signaling. However, FGF10 is claimed to antagonize SOX2 expression during mouse

development [148, 238]. One might argue that in GC progression SOX2 inhibition is followed by subsequent suppression of FGF10. Yet further experiments would be necessary to clarify if FGF10 down-regulation is an indirect effect of WNT up-regulation upon loss of SOX2 or if SOX2 can directly influence FGF10 expression in GC cells.

In summary, these results demonstrate that SOX2 is involved in several aspects of gastric carcinogenesis *in vitro* and *in vivo* by regulating the expression of genes implicated in cell proliferation, apoptosis and cell cycle regulation. Importantly, present data support a model of gastric carcinogenesis that involves the gastric stem cell and differentiation marker SOX2, indicating a novel pathway apart from activating mutations in the WNT signaling pathway or E-cadherin. The fact that SOX2 was specifically observed in over 40 % of intestinal tumors challenges the current doctrine that loss of SOX2 is implicated in development of intestinal metaplasia and intestinal type GC. Rather, these data support the hypothesis that tissue specific stem cells in the stomach are SOX2 proficient and involved in the carcinogenic process in the stomach, and that intestinal metaplasia – devoid of SOX2 expression – is not a direct precursor lesion of intestinal type GC.

4.2. The regulation of SOX2 in gastric cancer

To understand the influence of SOX2 on GC it is not only important to analyze potential target genes of SOX2 but also to understand how SOX2 is regulated in the stomach. Since it is known that STAT3 is regulating SOX2 expression during neuronal development [141] it was sought to investigate if it was also regulating SOX2 in GC. Co-localization of phosphorylated STAT3 (p-STAT3), Ki67 and SOX2 was obsereved in tumors derived from *Gp130* mutant mice with hyperactivated STAT3 signaling. These data are in line with the observations Foshay *et al.* made in neuronal development. They could detect co-localization of SOX2 and STAT3 *in vivo* and *in vitro* [141]. In a subset of GC cell lines analyzed a nuclear co-localization of both transcription factors could also be verified. These results again strengthen the assumption that STAT3 is regulating SOX2 in GC. Moreover, a correlation of p-STAT3 and SOX2 protein expression was observed in all

GC cell lines investigated, except AZ-521 cells. It has been shown before that STAT3 is not phosphorylated in AZ-521 cells. Okamoto *et al.* suggested that the lack of phosphorylation occurs due to a concomitant lack of expression of the proto-oncogene c-met [239]. It is therefore speculated that SOX2 as well might be indirectly regulated by c-met in GC. However, AZ-521 cells might use a different pathway modulating SOX2 expression. Additionally, SOX2 reporter assays in a subset of GC cell lines revealed an up-regulation of SOX2 transcriptional activity after transfection with STAT3. These results again indicate that there might be a direct link between SOX2 and STAT3 in GC cell lines.

IL-6 leads to the phosphorylation of STAT3 by binding to gp130 receptor activating JAK [240]. An up-regulation of the transcriptional activity of SOX2 upon IL-6 stimulation was observed in GC cell lines, suggesting an involvement of the IL-6/STAT3 pathway during regulation of SOX2 in gastric tumorigenesis. It has been shown recently that a suppression of SOX2 in ES cells as a result of sodium arsenite exposure is even further augmented due to simultaneous inhibition of IL-6/STAT3 dependent pathways [241]. However, the present data are the first hint of a regulation of SOX2 via the IL-6/STAT3 pathway in GC.

GC often is preceded by inflammation, usually triggered by *H. pylori* infection. It was observed that the infection of *H. pylori* had different effects on SOX2 transcriptional activity in several GC cell lines. In AGS cells it has been shown before that infection with *H. pylori* results in a decrease of SOX2 expression [242]. These results were confirmed and furthermore it could be observed that the decrease was dependent on the presence of the virulence factor CagA. This CagA dependent decrease was also verified at the protein level. Moreover, AGS cells showed a concomitant decrease of p-STAT3 protein expression after infection, strengthening the assumption that SOX2 is regulated by STAT3 in GC. A simultaneous decrease of SOX2 and p-STAT3 protein after *H. pylori* infection was also found in MKN45 cells. A decrease of SOX2 transcriptional activity was observed in AZ-521 cells, but interestingly was not dependent on the expression of CagA, leading to the conclusion that different cell lines have different susceptibility to *H. pylori* virulence factors, wherein the involvement of CagA is not fully understood so far. In addition, it has been shown before in MKN45 and MKN28 cells, that *H. pylori* down-regulates SOX2 expression independently from CagA expression [243]. St 3051 and St

2957 are both not expressing SOX2 endogenously. Stimulation of those cells with IL-4 resulted in an up-regulation of SOX2, as reported by Asonuma *et al.* [170]. Subsequent infection with *H. pylori* could significantly down-regulate SOX2 transcriptional activity, while infection with the CagA deficient mutant strain resulted in a less intense effect than infection with the wild type strain as seen in AGS cells. Thus, in these and in the previous cell lines, *H. pylori* is likely to influence cells via activation of one of the various pathways CagA can be involved in, like the RAS-mitogen activated protein kinase (MAPK) pathway [78, 79], the E-Cadherin/β-Catenin pathway [80] or activation of NF-κB [82]. In summary, it was observed that SOX2 is regulated by p-STAT3 in several GC cell lines and moreover infection with *H. pylori* is able to down-regulate SOX2 expression most likely via a down-regulation of p-STAT3, but the detailed involvement of this pathway still needs to be further elucidated.

4.3. The role of SOX2 during murine development

SOX2 is not only an important factor for pluripotency of embryonic stem cells but also known to be highly expressed in endoderm derived tissues of vertebrates, thus likely to contribute to stomach development and upholding of adult gastric stem cells. SOX2 expression during mouse embryonic development was confirmed in the central nervous system including the retina as reported by Stevanovic and Collignon [244, 245], in the hair follicles [246], in the lungs [247] as well as in stomach and esophagus [147, 148].

Deletion of *Sox2* using a conditional knock out mouse resulted in different phenotypes depending on dose and time point of injection of tamoxifen. Early *Sox2* depletion at day E 7.5 resulted in deformed embryos at day E 19.5 showing especially degeneration of the stomach. It was assumed that deletion of *Sox2* leads to an intestinalization of the stomach due to a shift in pathway regulation. *Sox2*, together with *Cdx2*, plays a crucial role during early gut development. It is important for regionalization of the primitive streak influencing anterior and posterior gut formation. Later on these compartments become subdivided into foregut, midgut and hindgut and then differentiate into stomach and esophagus where SOX2 is expressed, or duodenum,

jejunum, ileum caecum and colon which express CDX2 but not SOX2 [248]. In correlation to these data it has been shown recently that expression of *Sox2* in the prospective intestinal part redirects intestinal epithelium towards a premature gastric phenotype during gut development [249]. These data strengthen the findings of the present work that loss of *Sox2* in the gastric part of the gut on the other hand might induce an intestinal phenotype. A more severe macroscopic phenotype was observed with increasing dosage of tamoxifen injections. Injecting four times during pregnancy led to stagnation of development of knock-out embryos. This severe phenotype might occur due to several reasons. As mentioned above *Sox2* is not only important in development of the stomach but also contributing to formation of several other tissues. Therefore, degeneration of embryos could not further be curtailed to a specific reason and has to be investigated in more detail. Thus, to have a more specific view on *Sox2* deletion in embryonic stomach development and to avoid too many off-targets effects due to the expression of *Sox2* in other tissue we additionally worked with a *Sox17iCre* mouse. *Sox17* is expressed in the definite endoderm, which gives rise to lung, heart, liver, pancreas, stomach and GI tract. Since breeding generated no *Sox17iCre x Sox2*$^{fl/fl}$ offspring it is assumed that expression of *Sox2* in the definite endoderm is crucial for the viability of embryos. *Sox17iCre* is expressed as early as day E 6.0 in the extraembryonic visceral endoderm [198]. An expression of *Sox2* in the same part of the embryo was reported by Avilion *et al.* at day E 6.5 [127]. Hence the role of *Sox2* during this stage of embryonic development is pivotal, so that deletion leads to complete malformation and therefore resorption of k.o. embryos. Generation of a stomach specific Cre-mouse-line would be fundamental for further investigations of the role of *Sox2* during stomach development.

Furthermore, data of Avilion *et al.* could be confirmed by analyzing *Sox2*$^{\beta geo}$ mice. As reported earlier [144] no differences in number of blastocytes, late morulae or unfertilized eggs could be found, obtained from *Sox2*$^{\beta geo/wt}$ inbred compared to *Sox2*$^{wt/wt}$ inbred embryos, however, at E 7.5 about 1/4 of embryos of *Sox2*$^{\beta geo/wt}$ inbred mice where degenerated. These results confirm the essential role of *Sox2* during early mouse development since it is already expressed in the inner cell mass (ICM) of blastocytes, although only at the stage of implantation a disruption of *Sox2* leads to lethality. In the growing oocyte there is much maternal SOX2 protein left which might rescue the k.o.

blastocytes. This protein is diluted during rapid growth of the embryo and hence is not able to further rescue k.o. embryos after implantation.

During early postnatal development of mice, k.o. of *Sox2* by induction of CreER^{T2} via breast feeding led to retardation in development which was measurable after one week to 10 days. K.o. pups showed significantly lower weight compared to their wt littermates (Fig. 46), did not develop proper fur (Fig. 47) and died within two to three weeks of age. Since *Sox2* is an important factor in development of dermal papillae [246] this explains the improper growth of fur. However, further experiments should be conducted to investigate how *Sox2* regulates generation of fur. Furthermore, *Sox2* is important in stomach development of mice [148] and loss of *Sox2* leads to a growth defect and moreover might also influence hormonal regulation in the stomach [147]. Negative effects of *Sox2* k.o. in pups get more severe over time, since the daily uptake of breast milk increases the tamoxifen dosage, thus increasing the k.o. of *Sox2*, and resulted in weight loss and eventually in their death. Combined with the findings that SOX2 is co-expressed at proliferation sites in the assumed stem cell zones of the adult mouse stomach and *Sox2* k.o. in adult mice leads to less proliferating cells (Fig. 49) it is suggesed that *Sox2* k.o. in pups influences the growth of the stomach by suppression of proliferating cells, hence leading to malformations. Additionally, a hormonal dysregulation is also likely to influence weight loss in k.o. pups.

A recent publication by Arnold *et al.* showed that *Sox2* positive cells are indeed potential markers for adult stem cells in the stomach [37]. They could also observe that adult *Sox2* k.o. resulted in death of mice, which they explained by the development of gastric ulcer. However, in this study adult mice did not die after *Sox2* k.o. This might be due to the different k.o. systems. Excision of *Sox2* in the system applied in the present study is strongly dependent on the dosage of tamoxifen reaching the cells, thus this k.o. occurs generally only partially and leaves the cells alive. The ganciclovir/delta thmyidine kinase (DTK) system applied by Arnold *at al.* is not only more sensitive but also kills every cell expressing DTK, thus every *Sox2* positive cell, after ganciclovir induction. Hence, a more severe phenotype is expected. Nevertheless, findings of this study confirm the results of Arnold *et. al.* that *Sox2* is influencing proliferation of cells in the adult mouse

stomach. Besides, these data are consistent with earlier findings shown in this work that *Sox2* is also expressed at proliferation sites in gastric tumor samples, again accrediting *Sox2* an essential role in controlling proliferation.

The observed phenomenon that *Sox2*$^{fl/fl}$ *x Rosa-CreER*^{T2} mothers induced with tamoxifen were not able to do proper lactation might be due to the fact that SOX2, among other stem cell factors, was recently found to be expressed in breast milk cells which were derived from the breast epithelium during lactation [250]. Hence *Sox2* might be essential to initiate proper lactation of mothers.

In summary it can be concluded that *Sox2* is essential for proper murine development as early as gastrulation. Embryonic k.o. after E 7.5 led to malformations of the stomach and was becoming more severe in a dose-dependent manner influencing the whole embryo. During early post-natal mouse development, depletion of *Sox2* resulted in death of pups which may be due to problems in nutrient uptake and hormonal regulations in the stomach. *Sox2* was also seen to be essential to sustain proliferation in gastric crypts of adult mice, suggesting that it indeed might be a potential factor for the maintenance of adult gastric stem cells.

5. Summary

This work for verified a tumorigenic role for SOX2 in gastric cancer. It was observed that SOX2 is a major factor contributing to cell proliferation in GC cells and this was strengthened by observation of co-expression of SOX2 and KI67 in gastric tumor tissue. Moreover, a role for SOX2 in migration was found and it was seen that blocking SOX2 induced apoptosis and senescence in GC cells. Analysis of the cell cycle revealed that inhibition of SOX2 led to cell cycle arrest. Moreover, it was verified in a xenograft model that SOX2 also influenced tumor growth *in vivo*. These results indicate a crucial role of SOX2 during gastric tumorigenesis.

Additionally, for the first time target genes of SOX2 in gastric tumors could be assigned. The regulation of several genes related to the canonical WNT pathway was observed. An up-regulation of LEF1 and DKK4 was seen as well as a down-regulation of FGF10 after blocking SOX2 activity. These data indicate that SOX2 contributes to cancer development through regulation of WNT pathway genes. Moreover, SOX2 was found to influence the expression of p21 in gastric tumors. An up-regulation of p21 after blocking SOX2 was observed with a concomitant down-regulation of ΔNP63. These results suggest that SOX2 might induce the expression of ΔNP63, which in turn represses p21 expression, contributing in this way to proliferation and apoptosis of GC cells. These are the first data for SOX2 target genes in GC.

Two different mechanisms of SOX2 regulation were also observed. A co-expression of SOX2 and STAT3 was observed in gastric tumors and in GC cell lines. Furthermore, it was shown that STAT3 expression directly influenced SOX2 transcriptional activity as well as SOX2 protein expression in several GC cells, and that this effect might be regulated through IL-6. Additionally, a link between *H. pylori* infection and SOX2 expression was found in gastric tumor cells. A down-regulation of SOX2 upon infection with *H. pylori* was observed and, in a subset of cell lines, the virulence factor CagA was identified as a key player in this regulation. Furthermore, an additional decrease of p-STAT3 protein expression levels following *H. pylori* infection was found.

Finally, SOX2 was identified to be a major factor influencing mouse development. Depletion of *Sox2* in murine embryos resulted in malformations of several organs. A severe stomach phenotype was observed and therefore it is suggested that loss of *Sox2* leads to intestinalization of the stomach due to a shift in pathway regulations. Moreover, it was assumed that *Sox2* is important for a proper hormone regulation in early post-natal development of mice. It was also seen that *Sox2* expression in mothers influenced lactation. Furthermore, SOX2 protein was observed at the presumptive gastric stem cell zone and it was shown that inhibition of *Sox2* influenced proliferation in the adult mouse stomach, suggesting that *Sox2* might be a potential adult stem cell marker in the stomach.

