NUNO MIGUEL RAMALHO

FUNCTIONAL STUDIES ON ZEB TRANSCRIPTION FACTORS IN GLIOBLASTOMA

UNIVERSIDADE DO ALGARVE

Departamento de Ciências Biomédicas e Medicina

2020

NUNO MIGUEL RAMALHO

FUNCTIONAL STUDIES ON ZEB TRANSCRIPTION FACTORS IN GLIOBLASTOMA

Master in Biomedical Sciences – Disease Mechanisms

This work was done under the supervision of

Diogo S. Castro, Ph.D
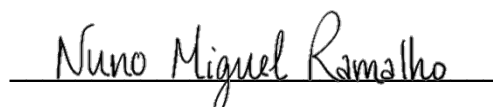
Carlos Matos, Ph.D

UNIVERSIDADE DO ALGARVE

Departamento de Ciências Biomédicas e Medicina

2020

FUNCTIONAL STUDIES ON ZEB TRANSCRIPTION FACTORS IN GLIOBLASTOMA

Declaração de autoria de trabalho

Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Nuno Miguel Ramalho

(Nuno Miguel Ramalho)

IV

"

However difficult life may seem, there is always something you can do, and succeed at. It matters that you don't just give up.

Stephen Hawking

Acknowledgements

First, I would like to start expressing my gratitude to my supervisor Diogo Castro. I cannot thank you enough all the support, knowledge shared, as well as friendship and empathy. You welcomed me into MNB, which then turned into SCN. You made possible the experience of working in Instituto Gulbenkian de Ciência and i3S - Instituto de Investigação e Inovação em Saúde. That was beyond my expectations. To my lab colleagues Vera Teixeira, Mário Soares, Abeer Heskol, and Lígia Tavares, similar to Diogo, I can't express my gratitude enough and thank you all the patience to teach me and help me turn into a better professional, a better scientist. To my dear Cristina Ferrás, I want to thank you for welcoming into i3S like no other, as well as for being there to motivate me whenever I needed. I also want to thank Carlos Matos and Isabel Duarte, two amazing people from Universidade do Algarve. Carlos for accepting the challenge of being my co-supervisor and motivating me to maintain my focus where it needed to be, and Isabel Duarte for always being kind to help in terms of R programming. Professor Clévio Nóbrega, thank you for always being there to help, motivate, and assure me that I had all the skills needed to successfully finish this thesis. Finally, I would like to share my gratitude to everyone involved in the making process of this study and not cited individually here.

An extensive and challenging project such as the one presented here could not be done without the support of my friends. Margarida Palhavã, Ricardo Nunes, Joana Rolo, Sofia Lopes, Alexandra Inverno, Catarina Martins and Francisco Alberto, I want to thank you once more for always being so patient when I desperate, and helping me stand up and face the reality – that I could do this. I owe you a big thank you.

Eu guardo este parágrafo às pessoas mais importantes na minha vida, a minha família. Aos meu pais, irmão e avó, os grandes pilares que sempre fizeram o possível (e impossível) para me ajudar não só nesta etapa, mas em todos os momentos da minha vida. À minha madrinha Céu, por estar sempre pronta a ajudar e apoiar-me, permitindo a minha formação e bem estar. À restante família, por nunca me deixar esquecer das minhas capacidades, estando sempre presente e permitir o meu sucesso.

Abstract

Glioblastoma (GBM) remains the deadliest primary brain tumour in adults, in part due to its highly invasive nature. Although not a classical model of epithelial-to-mesenchymal transition (EMT), recent work from our group has implicated the EMT transcription factor (TF) ZEB1 in regulating an "EMT-like" process that contributes to GBM tumour invasion. It is also known that ZEB1 works both as an activator and repressor of gene expression in various gene expression paradigms, including in GBM. Another member of the ZEB family, ZEB2, has also been implicated in GBM pathophysiology, but how its function differs from that of ZEB1, remains unclear. In this work, we focused on the role of ZEB2 in GBM, and how it compares with ZEB1. First, we compared the activity of ZEB proteins in transcriptional assays, performing reporter gene assays in transfected cells. Results show ZEB2 has repressive activity in two gene expression paradigms where ZEB1 functions as a transcriptional activator, revealing distinct features of these TFs. Next, we investigated how the two ZEB TFs compare in correlational studies using transcriptomics data from large cohorts of GBM tumours. We found both TFs to be differently expressed across different GBM subtypes. However, we discovered high inconsistency of results obtained across data sets, highlighting unexpected differences between transcriptomics databases. Last, we compared how ZEB TFs are recruited to regulatory regions of target genes in a cellular model of GBM using chromatin immunoprecipitation followed by quantitative polymerase chain reaction (ChIP-qPCR). This showed that the opposing transcriptional activities observed are both associated with binding of each TF to regulatory regions. Moreover, the ChIP protocol established was used for preparing a ChIP-sequencing sample to compare the genome-wide binding profiles of both ZEB TFs.

Resumo

O glioblastoma (GBM) é o tumor cerebral primário mais fatal em adultos, em parte devido à sua natureza altamente invasiva. Embora não seja um modelo clássico de transição epitélio-mesênquima (EMT), um estudo recente do nosso grupo associou o fator de transcrição (TF) ZEB1 à regulação de um processo "semelhante a EMT" que contribui para a invasão tumoral em GBM. Sabe-se ainda que ZEB1 funciona tanto como ativador como repressor em vários paradigmas de expressão génica, incluindo no GBM. Outro membro da família ZEB, ZEB2, também está envolvido na fisiopatologia do GBM, mas sem conhecimento acerca de que forma a sua função difere da de ZEB1. Neste trabalho, focámo-nos no papel do ZEB2 em GBM e como se compara ao ZEB1. Em primeiro lugar, comparámos a atividade das proteínas ZEB em ensaios de transcrição, realizando ensaios de genes-repórter em células transfetadas. Os resultados mostram que ZEB2 tem atividade repressiva em dois paradigmas onde ZEB1 atua como um ativador transcricional, revelando características distintas desses TFs. Em seguida, investigámos como ambos os ZEBs se comparam em estudos correlacionais usando dados transcriptómicos de grandes coortes de GBM. Ambos os TFs apresentaram diferentes níveis de expressão entre diferentes subtipos de GBM. Contudo, verificou-se alta inconsistência dos resultados obtidos entre os conjuntos de dados, destacando diferenças inesperadas nas bases de dados transcriptómicas. Por último, comparámos a forma como ambos os TFs são recrutados para regiões reguladoras de genes-alvo num modelo celular de GBM usando imunoprecipitação da cromatina seguida por reação em cadeia da polimerase (ChIP-qPCR). Esta técnica permitiu verificar que as atividades de transcrição antagonistas observadas estão associadas à ligação de cada TF a regiões reguladoras. Além disso, o protocolo ChIP estabelecido foi usado para preparar uma amostra de sequenciação ChIP para comparar os perfis de ligação de ambos os TFs ZEB no genoma.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

APS – ammonium persulphate

bFGF – basic fibroblast growth factor

BMPs – bone morphogenic proteins

BSA – bovine serum albumin

BSA – bovine serum albumin

CDKN2A or p16^{INK4a} – cyclin-dependent kinase inhibitor 2A

ChIP – chromatin immunoprecipitation

ChIP-seq – chromatin immunoprecipitation followed by sequencing

CID – CtBP interaction domain

CNS – central nervous system

CtBP – C-terminal-binding protein

CZF – C-terminus zinc finger domain

D-2HG – D-2-hydroxyglutarate

DMEM – Dulbecco's modified eagle's medium

DMSO – dimethyl sulfoxide

DNA – deoxyribonucleic acid

DSG – disuccinimidyl-glutarate

DSG – disuccinimidyl-glutarate

DTT – DL-dithiothreitol

E. coli – *Escherichia coli*

EDTA – ethylenediaminetetraacetic

EGF – epidermal growth factor

EGFR – epidermal growth factor receptor

EGTA – egtazic acid

EMT – epithelial-to-mesenchymal transition

Erk – extracellular-signal-regulated kinases

FBS – fetal bovine serum

GBM – glioblastoma

GEF – guanidine-exchange factor

GSCs – glioma stem cells

GSK3β – glycogen synthase kinase 3

GZ – germinal zone

h – hours

HD – homeodomain

HMG – high-motility group

HRP – horseradish peroxidase

IDH1 – Isocitrate dehydrogenase 1

I-Smads – Inhibitory Smads

KO – knock-out

LAMC2 – laminin Subunit Gamma 2

LB – Luria-Bertani

LEF – lymphoid enhancer-binding factor

LOH – loss of heterozygosity

MAPK – mitogen activated protein kinases

MET – mesenchymal-epithelial transition

min – minutes

MUT – mutated

Na-DOC – sodium deoxycholate

NADP+ – nicotinamide-adenine-dinucleotide phosphate

NADPH – dihydronicotinamide-adenine dinucleotide phosphate

NF1 – neurofibromin 1

NPCs – neural progenitor cells

NRP2 – neuropilin 2

NSCs – neuro stem cells

NZF – N-terminus zinc finger domain

XX

ºC – Degree Celsius

ON – Overnight

ONPG – Ortho-nitrophenyl-β-galactoside

OPCs – oligodendrocyte precursor cells

ORF – open reading frame

p – Shorter arm of chromosome

P/CAF – p300/CBP-associated factor

PBS – phosphate-buffered saline

PDGFRα – platelet-derived growth factor receptor-α ()

PEI – polyethylenimine

Prex1 – phosphatidylinositol-3,4,5-Trisphosphate Dependent Rac Exchange Factor 1

PTEN – phosphatase and tensin homolog

q – longer arm of chromosome

qPCR – Quantitative PCR

RACs – rho family of small GTP-binding proteins

RLU – relative light unit

rpm – revolutions per minute

RT – room temperature

RTK – receptor tyrosine kinase

SD – standard deviation

SDS – sodium dodecyl sulphate

SDS-PAGE – sodium dodecyl sulphate-polyacrylamide gel electrophoresis

Snail – snail family transcriptional repressor

ssGSEA – single sample gene set enrichment analysis

TBS – Tris-buffered saline

TCF – T cell factor

TE – tris-EDTA

TEMED – N,N,N′,N′-Tetramethylethylenediamine

TFs – transcription factors

TGF-β – transforming growth factor β

TP53 – tumour protein P53

Twist – twist family of basic helix-loop-helix transcription factor

uPA - urokinase-type plasminogen activator

V – volts

W – watts

WHO – World Health Organization

WT – wild-type

YAP1 – yes-associated protein 1

ZEB – zinc-finger E-box-binding

α-KG – α-ketoglutarate

CHAPTER ONE

GENERAL INTRODUCTION

1. Introduction

1.1. Gliomas

Gliomas are considered the most common primary malignant brain tumour type in adults, representing approximately 80 % of all malignant brain tumours. Despite its typical aggressive behaviour, not every glioma behaves in a malignant fashion (Ostrom *et al.*, 2015). According to the World Health Organization (WHO), gliomas are classified according to a scale that ranks their malignancy (grades I to IV), as well as by the presentation of glial histological features (i.e. markers of oligodendrocytes and astrocytes). Altogether, this subtyping system created by the WHO is based on specific histologic and phenotypic characteristics, such as proliferative behaviour, recurrence after surgical resection and therapy resistance (Louis *et al.*, 2007).

Gliomas malignancy scale (WHO grading) divides tumours into four distinct grades. Lesions with grade I present low proliferative potential and the possibility of cure following surgical resection alone. Grade II applies to infiltrative and, despite the low-level proliferative activity, often recurrent lesions. Some type II tumours tend to progress to higher grades of malignancy. The designation WHO grade III is generally reserved for lesions with histological evidence of malignancy, including nuclear atypia and brisk mitotic activity. In most cases, patients with grade III tumours receive adjuvant radiation and/or chemotherapy. Lastly, grade IV is assigned to highly proliferative and invasive, typically associated with the existence of necrotic regions and a very poor outcome post-tumour resection. Combination of these hallmarks for grade IV gliomas results in intra-tumour heterogeneity, which disturbs the treatment. This type of tumours tends to reoccur after resection even with standard treatment post-resection (Frosina, 2009; Seymour, Nowak and Kakulas, 2015). A typical example of grade IV neoplasm is glioblastoma (Wesseling and Capper, 2018).

Histological grading distinguishes gliomas in two major subtypes: oligodendroglial, including pure oligodendroglial and mixed oligoastrocytic tumours (both grade II), and astrocytic tumours, including pilocytic astrocytomas (grade I), astrocytomas (grade II, III and IV). Astrocytomas grade IV are most known as glioblastoma (GBM).

1.2. Glioblastoma multiforme

Glioblastoma, also known as grade IV astrocytoma, accounts for the majority of gliomas (55.1 %), making it the most frequent type of primary malignant tumour of the central nervous system (CNS) in adults (Ostrom *et al.*, 2015). It is also known as the most aggressive glioma subtype, presenting a very poor prognosis – median survival is only nine months without

proper treatment. This can be extended to 15-16 months for those receiving standard of care (Johnson, Leeper and Uhm, 2013; Gilbert *et al.*, 2014). Surgery remains a hallmark in the treatment of malignant brain tumours, GBM included. It is imperative in this type of tumour to have appropriate imaging, so the maximal safe resection is achieved. This is a particularly difficult task, given the highly infiltrative behaviour of GBM tumours. Complementary treatment is also common after surgery, specifically chemotherapy using temozolomide, and radiotherapy (Fernandes *et al.*, 2017). Regardless of many great advances in the GBM field, such as new and more advanced diagnostic modalities and multidisciplinary treatment, all have failed to improve its dismal prognosis. Novel therapeutic strategies are needed, namely strategies that may target its highly infiltrative behaviour (Deorah *et al.*, 2006; Reardon *et al.*, 2012; Ostrom *et al.*, 2014).

1.2.1. Primary and secondary glioblastomas

Glioblastoma was first distinguished as primary (*de novo)* or secondary by H. J. Scherer in 1940 (H. J. Scherer, 1940). Scherer referred that secondary glioblastomas developing in astrocytomas should be distinguished from primary glioblastomas and that they were probably responsible for most of the GBMs of long clinical duration. This was a remarkable observation at the time, although their origin was not well established until the introduction of immunochemistry. It is now known that more than 90 % of the GBMs are classified as primary since they arise in the absence of prior disease. These tumours are aggressive, highly invasive, and very rapidly developing neoplasms that are more commonly seen in elderly patients (around the age of 65). On the other hand, secondary GBMs develop from low-grade diffuse or anaplastic astrocytomas which, because of their tendency for diffuse infiltration of neighbouring brain structures, tend to recur, often with histologic and biologic characteristics of a more malignant grade, such as GBM. Secondary GBMs are much less common than primary tumours, representing only approximately 5 % of total cases of GBM, and are associated with a better prognosis. The longer survival of patients with secondary GBM tumours may in part relate to the fact that these typically affect younger patients (below the age of 45) (Kleihues and Ohgaki, 1999; Ohgaki and Kleihues, 2005, 2013).

Histopathologically primary and secondary GBMs remain undistinguishable but, in 1996, genetic alterations were attributed to each group (Watanabe *et al.*, 1996). Since then, the in-depth understanding of the genetic, epigenetic, and molecular profiles of these tumours allowed for the distinction to become clearer (Mansouri, Karamchandani and Das, 2017) (Figure 1.1).

At the population level, the most frequent genetic alterations in GBMs are loss of heterozygosity (LOH) of the longer (q), shorter (p), or both arms of chromosomes 10 and 19;

4

epidermal growth factor receptor (EGFR) amplification; tumour protein P53 (TP53) mutations; cyclin-dependent kinase inhibitor 2A (CDKN2A, also called p16[INK4a]) homozygous deletion; phosphatase and tensin homolog (PTEN) mutations. Both primary and secondary GBMs may have these genetic alterations, but what differs among them is the frequency. While EGFR amplification, PTEN mutations and p16[INK4a] homozygous deletion are much more typically seen in primary GBMs, TP53 mutations are more common to secondary GBMs. These TP53 mutations are often the earliest detectable alteration since they are already present in low-grade and anaplastic gliomas (Kleihues and Ohgaki, 1999; Ohgaki and Kleihues, 2005) (Figure 1.1).

Additionally, Verhaak and colleagues categorized GBM tumours according to the expression profile. This was later found to associate with distinct mutations so that GBM tumours were classified as classical if it presented predominantly EGFR mutations, mesenchymal if it presented mostly mutations in neurofibromin 1 (NF1), and proneural is characterized by presenting predominantly PDGFRA/IDH1 mutations (Verhaak *et al.*, 2010).

All these findings helped a lot in understanding the differences between primary and secondary GBMs, but their main genetic difference, a mutation in isocitrate dehydrogenase 1 (IDH1) gene, was discovered only more recently in 2008 (Parsons *et al.*, 2008). IDH1 is a metabolic enzyme responsible for the oxidative decarboxylation of isocitrate to α-ketoglutarate (α-KG). When mutated, IDH1 reverses the normal activity of converting NADP+ to NADPH. Instead of using isocitrate, mutated IDH1 can use the final product α-KG to generate the metabolite D-2-hydroxyglutarate (D-2HG) (Figure 1.1) (Dang *et al.*, 2009). D-2HG is a similar molecule to α-KG but acts as a weak competitive inhibitor of α-KG-dependent dioxygenases, which are involved in a wide range of cellular processes such as hypoxia, angiogenesis, and regulation of epigenetics (Xu *et al.*, 2011). Even though there is not a vast knowledge in the effects of D-2HG, there is a correlation between intracellular concentrations of D-2HG and the epigenetic effects in IDH1 mutant tumours. These epigenetic modifications are associated with altered expression of genes involved in various cellular processes and have been implicated in a block of cell differentiation characteristic of low-grade gliomas (Lu *et al.*, 2012; Turcan *et al.*, 2012).

Parsons and colleagues were the firsts to associate IDH1 gene mutations to secondary GBM (Parsons *et al.*, 2008). This observation was so critical that nowadays is still used as the main differentiator marker between primary and secondary GBMs.
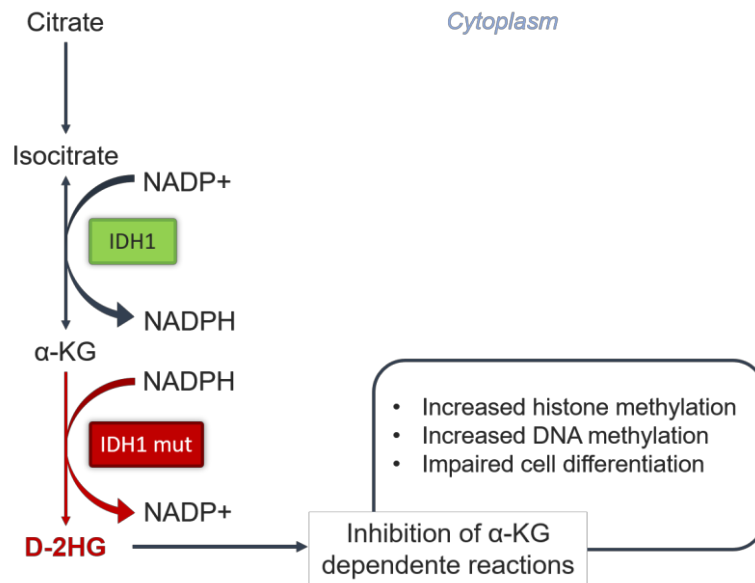
1.2.2. Pathophysiology of GBM

Figure 1.1 | Enzymatic activities of wild type and mutated IDH1 enzyme.

Notes: The IDH1 is an enzyme located in the cytoplasm and peroxysomes. Its function is characterized by catalyzing the reversible NADP+-dependent oxidative decarboxylation of isocitrate to αKG. When mutated, IDH1 gains neomorphic enzymatic activity, converting NADPH and αKG to NADP+ and D-2HG. This last metabolite acts as a weak competitive inhibitor of αKG-dependent dioxygenases, which, in turn, are involved in various cellular processes such as hypoxia, angiogenesis, and regulation of epigenetics. Excess of D-2HG is associated with increased histone and DNA methylation, altering the ability of cells to differentiate.

Abbreviations: α-KG, α-ketoglutarate; D-2HG, D-2-hydroxyglutarate; IDH1, isocitrate dehydrogenase 1; NADP, nicotinamide-adenine-dinucleotide phosphate; NADPH, dihydronicotinamide-adenine dinucleotide phosphate.

GBM tumours are highly heterogeneous and contain cells exhibiting various degrees of differentiation. Very importantly, these include cells with neural stem cell-like characteristics, also referred to as GBM stem cells (GSCs), which are very important for tumour development. One of the characteristics of GSCs is their tumour-inducing capacity, defined when injected into the brain of immuno-compromised mice (J. Lee *et al.*, 2006; Xie, Mittal and Berens, 2014; Jensen *et al.*, 2016). GSCs can be selected and maintained in culture from tumour biopsies under defined culture conditions (e.g. mitogens, absence of serum), and are considered a very useful in vitro model of GBM.