References

1. Wells, J.M. and D.A. Melton, *Vertebrate endoderm development.* Annu Rev Cell Dev Biol, 1999. **15**: p. 393-410.
2. de Santa Barbara, P., G.R. van den Brink, and D.J. Roberts, *Development and differentiation of the intestinal epithelium.* Cell Mol Life Sci, 2003. **60**(7): p. 1322-32.
3. Fukuda, K. and S. Yasugi, *The molecular mechanisms of stomach development in vertebrates.* Dev Growth Differ, 2005. **47**(6): p. 375-82.
4. Lickert, H., et al., *Expression patterns of WNT genes in mouse gut development.* Mechanisms of Development, 2001. **105**(1-2): p. 181-184.
5. Katoh, M., *Networking of WNT, FGF, Notch, BMP, and Hedgehog signaling pathways during carcinogenesis.* Stem Cell Rev, 2007. **3**(1): p. 30-8.
6. Peek, R.M., Jr. and M.J. Blaser, *Helicobacter pylori and gastrointestinal tract adenocarcinomas.* Nat Rev Cancer, 2002. **2**(1): p. 28-37.
7. Wheater, P.R., H.G. Burkitt, and V.G. Daniels, *Functional Histology: A colour atlas and text.* Churchill Livingston 1987: p. 203 - 284.
8. Lee, E.R., et al., *Division of the mouse gastric mucosa into zymogenic and mucous regions on the basis of gland features.* Am J Anat, 1982. **164**(3): p. 187-207.
9. Mills, J.C. and R.A. Shivdasani, *Gastric epithelial stem cells.* Gastroenterology, 2011. **140**(2): p. 412-24.
10. Quante, M. and T.C. Wang, *Stem cells in gastroenterology and hepatology.* Nat Rev Gastroenterol Hepatol, 2009. **6**(12): p. 724-37.
11. Canfield, V., et al., *Genetic ablation of parietal cells in transgenic mice: a new model for analyzing cell lineage relationships in the gastric mucosa.* Proc Natl Acad Sci U S A, 1996. **93**(6): p. 2431-5.
12. Nomura, S., et al., *Lineage and clonal development of gastric glands.* Dev Biol, 1998. **204**(1): p. 124-35.
13. Karam, S.M. and C.P. Leblond, *Dynamics of epithelial cells in the corpus of the mouse stomach. I. Identification of proliferative cell types and pinpointing of the stem cell.* Anat Rec, 1993. **236**(2): p. 259-79.
14. Karam, S.M. and C.P. Leblond, *Dynamics of epithelial cells in the corpus of the mouse stomach. V. Behavior of entero-endocrine and caveolated cells: general conclusions on cell kinetics in the oxyntic epithelium.* Anat Rec, 1993. **236**(2): p. 333-40.
15. Karam, S.M. and C.P. Leblond, *Dynamics of epithelial cells in the corpus of the mouse stomach. III. Inward migration of neck cells followed by progressive transformation into zymogenic cells.* Anat Rec, 1993. **236**(2): p. 297-313.
16. Karam, S.M. and C.P. Leblond, *Dynamics of epithelial cells in the corpus of the mouse stomach. II. Outward migration of pit cells.* Anat Rec, 1993. **236**(2): p. 280-96.
17. Karam, S.M., *Dynamics of epithelial cells in the corpus of the mouse stomach. IV. Bidirectional migration of parietal cells ending in their gradual degeneration and loss.* Anat Rec, 1993. **236**(2): p. 314-32.
18. Lee, E.R. and C.P. Leblond, *Dynamic histology of the antral epithelium in the mouse stomach: II. Ultrastructure and renewal of isthmal cells.* Am J Anat, 1985. **172**(3): p. 205-24.
19. Corpron, R.E., *The ultrastructure of the gastric mucosa in normal and hypophysectomized rats.* Am J Anat, 1966. **118**(1): p. 53-90.
20. Nomura, S., et al., *Clonal analysis of isolated single fundic and pyloric gland of stomach using X-linked polymorphism.* Biochem Biophys Res Commun, 1996. **226**(2): p. 385-90.

21. Qiao, X.T., et al., *Prospective identification of a multilineage progenitor in murine stomach epithelium.* Gastroenterology, 2007. **133**(6): p. 1989-98.
22. Barker, N., et al., *Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro.* Cell Stem Cell, 2010. **6**(1): p. 25-36.
23. Vries, R.G., M. Huch, and H. Clevers, *Stem cells and cancer of the stomach and intestine.* Mol Oncol, 2010. **4**(5): p. 373-84.
24. Barker, N., et al., *Identification of stem cells in small intestine and colon by marker gene Lgr5.* Nature, 2007. **449**(7165): p. 1003-7.
25. Sauer, B. and N. Henderson, *Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1.* Proc Natl Acad Sci U S A, 1988. **85**(14): p. 5166-70.
26. Feil, R., et al., *Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains.* Biochem Biophys Res Commun, 1997. **237**(3): p. 752-7.
27. Soriano, P., *Generalized lacZ expression with the ROSA26 Cre reporter strain.* Nat Genet, 1999. **21**(1): p. 70-1.
28. Barker, N. and H. Clevers, *Lineage tracing in the intestinal epithelium.* Curr Protoc Stem Cell Biol, 2010. **Chapter 5**: p. Unit5A 4.
29. Snippert, H.J. and H. Clevers, *Tracking adult stem cells.* EMBO Rep, 2011. **12**(2): p. 113-22.
30. Farrell, J.J., et al., *TFF2/SP-deficient mice show decreased gastric proliferation, increased acid secretion, and increased susceptibility to NSAID injury.* J Clin Invest, 2002. **109**(2): p. 193-204.
31. Quante, M., et al., *TFF2 mRNA transcript expression marks a gland progenitor cell of the gastric oxyntic mucosa.* Gastroenterology, 2010. **139**(6): p. 2018-2027 e2.
32. Nam, K.T., et al., *Mature chief cells are cryptic progenitors for metaplasia in the stomach.* Gastroenterology, 2010. **139**(6): p. 2028-2037 e9.
33. Takaishi, S., et al., *Identification of gastric cancer stem cells using the cell surface marker CD44.* Stem Cells, 2009. **27**(5): p. 1006-20.
34. Giannakis, M., et al., *Molecular properties of adult mouse gastric and intestinal epithelial progenitors in their niches.* J Biol Chem, 2006. **281**(16): p. 11292-300.
35. Zhao, P., Y. Li, and Y. Lu, *Aberrant expression of CD133 protein correlates with Ki-67 expression and is a prognostic marker in gastric adenocarcinoma.* BMC Cancer, 2010. **10**: p. 218.
36. Lengner, C.J., et al., *Oct4 expression is not required for mouse somatic stem cell self-renewal.* Cell Stem Cell, 2007. **1**(4): p. 403-15.
37. Arnold, K., et al., *Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice.* Cell Stem Cell, 2011. **9**(4): p. 317-29.
38. Montgomery, R.K., et al., *Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells.* Proc Natl Acad Sci U S A, 2011. **108**(1): p. 179-84.
39. Sangiorgi, E. and M.R. Capecchi, *Bmi1 is expressed in vivo in intestinal stem cells.* Nat Genet, 2008. **40**(7): p. 915-20.
40. Tian, H., et al., *A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable.* Nature, 2011. **478**(7368): p. 255-9.
41. Parkin, D.M., et al., *Global cancer statistics, 2002.* CA Cancer J Clin, 2005. **55**(2): p. 74-108.
42. Jemal, A., et al., *Cancer statistics, 2006.* CA Cancer J Clin, 2006. **56**(2): p. 106-30.
43. Crew, K.D. and A.I. Neugut, *Epidemiology of gastric cancer.* World J Gastroenterol, 2006. **12**(3): p. 354-62.
44. Nakamura, T., et al., *A clinicopathological study in young patients with gastric carcinoma.* J Surg Oncol, 1999. **71**(4): p. 214-9.

45. Shikata, K., et al., *Population-based prospective study of the combined influence of cigarette smoking and Helicobacter pylori infection on gastric cancer incidence: the Hisayama Study.* Am J Epidemiol, 2008. **168**(12): p. 1409-15.
46. Nomura, A., et al., *Cigarette smoking and stomach cancer.* Cancer Res, 1990. **50**(21): p. 7084.
47. Forman, D., et al., *Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation.* BMJ, 1991. **302**(6788): p. 1302-5.
48. Jemal, A., et al., *Cancer statistics, 2010.* CA Cancer J Clin, 2010. **60**(5): p. 277-300.
49. Correa, P., *Is gastric cancer preventable?* Gut, 2004. **53**(9): p. 1217-9.
50. Smith, M.G., et al., *Cellular and molecular aspects of gastric cancer.* World J Gastroenterol, 2006. **12**(19): p. 2979-90.
51. el-Rifai, W. and S.M. Powell, *Molecular and biologic basis of upper gastrointestinal malignancy. Gastric carcinoma.* Surg Oncol Clin N Am, 2002. **11**(2): p. 273-91, viii.
52. Ming, S.C., *Gastric carcinoma. A pathobiological classification.* Cancer, 1977. **39**(6): p. 2475-85.
53. Lauren, P., *The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification.* Acta Pathol Microbiol Scand, 1965. **64**: p. 31-49.
54. Catalano, V., et al., *Gastric cancer.* Crit Rev Oncol Hematol, 2005. **54**(3): p. 209-41.
55. Solcia, E., et al., *Intestinal and diffuse gastric cancers arise in a different background of Helicobacter pylori gastritis through different gene involvement.* Am J Surg Pathol, 1996. **20 Suppl 1**: p. S8-22.
56. Yasui, W., et al., *Genetic and epigenetic alterations in multistep carcinogenesis of the stomach.* J Gastroenterol, 2000. **35 Suppl 12**: p. 111-5.
57. Tahara, E., *Genetic pathways of two types of gastric cancer.* IARC Sci Publ, 2004(157): p. 327-49.
58. Correa, P., et al., *A model for gastric cancer epidemiology.* Lancet, 1975. **2**(7924): p. 58-60.
59. Correa, P. and J. Houghton, *Carcinogenesis of Helicobacter pylori.* Gastroenterology, 2007. **133**(2): p. 659-72.
60. Gutierrez-Gonzalez, L. and N.A. Wright, *Biology of intestinal metaplasia in 2008: more than a simple phenotypic alteration.* Dig Liver Dis, 2008. **40**(7): p. 510-22.
61. Reis, C.A., et al., *Intestinal metaplasia of human stomach displays distinct patterns of mucin (MUC1, MUC2, MUC5AC, and MUC6) expression.* Cancer Res, 1999. **59**(5): p. 1003-7.
62. Tsukamoto, T., T. Mizoshita, and M. Tatematsu, *Gastric-and-intestinal mixed-type intestinal metaplasia: aberrant expression of transcription factors and stem cell intestinalization.* Gastric Cancer, 2006. **9**(3): p. 156-66.
63. Nardone, G., A. Rocco, and P. Malfertheiner, *Review article: helicobacter pylori and molecular events in precancerous gastric lesions.* Aliment Pharmacol Ther, 2004. **20**(3): p. 261-70.
64. Silberg, D.G., et al., *Cdx1 and cdx2 expression during intestinal development.* Gastroenterology, 2000. **119**(4): p. 961-71.
65. Beck, F., et al., *Reprogramming of intestinal differentiation and intercalary regeneration in Cdx2 mutant mice.* Proc Natl Acad Sci U S A, 1999. **96**(13): p. 7318-23.
66. Silberg, D.G., et al., *Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice.* Gastroenterology, 2002. **122**(3): p. 689-96.
67. Judd, L.M., et al., *Gastric cancer development in mice lacking the SHP2 binding site on the IL-6 family co-receptor gp130.* Gastroenterology, 2004. **126**(1): p. 196-207.

68. Tanaka, A., et al., *Helicobacter pylori heat shock protein 60 antibodies are associated with gastric cancer.* Pathol Res Pract, 2009. **205**(10): p. 690-4.
69. Bornschein, J., et al., *H. pylori infection is a key risk factor for proximal gastric cancer.* Dig Dis Sci, 2010. **55**(11): p. 3124-31.
70. Lynch, H.T., et al., *Gastric cancer: new genetic developments.* J Surg Oncol, 2005. **90**(3): p. 114-33; discussion 133.
71. Pharoah, P.D., P. Guilford, and C. Caldas, *Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families.* Gastroenterology, 2001. **121**(6): p. 1348-53.
72. Brooks-Wilson, A.R., et al., *Germline E-cadherin mutations in hereditary diffuse gastric cancer: assessment of 42 new families and review of genetic screening criteria.* J Med Genet, 2004. **41**(7): p. 508-17.
73. Machado, J.C., et al., *E-cadherin gene mutations provide a genetic basis for the phenotypic divergence of mixed gastric carcinomas.* Lab Invest, 1999. **79**(4): p. 459-65.
74. Yokozaki, H., et al., *p53 point mutations in primary human gastric carcinomas.* J Cancer Res Clin Oncol, 1992. **119**(2): p. 67-70.
75. Feldman, R.A., A.J. Eccersley, and J.M. Hardie, *Epidemiology of Helicobacter pylori: acquisition, transmission, population prevalence and disease-to-infection ratio.* Br Med Bull, 1998. **54**(1): p. 39-53.
76. Dooley, C.P., et al., *Prevalence of Helicobacter pylori infection and histologic gastritis in asymptomatic persons.* N Engl J Med, 1989. **321**(23): p. 1562-6.
77. Odenbreit, S., et al., *Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion.* Science, 2000. **287**(5457): p. 1497-500.
78. Higashi, H., et al., *SHP-2 tyrosine phosphatase as an intracellular target of Helicobacter pylori CagA protein.* Science, 2002. **295**(5555): p. 683-686.
79. Churin, Y., et al., *Helicobacter pylori CagA protein targets the c-Met receptor and enhances the motogenic response.* Journal of Cell Biology, 2003. **161**(2): p. 249-255.
80. Murata-Kamiya, N., et al., *Helicobacter pylori CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells.* Oncogene, 2007. **26**(32): p. 4617-4626.
81. Bornschein, J. and P. Malfertheiner, *Gastric carcinogenesis.* Langenbecks Arch Surg, 2011. **396**(6): p. 729-42.
82. Kim, S.Y., et al., *Helicobacter pylori CagA transfection of gastric epithelial cells induces interleukin-8.* Cellular Microbiology, 2006. **8**(1): p. 97-106.
83. Douraghi, M., et al., *Multiple Gene Status in Helicobacter pylori Strains and Risk of Gastric Cancer Development.* Digestion, 2009. **80**(3): p. 200-207.
84. Basso, D., et al., *Clinical relevance of Helicobacter pylori cagA and vacA gene polymorphisms.* Gastroenterology, 2008. **135**(1): p. 91-99.
85. Molinari, M., et al., *Selective inhibition of Ii-dependent antigen presentation by Helicobacter pylori toxin VacA.* Journal of Experimental Medicine, 1998. **187**(1): p. 135-140.
86. Papini, E., et al., *Selective increase of the permeability of polarized epithelial cell monolayers by Helicobacter pylori vacuolating toxin.* J Clin Invest, 1998. **102**(4): p. 813-20.
87. Willhite, D.C. and S.R. Blanke, *Helicobacter pylori vacuolating cytotoxin enters cells, localizes to the mitochondria, and induces mitochondrial membrane permeability changes correlated to toxin channel activity.* Cell Microbiol, 2004. **6**(2): p. 143-54.
88. Nakayama, M., et al., *Helicobacter pylori VacA-induced Inhibition of GSK3 through the PI3K/Akt Signaling Pathway.* Journal of Biological Chemistry, 2009. **284**(3): p. 1612-1619.