GBM displays a highly invasive behaviour, infiltrating surrounding brain parenchyma, yet typically confined to the CNS and without metastasizing (Omuro and DeAngelis, 2013).

Tumour development is sustained by cycles of proliferation/invasion, which are driven by the highly plastic phenotype of GSCs (Xie, Mittal and Berens, 2014). Cell proliferation within the tumour results in necrosis due to lack of oxygen. In reaction to hypoxia, GSCs become infiltrative and move away from necrotic regions, originating so-called pseudopalidases, which represent a common histological characteristic of GBM. Cells under hypoxia also respond by producing angiogenic factors to induce new blood vessel formation. Once cells reach a perivascular zone, they shift from invasive and motile to a more proliferative phenotype. The shift of GSCs between proliferative and infiltrative states indicates these cells are highly plastic, and the acquisition of infiltrative behaviour has been compared to an epithelial-to-mesenchymal transition (EMT) like process.

1.3. Epithelial-to-mesenchymal transition

EMT is a biological process that allows polarized epithelial cells to undergo extensive biochemical and cellular changes capable of losing their epithelial characteristics while acquiring a mesenchymal phenotype. The pioneering work of Elizabeth Hay described an "epithelial mesenchymal transformation" using a model of chick primitive streak formation (Hay, 1995). This "transformation" term has since then been replaced with "transition" in part to reflect the reversibility of the process (Hay, 1995; J. M. Lee *et al.*, 2006). The major characteristics of EMT are alterations of cellular morphology, cellular architecture, cell adhesion and gene expression, which occur concomitantly with increased cell migration capacity and elevated resistance to apoptosis (Hay, 1995; Kalluri, Neilson and Kalluri, 2003; J. M. Lee *et al.*, 2006).

An EMT can occur under three distinct biological settings and is thus categorized into three types (Figure 1.2). Type 1 refers to the EMT that occurs during implantation, embryogenesis, and organ development. This transition is critical for normal development, since it occurs as soon as the implantation of the embryo, and also during placenta formation, gastrulation, and organogenesis as well (Hay, 1995; Vicovac and Aplin, 1996). In this type is also common for cells generated by EMT to be re-induced as secondary epithelial cells in mesodermal and endodermal organs by the opposite process, a so-called mesenchymal-epithelial transition (MET). Type 2 EMT is associated with inflammatory processes, such as tissue regeneration and organ fibrosis. For this, secondary epithelial or endothelial cells acquire mesenchymal characteristics frequently induced in response to inflammation, migrating to injury regions and transitioning to resident tissue fibroblasts. Lastly, type 3 EMT consists of the dramatic changes in epithelial carcinoma cells, turning into metastatic tumour cells. These cells, present in primary nodules, gain motility and invasive capacities, allowing

them to migrate through the bloodstream and form secondary nodules distant from the primary, by undergoing MET (Zeisberg and Neilson, 2009).

1.3.1. Regulation of EMT

The complex gene expression program associated with an EMT is to large extent regulated by important transcription factors, such as snail family transcriptional repressor (Snail), twist family of basic helix-loop-helix transcription factor (Twist), and zinc-finger E-box-binding (ZEB), expressed early in this process. These are both required and sufficient (when expressed in certain cellular contexts) to induce an EMT program and are therefore referred to as "classical EMT TFs".

Each EMT TF is expressed in different cellular contexts, being induced by distinct signalling pathways (Peinado, Olmeda and Cano, 2007; Lamouille, Xu and Derynck, 2014).
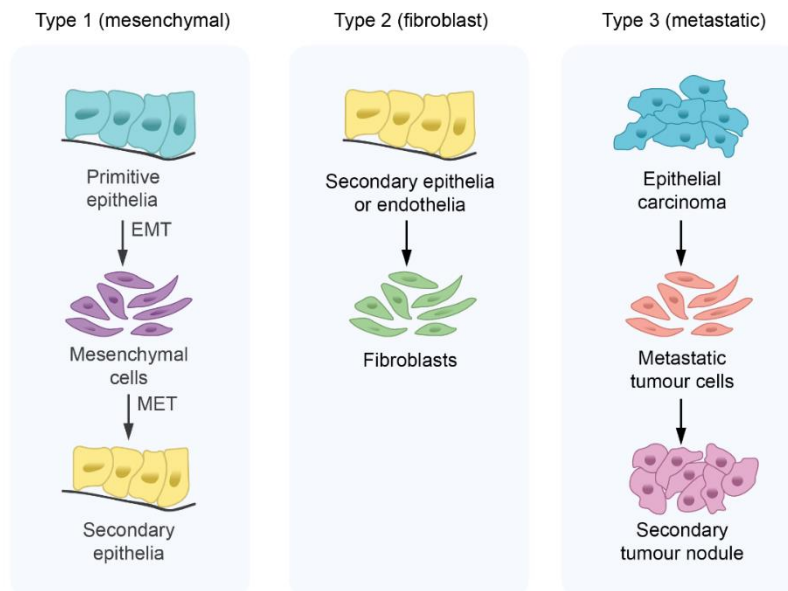
Figure 1.2 | Types of Epithelial to mesenchymal transitions.

Notes: EMTs can be viewed as cell plasticity. EMT is categorized into three different types depending on the phenotype of the cells. Type 1 EMT is associated with gastrulation or neural crest migration, where primitive epithelial cells transition into mesenchymal cells. These mesenchymal cells undergo MET to form secondary epithelial cells. Type 2 EMT is seen when secondary epithelial cells or endothelial cells populate interstitial spaces with resident or inflammation-induced fibroblasts, the latter during persistent injury. Type 3 EMT is part of the metastatic process, whereby epithelial tumour cells leave a primary tumour nodule, migrate to a new tissue site, and reform as a secondary tumour nodule.

Abbreviations: EMT, epithelial-to-mesenchymal transition; MET, mesenchymal-to-epithelial transition.

(Modified from (Zeisberg and Neilson, 2009)

These EMT pathways also often cooperate with the classical EMT TFs to regulate common target genes important for the EMT process. Despite much research in the field, it is not clear how much redundancy and/or specificity is there amongst EMT TFs. A strong possibility is that each EMT TF can activate a partially overlapping, partially distinct EMT program. Amongst EMT inducing pathways, the TGF-β and WNT pathways are particularly relevant.

1.3.1.1. TGF-β signalling pathway

TGF-β is a member of a large family of cytokines, including activins and bone morphogenic proteins (BMPs). These cytokines are responsible for regulating a wide variety of biological processes such as proliferation, differentiation, EMT and apoptosis. In order to be functional, TGF-β family receptors (TGF-βrI, TGF-βrII and TGF-βrIII), a type of serine/threonine receptors, dimerize and group with another dimer, forming a tetramer. This event is required to initiate cell signalling (Lamouille, Xu and Derynck, 2014). Receptor combination often binds to different ligands, resulting in the activation of different receptors and mediating different extrinsic signals. Once TGF-βrs are phosphorylated, phosphorylation of the cytoplasmic signalling molecules Smad2 and Smad3 for the TGF-β/activin pathway, or Smad1/5/9 for the BMP pathway occurs. This results ultimately in their translocation to the nucleus. Once Smads are activated, they partner with transcription factors, modulating gene expression. Inhibitory Smads (I-Smads) 6 and 7 antagonize activation of receptor-regulated R-Smads. The expression of these two I-Smads is induced by both activin/TGF-β and BMP signalling as part of a negative feedback loop (Schmierer and Hill, 2007). Moreover, in certain contexts, TGF-β signalling may be Smad-independent, inducing responses unrelated to transcription by the activation of Erk, and p38 MAPK pathways (Derynck and Zhang, 2003; Horbelt, Denkis and Knaus, 2012).

TGF-β signalling is associated with EMT and EMT-like events both during development and postnatally, this last in wound healing, fibrosis and cancer (Lamouille, Xu and Derynck, 2014). In several epithelial tumours (e.g. cervical, colorectal, oesophageal), TGF-β signalling pathway is known to have a dual role – tumour suppressor and promoter of tumour growth, invasion, and metastasis. This TGF-β contradictory effect is partly dependent on the tumour environment (Margadant and Sonnenberg, 2010). Specifically in glioma, TGF-β promotes invasion, growth and metastasis through activation of an EMT-like program (Margadant and Sonnenberg, 2010; Zhang *et al.*, 2017).

1.3.1.2. WNT signalling pathway

Canonical WNT signalling is an evolutionarily conserved pathway that involves the binding of WNT ligands (cysteine-rich proteins) to transmembrane Frizzled receptors.

Consequently, glycogen synthase kinase 3 (GSK3β) is inhibited, preventing β-catenin phosphorylation, ubiquitylation and degradation. As result, β-catenin enters the nucleus, forming a complex with T cell factor (TCF) and lymphoid enhancer-binding factor (LEF), in order to regulate the transcription of target genes, such as the ones involved in cell differentiation and proliferation (Eastman and Grosschedl, 1999). WNT pathway has been classified as either canonical (β-catenin-dependent and more common) or noncanonical (β-catenin-independent and less common) signalling pathways, depending on the cellular context and WNT receptors (Niehrs, 2012). WNT is known as a regulator of EMT both during embryogenesis and cancer. In development, WNT regulates EMT to promote endoderm and mesoderm formation, as well as neural crest delamination, notochord and somite formation (Zhang, Tian and Xing, 2016). During cancer progression, WNT modulates EMT target genes' transcription to promote invasion and migration of tumour cells, resulting in metastasis (Holland *et al.*, 2013). Some cancer types modulated by WNT signalling are ovarian, breast, and colorectal carcinomas (Vermeulen *et al.*, 2010; Wu *et al.*, 2012; Deng *et al.*, 2016). In addition to controlling directly transcription of EMT TFs, inhibition of GSK3β downstream WNT signalling can also promote EMT by increasing Snail stability (Zhou *et al.*, 2004). Lastly, LEF1 and ZEB1, downstream factors in the canonical WNT pathway, have shown to have a critical role in EMT process, promoting cell migration and invasion of several cancers, including GBM (Sánchez-Tilló *et al.*, 2015; Santiago *et al.*, 2017; Rosmaninho *et al.*, 2018).

1.3.2. ZEB1 and ZEB2 TFs

ZEB family of TFs is composed by ZEB1, also known as Zfhx1a or δEF1, and by ZEB2, also known as SIP1. Both ZEB TFs share a series of conserved protein domains. Both have a centrally located homeodomain that does not bind DNA, possibly promoting protein-protein interactions (Figure 1.3). In addition, ZEB proteins have two separate arrays of zinc-fingers (four at the N-terminus and three at the C-terminus). These show a high degree of homology between each ZEB protein, suggesting very similar DNA-binding specificities. ZEB TFs are known to repress transcription of target genes by directly binding to 5'-CACCT sequences, also known as E-boxes, located at gene regulatory regions. C-terminal-binding protein (CtBP) is a co-repressor considered as a major factor involved in the molecular mechanism of action of ZEB TFs. The CtBP interaction domain (CID) located in between the zinc-finger clusters is critical for the recruitment of this co-factor and consequently for the repression activity of ZEB proteins. Although CID is critical in repressing target genes, another repressor domain (identified as important for lymphoid differentiation) is located closer to the N-terminus (Verschueren *et al.*, 1999). Another important domain for both ZEB TFs' function is a Smad-binding domain (SBD). This domain interacts with Smads, establishing a cross-talk between ZEB proteins and the TGF-β/BMP signalling pathway (Postigo, 2003). Domains unique to

10

ZEB1 are two activation domains: one closer to N-terminus, responsible for the recruitment of histone acetyltransferase p300 and its associated protein, p300/CBP-associated factor (P/CAF), and another closer to C-terminus, less characterized (Sánchez-Tilló *et al.*, 2015). In line with this, in addition to repression, various examples show ZEB1 can also serve as a transcriptional activator, usually in cross-talk with other pathways. The activation molecular mechanisms of ZEB1 were firstly described by Postigo, in the context of TGF-β/BMP signalling (Postigo, 2003). Postigo described a synergistic ZEB1 interaction with Smad proteins, promoting the activation of target genes of TGF-β/BMP signalling pathway. In another example, ZEB1 promotes the expression of laminin subunit gamma 2 (LAMC2) and urokinase-type plasminogen activator (uPA) in the context of Wnt signalling (Sánchez-Tilló *et al.*, 2015). In this case, ZEB1 interaction with the TCF4/β-catenin complex replaces binding of CtBP in favour of p300 (Sánchez-Tilló *et al.*, 2015). Moreover, Lehmann and colleagues showed that ZEB1 interaction with yes-associated protein 1 (YAP1) turns it into an activator, cooperating with the hippo pathway (Lehmann *et al.*, 2016). Lastly, in a previous work from our lab focused on a GSC model, Rosmaninho and colleagues described the indirect recruitment of ZEB1 by LEF/TCF factors to regulatory regions of target genes, resulting in transcriptional activation in a Wnt-independent manner (Rosmaninho *et al.*, 2018). Strikingly, ZEB2 lacks the activation domains found in ZEB1, being unclear from the literature if it can promote gene activation (Postigo *et al.*, 2003).
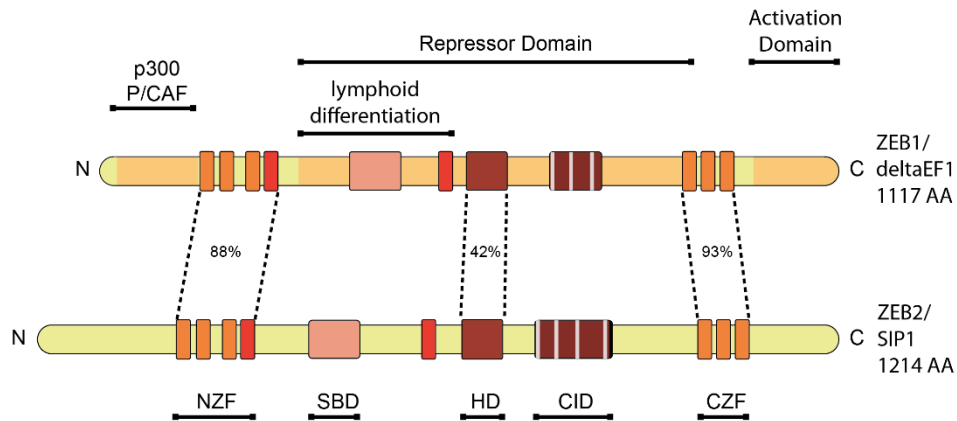
Figure 1.3 | Schematic representation of the two members of the ZEB family of transcription factors and respective domains.

Notes: The ZEB family of TFs contains two members, ZEB1 and ZEB2, which share several well conserved protein domains. These include clusters of zinc-fingers that mediate DNA-binding, a central located homeo-domain that mediates protein-protein interactions, two important domains in the repression of ZEB target genes – SBD and CID –, as well as a domain involved in lymphoid differentiation, located closer to N-terminus. ZEB1 also has two activation domains – one closer to N-terminus, responsible for recruiting p300 and P/CAF, and another closer to C-terminus. Percentage refers the degree of protein sequence homology.

ZEB proteins play important roles in embryonic development. ZEB TFs are expressed and required for cartilage, bone, and muscle formation as well as in the development of hematopoietic cells. For this, there is a need for proper spatiotemporal gene regulation of ZEB genes, which are often expressed in the same tissues but in complementary domains (Vandewalle, Van Roy and Berx, 2009). Although in many cases ZEB TFs are not expressed in the context of a classical EMT, they are thought in such cases to regulate processes that display similarities with EMT ("EMT-like") (Depner *et al.*, 2016; Rosmaninho *et al.*, 2018).

During early development, ZEB1 is expressed in the notochord, somites, limb, neural crest derivatives and a few restricted sites of the brain and spinal cord (Takagi et al., 1998). Therefore, disturbances of ZEB1 expression during development (i.e. knock-out (KO) and mutations) are highly associated with defects such as T cell deficiency of the thymus, cleft secondary palate, defective nasal formation, and other craniofacial abnormalities (Higashi et al., 1997; Takagi et al., 1998). In addition to craniofacial defects, null mice have presented musculoskeletal abnormalities including shortened limbs and digits, fusion and curvatures in the skeleton and tail, as well as defects in smooth muscle (Takagi et al., 1998; Nishimura et al., 2006). More severe CNS defects including failure of neural tube closure at both cranial and caudal neuropores were also seen in a subset of null mice (Takagi et al., 1998). ZEB1 KO homozygotes develop to term but do not survive postnatally (Liu et al., 2008).

12

Similar to ZEB1 KO, development of ZEB2 KO embryos is arrested at E8.5, with failure of neural tube closure, defects in cranial neural crest migration, and production of short somites. This reflects the prominent expression of ZEB2 in neural epithelium, neural crest and presomitic mesoderm (Putte et al., 2003; Maruhashi et al., 2005). Complementary expression study suggests cross-regulatory interactions (e.g. ZEB1 repressing ZEB2 and vice versa). Analysis of single mutants shows indeed that removing one ZEB TF is often associated with upregulation of the other (Miyoshi et al., 2006). Altogether, the complex expression patterns and cross-regulatory interactions observed, make it difficult to conclude how the activity of both ZEB TFs compare at the molecular level.

1.3.2.1. ZEB TFs and neurogenesis

ZEB transcription factors also play important roles in neural development, namely during neurogenesis. In order to form the CNS circuitry, newborn neurons need to exit their germinal zone (GZ), develop axons and dendrites, migrate to their final position and synaptically engage with other neurons (Singh *et al.*, 2016). In most neuronal lineages along the anterior/posterior axis of the developing CNS, ZEB1 expression in neural/stem progenitor cells is followed by expression of ZEB2 in new-born neurons. A study in the developing cerebellum has shown ZEB1 to keep neural progenitors in an immature state, by stopping them from becoming polarized and thereby retaining progenitors in the GZ. The cell polarization step requires the down-regulation of ZEB1 expression that occurs concomitant with differentiation. A similar process occurs during the development of the neocortex. To have proper development of the neocortex, cortical neural stem/progenitor cells need to switch differentiation/migration programs in a coordinated fashion. At onset of differentiation, neurons exit the GZs with a short bipolar morphology, which then changes to multipolar and finally to bipolar once again. Like in the cerebellum, this requires the down-regulation of ZEB1 expression at onset of differentiation. Overall, both studies suggest ZEB1 controls the rate of neurogenesis in different embryonic regions by an EMT-like process that impacts on cell polarity and morphology (Wang *et al.*, 2019).

ZEB2 also plays an important function in neurogenesis. However, its role reflects its later expression in post-mitotic neurons, being for example required for proper differentiation and migration of interneurons in the developing neocortex. These processes need to be well orchestrated since the dysregulation leads to neurodevelopmental disorders (Levitt, Eagleson and Powell, 2004; Berghe *et al.*, 2013).

1.3.2.2. ZEB TFs and GBM

Amongst classical EMT transcription factors, members of the ZEB family (ZEB1 and ZEB2) are major candidates to control an EMT-like process in GBM. ZEB1 expression has been found at both tumour core and periphery, although with heterogeneity in protein levels. Immune infiltrative cells (microglia and tumour-associated macrophages) have been shown to account for up to 30 % of cells in glioma and are ZEB1 negative, therefore likely contributing to this heterogeneity (Euskirchen *et al.*, 2017). The fact that ZEB1 expression is not restricted to the infiltrating tumour rim as previously suggested (Siebzehnrubl *et al.*, 2013; Zhang *et al.*, 2016) is in agreement with the view that GBM growth occurs concomitantly with waves of cell invasion in various regions across the tumour, including the tumour core (Kahlert *et al.*, 2015; Zhang *et al.*, 2016). By contrast, ZEB2 expression was suggested to predominate at tumour border, although its expression in GBM and how it relates to that of ZEB1 remains to be properly characterized (Depner *et al.*, 2016).

Gene knock-down studies confirmed the key roles of ZEB factors in GBM – in cultured GSCs, both ZEB factors have been implicated in their proliferation, migration, and survival (Qi *et al.*, 2012; Siebzehnrubl *et al.*, 2013). In in vivo studies, using mouse brain xenografts from patient-derived GSCs, both ZEB1 and ZEB2 proteins were shown to contribute to tumour growth and invasiveness, increasing its malignancy. ZEB1 was shown to interact with microRNA-200, modulating the expression of several target genes such as c-MYB, ROBO1, OLIG2 and MGMT, regulating cell invasion, migration and chemoresistance (Siebzehnrubl *et al.*, 2013). ZEB2 was shown to repress ephrinB2 within the cellular context of hypoxia, thus promoting cell invasion of surrounding brain tissue (Depner *et al.*, 2016).