89. Manente, L., et al., *The Helicobacter pylori's protein VacA has direct effects on the regulation of cell cycle and apoptosis in gastric epithelial cells.* Journal of Cellular Physiology, 2008. **214**(3): p. 582-587.
90. Boren, T., et al., *Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens.* Science, 1993. **262**(5141): p. 1892-5.
91. Gerhard, M., et al., *Clinical relevance of the Helicobacter pylori gene for blood-group antigen-binding adhesin.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12778-83.
92. El-Omar, E.M., *The importance of interleukin 1beta in Helicobacter pylori associated disease.* Gut, 2001. **48**(6): p. 743-7.
93. Furuta, T., et al., *Interleukin 1beta polymorphisms increase risk of hypochlorhydria and atrophic gastritis and reduce risk of duodenal ulcer recurrence in Japan.* Gastroenterology, 2002. **123**(1): p. 92-105.
94. Vilaichone, R.K., et al., *Gastric mucosal cytokine levels in relation to host interleukin-1 polymorphisms and Helicobacter pylori cagA genotype.* Scandinavian Journal of Gastroenterology, 2005. **40**(5): p. 530-539.
95. Machado, J.C., et al., *A proinflammatory genetic profile increases the risk for chronic atrophic gastritis and gastric carcinoma.* Gastroenterology, 2003. **125**(2): p. 364-71.
96. Taguchi, A., et al., *Interleukin-8 promoter polymorphism increases the risk of atrophic gastritis and gastric cancer in Japan.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(11 Pt 1): p. 2487-93.
97. Tu, S., et al., *Overexpression of interleukin-1beta induces gastric inflammation and cancer and mobilizes myeloid-derived suppressor cells in mice.* Cancer Cell, 2008. **14**(5): p. 408-19.
98. Crabtree, J.E., et al., *Mucosal tumour necrosis factor alpha and interleukin-6 in patients with Helicobacter pylori associated gastritis.* Gut, 1991. **32**(12): p. 1473-7.
99. Furukawa, K., et al., *Enhanced mucosal expression of interleukin-6 mRNA but not of interleukin-8 mRNA at the margin of gastric ulcer in Helicobacter pylori-positive gastritis.* J Gastroenterol, 1998. **33**(5): p. 625-33.
100. Jackson, C.B., et al., *Augmented gp130-mediated cytokine signalling accompanies human gastric cancer progression.* J Pathol, 2007. **213**(2): p. 140-51.
101. Gong, W., et al., *Expression of activated signal transducer and activator of transcription 3 predicts expression of vascular endothelial growth factor in and angiogenic phenotype of human gastric cancer.* Clin Cancer Res, 2005. **11**(4): p. 1386-93.
102. Giraud, A.S., et al., *Differentiation of the Gastric Mucosa IV. Role of trefoil peptides and IL-6 cytokine family signaling in gastric homeostasis.* Am J Physiol Gastrointest Liver Physiol, 2007. **292**(1): p. G1-5.
103. Polyak, K. and W.C. Hahn, *Roots and stems: stem cells in cancer.* Nature Medicine, 2006. **12**(3): p. 296-300.
104. Eaves, C.J., *CANCER STEM CELLS Here, there, everywhere?* Nature, 2008. **456**(7222): p. 581-582.
105. Jordan, C.T., M.L. Guzman, and M. Noble, *Cancer stem cells.* N Engl J Med, 2006. **355**(12): p. 1253-61.
106. Al-Hajj, M., et al., *Prospective identification of tumorigenic breast cancer cells.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3983-8.
107. Zhang, S., et al., *Identification and characterization of ovarian cancer-initiating cells from primary human tumors.* Cancer Research, 2008. **68**(11): p. 4311-4320.
108. Singh, S.K., et al., *Identification of a cancer stem cell in human brain tumors.* Cancer Research, 2003. **63**(18): p. 5821-8.
109. Li, C., et al., *Identification of pancreatic cancer stem cells.* Cancer Research, 2007. **67**(3): p. 1030-7.

110. Prince, M.E., et al., *Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 973-8.
111. O'Brien, C.A., et al., *A human colon cancer cell capable of initiating tumour growth in immunodeficient mice.* Nature, 2007. **445**(7123): p. 106-110.
112. Lapidot, T., et al., *A Cell Initiating Human Acute Myeloid-Leukemia after Transplantation into Scid Mice.* Nature, 1994. **367**(6464): p. 645-648.
113. Bonnet, D. and J.E. Dick, *Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell.* Nature Medicine, 1997. **3**(7): p. 730-737.
114. Rocco, A., et al., *CD133 and CD44 cell surface markers do not identify cancer stem cells in primary human gastric tumours.* J Cell Physiol, 2011.
115. Gubbay, J., et al., *A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes.* Nature, 1990. **346**(6281): p. 245-50.
116. Bowles, J., G. Schepers, and P. Koopman, *Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators.* Dev Biol, 2000. **227**(2): p. 239-55.
117. Harley, V.R., R. Lovell-Badge, and P.N. Goodfellow, *Definition of a consensus DNA binding site for SRY.* Nucleic Acids Res, 1994. **22**(8): p. 1500-1.
118. Soullier, S., et al., *Diversification pattern of the HMG and SOX family members during evolution.* J Mol Evol, 1999. **48**(5): p. 517-27.
119. Wright, E.M., B. Snopek, and P. Koopman, *Seven new members of the Sox gene family expressed during mouse development.* Nucleic Acids Res, 1993. **21**(3): p. 744.
120. Kiefer, J.C., *Back to basics: Sox genes.* Dev Dyn, 2007. **236**(8): p. 2356-66.
121. Wegner, M., *All purpose Sox: The many roles of Sox proteins in gene expression.* International Journal of Biochemistry & Cell Biology, 2010. **42**(3): p. 381-390.
122. Dong, C., D. Wilhelm, and P. Koopman, *Sox genes and cancer.* Cytogenet Genome Res, 2004. **105**(2-4): p. 442-7.
123. Uchikawa, M., Y. Kamachi, and H. Kondoh, *Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken.* Mech Dev, 1999. **84**(1-2): p. 103-20.
124. Kamachi, Y., M. Uchikawa, and H. Kondoh, *Pairing SOX off with partners in the regulation of embryonic development.* Trends in Genetics, 2000. **16**(4): p. 182-187.
125. Wegner, M., *From head to toes: the multiple facets of Sox proteins.* Nucleic Acids Res, 1999. **27**(6): p. 1409-20.
126. Pevny, L.H. and R. Lovell-Badge, *Sox genes find their feet.* Curr Opin Genet Dev, 1997. **7**(3): p. 338-44.
127. Avilion, A.A., et al., *Multipotent cell lineages in early mouse development depend on SOX2 function.* Genes Dev, 2003. **17**(1): p. 126-40.
128. Katoh, Y. and M. Katoh, *Comparative genomics on SOX2 orthologs.* Oncol Rep, 2005. **14**(3): p. 797-800.
129. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.
130. Masui, S., et al., *Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.* Nat Cell Biol, 2007. **9**(6): p. 625-35.
131. Tomioka, M., et al., *Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex.* Nucleic Acids Res, 2002. **30**(14): p. 3202-13.
132. Boyer, L.A., et al., *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. **122**(6): p. 947-56.
133. Wang, J.L., et al., *A protein interaction network for pluripotency of embryonic stem cells.* Nature, 2006. **444**(7117): p. 364-368.

134. Chickarmane, V., et al., *Transcriptional dynamics of the embryonic stem cell switch.* Plos Computational Biology, 2006. **2**(9): p. 1080-1092.
135. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
136. Hochedlinger, K. and K. Plath, *Epigenetic reprogramming and induced pluripotency.* Development, 2009. **136**(4): p. 509-523.
137. Amabile, G. and A. Meissner, *Induced pluripotent stem cells: current progress and potential for regenerative medicine.* Trends in Molecular Medicine, 2009. **15**(2): p. 59-68.
138. Trounson, A., *Rats, Cats, and Elephants, but Still No Unicorn: Induced Pluripotent Stem Cells from New Species.* Cell Stem Cell, 2009. **4**(1): p. 3-4.
139. Kondoh, H. and Y. Kamachi, *SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms.* Int J Biochem Cell Biol, 2010. **42**(3): p. 391-9.
140. Inoue, M., et al., *PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in embryonic visual system development.* Genes to Cells, 2007. **12**(9): p. 1049-1061.
141. Foshay, K.M. and G.I. Gallicano, *Regulation of Sox2 by STAT3 initiates commitment to the neural precursor cell fate.* Stem Cells Dev, 2008. **17**(2): p. 269-78.
142. Wood, H.B. and V. Episkopou, *Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages.* Mech Dev, 1999. **86**(1-2): p. 197-201.
143. Ivanova, N., et al., *Dissecting self-renewal in stem cells with RNA interference.* Nature, 2006. **442**(7102): p. 533-8.
144. Kelberman, D., et al., *Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans.* J Clin Invest, 2006. **116**(9): p. 2442-55.
145. Kiernan, A.E., et al., *Sox2 is required for sensory organ development in the mammalian inner ear.* Nature, 2005. **434**(7036): p. 1031-5.
146. Shah, D., et al., *Bilateral microphthalmia, esophageal atresia, and cryptorchidism: the anophthalmia-esophageal-genital syndrome.* Am J Med Genet, 1997. **70**(2): p. 171-3.
147. Williamson, K.A., *Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome.* Human Molecular Genetics, 2006. **15**(9): p. 1413-1422.
148. Que, J., et al., *Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm.* Development, 2007. **134**(13): p. 2521-2531.
149. Ellis, P., et al., *SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult.* Dev Neurosci, 2004. **26**(2-4): p. 148-65.
150. Taranova, O.V., *SOX2 is a dose-dependent regulator of retinal neural progenitor competence.* Genes & Development, 2006. **20**(9): p. 1187-1202.
151. Fauquier, T., et al., *SOX2-expressing progenitor cells generate all of the major cell types in the adult mouse pituitary gland.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 2907-12.
152. Que, J., et al., *Multiple roles for Sox2 in the developing and adult mouse trachea.* Development, 2009. **136**(11): p. 1899-907.
153. Okubo, T., C. Clark, and B.L. Hogan, *Cell lineage mapping of taste bud cells and keratinocytes in the mouse tongue and soft palate.* Stem Cells, 2009. **27**(2): p. 442-50.
154. Driskell, R.R., et al., *Sox2-positive dermal papilla cells specify hair follicle type in mammalian epidermis.* Development, 2009. **136**(16): p. 2815-23.
155. Sattler, H.P., et al., *Novel amplification unit at chromosome 3q25-q27 in human prostate cancer.* Prostate, 2000. **45**(3): p. 207-15.
156. Sanada, Y., et al., *Histopathologic evaluation of stepwise progression of pancreatic carcinoma with immunohistochemical analysis of gastric epithelial transcription factor*

SOX2: comparison of expression patterns between invasive components and cancerous or nonneoplastic intraductal components. Pancreas, 2006. **32**(2): p. 164-70.

157. 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.
158. Chen, Y., et al., *The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer.* J Biol Chem, 2008. **283**(26): p. 17969-78.
159. 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.
160. Bass, A.J., et al., *SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas.* Nat Genet, 2009. **41**(11): p. 1238-42.
161. Hussenet, T., et al., *SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas.* PLoS One, 2010. **5**(1): p. e8960.
162. Xiang, R., et al., *Downregulation of transcription factor SOX2 in cancer stem cells suppresses growth and metastasis of lung cancer.* British Journal of Cancer, 2011. **104**(9): p. 1410-1417.
163. Du, L., et al., *Sox2 nuclear expression is closely associated with poor prognosis in patients with histologically node-negative oral tongue squamous cell carcinoma.* Oral Oncol, 2011.
164. Laga, A.C., et al., *Expression of the embryonic stem cell transcription factor SOX2 in human skin: relevance to melanocyte and merkel cell biology.* Am J Pathol, 2010. **176**(2): p. 903-13.
165. Girouard, S.D., et al., *SOX2 contributes to melanoma cell invasion.* Lab Invest, 2011.
166. Saigusa, S., et al., *Correlation of CD133, OCT4, and SOX2 in Rectal Cancer and Their Association with Distant Recurrence After Chemoradiotherapy.* Annals of Surgical Oncology, 2009. **16**(12): p. 3488-3498.
167. Neumann, J., et al., *SOX2 expression correlates with lymph-node metastases and distant spread in right-sided colon cancer.* BMC Cancer, 2011. **11**(1): p. 518.
168. Kormish, J.D., D. Sinner, and A.M. Zorn, *Interactions Between SOX Factors and WNT/beta-Catenin Signaling in Development and Disease.* Developmental Dynamics, 2010. **239**(1): p. 56-68.
169. Fang, X., et al., *ChIP-seq and functional analysis of the SOX2 gene in colorectal cancers.* OMICS, 2010. **14**(4): p. 369-84.
170. Asonuma, S., et al., *Helicobacter pylori induces gastric mucosal intestinal metaplasia through the inhibition of interleukin-4-mediated HMG box protein Sox2 expression.* Am J Physiol Gastrointest Liver Physiol, 2009. **297**(2): p. G312-22.
171. Otsubo, T., et al., *SOX2 is frequently downregulated in gastric cancers and inhibits cell growth through cell-cycle arrest and apoptosis.* Br J Cancer, 2008. **98**(4): p. 824-31.
172. Tsukamoto, T., et al., *Sox2 expression in human stomach adenocarcinomas with gastric and gastric-and-intestinal-mixed phenotypes.* Histopathology, 2005. **46**(6): p. 649-58.
173. Tsukamoto, T., et al., *Down-regulation of a gastric transcription factor, Sox2, and ectopic expression of intestinal homeobox genes, Cdx1 and Cdx2: inverse correlation during progression from gastric/intestinal-mixed to complete intestinal metaplasia.* J Cancer Res Clin Oncol, 2004. **130**(3): p. 135-45.
174. Matsuoka, J., et al., *Role of the Stemness Factors Sox2, Oct3/4, and Nanog in Gastric Carcinoma.* J Surg Res, 2010.
175. Nockel, J., et al., *Characterization of gastric adenocarcinoma cell lines established from CEA424/SV40 T antigen-transgenic mice with or without a human CEA transgene.* BMC Cancer, 2006. **6**: p. 57.
176. Gossen, M. and H. Bujard, *Tight control of gene expression in mammalian cells by tetracycline-responsive promoters.* Proc Natl Acad Sci U S A, 1992. **89**(12): p. 5547-51.

177. Gossen, M., et al., *Transcriptional activation by tetracyclines in mammalian cells.* Science, 1995. **268**(5218): p. 1766-9.
178. Morin, P.J., et al., *Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC.* Science, 1997. **275**(5307): p. 1787-90.
179. Vermes, I., et al., *A novel assay for apoptosis. Flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled Annexin V.* J Immunol Methods, 1995. **184**(1): p. 39-51.
180. Baltrus, D.A., et al., *The complete genome sequence of Helicobacter pylori strain G27.* J Bacteriol, 2009. **191**(1): p. 447-8.
181. Figura, N., et al., *Inactivation of Helicobacter pylori cagA gene affects motility.* Helicobacter, 2004. **9**(3): p. 185-93.
182. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
183. Wettenhall, J.M. and G.K. Smyth, *limmaGUI: a graphical user interface for linear modeling of microarray data.* Bioinformatics, 2004. **20**(18): p. 3705-6.
184. Hayflick, L. and P.S. Moorhead, *The serial cultivation of human diploid cell strains.* Exp Cell Res, 1961. **25**: p. 585-621.
185. Takemura, K., et al., *Expression of vimentin in gastric cancer: a possible indicator for prognosis.* Pathobiology, 1994. **62**(3): p. 149-54.
186. Fuyuhiro, Y., et al., *Clinical significance of vimentin-positive gastric cancer cells.* Anticancer Res, 2010. **30**(12): p. 5239-43.
187. Abbas, T. and A. Dutta, *p21 in cancer: intricate networks and multiple activities.* Nat Rev Cancer, 2009. **9**(6): p. 400-14.
188. Liggett, W.H., Jr. and D. Sidransky, *Role of the p16 tumor suppressor gene in cancer.* J Clin Oncol, 1998. **16**(3): p. 1197-206.
189. Kajita, M., K.N. McClinic, and P.A. Wade, *Aberrant expression of the transcription factors snail and slug alters the response to genotoxic stress.* Mol Cell Biol, 2004. **24**(17): p. 7559-66.
190. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
191. Frijters, R., et al., *CoPub: a literature-based keyword enrichment tool for microarray data analysis.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W406-10.
192. Zhang, J.Y., et al., *delta-Catenin promotes malignant phenotype of non-small cell lung cancer by non-competitive binding to E-cadherin with p120ctn in cytoplasm.* J Pathol, 2010. **222**(1): p. 76-88.
193. Wang, M., et al., *Expression of delta-catenin is associated with progression of human astrocytoma.* BMC Cancer, 2011. **11**: p. 514.
194. Pennisi, D., et al., *Mutations in Sox18 underlie cardiovascular and hair follicle defects in ragged mice.* Nat Genet, 2000. **24**(4): p. 434-7.
195. Nicolas, M., et al., *Notch1 functions as a tumor suppressor in mouse skin.* Nat Genet, 2003. **33**(3): p. 416-21.
196. Lobry, C., P. Oh, and I. Aifantis, *Oncogenic and tumor suppressor functions of Notch in cancer: it's NOTCH what you think.* J Exp Med, 2011. **208**(10): p. 1931-5.
197. Vooijs, M., J. Jonkers, and A. Berns, *A highly efficient ligand-regulated Cre recombinase mouse line shows that LoxP recombination is position dependent.* EMBO Rep, 2001. **2**(4): p. 292-7.
198. Engert, S., et al., *Sox17-2A-iCre: A knock-in mouse line expressing Cre recombinase in endoderm and vascular endothelial cells.* genesis, 2009. **47**(9): p. 603-610.
199. Kurayoshi, M., et al., *Expression of WNT-5a is correlated with aggressiveness of gastric cancer by stimulating cell migration and invasion.* Cancer Res, 2006. **66**(21): p. 10439-48.