Mechanistically, ZEB1 was shown to be part of a transcriptional network that drives gliomagenesis (Singh *et al.*, 2017). Our group led the first study to incorporate a genomic approach to characterize ZEB1 transcriptional program in a GSC model (NCH421k cells) combining chromatin immunoprecipitation (ChIP) followed by sequencing (ChIP-seq) with expression profiling upon ZEB1 KD (Rosmaninho *et al.*, 2018). Evidence from ChIP-seq converged on LEF/TCF TFs as important factors in mediating indirect recruitment of ZEB1 (via HMG motifs) to its target genes in a genome-wide scale. This results in ZEB1 promoting transcriptional activation in synergy with LEF/TCF TFs of target genes, such as phosphatidylinositol-3,4,5-Trisphosphate Dependent Rac Exchange Factor 1 (Prex1) and neuropilin 2 (Nrp2). Prex1 gene encodes a guanidine-exchange factor (GEF) for the Rho family of small GTP-binding proteins (RACs) and was further shown to promote GBM cell invasion (Rosmaninho *et al.*, 2018). In line with that, Prex1 and ZEB1 expression levels are highly correlated in GBM patient tumour samples, with Prex1 levels being highest and indicative of poor patient prognosis in the classical GBM subtype. In conclusion from this study, ZEB1 was

14

found to be recruited to regulatory regions genes directly, binding to E-boxes and promoting transcriptional repression, and indirectly, via HMG motifs and promoting gene activation. When combined the ChIP-seq with expression profiling upon ZEB1 KD, 60 genes were bound to ZEB1 and downregulated, suggesting being possible target genes for the above-described activation function of ZEB1. Oppositely, 42 genes were bound to ZEB1 and upregulated upon ZEB1 KD, suggesting being possible target genes for the classical repressor activity of ZEB1.

Summarizing, studies based on in vitro and in vivo models of GBM implicated both ZEB factors in tumour growth and invasiveness. Even though ZEB1 has been extensively studied in terms of transcriptional mechanism and GBM pathophysiology, ZEB2 is poorly understood in this subject since the focus had been in other cancer types (Rosivatz et al., 2002; Elloul et al., 2005; Imamichi et al., 2007).

2. Aims

The main aim of this work was to provide insights into how the transcriptional activities of the two ZEB TFs compare. The specific goals were to:

1. Use transcriptional assays to compare the activities of both ZEB TFs. The focus was on gene expression paradigms where ZEB1 had been shown to function as a transcriptional activator, and where the activity of ZEB2 had never been tested;

2. Provide evidence of gene regulation by ZEB TFs, by performing correlational studies using transcriptomics data from large cohorts of GBM tumours. This had as starting point the previous identification of ZEB1 target genes described by Rosmaninho and colleagues (Rosmaninho *et al.*, 2018), and included both ZEB1 and ZEB2;

3. To prepare a ChIP-seq sample from ZEB2 in a cellular model of GBM, with the aim of comparing the genomic binding profiles of both ZEB TFs. This required to test a chromatin immunoprecipitation assay for ZEB2, which had not been previously described.

CHAPTER TWO

Comparison of ZEB factors in transcriptional assays

1. Materials and methods

1.1. Expression vectors

Expression vectors used both for reporter gene assays and Western blot analysis are listed in Table 2.1.

Table 2.1 | Expression vectors

Vector	Reference
pCAGGS-IRES- GFP	Gift from James Briscoe
pCAGGS-hZEB1-IRES-GFP	(Rosmaninho et al., 2018)
pCAGGS-hZEB2-IRES-GFP	This study
pME-18F-LEF1-Flag	(Billin, Thirlwell and Ayer, 2000)
CMV-Flag-YAP1	Gift from Florence Janody
pcDNA4/His-Max-C	Gift from Janet E. Mertz
pcDNA4/His-Max-C-hZEB2	Gift from Janet E. Mertz
pCAGGS-IRES-GFP-mZEB2	Francisca Vasconcelos (unpublished)
pcDNA3-HA-TCF4	Gift from Frank McCormick
pcDNA 3.1 β-catenin S33Y	(Kolligs et al., 1999)
pME-18F	Gift from Ryoichiro Kageyama
pPYCAG-MCS-V5-FLAG2	Gift from Debbie van den Berg
pPyCAG-hZEB1-MCSV5	Vera Teixeira (unpublished)
pPyCAG-hZEB1-MCS-V5 dN	Vera Teixeira (unpublished)
pPyCAG-hZEB1-MCS-V5 dC	Vera Teixeira (unpublished)

Vector	Reference
pPyCAG-hZEB1-MCS-V5 dNdC	Vera Teixeira (unpublished)

1.2. Luciferase vectors

Luciferase vectors used for reporter gene assays are listed in Table 2.2.

Table 2.2 | Luciferase vectors

Vector	Reference
β-globin::*luc*	(Castro *et al.*, 2006)
Nrp2::*luc*	(Rosmaninho *et al.*, 2018)
Prex1::*luc*	(Rosmaninho *et al.*, 2018)
(TEAD BS)×4::*luc*	This Study
(TEAD BS mut)×4::*luc*	This study
LAMC2::*luc*	(Sánchez-Tilló *et al.*, 2015)
uPA::*luc*	(Sánchez-Tilló *et al.*, 2015)
pGL3 ZEB1 prom::*luc*	Pedro Rosmaninho (unpublished)

1.3. Cloning

1.3.1. Annealing of oligonucleotides

For cloning of (TEAD BS)×4, the required oligonucleotides were resuspended in RNase free water to a final concentration of 100 μM. The reverse complements of each oligonucleotide were joined in freshly made annealing buffer (10 mM of Tris pH 7.5, 1 mM of Ethylenediaminetetraacetic acid (EDTA) and 50 mM of NaCl) at a final concentration of 10 μM, briefly boiled at 95 degree Celsius (ºC) and slowly cooled down to room temperature (RT).

20

1.3.2. DNA restriction digestion

Preparative digestion of β-globin::*luc* plasmid for cloning of (TEAD BS)×4 was performed in 50 µL total volume with 3 µg of DNA and 40 units of each SalI-HF (NEB) and NheI-HF (NEB) overnight (ON) at 37 ºC.

Preparative digestion of pCAGGS-IRES-GFP-linkerA for cloning of human ZEB2 was performed in 50 µL total volume with 0.3 µg of DNA and 40 units of each EcoRV-HF (NEB) and XbaI-HF (NEB) ON at 37 ºC.

1.4.1. DNA purification

Plasmids were isolated from *Escherichia coli* (*E. coli*) DH5α using Mini (ZYMO Research) or Midi-Prep (QIAGEN) kits. DNA bands from agarose gels were purified with the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare). All steps were performed as recommended by the supplier.

1.4.2. Ligation

To ligate linearized plasmid with DNA fragment encompassing the hZEB2 cDNA, ligations were performed in a 10:1 molar ratio of insert:backbone, with 5 units of T4 DNA ligase (Thermo Fisher Scientific) and 1x T4 ligase buffer. The samples were then incubated at 16 ºC ON and transformed the next day with chemically competent *E.coli* DH5α prepared using CaCl$_2$. Colonies were selected and inoculated in Luria-Bertani (LB) medium with ampicillin at 37 ºC ON, 220 revolutions per minute (rpm). To confirm the correct insertion of the insert into the backbone vector, digestion was performed at 37 ºC for 1.5 h and the resulting products were analysed on a 1 % agarose gel.

1.4.3. Transformation into chemically competent *E.coli*

Approximately 100 µL of chemically competent *E.coli* DH5α were used per transformation reaction. Bacteria were incubated with 500 ng of vector DNA for 30 minutes (min) on ice. After a heat-shock of 60 seconds at 37 ºC, the bacteria were chilled on ice for at least 2 min and 240 µL of LB was added. The bacteria were then incubated for approximately 45 min at 37 ºC on a shaker incubator set for 220 rpm and subsequently plated on LB-Ampicillin Agar plates and placed ON at 37 ºC.

1.8. Subcloning

Table 2.3 | Oligonucleotides used for the construction of (TEAD BS)×4*::luc* reporters

(TEAD BS)×4*::luc*	FW	TCGACGAATTCGGCCAGTGCCAAGTTGAGACACATT CCACACATTCCACTGCAAGCTTGAGACACATT CCACACATTCCACTGCG
	RV	CTAGCGCAGTGGAATGTGTGGAATGTGTCTCAAGCT TGCAGTGGAATGTGTGGAATGTGTCTCAACTTGGCA CTGGCCGAATTCG
(TEAD BS mut)×4*::luc*	FW	TCGACGAATTCGGCCAGTGCCAAGTTGAGACACcgca CACACCcgcaCACTGCAAGCTTGAGACACcgcaCACACc gcaCACTGCG
	RV	CTAGCGCAGTGtgcgGTGTGtgcgGTGTCTCAAGCTTG CAGTGtgcgGTGTGtgcgGTGTCTCAACTTGGCACTGG CCGAATTCG

(TEAD BS)×4::*luc* and (TEAD BS mut)×4::*luc*

Oligonucleotides to generate the (TEAD BS)×4::*luc* reporters were designed based on a previous study (Lehmann *et al.*, 2016), and contained an EcoRI restriction site and cohesive ends compatible with Sal1 and NheI sites upon annealing (Table 2.3). Oligonucleotides were then annealed as described above.

The β-globin vector was digested with 40 units of SalI-HF (NEB) and 40 units of NheI-HF (NEB). The linearized backbone was purified via agarose gel. The (TEAD BS)×4 oligonucleotide was ligated, according to section 2.7, into the β-globin vector upstream of the luciferase gene. Bacteria were transformed and positive colonies were screened by digesting the purified DNA with EcoRI-HF (NEB) and by subsequent analysis of the digestion pattern in agarose gel.

pCAGGS-ZEB2-IRES-GFP

The full-length cDNA of human ZEB2 was excised from the pcDNA4hismaxC_hZEB2 vector using 40 units of each EcoRV-HF (NEB) and XbaI-HF (NEB) and subcloned into pCAGGS-IRES-GFP vector using 40 units of each EcoRV-HF (NEB) and NheI-HF (NEB).

1.9. Western blot analysis of P19 transfected cells

1.9.1. P19 cell culture

P19 embryonic carcinoma cells were maintained in Dulbecco's modified eagle's medium (DMEM)/High glucose (Gibco) supplemented with fetal bovine serum (FBS) heat-

inactivated (10 %, PAA Laboratories, GE Healthcare), Penicillin-Streptomycin (100 U/mL, Gibco) and L-Glutamine (2 mM, Gibco) in T- flasks, plates or well plates (Corning).

1.9.2. Transfection and preparation of lysates from P19 cells

On the previous day, approximately 230,000 P19 cells were plated to obtain a 75-80 % confluency on the day of the transfection. Transfection was carried out in 6-well plates, with linear polyethylenimine (PEI, Sigma-Aldrich) in the proportion of DNA:PEI (w/w) of 1:2.5 mixed in serum-free medium (plain DMEM). The total amount of DNA/cm^2 was 500 ng. The medium was replaced with fresh complete medium (DMEM) 4-6 hours (h) after transfection.