200. Clements, W.M., et al., *beta-Catenin mutation is a frequent cause of WNT pathway activation in gastric cancer.* Cancer Res, 2002. **62**(12): p. 3503-6.
201. Kang, M.H., et al., *BMP2 accelerates the motility and invasiveness of gastric cancer cells via activation of the phosphatidylinositol 3-kinase (PI3K)/Akt pathway.* Exp Cell Res, 2010. **316**(1): p. 24-37.
202. Park, S.B., et al., *SOX2 has a crucial role in the lineage determination and proliferation of mesenchymal stem cells through Dickkopf-1 and c-MYC.* Cell Death Differ, 2012. **19**(3): p. 534-45.
203. Mansukhani, A., *Sox2 induction by FGF and FGFR2 activating mutations inhibits WNT signaling and osteoblast differentiation.* The Journal of Cell Biology, 2005. **168**(7): p. 1065-1076.
204. Van Raay, T.J., et al., *Frizzled 5 signaling governs the neural potential of progenitors in the developing Xenopus retina.* Neuron, 2005. **46**(1): p. 23-36.
205. Korinek, V., et al., *Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma.* Science, 1997. **275**(5307): p. 1784-7.
206. Tsuruzoe, S., et al., *Inhibition of DNA binding of Sox2 by the SUMO conjugation.* Biochem Biophys Res Commun, 2006. **351**(4): p. 920-6.
207. Jeong, C.H., et al., *Phosphorylation of Sox2 cooperates in reprogramming to pluripotent stem cells.* Stem Cells, 2010. **28**(12): p. 2141-50.
208. Baltus, G.A., et al., *Acetylation of sox2 induces its nuclear export in embryonic stem cells.* Stem Cells, 2009. **27**(9): p. 2175-84.
209. Oppel, F., et al., *SOX2-RNAi attenuates S-phase entry and induces RhoA-dependent switch to protease-independent amoeboid migration in human glioma cells.* Mol Cancer, 2011. **10**: p. 137.
210. Alonso, M.M., et al., *Genetic and epigenetic modifications of Sox2 contribute to the invasive phenotype of malignant gliomas.* PLoS One, 2011. **6**(11): p. e26740.
211. Collado, M., M.A. Blasco, and M. Serrano, *Cellular senescence in cancer and aging.* Cell, 2007. **130**(2): p. 223-33.
212. te Poele, R.H., et al., *DNA damage is able to induce senescence in tumor cells in vitro and in vivo.* Cancer Res, 2002. **62**(6): p. 1876-83.
213. Roberson, R.S., et al., *Escape from therapy-induced accelerated cellular senescence in p53-null lung cancer cells and in human lung cancers.* Cancer Res, 2005. **65**(7): p. 2795-803.
214. Yoon, D.S., et al., *Importance of Sox2 in maintenance of cell proliferation and multipotency of mesenchymal stem cells in low-density culture.* Cell Prolif, 2011. **44**(5): p. 428-40.
215. Raabe, E.H., et al., *BRAF activation induces transformation and then senescence in human neural stem cells: a pilocytic astrocytoma model.* Clin Cancer Res, 2011. **17**(11): p. 3590-9.
216. Lin, F., et al., *Sox2 targets cyclinE, p27 and survivin to regulate androgen-independent human prostate cancer cell proliferation and apoptosis.* Cell Prolif, 2012. **45**(3): p. 207-16.
217. Vita, M. and M. Henriksson, *The Myc oncoprotein as a therapeutic target for human cancer.* Semin Cancer Biol, 2006. **16**(4): p. 318-30.
218. Mollah, M.L., D.K. Park, and H.J. Park, *Cordyceps militaris Grown on Germinated Soybean Induces G2/M Cell Cycle Arrest through Downregulation of Cyclin B1 and Cdc25c in Human Colon Cancer HT-29 Cells.* Evid Based Complement Alternat Med, 2012. **2012**: p. 249217.
219. Xu, Y.N., et al., *Autophagy Influences Maternal mRNA Degradation and Apoptosis in Porcine Parthenotes Developing In Vitro.* J Reprod Dev, 2012.

220. Taki, T., et al., *Immunological analysis of glycolipids on cell surfaces of cultured human tumor cell lines: expression of lactoneotetraosylceramide on tumor cell surfaces.* J Biochem, 1985. **98**(4): p. 887-95.
221. Tani, Y., et al., *Transcription factor SOX2 up-regulates stomach-specific pepsinogen A gene expression.* J Cancer Res Clin Oncol, 2007. **133**(4): p. 263-9.
222. Seo, Y.H., et al., *Prognostic significance of p21 and p53 expression in gastric cancer.* Korean J Intern Med, 2003. **18**(2): p. 98-103.
223. Kapranos, N., et al., *p53, p21 and p27 protein expression in head and neck cancer and their prognostic value.* Anticancer Res, 2001. **21**(1B): p. 521-8.
224. Martin-Caballero, J., et al., *Tumor susceptibility of p21(Waf1/Cip1)-deficient mice.* Cancer Res, 2001. **61**(16): p. 6234-8.
225. Long, K.B. and J.L. Hornick, *SOX2 is highly expressed in squamous cell carcinomas of the gastrointestinal tract.* Hum Pathol, 2009. **40**(12): p. 1768-73.
226. Hussenet, T. and S. du Manoir, *SOX2 in squamous cell carcinoma: Amplifying a pleiotropic oncogene along carcinogenesis.* Cell Cycle, 2010. **9**(8).
227. Barbieri, C.E., et al., *IGFBP-3 is a direct target of transcriptional regulation by DeltaNp63alpha in squamous epithelium.* Cancer Res, 2005. **65**(6): p. 2314-20.
228. Ramsey, M.R., et al., *Physical association of HDAC1 and HDAC2 with p63 mediates transcriptional repression and tumor maintenance in squamous cell carcinoma.* Cancer Res, 2011. **71**(13): p. 4373-9.
229. Wu, G., et al., *DeltaNp63alpha up-regulates the Hsp70 gene in human cancer.* Cancer Res, 2005. **65**(3): p. 758-66.
230. Keyes, W.M., et al., *DeltaNp63alpha is an oncogene that targets chromatin remodeler Lsh to drive skin stem cell proliferation and tumorigenesis.* Cell Stem Cell, 2011. **8**(2): p. 164-76.
231. Westfall, M.D., et al., *The Delta Np63 alpha phosphoprotein binds the p21 and 14-3-3 sigma promoters in vivo and has transcriptional repressor activity that is reduced by Hay-Wells syndrome-derived mutations.* Mol Cell Biol, 2003. **23**(7): p. 2264-76.
232. Fatima, S., et al., *Dickkopf 4 (DKK4) acts on WNT/beta-catenin pathway by influencing beta-catenin in hepatocellular carcinoma.* Oncogene, 2012. **31**(38): p. 4233-44.
233. Bazzi, H., et al., *The WNT inhibitor, Dickkopf 4, is induced by canonical WNT signaling during ectodermal appendage morphogenesis.* Dev Biol, 2007. **305**(2): p. 498-507.
234. Dean, C.H., et al., *Canonical WNT signaling negatively regulates branching morphogenesis of the lung and lacrimal gland.* Dev Biol, 2005. **286**(1): p. 270-86.
235. Raucci, A., et al., *Osteoblast proliferation or differentiation is regulated by relative strengths of opposing signaling pathways.* J Cell Physiol, 2008. **215**(2): p. 442-51.
236. Adnane, J., et al., *BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers.* Oncogene, 1991. **6**(4): p. 659-63.
237. Katoh, M. and M. Katoh, *FGFR2 and WDR11 are neighboring oncogene and tumor suppressor gene on human chromosome 10q26.* Int J Oncol, 2003. **22**(5): p. 1155-9.
238. Nyeng, P., et al., *FGF10 signaling controls stomach morphogenesis.* Dev Biol, 2007. **303**(1): p. 295-310.
239. Okamoto, W., et al., *Differential roles of STAT3 depending on the mechanism of STAT3 activation in gastric cancer cells.* Br J Cancer, 2011. **105**(3): p. 407-12.
240. Heinrich, P.C., et al., *Principles of interleukin (IL)-6-type cytokine signalling and its regulation.* Biochem J, 2003. **374**(Pt 1): p. 1-20.
241. Ivanov, V.N., G. Wen, and T.K. Hei, *Sodium arsenite exposure inhibits AKT and Stat3 activation, suppresses self-renewal and induces apoptotic death of embryonic stem cells.* Apoptosis, 2012.
242. Camilo, V., et al., *Helicobacter pylori and the BMP pathway regulate CDX2 and SOX2 expression in gastric cells.* Carcinogenesis, 2012. **33**(10): p. 1985-92.

243. Zhang, X., et al., *H. pylori induces the expression of Hath1 in gastric epithelial cells via interleukin-8/STAT3 phosphorylation while suppressing Hes1.* J Cell Biochem, 2012.
244. Stevanovic, M., et al., *The cDNA sequence and chromosomal location of the human SOX2 gene.* Mamm Genome, 1994. **5**(10): p. 640-2.
245. Collignon, J., et al., *A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2.* Development, 1996. **122**(2): p. 509-20.
246. Rendl, M., L. Lewis, and E. Fuchs, *Molecular dissection of mesenchymal-epithelial interactions in the hair follicle.* PLoS Biol, 2005. **3**(11): p. e331.
247. Ishii, Y., et al., *Region-specific expression of chicken Sox2 in the developing gut and lung epithelium: regulation by epithelial-mesenchymal interactions.* Dev Dyn, 1998. **213**(4): p. 464-75.
248. Sherwood, R.I., T.Y. Chen, and D.A. Melton, *Transcriptional dynamics of endodermal organ formation.* Dev Dyn, 2009. **238**(1): p. 29-42.
249. Raghoebir, L., et al., *SOX2 redirects the developmental fate of the intestinal epithelium toward a premature gastric phenotype.* J Mol Cell Biol, 2012.
250. Hassiotou, F., et al., *Breastmilk is a novel source of stem cells with multilineage differentiation potential.* Stem Cells, 2012. **30**(10): p. 2164-74.

Appendix

1. List of differentially expressed genes (limmatable p 0.01)

ID	gene assignment	gene name	
Down-regulated			
7943413	NM_001165	BIRC3	baculoviral IAP repeat-containing 3
8106730	NM_022550	XRCC4	X-ray repair complementing defective repair in Chinese hamster cells
8175252	---		
8102792	NM_019035	PCDH18	protocadherin 18
8111993	NM_004465	FGF10	fibroblast growth factor 10
7971565	NM_005767	P2RY5	purinergic receptor P2Y, G-protein coupled, 5
8128383	NM_017421	COQ3	coenzyme Q3 homolog, methyltransferase (S. cerevisiae)
7951259	NM_002425	MMP10	matrix metallopeptidase 10 (stromelysin 2)
8175250	---		
8091103	NM_006286	TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)
8108370	NM_001964	EGR1	early growth response 1
7919780	NM_018178	GOLPH3L	golgi phosphoprotein 3-like
8096160	NM_001025616	ARHGAP24	Rho GTPase activating protein 24
7923516	NM_016243	CYB5R1	cytochrome b5 reductase 1
8046922	NM_000090	COL3A1	collagen, type III, alpha 1
8155754	NM_153267	MAMDC2	MAM domain containing 2
8057620	NM_000393	COL5A2	collagen, type V, alpha 2
8078330	NM_001003793	RBMS3	RNA binding motif, single stranded interacting protein
8174513	NM_145234	CHRDL1	chordin-like 1
7968029	NM_001014442	PCOTH	prostate collagen triple helix
7988260	NM_032892	FRMD5	FERM domain containing 5
8169272	NM_052936	ATG4A	ATG4 autophagy related 4 homolog A (S. cerevisiae)
8135915	NM_013332	HIG2	hypoxia-inducible protein 2
8095728	NM_001432	EREG	epiregulin
8082075	NM_138287	DTX3L	deltex 3-like (Drosophila)
7965335	NM_001946	DUSP6	dual specificity phosphatase 6
8117219	NM_018473	THEM2	thioesterase superfamily member 2
7916727	NM_014288	ITGB3BP	integrin beta 3 binding protein (beta3-endonexin)
8160431	BC021861	LOC554202	hypothetical LOC554202
8144742	NM_181723	EFHA2	EF-hand domain family, member A2
8120602	NM_024576	OGFRL1	opioid growth factor receptor-like 1
7980403	BC015054	C14orf133	chromosome 14 open reading frame 133
8057689	NM_022353	OSGEPL1	O-sialoglycoprotein endopeptidase-like 1
7935660	NM_015221	DNMBP	dynamin binding protein
7918034	NM_001077394	DPH5	DPH5 homolog (S. cerevisiae)

7938563	NM_001178	ARNTL	aryl hydrocarbon receptor nuclear translocator-like
7984686	NM_033028	BBS4	Bardet-Biedl syndrome 4
8085608	NM_012260	HACL1	2-hydroxyacyl-CoA lyase 1
7985119	NM_018200	HMG20A	high-mobility group 20A
8134091	NM_012129	CLDN12	claudin 12
8150014	NM_018250	INTS9	integrator complex subunit 9
8135576	NM_015641	TES	testis derived transcript (3 LIM domains)
8069744	NM_016940	RWDD2B	RWD domain containing 2B
7982597	NM_003246	THBS1	thrombospondin 1
8085984	NM_017784	OSBPL10	oxysterol binding protein-like 10
8122265	NM_006290	TNFAIP3	tumor necrosis factor, alpha-induced protein 3
8094301	NM_004787	SLIT2	slit homolog 2 (Drosophila)
8149877	NM_007257	PNMA2	paraneoplastic antigen MA2
8129497	NM_001431	EPB41L2	erythrocyte membrane protein band 4.1-like 2
7959025	NM_001109903	RNFT2	ring finger protein, transmembrane 2
8089701	NM_015642	ZBTB20	zinc finger and BTB domain containing 20
7928671	BC022252	C10orf57	chromosome 10 open reading frame 57
7975268	NM_001172	ARG2	arginase, type II
8113726	NM_000943	PPIC	peptidylprolyl isomerase C (cyclophilin C)
8019954	BC137437	FLJ35776	hypothetical LOC649446
7966026	NM_014840	NUAK1	NUAK family, SNF1-like kinase, 1
7901363	NM_001262	CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)
7913883	NM_000437	PAFAH2	platelet-activating factor acetylhydrolase 2, 40kDa
8002941	NM_199355	ADAMTS18	ADAM metallopeptidase with thrombospondin type 1 motif, 18
7991485	NM_001040655	TTC23	tetratricopeptide repeat domain 23
7964834	NM_001874	CPM	carboxypeptidase M
8136341	NM_001724	BPGM	2,3-bisphosphoglycerate mutase
8068200	BC007928	C21orf119	chromosome 21 open reading frame 119
8096314	NM_000297	PKD2	polycystic kidney disease 2 (autosomal dominant)
8120860	NM_183050	BCKDHB	branched chain keto acid dehydrogenase E1, beta polypeptide
8081241	BC006512	C3orf26	chromosome 3 open reading frame 26
7965403	NM_002345	LUM	lumican
8113800	NM_001999	FBN2	fibrillin 2
8124510	NM_003519	HIST1H2BL	histone cluster 1, H2bl
8116921	NM_001955	EDN1	endothelin 1
8106776	NM_001867	COX7C	cytochrome c oxidase subunit VIIc
8041553	NM_024775	GEMIN6	gem (nuclear organelle) associated protein 6
8150364	NM_031940	TM2D2	TM2 domain containing 2
8002904	NM_012091	ADAT1	adenosine deaminase, tRNA-specific 1
8088065	NM_001005159	SFMBT1	Scm-like with four mbt domains 1
7979813	NM_004926	ZFP36L1	zinc finger protein 36, C3H type-like 1
8003922	NM_001001683	MED11	mediator complex subunit 11
8103853	BC033172	MGC45800	hypothetical protein LOC90768
8126360	NM_004275	MED20	mediator complex subunit 20