Approximately 24 h after stopping transfection, cells were washed with ice-cold phosphate-buffered saline (PBS) and harvested by scraping in ice-cold lysis buffer (20 mM Tris HCl pH 8.0, 140 mM NaCl, 10 % Glycerol, 1 % NP-40, 2 mM EDTA and protease inhibitors (Roche)). For protein quantification, a standard curve was generated using BSA. Protein quantification was carried out using the Bradford method, in which protein assay dye reagent (Bio-Rad) was diluted 5x, and 1 µL of the sample was added to each condition. Absorbance was measured at 595 nm in the UVmini-1240 spectrophotometer (Shimadzu).

1.9.3. Western Blot

A 10 % sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel was prepared for each Western blot. The stacking gel was made up of 5 % of acrylamide/bis-acrylamide (Bio-Rad), 126 mM of Tris-HCl pH 6.8, 0.1 % of sodium dodecyl sulphate (SDS), 0.1 % of ammonium persulphate (APS, Sigma-Aldrich) and 0.1 % of N,N,N′,N′-Tetramethylethylenediamine (TEMED, Sigma-Aldrich). The resolving gel was made up of 10 % acrylamide, 375 mM of Tris-HCl pH 8.8, 0.1 % of SDS, 0.1 % of APS and 4 % of TEMED. Cell lysates were diluted in 2 × Laemmli buffer (Sigma-Aldrich), denatured for 5 min at 95 ºC and loaded onto the gel with Precision Plus Protein Kaleidoscope (Bio-Rad) as molecular weight marker. The gel was run at 150 volts (V) in running buffer (25 mM of Tris base, 200 mM of glycine and 0.1 % SDS).

To transfer proteins to a nitrocellulose membrane (GE Healthcare) the gel was soaked in transfer buffer (39 mM of Glycine and 48 mM of Tris) with 20 % of methanol (Sigma-Aldrich) and 0.037 % of SDS, together with the nitrocellulose membrane and four Whatman filter papers (Sigma-Aldrich), for 10 min. The transfer apparatus was assembled in the following order from bottom to top: sponge, Whatman filter paper, SDS-PAGE gel, nitrocellulose membrane, Whatman filter paper, sponge. After setting every component in place and removing the excess buffer, proteins were electrophoretically transferred at 100 V for 65 min.

Once the transfer was complete, the membrane was stained with Ponceau staining solution (1.3 mM of Ponceau S (Sigma-Aldrich) and 5 % of acetic acid (Sigma-Aldrich)), to visualize proteins. The membrane was washed with water and then blocked in 5 % Molico low-fat milk (Nestlé) in 1 × Tris-buffered saline (TBS) with 0.1 % tween for 1 h at RT. The membrane was incubated ON in agitation, at 4 °C, with primary antibodies (Table 2.4) diluted in 1 % milk in 1 × TBS with 0.1 % tween. In the following day, the membrane was washed 3 x 15 min with 1 × TBS with 0.1 % tween, in agitation. After incubating, also in agitation, for 1 h at RT with the horseradish peroxidase (HRP) conjugated secondary antibodies (Table 2.5) diluted in 1 % milk in 1 × TBS with 0.1 % tween, the membrane was washed 3 × for 15 min each as described above. Lastly, protein detection was performed by covering the membrane with Pierce ECL Plus Western Blotting Substrate (Thermo Fisher Scientific) for 1 min and imaged obtained by ChemiDoc XRS+ (Bio-Rad).

Table 2.4 | Primary antibodies used in Western Blot

Antigen (Species)	Working dilution in WB	Catalogue number	Company
Anti-ZEB2 (Rabbit)	1:250	ab223688	ABCAM
Anti-α-Tubulin (Mouse)	1:3000	T6074	Sigma-Aldrich

Table 2.5 | Secondary antibodies used in Western Blot

Antigen (Species)	Working dilution in WB	Company/Source
Goat Anti-Rabbit IgG (H+L) Poly-HRP	1:4000	Jackson ImmunoResearch
Goat Anti-Mouse IgG (H+L) Poly-HRP	1:4000	Jackson ImmunoResearch

1.9.4. Transcriptional assays in P19 cells

P19 cells were seeded into 48-well plates (50,000/well) 18 h before transfection. On the day of transfection, master mixes containing DNA and PEI (1:2.5, respectively) were prepared, so that each well was co-transfected with 200 ng of expression plasmids (Table 2.1), 100 ng of firefly luciferase reporter plasmid (Table 2.2) and 200 ng of pCMV-LacZ plasmid as an internal control. 24-36 h after transfection, cells were lysed with Glo Lysis Buffer, 1 × (VWR) and frozen at -80 °C. For reporter gene assays, 50 µL of cell lysates were distributed in two separate 96-well plates and four biological replicates per condition to proceed to assay the luciferase and β-galactosidase activities.

For luciferase, a freshly made luciferase assay buffer (25 mM of gly-gly pH 7.8 (Sigma-Aldrich), 15 mM of KPO_4 (Sigma-Aldrich) buffer, 15 mM of $MgSO_4$ (Sigma-Aldrich) and 4 mM of egtazic acid (EGTA) pH 7.8 (Sigma-Aldrich)) was mixed with 2 mM of ATP, 1 mM of DL-dithiothreitol (DTT, Promega) and 0.07 mM of firefly D-luciferin. 75 µls were added to each well containing cell lysate. Luminescence from luciferase activity was assessed with Synergy 2 Multi-Mode Microplate Reader (BioTek), set at RT and end-point as reading method.

For β-galactosidase activity, a freshly made solution was prepared each time by adding 1 mM of DTT (Promega) and 8.3 mM of ortho-nitrophenyl-β-galactoside (ONPG, Sigma-Aldrich) to a buffer composed of 60 mM of Na_2HPO_4 (Sigma-Aldrich), 10 mM of KCl (Sigma-Aldrich), 1 mM of $MgCl_2$ (Sigma-Aldrich) and NaH_2PO_4 (Sigma-Aldrich). 150 µls of this solution was added to each well containing cell lysate, and the plate was incubated at 37 ºC until a yellow colour develops. UV-visible absorbance from β-galactosidase activity was assessed with Synergy 2 Multi-Mode Microplate Reader, set at RT and 410 nm.

Data are presented as mean ± standard deviation (SD) of four biological replicates and One-Way or Two-Way ANOVA with Bonferroni correction for multiple testing was applied for statistical significance and graphic visualization was obtained with GraphPad Prism (version 8.4.3). Data shown correspond to one representative of at least three experiments (except when explicitly stated in the text).

2. Results

Although EMT TFs are usually considered transcriptional repressors, ZEB1 has been shown to promote gene activation in various contexts. In most of these cases, it is not known how the activity of ZEB2 compares to that of ZEB1.

One way to study the activity of TFs on their target genes is to perform transcriptional assays, which use reporter genes such as luciferase. In this approach, the regulatory region of the gene of interest is sub-cloned in an adequate vector upstream a minimal promoter (e.g. β-globin) and the luciferase gene. This luciferase containing plasmid is co-transfected with expression vectors for TFs being studied, in a relevant cell line. In order to normalize the number of transfected cells, a third plasmid expressing a different reporter gene (e.g. lacZ) is co-transfected. The result is given as the ratio of the activity of both reporters (luciferase/β-galactosidase). P19 cells were chosen for this study since they have been a useful model to study cellular and molecular events (e.g. transcriptional regulation) underlying neurogenesis (McBurney, 1993; Bressler *et al.*, 2011). Moreover, P19 cells are easily transfected and were used by Rosmaninho and colleagues in previous transcriptional assays that are the basis for the current study.

In order to generate a plasmid driving human ZEB2 expression to be used in transcriptional assays, the cDNA of this gene was sub-cloned into the CAGGS-IRES-GFP plasmid (see Materials and methods for details). Expression of human ZEB2 was confirmed by Western blot analysis of whole-cell lysates of P19 cells transfected with the newly generated expression vector (Figure 2.1), using an anti-ZEB2 antibody. In parallel, lysate of P19 cells
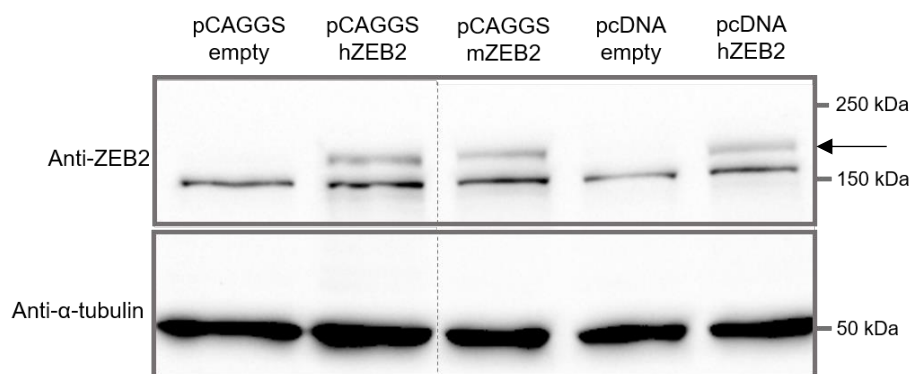

Figure 2.1 | Expression of ZEB2 in P19 cells transfected with ZEB2 expression constructs, assessed by Western blot analysis

Arrow indicates specific band corresponding to plasmid-driven ZEB2 expression. α-tubulin is shown as loading control.

26

transfected with human ZEB1 expression plasmid was also analysed, demonstrating the specificity of the ZEB2 antibody used.

Next, the activity of both ZEB TFs was compared in previously established paradigms, where ZEB1 was shown to activate gene transcription. First, the ZEB1/LEF1 synergy model, whereby ZEB1 co-activates LEF1 target genes (e.g. Nrp2 and Prex1) (Rosmaninho *et al.*, 2018), was used to test ZEB2 TF activity in the same cellular context. For this, P19 cells were co-transfected with expression plasmids of human ZEB1, ZEB2, LEF1 (or control empty plasmid) (Table 2.1) and a reporter plasmid containing the previously characterized gene regulatory regions of Nrp2 or Prex1 (Table 2.2).