7967230	NM_019887	DIABLO	diablo homolog (Drosophila)
8113761	NM_020747	ZNF608	zinc finger protein 608
8132515	NM_000712	BLVRA	biliverdin reductase A
8044133	NM_003581	NCK2	NCK adaptor protein 2
8120411	NM_000947	PRIM2	primase, DNA, polypeptide 2 (58kDa)
7904907	NM_004326	BCL9	B-cell CLL
7952339	NR_001453	LOC85389	RNA, small nucleolar
7954036	NM_016355	DDX47	DEAD (Asp-Glu-Ala-Asp) box polypeptide 47
8083272	NM_004130	GYG1	glycogenin 1
8111101	NM_054027	ANKH	ankylosis, progressive homolog (mouse)
8160036	AK292632	C9orf123	chromosome 9 open reading frame 123
8099570	BC050697	C4orf30	chromosome 4 open reading frame 30
7997332	NM_001105663	NUDT7	nudix (nucleoside diphosphate linked moiety X)-type motif 7
8064978	NM_000214	JAG1	jagged 1 (Alagille syndrome)
8078834	NM_020839	WDR48	WD repeat domain 48
8107814	NM_016048	ISOC1	isochorismatase domain containing 1
8173531	NM_018486	HDAC8	histone deacetylase 8
8068375	NM_058182	FAM165B	family with sequence similarity 165, member B
7940079	NM_001085458	CTNND1	catenin (cadherin-associated protein), delta 1
8023312	---		
8111153	NM_012334	MYO10	myosin X

Up-regulated

8059712	NR_004398	SNORD82	small nucleolar RNA, C
8110022	NM_003945	ATP6V0E1	ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e1
8067798	NM_018419	SOX18	SRY (sex determining region Y)-box 18
8115831	NM_004417	DUSP1	dual specificity phosphatase 1
7992402	NM_004548	NDUFB10	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10, 22kDa
8024170	NM_012292	HMHA1	histocompatibility (minor) HA-1
7977933	NM_012244	SLC7A8	solute carrier family 7 (cationic amino acid transporter, y+ system)
8165508	NM_001004354	NRARP	Notch-regulated ankyrin repeat protein
8024120	NM_019112	ABCA7	ATP-binding cassette, sub-family A (ABC1), member 7
8007471	NM_031858	NBR1	neighbor of BRCA1 gene 1
7988212	NM_025165	ELL3	elongation factor RNA polymerase II-like 3
7899821	NM_020888	KIAA1522	KIAA1522
8016463	NM_018952	HOXB6	homeobox B6
8138045	NM_014413	EIF2AK1	eukaryotic translation initiation factor 2-alpha kinase 1
8025368	NM_003083	SNAPC2	small nuclear RNA activating complex, polypeptide 2, 45kDa
8081838	NM_020754	CDGAP	Cdc42 GTPase-activating protein
8025053	NM_003811	TNFSF9	tumor necrosis factor (ligand) superfamily, member 9
8165217	NM_017617	NOTCH1	Notch homolog 1, translocation-associated (Drosophila)
7980338	NM_024496	C14orf4	chromosome 14 open reading frame 4
8068866	NM_004571	PKNOX1	PBX
8145942	NM_144652	LETM2	leucine zipper-EF-hand containing transmembrane protein 2
7948741	NM_012200	B3GAT3	beta-1,3-glucuronyltransferase 3 (glucuronosyltransferase I)
8008922	NM_003620	PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform

7998063	NM_006086	TUBB3	tubulin, beta 3
8072160	NM_032173	ZNRF3	zinc and ring finger 3
7905918	NM_004952	EFNA3	ephrin-A3
8030220	NM_003660	PPFIA3	protein tyrosine phosphatase
7920707	NM_006589	C1orf2	chromosome 1 open reading frame 2
7945371	NM_021034	IFITM3	interferon induced transmembrane protein 3 (1-8U)
8007620	NM_002087	GRN	granulin
8016457	NM_002147	HOXB5	homeobox B5
8108080	NM_015288	PHF15	PHD finger protein 15
8032839	NM_032108	SEMA6B	sema domain, transmembrane domain (TM)
8067140	NM_000782	CYP24A1	cytochrome P450, family 24, subfamily A, polypeptide 1
8119088	NM_078467	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
8104107	NM_173553	TRIML2	tripartite motif family-like 2
8010737	NM_005052	RAC3	ras-related C3 botulinum toxin substrate 3
8007008	NM_003250	THRA	thyroid hormone receptor, alpha
7963061	NM_001008223	C1QL4	complement component 1, q subcomponent-like 4
7949021	NM_173587	RCOR2	REST corepressor 2
7918694	NM_001010922	BCL2L15	BCL2-like 15
8034130	NM_015493	KANK2	KN motif and ankyrin repeat domains 2
7937330	NM_006435	IFITM2	interferon induced transmembrane protein 2 (1-8D)
7915472	NM_006516	SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1
8165538	NM_001033113	ENTPD8	ectonucleoside triphosphate diphosphohydrolase 8
8032392	NM_199054	MKNK2	MAP kinase interacting serine
7995783	NM_005953	MT2A	metallothionein 2A
8038367	NM_020309	SLC17A7	solute carrier family 17
8147351	NM_017697	RBM35A	RNA binding motif protein 35A
8158771	NM_031426	C9orf58	chromosome 9 open reading frame 58
8166469	NM_002970	SAT1	spermidine
7938295	NM_000990	RPL27A	ribosomal protein L27a
8016523	NM_004123	GIP	gastric inhibitory polypeptide
8038126	NM_001217	CA11	carbonic anhydrase XI
7981514	NM_138420	AHNAK2	AHNAK nucleoprotein 2
7945680	NR_002196	H19	H19, imprinted maternally expressed transcript
8016745	NM_003971	SPAG9	sperm associated antigen 9
8030002	NM_153608	ZNF114	zinc finger protein 114
8029530	NM_000041	APOE	apolipoprotein E
8165552	NM_015537	NELF	nasal embryonic LHRH factor
7974870	NM_003082	SNAPC1	small nuclear RNA activating complex, polypeptide 1, 43kDa
8022404	NM_001098801	C18orf19	chromosome 18 open reading frame 19
7899265	NM_006142	SFN	stratifin
8004510	NM_001251	CD68	CD68 molecule
8027002	NM_004864	GDF15	growth differentiation factor 15
8102232	NM_016269	LEF1	lymphoid enhancer-binding factor 1
8026971	NM_006332	IFI30	interferon, gamma-inducible protein 30
8034263	NM_001420	ELAVL3	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 3
8150529	NM_014420	DKK4	dickkopf homolog 4 (Xenopus laevis)

7937335	NM_003641	IFITM1	interferon induced transmembrane protein 1 (9-27)
7921099	NM_001878	CRABP2	cellular retinoic acid binding protein 2

2. List of DEGs assigned in Cluster

Cluster 1

ID	Symbol	Name	EntrezID
7921099	CRABP2	cellular retinoic acid binding protein 2	1382
8034263	ELAVL3	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 3 (Hu antigen C)	1995
8150529	DKK4	dickkopf homolog 4 (Xenopus laevis)	27121
8038126	CA11	carbonic anhydrase XI	770
8072160	ZNRF3	zinc and ring finger 3	84133
8032392	MKNK2	MAP kinase interacting serine	2872
7998063	TUBB3	tubulin, beta 3	10381
8034130	KANK2	KN motif and ankyrin repeat domains 2	25959
8026971	IFI30	interferon, gamma-inducible protein 30	10437
8102232	LEF1	lymphoid enhancer-binding factor 1	51176
8030220	PPFIA3	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 3	8541
8029530	APOE	apolipoprotein E	348
8025053	TNFSF9	tumor necrosis factor (ligand) superfamily, member 9	8744
8165508	NRARP	Notch-regulated ankyrin repeat protein	441478
7963061	C1QL4	complement component 1, q subcomponent-like 4	338761
8063078	CTSA	cathepsin A	5476
7937485	PNPLA2	patatin-like phospholipase domain containing 2	57104
8101376	SEC31A	SEC31 homolog A (S. cerevisiae)	22872
7949021	RCOR2	REST corepressor 2	283248
8064808	SLC23A2	solute carrier family 23 (nucleobase transporters), member 2	9962
8038367	SLC17A7	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 7	57030
8133459	CLIP2	CAP-GLY domain containing linker protein 2	7461
7899265	SFN	stratifin	2810
7937330	IFITM2	interferon induced transmembrane protein 2 (1-8D)	10581
8158771	C9orf58	chromosome 9 open reading frame 58	83543
8138045	EIF2AK1	eukaryotic translation initiation factor 2-alpha kinase 1	27102
7937335	IFITM1	interferon induced transmembrane protein 1 (9-27)	8519
8133721	HSPB1	heat shock 27kDa protein 1	3315
7956524	PIP4K2C	phosphatidylinositol-5-phosphate 4-kinase, type II, gamma	79837
7967544	SCARB1	scavenger receptor class B, member 1	949
8032839	SEMA6B	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6B	10501
8010737	RAC3	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)	5881
7906085	LMNA	lamin A	4000
7997427	CMIP	c-Maf-inducing protein	80790
8104107	TRIML2	tripartite motif family-like 2	205860

8092002	ARPM1	actin related protein M1	84517
8016457	HOXB5	homeobox B5	3215
7955858	HOXC10	homeobox C10	3226
8145942	LETM2	leucine zipper-EF-hand containing transmembrane protein 2	137994
7999102	TFAP4	transcription factor AP-4 (activating enhancer binding protein 4)	7023
7945680	H19	H19, imprinted maternally expressed transcript (non-protein coding)	283120
7920707	C1orf2	chromosome 1 open reading frame 2	10712
8030002	ZNF114	zinc finger protein 114	163071
7959312	NA		
8119088	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	1026
8003903	ARRB2	arrestin, beta 2	409
8007008	THRA	thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian)	7067
7940924	PLCB3	phospholipase C, beta 3 (phosphatidylinositol-specific)	5331
7956242	COQ10A	coenzyme Q10 homolog A (S. cerevisiae)	93058
8008922	PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform	8493
8000890	SEPHS2	selenophosphate synthetase 2	22928
7898957	RCAN3	RCAN family member 3	11123
8009568	TTYH2	tweety homolog 2 (Drosophila)	94015
7899821	KIAA1522	KIAA1522	57648
8016523	GIP	gastric inhibitory polypeptide	2695
7926983	CREM	cAMP responsive element modulator	1390
7994518	SPNS1	spinster homolog 1 (Drosophila)	83985
8022404	C18orf19	chromosome 18 open reading frame 19	125228
8002121	CTRL	chymotrypsin-like	1506
8076331	TOB2	transducer of ERBB2, 2	10766
7938295	RPL27A	ribosomal protein L27a	6157
7980338	C14orf4	chromosome 14 open reading frame 4	64207
7973067	NP	nucleoside phosphorylase	4860
7999079	ADCY9	adenylate cyclase 9	115
7945371	IFITM3	interferon induced transmembrane protein 3 (1-8U)	10410
8016745	SPAG9	sperm associated antigen 9	9043
8007620	GRN	granulin	2896
8032410	MOBKL2A	MOB1, Mps One Binder kinase activator-like 2A (yeast)	126308
7996772	SLC7A6	solute carrier family 7 (cationic amino acid transporter, y+ system), member 6	9057
7909285	PFKFB2	6-phosphofructo-2-kinase	5208
8159554	UAP1L1	UDP-N-acteylglucosamine pyrophosphorylase 1-like 1	91373
8165217	NOTCH1	Notch homolog 1, translocation-associated (Drosophila)	4851
8058091	SATB2	SATB homeobox 2	23314
8026991	PGPEP1	pyroglutamyl-peptidase I	54858
7937135	C10orf93	chromosome 10 open reading frame 93	255352
7992518	RAB26	RAB26, member RAS oncogene family	25837
7948399	PATL1	protein associated with topoisomerase II homolog 1 (yeast)	219988
7934393	PPP3CB	protein phosphatase 3 (formerly 2B), catalytic subunit, beta	5532

		isoform	
7952116	BCL9L	B-cell CLL	283149
8040742	MAPRE3	microtubule-associated protein, RP	22924
7981514	AHNAK2	AHNAK nucleoprotein 2	113146
7960933	M6PR	mannose-6-phosphate receptor (cation dependent)	4074
8037537	ERCC2	excision repair cross-complementing rodent repair deficiency, complementation group 2	2068
8034521	HOOK2	hook homolog 2 (Drosophila)	29911
8037525	CKM	creatine kinase, muscle	1158
7983228	MAP1A	microtubule-associated protein 1A	4130
8172088	BCOR	BCL6 co-repressor	54880
8166469	SAT1	spermidine	6303
7949264	EHD1	EH-domain containing 1	10938
7992744	FLYWCH1	FLYWCH-type zinc finger 1	84256
8131339	FSCN1	fascin homolog 1, actin-bundling protein (Strongylocentrotus purpuratus)	6624
8169598	ZCCHC12	zinc finger, CCHC domain containing 12	170261
8035318	UNC13A	unc-13 homolog A (C. elegans)	23025
7992402	NDUFB10	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10, 22kDa	4716
7986383	IGF1R	insulin-like growth factor 1 receptor	3480
7905918	EFNA3	ephrin-A3	1944
8017421	CCDC47	coiled-coil domain containing 47	57003
8029856	GRLF1	glucocorticoid receptor DNA binding factor 1	2909
8080714	FLNB	filamin B, beta (actin binding protein 278)	2317
8039362	C19orf51	chromosome 19 open reading frame 51	352909
7990253	NPTN	neuroplastin	27020
8024120	ABCA7	ATP-binding cassette, sub-family A (ABC1), member 7	10347
8118682	PHF1	PHD finger protein 1	5252
8093852	MSX1	msh homeobox 1	4487
8151816	GEM	GTP binding protein overexpressed in skeletal muscle	2669
8066214	TGM2	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	7052
8170921	PLXNA3	plexin A3	55558
8003116	HSDL1	hydroxysteroid dehydrogenase like 1	83693
7991234	MFGE8	milk fat globule-EGF factor 8 protein	4240
8011093	SLC43A2	solute carrier family 43, member 2	124935
8138728	HOXA4	homeobox A4	3201
8026687	OCEL1	occludin	79629
7940191	STX3	syntaxin 3	6809
8086330	AXUD1	AXIN1 up-regulated 1	64651
8034783	LPHN1	latrophilin 1	22859
8095870	CCNG2	cyclin G2	901
7948995	ATL3	atlastin 3	25923
7953442	GPR162	G protein-coupled receptor 162	27239
8028084	APLP1	amyloid beta (A4) precursor-like protein 1	333

8164343	FAM102A	family with sequence similarity 102, member A	399665
8086961	PFKFB4	6-phosphofructo-2-kinase	5210
7941260	FRMD8	FERM domain containing 8	83786
7918694	BCL2L15	BCL2-like 15	440603
7949067	BAD	BCL2-associated agonist of cell death	572
8165538	ENTPD8	ectonucleoside triphosphate diphosphohydrolase 8	377841
7962185	AMN1	antagonist of mitotic exit network 1 homolog (S. cerevisiae)	196394
8095341	NA		
7937341	B4GALNT4	beta-1,4-N-acetyl-galactosaminyl transferase 4	338707
8046276	SP5	Sp5 transcription factor	389058
7963590	CSAD	cysteine sulfinic acid decarboxylase	51380
8174610	LRCH2	leucine-rich repeats and calponin homology (CH) domain containing 2	57631
7959330	WDR66	WD repeat domain 66	144406
7974870	SNAPC1	small nuclear RNA activating complex, polypeptide 1, 43kDa	6617
8164293	AK1	adenylate kinase 1	203
7966851	TAOK3	TAO kinase 3	51347
8001531	MT1G	metallothionein 1G	4495
8147351	RBM35A	RNA binding motif protein 35A	54845
7949412	LTBP3	latent transforming growth factor beta binding protein 3	4054
7945536	CEND1	cell cycle exit and neuronal differentiation 1	51286
8015769	BRCA1	breast cancer 1, early onset	672
7943803	DIXDC1	DIX domain containing 1	85458
8116780	DSP	desmoplakin	1832
8084219	KLHL24	kelch-like 24 (Drosophila)	54800
8077804	TATDN2	TatD DNase domain containing 2	9797
8036865	NA		
8175871	L1CAM	L1 cell adhesion molecule	3897
8048171	DKFZp434H1419	hypothetical protein DKFZp434H1419	150967
8038117	DBP	D site of albumin promoter (albumin D-box) binding protein	1628
8126402	TRERF1	transcriptional regulating factor 1	55809
7953409	PTMS	parathymosin	5763
8024170	HMHA1	histocompatibility (minor) HA-1	23526
7977933	SLC7A8	solute carrier family 7 (cationic amino acid transporter, y+ system), member 8	23428
8036473	PPP1R14A	protein phosphatase 1, regulatory (inhibitor) subunit 14A	94274
8073548	Sep-03	septin 3	55964
7959016	FLJ42957	FLJ42957 protein	400077
8029219	TMEM145	transmembrane protein 145	284339
8126428	TRERF1	transcriptional regulating factor 1	55809
8093494	CRIPAK	cysteine-rich PAK1 inhibitor	285464
7939559	TSPAN18	tetraspanin 18	90139
7937314	ATHL1	ATH1, acid trehalase-like 1 (yeast)	80162
8001946	LOC65331	hypothetical protein LOC653319	653319

9

Cluster 2

ID	Symbol	Name	EntrezID
7915472	SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	6513
8004510	CD68	CD68 molecule	968
8165552	NELF	nasal embryonic LHRH factor	26012
8037594	RTN2	reticulon 2	6253
8025368	SNAPC2	small nuclear RNA activating complex, polypeptide 2, 45kDa	6618
8010804	NARF	nuclear prelamin A recognition factor	26502
8037079	ATP1A3	ATPase, Na+	478
8100532	C4orf14	chromosome 4 open reading frame 14	84273
8067140	CYP24A1	cytochrome P450, family 24, subfamily A, polypeptide 1	1591
8068866	PKNOX1	PBX	5316
7994769	CORO1A	coronin, actin binding protein, 1A	11151
8108080	PHF15	PHD finger protein 15	23338
7996137	KATNB1	katanin p80 (WD repeat containing) subunit B 1	10300
8008664	AKAP1	A kinase (PRKA) anchor protein 1	8165
8018305	HN1	hematological and neurological expressed 1	51155
8017718	AXIN2	axin 2	8313
8131387	USP42	ubiquitin specific peptidase 42	84132
8110022	ATP6V0E1	ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e1	8992
8000884	XTP3TPA	XTP3-transactivated protein A	79077
8056784	DLX2	distal-less homeobox 2	1746
8172471	PIM2	pim-2 oncogene	11040
8029688	CD3EAP	CD3e molecule, epsilon associated protein	10849
8012054	DLG4	discs, large homolog 4 (Drosophila)	1742
8023995	FSTL3	follistatin-like 3 (secreted glycoprotein)	10272
7948741	B3GAT3	beta-1,3-glucuronyltransferase 3 (glucuronosyltransferase I)	26229
8155707	TJP2	tight junction protein 2 (zona occludens 2)	9414
7920642	MUC1	mucin 1, cell surface associated	4582
7912994	RP1-93P18.1	hypothetical protein LOC126917	126917
8016463	HOXB6	homeobox B6	3216
7977397	CRIP2	cysteine-rich protein 2	1397
7953564	C12orf57	chromosome 12 open reading frame 57	113246
8014825	FBXL20	F-box and leucine-rich repeat protein 20	84961
7930194	CNNM2	cyclin M2	54805
8108447	CXXC5	CXXC finger 5	51523
8034202	RAB3D	RAB3D, member RAS oncogene family	9545
8039340	TNNT1	troponin T type 1 (skeletal, slow)	7138
7998055	MC1R	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)	4157
7905329	MLLT11	myeloid	10962
8014956	NR1D1	nuclear receptor subfamily 1, group D, member 1	9572

8007745	HEXIM1	hexamethylene bis-acetamide inducible 1	10614
8149330	CTSB	cathepsin B	1508
8073890	GRAMD4	GRAM domain containing 4	23151
8035249	NR2F6	nuclear receptor subfamily 2, group F, member 6	2063
7913644	E2F2	E2F transcription factor 2	1870
8070269	DSCR3	Down syndrome critical region gene 3	10311
8027002	GDF15	growth differentiation factor 15	9518
8115831	DUSP1	dual specificity phosphatase 1	1843
7967202	RHOF	ras homolog gene family, member F (in filopodia)	54509
7948839	NXF1	nuclear RNA export factor 1	10482
8067798	SOX18	SRY (sex determining region Y)-box 18	54345
7930106	TMEM180	transmembrane protein 180	79847
7997533	OSGIN1	oxidative stress induced growth inhibitor 1	29948
8037579	ERCC1	excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)	2067
7995783	MT2A	metallothionein 2A	4502
7914094	WASF2	WAS protein family, member 2	10163
7906061	SYT11	synaptotagmin XI	23208
7955873	HOXC6	homeobox C6	3223
8066117	SAMHD1	SAM domain and HD domain 1	25939
8065889	UQCC	ubiquinol-cytochrome c reductase complex chaperone	55245
8037408	KCNN4	potassium intermediate	3783
8093130	RNF168	ring finger protein 168	165918
8081838	CDGAP	Cdc42 GTPase-activating protein	57514
8016438	HOXB2	homeobox B2	3212
7936041	ARL3	ADP-ribosylation factor-like 3	403
7956009	METTL7B	methyltransferase like 7B	196410
7988212	ELL3	elongation factor RNA polymerase II-like 3	80237
8007471	NBR1	neighbor of BRCA1 gene 1	4077
8131091	MAFK	v-maf musculoaponeurotic fibrosarcoma oncogene homolog K (avian)	7975
8133728	ZP3	zona pellucida glycoprotein 3 (sperm receptor)	7784
8076547	TTLL1	tubulin tyrosine ligase-like family, member 1	25809
7993713	IQCK	IQ motif containing K	124152
8155268	POLR1E	polymerase (RNA) I polypeptide E, 53kDa	64425
8101411	LIN54	lin-54 homolog (C. elegans)	132660
7915160	RRAGC	Ras-related GTP binding C	64121
8102482	SEC24D	SEC24 related gene family, member D (S. cerevisiae)	9871
8141728	RABL5	RAB, member RAS oncogene family-like 5	64792
8016452	HOXB4	homeobox B4	3214
7999553	FLJ11151	hypothetical protein FLJ11151	55313
8114526	DNAJC18	DnaJ (Hsp40) homolog, subfamily C, member 18	202052
7930454	PDCD4	programmed cell death 4 (neoplastic transformation inhibitor)	27250
8040103	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	3398

8074364	DGCR14	DiGeorge syndrome critical region gene 14	8220
8021653	SERPINB8	serpin peptidase inhibitor, clade B (ovalbumin), member 8	5271
8058388			
8055952	NR4A2	nuclear receptor subfamily 4, group A, member 2	4929
8034390	ZNF799	zinc finger protein 799	90576

Cluster 3

ID	Symbol
8029347	ZNF230
7990429	C15orf17
8125818	C6orf1
8084880	HES1
7953695	NA
7920875	SCARNA4
8063345	SNORD12C
7928821	NA
8055965	NA

Cluster 4

ID	Symbol	Name	EntrezID
8155754	MAMDC2	MAM domain containing 2	256691
8135576	TES	testis derived transcript (3 LIM domains)	26136
8085608	HACL1	2-hydroxyacyl-CoA lyase 1	26061
8116921	EDN1	endothelin 1	1906
8117219	THEM2	thioesterase superfamily member 2	55856
8078544	MLH1	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	4292
8096635	NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	4790
8150014	INTS9	integrator complex subunit 9	55756
8120411	PRIM2	primase, DNA, polypeptide 2 (58kDa)	5558
8132515	BLVRA	biliverdin reductase A	644
7988687	GABPB1	GA binding protein transcription factor, beta subunit 2	2553
8129497	EPB41L2	erythrocyte membrane protein band 4.1-like 2	2037
7928107	H2AFY2	H2A histone family, member Y2	55506
7936734	FGFR2	fibroblast growth factor receptor 2	2263
8068200	C21orf119	chromosome 21 open reading frame 119	84996
8106709	ATG10	ATG10 autophagy related 10 homolog (S. cerevisiae)	83734
8068375	FAM165B	family with sequence similarity 165, member B	54065
8164587	TOR1A	torsin family 1, member A (torsin A)	1861
8175250			
7928671	C10orf57	chromosome 10 open reading frame 57	80195
7924388	BPNT1	3'(2'), 5'-bisphosphate nucleotidase 1	10380
7977841	C14orf94	chromosome 14 open reading frame 94	54930
7971184	MRPS31	mitochondrial ribosomal protein S31	10240
8053496	POLR1A	polymerase (RNA) I polypeptide A, 194kDa	25885
8157843	RABEPK	Rab9 effector protein with kelch motifs	10244

8069676	ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif, 1	9510
8150276	PPAPDC1B	phosphatidic acid phosphatase type 2 domain containing 1B	84513
8124416	HIST1H3D	histone cluster 1, H3d	8351
7943919	TTC12	tetratricopeptide repeat domain 12	54970
8050060	TSSC1	tumor suppressing subtransferable candidate 1	7260
7921970	ALDH9A1	aldehyde dehydrogenase 9 family, member A1	223
8173531	HDAC8	histone deacetylase 8	55869
7913858	PAQR7	progestin and adipoQ receptor family member VII	164091
8148955	C8orf33	chromosome 8 open reading frame 33	65265
7991216	DET1	de-etiolated homolog 1 (Arabidopsis)	55070
7991209	MRPL46	mitochondrial ribosomal protein L46	26589
7930148	SFXN2	sideroflexin 2	118980
7904695	SEC22B	SEC22 vesicle trafficking protein homolog B (S. cerevisiae)	9554
7926661	MSRB2	methionine sulfoxide reductase B2	22921
7967287	ZCCHC8	zinc finger, CCHC domain containing 8	55596
8149877	PNMA2	paraneoplastic antigen MA2	10687
7925480	FH	fumarate hydratase	2271
8115210	TNIP1	TNFAIP3 interacting protein 1	10318
8169240	PRPS1	phosphoribosyl pyrophosphate synthetase 1	5631
7938528	PARVA	parvin, alpha	55742
8169028	NGFRAP1	nerve growth factor receptor (TNFRSF16) associated protein 1	27018
8126058	PPIL1	peptidylprolyl isomerase (cyclophilin)-like 1	51645
8103188	PET112L	PET112-like (yeast)	5188
7952339	LOC85389	RNA, small nucleolar	85389
7905088	HIST2H2AC	histone cluster 2, H2ac	8338
8122242	PEX7	peroxisomal biogenesis factor 7	5191
8021727	CNDP2	CNDP dipeptidase 2 (metallopeptidase M20 family)	55748
8054377	FHL2	four and a half LIM domains 2	2274
8002266	CTF8	chromosome transmission fidelity factor 8 homolog (S. cerevisiae)	54921
8054077	ACTR1B	ARP1 actin-related protein 1 homolog B, centractin beta (yeast)	10120
8118319			
8059249	OBSL1	obscurin-like 1	23363
8160036	C9orf123	chromosome 9 open reading frame 123	90871
8108370	EGR1	early growth response 1	1958
8169272	ATG4A	ATG4 autophagy related 4 homolog A (S. cerevisiae)	115201
7921526	PIGM	phosphatidylinositol glycan anchor biosynthesis, class M	93183
7944869	SPA17	sperm autoantigenic protein 17	53340
8150364	TM2D2	TM2 domain containing 2	83877
7902290	CTH	cystathionase (cystathionine gamma-lyase)	1491
8043848	MRPL30	mitochondrial ribosomal protein L30	51263

7935058	FER1L3	fer-1-like 3, myoferlin (C. elegans)	26509
8124423	HIST1H2BG	histone cluster 1, H2bg	8339
8049532	LRRFIP1	leucine rich repeat (in FLII) interacting protein 1	9208
7978644	NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	4792
8156295	SECISBP2	SECIS binding protein 2	79048
7953594	EMG1	EMG1 nucleolar protein homolog (S. cerevisiae)	10436
8044133	NCK2	NCK adaptor protein 2	8440
8081241	C3orf26	chromosome 3 open reading frame 26	84319
7985757	MRPS11	mitochondrial ribosomal protein S11	64963
8062312	MYL9	myosin, light chain 9, regulatory	10398
8138602	DFNA5	deafness, autosomal dominant 5	1687
7933186	ZNF32	zinc finger protein 32	7580
7899436	SESN2	sestrin 2	83667
8117426	HIST1H2BH	histone cluster 1, H2bh	8345
7942315	LRRC51	leucine rich repeat containing 51	220074
7909561	C1orf97	chromosome 1 open reading frame 97	84791
7932209			
8123678	C6orf145	chromosome 6 open reading frame 145	221749
7965403	LUM	lumican	4060
8008933	BCAS3	breast carcinoma amplified sequence 3	54828
8160431	LOC554202	hypothetical LOC554202	554202
7903203	SNX7	sorting nexin 7	51375
7925033	C1orf131	chromosome 1 open reading frame 131	128061
8055668			
7964248	SNORD59A	small nucleolar RNA, C	26789
8115164			
8005951	SNORD42B	small nucleolar RNA, C	26808
8114468	SNORD63	small nucleolar RNA, C	26785
8001748	SNORA50	small nucleolar RNA, H	677830
7922412	SNORD77	small nucleolar RNA, C	692197
8059712	SNORD82	small nucleolar RNA, C	25826
8062890	PABPC1L	poly(A) binding protein, cytoplasmic 1-like	80336
7899392	SCARNA1	small Cajal body-specific RNA 1	677774
8174507	SNORD96B	small nucleolar RNA, C	692226

Cluster 5

ID	**Symbol**	**Name**	**EntrezID**

7985119	HMG20A	high-mobility group 20A	10363
8085984	OSBPL10	oxysterol binding protein-like 10	114884
7979813	ZFP36L1	zinc finger protein 36, C3H type-like 1	677
8113761	ZNF608	zinc finger protein 608	57507
7967230	DIABLO	diablo homolog (Drosophila)	56616
8103853	MGC45800	hypothetical protein LOC90768	90768
8046922	COL3A1	collagen, type III, alpha 1	1281
8102214	PAPSS1	3'-phosphoadenosine 5'-phosphosulfate synthase 1	9061
8107814	ISOC1	isochorismatase domain containing 1	51015
7918034	DPH5	DPH5 homolog (S. cerevisiae)	51611
8128383	COQ3	coenzyme Q3 homolog, methyltransferase (S. cerevisiae)	51805
8136341	BPGM	2,3-bisphosphoglycerate mutase	669
8069744	RWDD2B	RWD domain containing 2B	10069
8099570	C4orf30	chromosome 4 open reading frame 30	54876
7975268	ARG2	arginase, type II	384
8003922	MED11	mediator complex subunit 11	400569
8111101	ANKH	ankylosis, progressive homolog (mouse)	56172
7935660	DNMBP	dynamin binding protein	23268
7959025	RNFT2	ring finger protein, transmembrane 2	84900
8123961	TBC1D7	TBC1 domain family, member 7	51256
7938301	C11orf17	chromosome 11 open reading frame 17	56672
8078834	WDR48	WD repeat domain 48	57599
7903803	AHCYL1	S-adenosylhomocysteine hydrolase-like 1	10768
8106776	COX7C	cytochrome c oxidase subunit VIIc	1350
8082816	SRPRB	signal recognition particle receptor, B subunit	58477
8057689	OSGEPL1	O-sialoglycoprotein endopeptidase-like 1	64172
7968029	PCOTH	prostate collagen triple helix	542767
7951781	C11orf71	chromosome 11 open reading frame 71	54494
7923516	CYB5R1	cytochrome b5 reductase 1	51706
8131583	BZW2	basic leucine zipper and W2 domains 2	28969
8124510	HIST1H2BL	histone cluster 1, H2bl	8340
8177232	JARID1D	jumonji, AT rich interactive domain 1D	8284
7964834	CPM	carboxypeptidase M	1368
8143188	CREB3L2	cAMP responsive element binding protein 3-like 2	64764
8084092	NDUFB5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa	4711
8025414	RAB11B	RAB11B, member RAS oncogene family	9230
8041179	CLIP4	CAP-GLY domain containing linker protein family, member 4	79745
8023261	ACAA2	acetyl-Coenzyme A acyltransferase 2	10449
8045349	MGAT5	mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase	4249
8059319	FARSB	phenylalanyl-tRNA synthetase, beta subunit	10056
8144802	PDGFRL	platelet-derived growth factor receptor-like	5157