In both Nrp2 and Prex1 regulatory regions (Figure 2.2), ZEB1 and ZEB2 on their own do not promote luciferase expression. As previously shown (Rosmaninho *et al.*, 2018), LEF1 promotes transactivation of Nrp2 and Prex1 regulatory regions, whereas co-expression of LEF1 and ZEB1 results in transcriptional synergy. By contrast to ZEB1, ZEB2 does not function
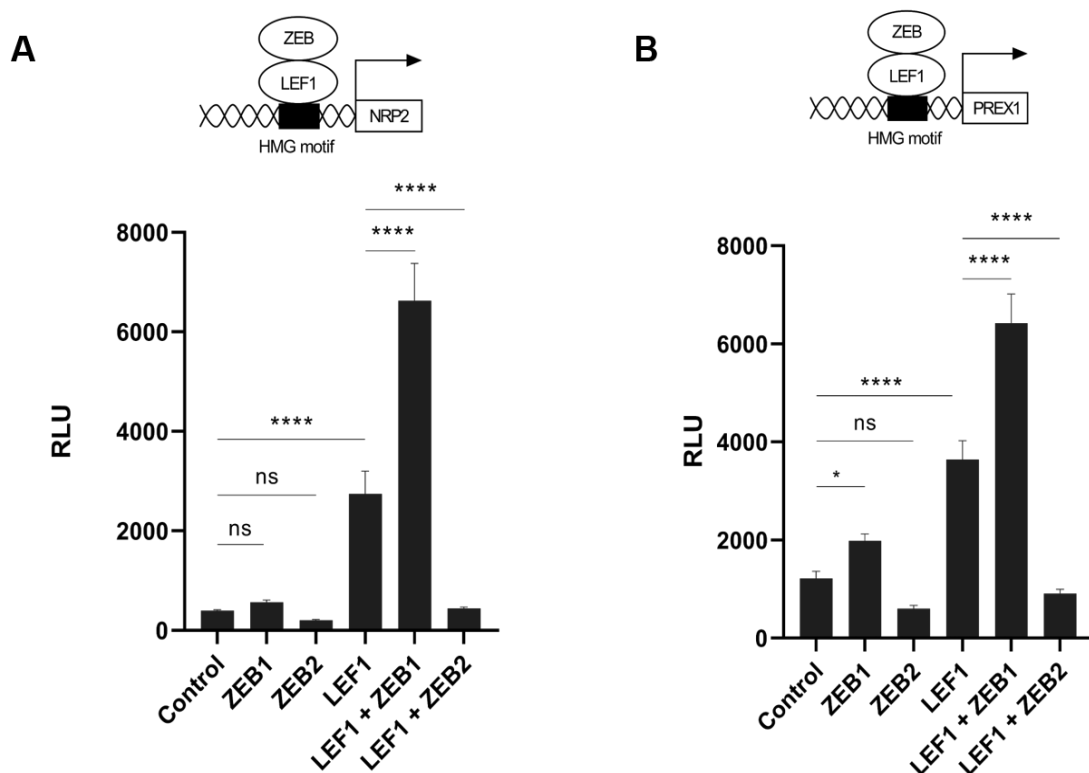
Figure 2.2 | ZEB2 does not function in synergy with LEF1 to transactivate regulatory regions of Nrp2 and Prex1 genes.

(A,B) Reporter gene assay using Nrp2 (left) or Prex1 (right) enhancer constructs, each containing high-motility group (HMG) motifs recognized by LEF1/TCF1 factors. P19 cells were co-transfected with either of the luciferase constructs and expression plasmids for human ZEB1, ZEB2 and LEF1, as indicated in figure. Data shown are representative of two independent experiments. Relative light unit (RLU) is shown as mean ± SD of four biological replicates (statistical significance determined by one-way ANOVA

in synergy with LEF1 to promote the activation of Nrp2 or Prex1 regulatory regions. Instead, consistent repression of LEF1 activity by ZEB2 is observed, when using both reporter gene constructs.

As control, P19 cells were co-transfected with increasing amounts of ZEB1 and ZEB2 expression plasmids, together with a luciferase reporter plasmid containing the upstream proximal promoter region (0.8 Kb) of mouse ZEB1 gene. ZEB1 was found to bind this

0.8 Kb ZEB1 promoter::*luc*
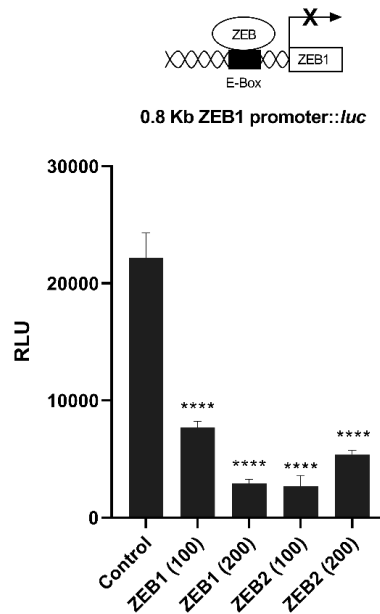
Figure 2.3 | ZEB2 represses ZEB1 promoter.

Reporter assay using a luciferase construct containing 0.8 Kb of ZEB1 mouse proximal promoter region spanning four E-box sequences. P19 cells were co-transfected with the luciferase construct and various amounts of expression plasmids for human ZEB1 or ZEB2, as indicated in figure. Data shown are representative of two independent experiments. Data are shown as mean ± SD of four biological replicates (statistical significance determined by one-way ANOVA with Bonferroni correction). In all cases: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

regulatory region in mouse neural stem cells (Singh *et al.*, 2016), which contains four E-boxes. As expected, both ZEB TFs were shown to repress reporter gene expression from this plasmid, to comparable levels (Figure 2.3).

Next, transcriptional assays were used in the context of a second paradigm whereby ZEB1 activates gene expression, this time in the context of the Hippo pathway. With that aim, luciferase vectors containing four consensus binding sites for TEAD TFs in tandem (or mutated versions) were generated, similar to what was described in a previous study (Lehmann *et al.*, 2016). This was done by subcloning annealed oligonucleotides upstream the β-globin minimal promoter, and the luciferase gene (see Materials and methods for details). To study the hippo pathway paradigm, P19 cells were co-transfected with expression plasmids for human ZEB1,

ZEB2, the Hippo pathway component YAP1, (or control empty plasmid) (Table 2.1) and the previously cloned reporter plasmid containing tandem repeats of wild-type (WT) or mutated (MUT) TEAD binding sites (Table 2.2).

As expected, YAP1 expression promotes transcription of the WT reporter gene (Figure 2.4, WT), presumably in combination with endogenously expressed TEAD proteins. While ZEB1 on its own does not promote reporter gene expression, it results in transcriptional synergy when co-expressed with YAP1. By contrast, the expression of ZEB2 inhibits the activity of YAP1 in this luciferase construct. As expected, mutation of TEAD consensus binding sites abolished almost completely transactivation of the luciferase construct.
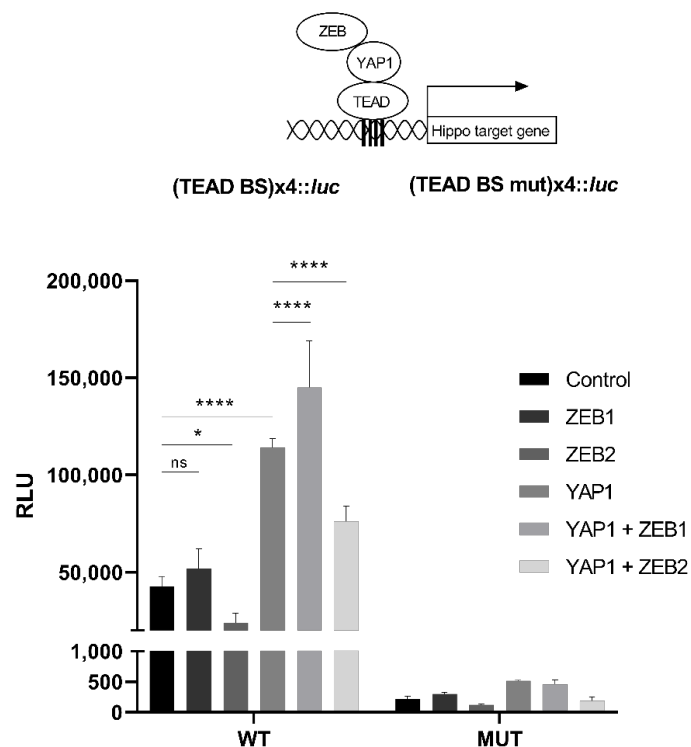
Figure 2.4 | ZEB2 does not function in synergy with YAP1 to transactivate a hippo pathway signalling sensor construct.

Reporter gene assays using constructs containing four tandem repeats of wild-type (WT) or mutated (MUT) TEAD binding sites upstream of a minimal promoter and the luciferase gene. P19 cells were co-transfected with either of the luciferase plasmids and expression plasmids for ZEB1, ZEB2 and YAP expression constructs. Data shown are representative of two independent experiments. Data are shown as mean ± SD of four biological replicates (statistical significance determined by two-way analysis of variance with Bonferroni multiple comparisons test). In all cases: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Finally, we tested the activity of ZEB2 in transcription regulation of WNT signalling target genes LAMC2 and uPA. ZEB1 was previously shown to activate expression of these two genes by binding to E-boxes in their promoters, when in the presence of TCF4 and nuclear

β-catenin (Sánchez-Tilló *et al.*, 2015). Transcriptional assays were performed using the luciferase constructs containing regulatory regions of LAMC2 and uPA genes as previously described. However, we were unable to reproduce in P19 cells the observations reported by (Sánchez-Tilló *et al.*, 2015), and this line of research was not further pursued (see Discussion).