7952069	C11orf60	chromosome 11 open reading frame 60	56912
7943413	BIRC3	baculoviral IAP repeat-containing 3	330
7934178	PCBD1	pterin-4 alpha-carbinolamine dehydratase	5092
8117408	HIST1H2AE	histone cluster 1, H2ae	3012
8081358	LOC100009676	Putative uncharacterized protein LOC100009676	100009676
8154100	VLDLR	very low density lipoprotein receptor	7436
8117580	HIST1H2AI	histone cluster 1, H2ai	8329
7906662	UFC1	ubiquitin-fold modifier conjugating enzyme 1	51506
7939411	C11orf74	chromosome 11 open reading frame 74	119710
8126360	MED20	mediator complex subunit 20	9477
7900167	CDCA8	cell division cycle associated 8	55143
8123695	PECI	peroxisomal D3,D2-enoyl-CoA isomerase	10455
7903786	CSF1	colony stimulating factor 1 (macrophage)	1435
8089993	WDR5B	WD repeat domain 5B	54554
8145470	DPYSL2	dihydropyrimidinase-like 2	1808
8043187	MAT2A	methionine adenosyltransferase II, alpha	4144
8149942	CCDC25	coiled-coil domain containing 25	55246
8123802	TXNDC5	thioredoxin domain containing 5	81567
8114425	CDC25C	cell division cycle 25 homolog C (S. pombe)	995
8124385	HIST1H4B	histone cluster 1, H4b	8366
7918925	TRIM45	tripartite motif-containing 45	80263
7961483	HIST4H4	histone cluster 4, H4	121504
8173261	KIAA1166	KIAA1166	55906
7982723	IVD	isovaleryl Coenzyme A dehydrogenase	3712
8113773	ALDH7A1	aldehyde dehydrogenase 7 family, member A1	501
8011407	TAX1BP3	Tax1 (human T-cell leukemia virus type I) binding protein 3	30851
8001537	NIP30	NEFA-interacting nuclear protein NIP30	80011
8124537	HIST1H3J	histone cluster 1, H3j	8356

Cluster 6

ID	Symbol	Name	EntrezID
7965335	DUSP6	dual specificity phosphatase 6	1848
8174513	CHRDL1	chordin-like 1	91851
7988260	FRMD5	FERM domain containing 5	84978
7966026	NUAK1	NUAK family, SNF1-like kinase, 1	9891
7984686	BBS4	Bardet-Biedl syndrome 4	585
7982597	THBS1	thrombospondin 1	7057
7919780	GOLPH3L	golgi phosphoprotein 3-like	55204
7954036	DDX47	DEAD (Asp-Glu-Ala-Asp) box polypeptide 47	51202
8096160	ARHGAP24	Rho GTPase activating protein 24	83478
8103859	DCTD	dCMP deaminase	1635

8096314	PKD2	polycystic kidney disease 2 (autosomal dominant)	5311
8088065	SFMBT1	Scm-like with four mbt domains 1	51460
8120602	OGFRL1	opioid growth factor receptor-like 1	79627
7971565	P2RY5	purinergic receptor P2Y, G-protein coupled, 5	10161
8106730	XRCC4	X-ray repair complementing defective repair in Chinese hamster cells 4	7518
7951259	MMP10	matrix metallopeptidase 10 (stromelysin 2)	4319
8002904	ADAT1	adenosine deaminase, tRNA-specific 1	23536
7913883	PAFAH2	platelet-activating factor acetylhydrolase 2, 40kDa	5051
8102792	PCDH18	protocadherin 18	54510
8019842	TYMS	thymidylate synthetase	7298
8144742	EFHA2	EF-hand domain family, member A2	286097
8019954	FLJ35776	hypothetical LOC649446	649446
8094301	SLIT2	slit homolog 2 (Drosophila)	9353
8160602	APTX	aprataxin	54840
7972217	SPRY2	sprouty homolog 2 (Drosophila)	10253
8083272	GYG1	glycogenin 1	2992
8091103	TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)	7029
7901363	CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	1031
8117120	ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	3400
7932616	ABI1	abl-interactor 1	10006
8045860	PKP4	plakophilin 4	8502
8134091	CLDN12	claudin 12	9069
8058373	WDR12	WD repeat domain 12	55759
8113726	PPIC	peptidylprolyl isomerase C (cyclophilin C)	5480
8078330	RBMS3	RNA binding motif, single stranded interacting protein	27303
7980403	C14orf133	chromosome 14 open reading frame 133	63894
8057620	COL5A2	collagen, type V, alpha 2	1290
7918323	SORT1	sortilin 1	6272
8122265	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	7128
8120860	BCKDHB	branched chain keto acid dehydrogenase E1, beta polypeptide	594
8116983	CD83	CD83 molecule	9308
8041553	GEMIN6	gem (nuclear organelle) associated protein 6	79833
8075673	RBM9	RNA binding motif protein 9	23543
7925904	AKR1CL2	aldo-keto reductase family 1, member C-like 2	83592
8092328	MCCC1	methylcrotonoyl-Coenzyme A carboxylase 1 (alpha)	56922
8058258	ALS2CR4	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 4	65062
8147573	OSR2	odd-skipped related 2 (Drosophila)	116039
7940079	CTNND1	catenin (cadherin-associated protein), delta 1	1500
7979551	PPP2R5E	protein phosphatase 2, regulatory subunit B', epsilon isoform	5529
7938563	ARNTL	aryl hydrocarbon receptor nuclear translocator-like	406
7964271	PRIM1	primase, DNA, polypeptide 1 (49kDa)	5557
8175252			

8097867	KIAA0922	KIAA0922	23240
8084766	TP63	tumor protein p63	8626
8121685	DCBLD1	discoidin, CUB and LCCL domain containing 1	285761
8111153	MYO10	myosin X	4651
8123274	MAP3K4	mitogen-activated protein kinase kinase kinase 4	4216
8156199	DAPK1	death-associated protein kinase 1	1612
7997332	NUDT7	nudix (nucleoside diphosphate linked moiety X)-type motif 7	283927
8111430	AMACR	alpha-methylacyl-CoA racemase	23600
7916727	ITGB3BP	integrin beta 3 binding protein (beta3-endonexin)	23421
7951614	PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A, beta isoform	5519
8084165	SOX2	SRY (sex determining region Y)-box 2	6657
8171205	NLGN4X	neuroligin 4, X-linked	57502
8138474	SP8	Sp8 transcription factor	221833
8042259	MDH1	malate dehydrogenase 1, NAD (soluble)	4190
8109732	MAT2B	methionine adenosyltransferase II, beta	27430
8135392	HBP1	HMG-box transcription factor 1	26959
8114320	HNRNPA0	heterogeneous nuclear ribonucleoprotein A0	10949
7991485	TTC23	tetratricopeptide repeat domain 23	64927
7912166	RERE	arginine-glutamic acid dipeptide (RE) repeats	473
8081953	GTF2E1	general transcription factor IIE, polypeptide 1, alpha 56kDa	2960
8095834	SHROOM3	shroom family member 3	57619
7935403	ARHGAP19	Rho GTPase activating protein 19	84986
8015039	SMARCE1	SWI	6605
8122409	PEX3	peroxisomal biogenesis factor 3	8504
8082229	UMPS	uridine monophosphate synthetase	7372
8042495	RY1	putative nucleic acid binding protein RY-1	11017
8145736	NRG1	neuregulin 1	3084
8022902	INO80C	INO80 complex subunit C	125476
8103911	IRF2	interferon regulatory factor 2	3660
8095728	EREG	epiregulin	2069
7922646	IFRG15	interferon responsive gene 15	64163
7932433	NSUN6	NOL1	221078
8088151	ACTR8	ARP8 actin-related protein 8 homolog (yeast)	93973
8097692	EDNRA	endothelin receptor type A	1909
7922268	KIFAP3	kinesin-associated protein 3	22920
7985349	MESDC1	mesoderm development candidate 1	59274
8113800	FBN2	fibrillin 2	2201
7904018 8175248	CTTNBP2NL	CTTNBP2 N-terminal like	55917
8111993	FGF10	fibroblast growth factor 10	2255
8015969	UBTF	upstream binding transcription factor, RNA polymerase I	7343

7975238	PLEKHH1	pleckstrin homology domain containing, family H (with MyTH4 domain) member 1	57475
8082075	DTX3L	deltex 3-like (Drosophila)	151636
8078147	BTD	biotinidase	686
8105040	OSMR	oncostatin M receptor	9180
8113413	NUDT12	nudix (nucleoside diphosphate linked moiety X)-type motif 12	83594
8088803	EIF4E3	eukaryotic translation initiation factor 4E family member 3	317649
8099235	MRFAP1L1	Morf4 family associated protein 1-like 1	114932
8069880	TIAM1	T-cell lymphoma invasion and metastasis 1	7074
8054364	TGFBRAP1	transforming growth factor, beta receptor associated protein 1	9392
8108683	PCDHB2	protocadherin beta 2	56133
7901592	C1orf83	chromosome 1 open reading frame 83	127428
8117537	HIST1H4I	histone cluster 1, H4i	8294
8121749	GJA1	gap junction protein, alpha 1, 43kDa	2697
8097307	INTU	inturned planar cell polarity effector homolog (Drosophila)	27152
7929032	FAS	Fas (TNF receptor superfamily, member 6)	355
8089701	ZBTB20	zinc finger and BTB domain containing 20	26137
8064978	JAG1	jagged 1 (Alagille syndrome)	182
8128284	EPHA7	EPH receptor A7	2045
8112709	C5orf37	chromosome 5 open reading frame 37	134359
8135915	HIG2	hypoxia-inducible protein 2	29923
8134201	GATAD1	GATA zinc finger domain containing 1	57798
8046792	DUSP19	dual specificity phosphatase 19	142679
8158725	ABL1	c-abl oncogene 1, receptor tyrosine kinase	25
8069532	HSPA13	heat shock protein 70kDa family, member 13	6782
8102311	CASP6	caspase 6, apoptosis-related cysteine peptidase	839
8060977	C20orf94	chromosome 20 open reading frame 94	128710
8021565	PHLPP	PH domain and leucine rich repeat protein phosphatase	23239
8102468	PRSS12	protease, serine, 12 (neurotrypsin, motopsin)	8492
7959002	SDSL	serine dehydratase-like	113675
7913864	C1orf215	chromosome 1 open reading frame 215	149421
8147439	PLEKHF2	pleckstrin homology domain containing, family F (with FYVE domain) member 2	79666
8143327	PARP12	poly (ADP-ribose) polymerase family, member 12	64761
8002941	ADAMTS18	ADAM metallopeptidase with thrombospondin type 1 motif, 18	170692
7969533	SLAIN1	SLAIN motif family, member 1	122060
7995552	CYLD	cylindromatosis (turban tumor syndrome)	1540
7904907	BCL9	B-cell CLL	283149
8100085	GNPDA2	glucosamine-6-phosphate deaminase 2	132789
8087201	IHPK2	inositol hexaphosphate kinase 2	51447
7956826	TBC1D30	TBC1 domain family, member 30	23329
7938390	ADM	adrenomedullin	133

8030391			
8146198	POLB	polymerase (DNA directed), beta	5423
8155824	TMC1	transmembrane channel-like 1	117531
8174086	ARMCX6	armadillo repeat containing, X-linked 6	54470
8107044	ERAP2	endoplasmic reticulum aminopeptidase 2	64167

3. CoPub gene list

cell differentiation - 25 genes

name	symbol	abstract count
cysteine rich protein 2	CRIP2	6
SRY box 18	SOX18	4
coronin,actin binding protein,1A	CORO1A	4
inhibitor of DNA binding 2	ID2A	3
v maf musculoaponeurotic fibrosarcoma oncogene homolog K	MAFK	12
hexamethylene bis acetamide inducible 1	HEXIM1	3
follistatin like 3	FSTL3	3
homeobox B2	HOXB2	8
RAB3D,member RAS oncogene family	RAB3D	4
homeobox C6	HOXC6	4
homeobox B4	HOXB4	9
nuclear receptor subfamily 4,group A,member 2	NR4A2	12
homeobox B6	HOXB6	4
E2F transcription factor 2	E2F2	4
cytochrome P450,family 24,subfamily A,polypeptide 1	CYP24A1	7
mucin 1,cell surface associated	MUC1	45
potassium intermediate/small conductance calcium activated channel	KCNN4	5
dual specificity phosphatase 1	DUSP1	8
solute carrier family 2,member 1	SLC2A1	29
growth differentiation factor 15	GDF15	4
CD68 molecule	CD68	40
cathepsin B	CTSB	22
zona pellucida glycoprotein 3	ZP3	3
discs,large homolog 4	DLG4	5

Cell growth – 39 genes

name	symbol	abstract count
interferon regulatory factor 2	IRF2	40
transforming growth factor,beta receptor associated protein 1	TGFBRAP 1	3
PH domain and leucine rich repeat protein phosphatase	PHLPP	4
oncostatin M receptor	OSMR	5
transcription factor Dp 2	TFDP2	4

epiregulin	EREG	11
cyclin dependent kinase inhibitor 2C	CDKN2C	27
neuregulin 1	NRG1	74
upstream binding transcription factor,RNA polymerase I	UBTF	16
T cell lymphoma invasion and metastasis 1	TIAM1	11
dual specificity phosphatase 6	DUSP6	11
thymidylate synthetase	TYMS	163
tumor necrosis factor,alpha induced protein 3	TNFAIP3	4
HMG box transcription factor 1	HBP1	3
jagged 1	JAG1	24
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	249
gap junction protein,alpha 1,43kDa	GJA1	107
thrombospondin 1	THBS1	77
RNA binding motif protein 9	RBM9	4
sprouty homolog 2	SPRY2	3
dCMP deaminase	DCTD	6
cylindromatosis	CYLD	4
catenin,delta 1	CTNND1	15
Fas	APT1	409
polycystic kidney disease 2	PKD2	12
tumor protein p73 like	TP73L	26
death associated protein kinase 1	DAPK1	6
shroom family member 3	SHROOM 3	4
abl interactor 1	ABI1	4
adrenomedullin	ADM	38
caspase 6,apoptosis related cysteine peptidase	CASP6	10
polymerase,beta	POLB	25
endothelin receptor type A	EDNRA	34
SRY box 2	SOX2	15
fibroblast growth factor 10	FGF10	6
uridine monophosphate synthetase	UMPS	4
aryl hydrocarbon receptor nuclear translocator like	ARNTL	4
matrix metallopeptidase 10	MMP10	6
CD83 molecule	CD83	9

apoptosis– 49 genes

name	**symbol**	**abstract count**
Fas	APT1	11734
caspase 6,apoptosis related cysteine peptidase	CASP6	332
inositol hexaphosphate kinase 2	IHPK2	7
death associated protein kinase 1	DAPK1	80
tumor protein p73 like	TP73L	231
tumor necrosis factor,alpha induced protein 3	TNFAIP3	24
transcription factor Dp 2	TFDP2	10

transforming growth factor,beta receptor associated protein 1	TGFBRAP1	4
mitogen activated protein kinase kinase kinase 4	MAP3K4	11
Rho GTPase activating protein 24	ARHGAP24	3
hypoxia inducible protein 2	HIG2	3
cyclin dependent kinase inhibitor 2C	CDKN2C	52
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	889
dual specificity phosphatase 6	DUSP6	28
interferon regulatory factor 2	IRF2	30
oncostatin M receptor	OSMR	5
methionine adenosyltransferase II,beta	MAT2B	4
thrombospondin 1	THBS1	168
cylindromatosis	CYLD	12
PH domain and leucine rich repeat protein phosphatase	PHLPP	3
CD83 molecule	CD83	87
jagged 1	JAG1	38
X ray repair complementing defective repair in Chinese hamster cells 4	XRCC4	19
T cell lymphoma invasion and metastasis 1	TIAM1	13
HMG box transcription factor 1	HBP1	4
sprouty homolog 2	SPRY2	6
polycystic kidney disease 2	PKD2	25
matrix metallopeptidase 10	MMP10	36
neuregulin 1	NRG1	66
fibroblast growth factor 10	FGF10	28
sortilin 1	SORT1	3
gap junction protein,alpha 1,43kDa	GJA1	137
upstream binding transcription factor,RNA polymerase I	UBTF	10
thymidylate synthetase	TYMS	143
EPH receptor A7	EPHA7	3
adrenomedullin	ADM	73
aprataxin	APTX	5
epiregulin	EREG	5
RNA binding motif protein 9	RBM9	4
catenin,delta 1	CTNND1	16
peptidylprolyl isomerase C	PPIC	3
polymerase,beta	POLB	50
slit homolog 2	SLIT2	3
Sp8 transcription factor	SP8	3
alpha methylacyl CoA racemase	AMACR	5
fibrillin 2	FBN2	3
SRY box 2	SOX2	21
aryl hydrocarbon receptor nuclear translocator like	ARNTL	9
abl interactor 1	ABI1	3

cell adhesion – 27 genes

name	**symbol**	**abstract count**

catenin,delta 1	CTNND1	178
neuroligin 4,X linked	NLGN4X	8
plakophilin 4	PKP4	6
myosin X	MYO10	6
T cell lymphoma invasion and metastasis 1	TIAM1	16
thrombospondin 1	THBS1	143
shroom family member 3	SHROOM3	13
tumor necrosis factor,alpha induced protein 3	TNFAIP3	3
EPH receptor A7	EPHA7	3
fibrillin 2	FBN2	4
epiregulin	EREG	4
gap junction protein,alpha 1,43kDa	GJA1	67
abl interactor 1	ABI1	5
death associated protein kinase 1	DAPK1	6
polycystic kidney disease 2	PKD2	8
dual specificity phosphatase 6	DUSP6	4
SRY box 2	SOX2	15
jagged 1	JAG1	9
tumor protein p73 like	TP73L	16
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	89
matrix metallopeptidase 10	MMP10	10
Fas	APT1	166
interferon regulatory factor 2	IRF2	3
fibroblast growth factor 10	FGF10	5
CD83 molecule	CD83	13
caspase 6,apoptosis related cysteine peptidase	CASP6	4
neuregulin 1	NRG1	11

growth – 61 genes

name	**symbol**	**abstract count**
transforming growth factor,beta receptor associated protein 1	TGFBRAP 1	19
fibroblast growth factor 10	FGF10	390
epiregulin	EREG	127
neuregulin 1	NRG1	837
sprouty homolog 2	SPRY2	61
nudix type motif 7	NUDT7	3
thrombospondin 1	THBS1	827
protein phosphatase 2,regulatory subunit A,beta isoform	PPP2R1B	6
transcription factor Dp 2	TFDP2	14
dual specificity phosphatase 6	DUSP6	73
slit homolog 2	SLIT2	35
oncostatin M receptor	OSMR	14
cyclin dependent kinase inhibitor 2C	CDKN2C	103

interferon regulatory factor 2	IRF2	71
jagged 1	JAG1	162
T cell lymphoma invasion and metastasis 1	TIAM1	50
abl interactor 1	ABI1	57
B cell CLL/lymphoma 9 like	BCL9L	4
chordin like 1	CHRDL1	7
WD repeat domain 12	WDR12	3
inositol hexaphosphate kinase 2	IHPK2	4
mitogen activated protein kinase kinase kinase 4	MAP3K4	12
hypoxia inducible protein 2	HIG2	3
upstream binding transcription factor,RNA polymerase I	UBTF	54
thymidylate synthetase	TYMS	701
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	1431
PH domain and leucine rich repeat protein phosphatase	PHLPP	7
fibrillin 2	FBN2	26
purinergic receptor P2Y,G protein coupled,5	P2RY5	3
EPH receptor A7	EPHA7	16
HMG box transcription factor 1	HBP1	12
gap junction protein,alpha 1,43kDa	GJA1	545
SRY box 2	SOX2	156
NUAK family,SNF1 like kinase,1	NUAK1	3
tumor protein p73 like	TP73L	161
catenin,delta 1	CTNND1	82
matrix metallopeptidase 10	MMP10	135
plakophilin 4	PKP4	5
uridine monophosphate synthetase	UMPS	47
Fas	APT1	2598
methionine adenosyltransferase II,beta	MAT2B	6
adrenomedullin	ADM	288
primase,polypeptide 1,49kDa	PRIM1	3
Sp8 transcription factor	SP8	15
death associated protein kinase 1	DAPK1	32
cylindromatosis	CYLD	18
shroom family member 3	SHROOM 3	24
methylcrotonoyl Coenzyme A carboxylase 1	MCCC1	4
myosin X	MYO10	7
collagen,type V,alpha 2	COL5A2	6
caspase 6,apoptosis related cysteine peptidase	CASP6	50
RNA binding motif protein 9	RBM9	13
sortilin 1	SORT1	6
peptidylprolyl isomerase C	PPIC	11
polycystic kidney disease 2	PKD2	48
peroxisomal biogenesis factor 3	PEX3	5
glycogenin 1	GYG1	8
tumor necrosis factor,alpha induced protein 3	TNFAIP3	7

endothelin receptor type A	EDNRA	219
dCMP deaminase	DCTD	18
protease,serine,12	PRSS12	4

cell proliferation - 41 genes

name	**symbol**	**abstract count**
WD repeat domain 12	WDR12	5
sprouty homolog 2	SPRY2	16
transcription factor Dp 2	TFDP2	9
fibroblast growth factor 10	FGF10	71
cyclin dependent kinase inhibitor 2C	CDKN2C	46
CD83 molecule	CD83	193
epiregulin	EREG	18
primase,polypeptide 1,49kDa	PRIM1	3
oncostatin M receptor	OSMR	5
dual specificity phosphatase 6	DUSP6	21
polycystic kidney disease 2	PKD2	31
neuregulin 1	NRG1	111
thrombospondin 1	THBS1	138
HMG box transcription factor 1	HBP1	5
PH domain and leucine rich repeat protein phosphatase	PHLPP	3
jagged 1	JAG1	33
tumor protein p73 like	TP73L	56
upstream binding transcription factor,RNA polymerase I	UBTF	15
tumor necrosis factor,alpha induced protein 3	TNFAIP3	5
catenin,delta 1	CTNND1	25
gap junction protein,alpha 1,43kDa	GJA1	139
abl interactor 1	ABI1	7
Fas	APT1	709
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	281
T cell lymphoma invasion and metastasis 1	TIAM1	7
cylindromatosis	CYLD	5
SRY box 2	SOX2	36
caspase 6,apoptosis related cysteine peptidase	CASP6	17
interferon regulatory factor 2	IRF2	10
adrenomedullin	ADM	66
endothelin receptor type A	EDNRA	74
aryl hydrocarbon receptor nuclear translocator like	ARNTL	12
thymidylate synthetase	TYMS	88
Sp8 transcription factor	SP8	3
dCMP deaminase	DCTD	4
death associated protein kinase 1	DAPK1	4
polymerase,beta	POLB	23
matrix metallopeptidase 10	MMP10	10
X ray repair complementing defective repair in Chinese hamster cells	XRCC4	3

4		
uridine monophosphate synthetase	UMPS	4
biotinidase	BTD	4

tissue regeneration – 5 genes

name	**symbol**	**abstract count**
SRY box 2	SOX2	7
fibroblast growth factor 10	FGF10	3
jagged 1	JAG1	3
thrombospondin 1	THBS1	3
gap junction protein,alpha 1,43kDa	GJA1	4

cell migration – 23 genes

name	**symbol**	**abstract count**
slit homolog 2	SLIT2	18
T cell lymphoma invasion and metastasis 1	TIAM1	30
shroom family member 3	SHROOM3	13
sprouty homolog 2	SPRY2	7
myosin X	MYO10	3
catenin,delta 1	CTNND1	27
abl interactor 1	ABI1	9
thrombospondin 1	THBS1	81
sortilin 1	SORT1	3
EPH receptor A7	EPHA7	3
matrix metallopeptidase 10	MMP10	13
neuregulin 1	NRG1	28
jagged 1	JAG1	7
cyclin dependent kinase inhibitor 2C	CDKN2C	3
gap junction protein,alpha 1,43kDa	GJA1	30
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	44
fibroblast growth factor 10	FGF10	3
tumor protein p73 like	TP73L	5
adrenomedullin	ADM	10
CD83 molecule	CD83	6
Fas	APT1	54
SRY box 2	SOX2	3
endothelin receptor type A	EDNRA	6

Phosphorylation – 50 genes

name	**symbol**	**abstract count**
NUAK family,SNF1 like kinase,1	NUAK1	7
B cell CLL/lymphoma 9 like	BCL9L	5
oncostatin M receptor	OSMR	11
dual specificity phosphatase 6	DUSP6	50

PH domain and leucine rich repeat protein phosphatase	PHLPP	9
sprouty homolog 2	SPRY2	23
catenin,delta 1	CTNND1	102
mitogen activated protein kinase kinase kinase 4	MAP3K4	11
epiregulin	EREG	25
T cell lymphoma invasion and metastasis 1	TIAM1	32
cyclin dependent kinase inhibitor 2C	CDKN2C	61
neuregulin 1	NRG1	210
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	914
gap junction protein,alpha 1,43kDa	GJA1	397
upstream binding transcription factor,RNA polymerase I	UBTF	36
chordin like 1	CHRDL1	4
transcription factor Dp 2	TFDP2	6
abl interactor 1	ABI1	28
X ray repair complementing defective repair in Chinese hamster cells 4	XRCC4	31
shroom family member 3	SHROOM3	20
EPH receptor A7	EPHA7	6
death associated protein kinase 1	DAPK1	24
glycogenin 1	GYG1	8
Sp8 transcription factor	SP8	10
methionine adenosyltransferase II,beta	MAT2B	3
sortilin 1	SORT1	4
peptidylprolyl isomerase C	PPIC	8
interferon regulatory factor 2	IRF2	16
HMG box transcription factor 1	HBP1	4
RNA binding motif protein 9	RBM9	7
dCMP deaminase	DCTD	13
caspase 6,apoptosis related cysteine peptidase	CASP6	28
uridine monophosphate synthetase	UMPS	16
fibroblast growth factor 10	FGF10	25
aryl hydrocarbon receptor nuclear translocator like	ARNTL	21
Fas	APT1	738
aprataxin	APTX	6
thrombospondin 1	THBS1	79
polycystic kidney disease 2	PKD2	18
CD83 molecule	CD83	41
tumor protein p73 like	TP73L	31
jagged 1	JAG1	18
matrix metallopeptidase 10	MMP10	21
slit homolog 2	SLIT2	3
adrenomedullin	ADM	44
thymidylate synthetase	TYMS	87
endothelin receptor type A	EDNRA	58
polymerase,beta	POLB	28
cylindromatosis	CYLD	3

malate dehydrogenase 1,NAD	MDH1	5

Metaplasia – 8 genes

name	symbol	abstract count
alpha methylacyl CoA racemase	AMACR	6
tumor protein p73 like	TP73L	16
SRY box 2	SOX2	9
catenin,delta 1	CTNND 1	3
jagged 1	JAG1	3
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	22
thrombospondin 1	THBS1	5
CD83 molecule	CD83	3

WNT signaling pathway – 5 genes

name	symbol	abstract count
fibroblast growth factor 10	FGF10	3
catenin,delta 1	CTNND 1	3
jagged 1	JAG1	4
SRY box 2	SOX2	3
gap junction protein,alpha 1,43kDa	GJA1	3

cell motility – 11 genes

name	symbol	abstract count
myosin X	MYO10	5
catenin,delta 1	CTNND1	21
shroom family member 3	SHROOM3	9
slit homolog 2	SLIT2	4
T cell lymphoma invasion and metastasis 1	TIAM1	6
abl interactor 1	ABI1	4
thrombospondin 1	THBS1	17
neuregulin 1	NRG1	10
gap junction protein,alpha 1,43kDa	GJA1	11
v abl Abelson murine leukemia viral oncogene homolog 1	ABL1	26
Fas	APT1	11

wound healing – 12 genes

name	symbol	abstract count
epiregulin	EREG	7
collagen,type V,alpha 2	COL5A2	3
fibroblast growth factor 10	FGF10	19
thrombospondin 1	THBS1	62

matrix metallopeptidase 10	MMP10	19
fibrillin 2	FBN2	3
T cell lymphoma invasion and metastasis 1	TIAM1	3
gap junction protein,alpha 1,43kDa	GJA1	36
cyclin dependent kinase inhibitor 2C	CDKN2C	3
jagged 1	JAG1	5
catenin,delta 1	CTNND 1	3
tumor protein p73 like	TP73L	3
neuregulin 1	NRG1	4

Acknowledgements

An erster Stelle möchte ich mich bei **Prof. Markus Gerhard** für die Überlassung des spannenden und vielseitigen Themas bedanken, aber auch für die Einarbeitung in die Materie und dafür, dass er mir die Möglichkeit gegeben hat ein so breites Spektrum an Wissen und Techniken zu erlernen.

Vielen Dank an **Prof. Thomas Cremer**, dass er sich dazu bereit erklärt hat als Doktorvater meine Dissertation vor der Fakultät für Biologie der LMU München zu vertreten.

Mein Dank geht auch an **Prof. Dirk Busch**, der es mir ermöglicht hat den letzten Teil meiner Doktorarbeit in seinem Institut anfertigen zu können.

Ich bedanke mich darüber hinaus bei der **Universität Bayern e.V**. und bei der **Schering Stiftung** für die finanzielle Unterstützung meines Projekts.

Herzlichen Dank an **Prof. Eckhard Wolf** und seine gesamte Truppe des Genzentrums München und der German Mouse Clinic in Neuherberg, v.a. **Dr. Marlon Schneider**, **Dr. Maik Dahlhoff**, **Dr. Ingrid Renner-Müller**, **Petra Renner** (mit dem gesamten Mausstall) und **Dr. Birgit Rathkolb**, für die Unterstützung meiner Arbeit und dass ich dort unter fachmännischer Anleitung meine ersten Mauserfahrungen sammeln durfte.

Großer Dank geht an **Prof. Ernst Wagner** und **Dr. Manfred Ogris** und seine Arbeitsgruppe am Institut für Chemie und Pharmazie der LMU München. Ich danke besonders **Katarina Farkasova** für die Hilfe bei der Durchführung der ersten Maus *in vivo* Experimente und **Dr. Joana Viola** für die tolle, kompetente und nette Zusammenarbeit bei der zweiten *in vivo* Runde.

Danke auch an **Dr. Andreas Jung** und sein Team in der Pathologie der LMU München für die Durchführung vieler Immunhistofärbungen.

Großer Dank geht natürlich an die **gesamte AG Gerhard**. Ich danke den Ersten der Gruppe, **Florian**, **Behnam**, **Michael** und **Christian**, dass ich so nett aufgenommen wurde. Außerdem vielen lieben Dank an **Anke** für alle Runden Kaffee, kompetente Hilfe in Fachfragen und auch mal Nicht-Fachfragen. **Jeannette** danke ich, dass sie mir so oft helfend unter die Arme gegriffen hat und sich um meine Zellen und Mäuse gekümmert hat und für die vielen lustigen Zeiten in- und außerhalb des Labors. **Zohra** möchte ich auch besonders danken für moralische Unterstützung und sonnige Isar-Momente. Ein riesengroßes Dankeschön geht ganz speziell an die gute Seele im Labor, **Raquel**, für unglaublich viel Hilfe, Geduld, Zeit, Wissen, Aufopferung, Unterstützung und Unterhaltung.

Kathrin und **Sarah** danke ich für die Zeit die wir verbracht haben und für alle lustigen, ernsten, hilfreichen oder blödsinnigen Diskussionen die wir geführt haben.

Der größte Dank geht an **meine Freunde** und ganz speziell an **meine Eltern** und **meine beiden Schwestern** dafür, dass sie immer hinter mir stehen.