Overall, transcriptional assays show opposing activities of both ZEB TFs in two distinct gene regulation paradigms. ZEB1 has two previously characterised activation domains – one situated closer to the N-terminus (p300 and P/CAF binding domains), and another closer to the C-terminus (Sánchez-Tilló *et al.*, 2015). Therefore, we next tested plasmids expressing N- and C-terminus deletion mutants of ZEB1, previously generated in the lab (Vera Teixeira, unpublished), for their ability to transactivate the Nrp2 promoter construct. As previously observed (Figure 2.2A), ZEB1 and LEF1 activate in synergy the Nrp2 regulatory region. Deletion of each ZEB1 terminal domain significantly reduced ZEB1 mediated activation of luciferase, with the double deletion resulting in almost complete abrogation of ZEB1 activity (Figure 2.5). Altogether, these results suggest that full synergy between ZEB1 and LEF1 requires the integrity of both N- and C-terminus regions. These results await further validation, given that for time constraints, only one experiment was performed.
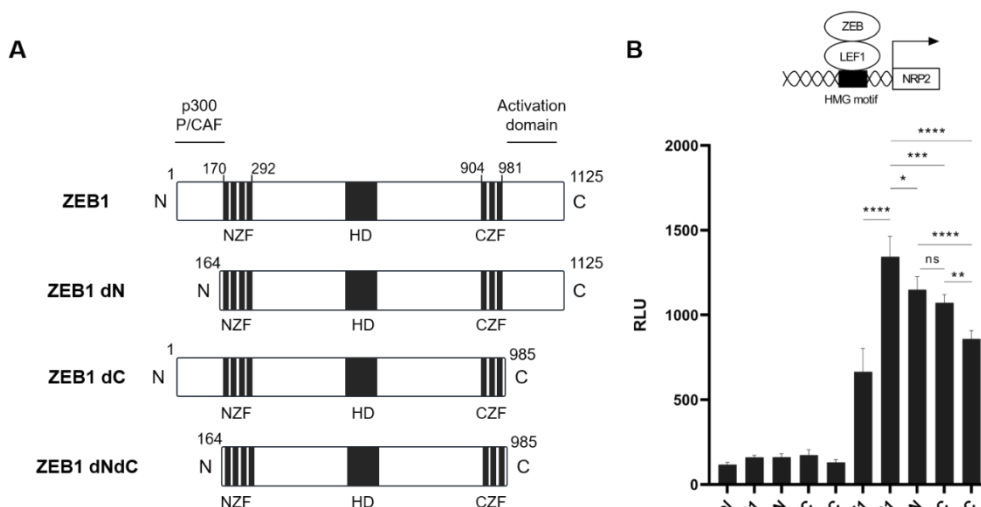
Figure 2.5 | ZEB1 seems to need both N- and C-terminus regions to activate the expression of its target genes

(A) Scheme of deletions of ZEB1 transcription factor N- (dN), C- (dC), or both (dNdC) terminal regions containing an activation domain close to C-terminus and another one containing a binding site for the co-activators p300 and P/CAF, close to N-terminus. NZF, N-terminus zinc finger domain; HD, homeodomain; CZF, C-terminus zinc finger domain. **(B)** Reporter gene assay using the Nrp2 enhancer construct containing high-motility group (HMG) motifs. P19 cells were co-transfected with the luciferase plasmid and expression plasmids for human LEF1, and various ZEB1 derivatives, as described in figure. Only one independent experiment was performed. Relative light unit (RLU) is shown as mean ± SD of four biological replicates (statistical significance determined by one-way ANOVA with Bonferroni correction). In all cases: ns, not significant, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

CHAPTER THREE

Correlational studies using transcriptomic data sets of glioblastoma tumours

1. Materials and methods

1.1. Programming in R

R is a programming language and an environment for statistical computing and graphic visualization. The environment interacts with other repositories with freely available R packages containing data, codes, documentation, tests and graphics.

In this chapter, all graphics, plots and statistical tests were constructed using RStudio 4.0.0 (RStudio Team, 2020), an integrated development environment for R that includes a code editor, debugging and visualization tools. RStudio was run under platform arch x86_64-w64-mingw32.

1.2. Data set characterization

All glioma data sets used in this study are publicly available in GlioVis portal (http://gliovis.bioinfo.cnio.es/). In order to choose the data sets for this study, phenotype data of each data set was downloaded and assessed in RStudio. First, the number of GBM and non-tumour samples were obtained, and data sets with less than 150 GBM samples were excluded (reducing the number of data sets from 27 to 7). Second, two data sets were also excluded since they missed non-tumour samples, reducing from 7 to 5 data sets. Third and last, only data sets using a similar gene expression platform (Affymetrix HG-U133 arrays) were selected, for comparative purposes with the work from Rosmaninho and colleagues (Rosmaninho *et al.*, 2018). As result, the three data sets used were Gravendeel (GEO Accession number GSE16011) (Gravendeel *et al.*, 2009), The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) and The Repository of Molecular Brain Neoplasia Data (REMBRANDT, GEO Accession number GSE108474) (Gusev *et al.*, 2018), referred during this work as Gravendeel, TCGA and REMBRANDT data sets, respectively.

1.3. Transcriptomics

1.3.1. Treatment of data

First, and using the phenotype data, GBM and non-tumour samples were isolated, discarding other glioma types that may exist in each data set. Second, and since primary GBM tumours are characterized by carrying low G-CIMP epigenetic profile (Malta *et al.*, 2018), "G-CIMP" GBM samples (corresponding to secondary GBM) were excluded from the analysis. Third and last, GBM subtypes were assessed following the single sample gene set enrichment analysis (ssGSEA) subtyping method provided by *Bioconductor* package for R (Hänzelmann, Castelo and Guinney, 2013).

Expression data for all samples to be analysed was downloaded also from GlioVis portal, and intersected with the treated phenotype data, resulting in the expression data to be subsequently studied.

1.3.2. Correlation studies

List of genes up- and downregulated upon ZEB1 knock-down and associated with at least one ZEB1 binding event following the nearest gene annotation from Rosmaninho and colleagues' study (Rosmaninho *et al.*, 2018) was intersected with expression data from each data set, in order to obtain exclusively the expression levels of these genes in primary GBM and non-tumour samples. Pearson's correlation analysis was performed using as base the R function *cor.test*. Statistical tests were obtained also in RStudio following a student's t-distribution under the null hypothesis, most known as *t*-test. Correlation plots were performed using *ggplot2* (version 3.3.2) (Wickham, 2016) and *ggpubr* (version 0.4.0) (Kassambara, 2020) packages.

1.3.3. Expression of ZEB TFs in GBM

The mean of expression levels of ZEB genes among the different data sets was obtained in all primary GBM samples, GBM samples categorized into subtype, and normal brain (non-tumour samples). Plots were obtained using the default R functions *boxplot* and *stripchart.* All statistical tests were obtained in RStudio following a *t*-test.

2. Results

In order to perform correlative studies using transcriptomic data sets of large cohorts of GBM patients, all the data sets available in GlioVis portal (http://gliovis.bioinfo.cnio.es/) were evaluated to select which ones to use for further treatment. Three data sets were selected, based on several aspects, such as the amount of GBM samples (more than 150), the existence of both tumour and non-tumour samples and the method used to quantify gene expression levels (Affymetrix HG U133) (see Materials and methods). This filtering resulted in three data sets: the Gravendeel (GEO Accession number GSE16011) (Gravendeel et al., 2009), TCGA (http://cancergenome.nih.gov/) and REMBRANDT (Gusev et al., 2018) (GEO Accession number GSE108474) data sets (Table 3.1). Two of these (Gravendeel and TCGA) had been previously used by Rosmaninho and colleagues (Rosmaninho et al., 2018).

Table 3.1 | Number of GBM samples among the different data sets used in this study

	Total GBMs	Primary GBMs	Mesenchymal Primary GBMs	Classical Primary GBMs	Proneural Primary GBMs	Non-tumour
Gravendeel	159	136	52	47	37	8
TCGA	528	482	164	197	121	10
REMBRANDT	219	208	71	71	66	28

In Rosmaninho and colleagues' study (Rosmaninho *et al.*, 2018), a combination of ZEB1 knockdown and ZEB1 ChIP-seq in a model of GSCs (NCH421k cells) resulted in a high-confidence list of 60 genes directly activated by ZEB1. Subsequently, validation of these results using transcriptomics data from Gravendeel data set resulted in the identification of 23 genes that were positively correlated with ZEB1 expression in primary GBM samples (Figure 3.1).

To further compare the gene regulation activities of both ZEB TFs, we started by correlating the expression of ZEB2 and each of those 23 genes. As perhaps expected, the number of genes positively correlated with ZEB2 was much lower (n= 7). However, when taken into consideration all 60 ZEB1 activated genes in NCH421k cells, 34 genes were positively correlated with ZEB2 (Figure 3.2A). This result was not in line with the opposing activities of ZEB TFs in transcriptional assays and was further expanded to the two additional data sets (TCGA and REMBRANDT).
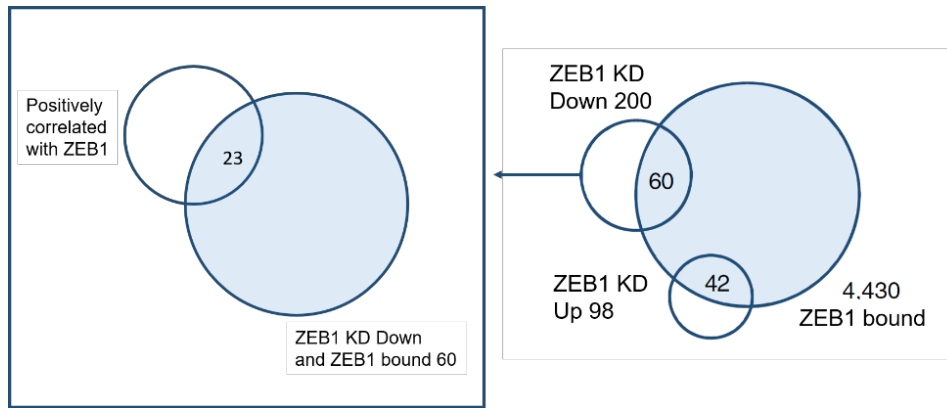
Figure 3.1 | Right: Venn diagram depicting the number of genes up and downregulated upon ZEB1 knockdown in NCH421k cells and unique genes associated with at least one ZEB1 binding event (ZEB1 bound). **Left:** Ven diagram depicting how many of the genes directly activated by ZEB1 in NCH421k cells are and positively correlated with ZEB1 expression in GBM tumours from the Gravendeel data set.

Abbreviations: KD, knockdown; Down, downregulated; Up, upregulated

(Modified from Rosmaninho et al., 2018)

When ZEB1 expression was compared with that of the 60 putative target genes using REMBRANDT data set, the number of positive (n= 22) and negative (n= 18) correlations found was almost equal. When the same analyses were extended to ZEB2, the number of positive and negative correlations was the same (n= 18). Strikingly, only one gene was found to be positively correlated simultaneously by both ZEB1 and ZEB2, while many (n= 27) were oppositely regulated by both ZEB TFs (Figure 3.2B). When the same analysis was extended to the TCGA data set, more genes were found to be positively correlated with ZEB1 (n= 25), than negatively correlated (n= 12). Likewise, more genes were positively correlated with ZEB2 (n= 22) than negatively correlated (n= 13). Strikingly, in this data set, a large number of genes was found to be equally correlated with both ZEB TFs, either positively (n= 18) or negatively (n= 5) (Figure 3.2C).

Next, correlative studies were extended to the list of 42 genes found to be repressed by ZEB1 in NCH421k cells, since Rosmaninho and colleagues did not investigate this point. Surprisingly, in Gravendeel data set, both ZEB1 and ZEB2 were shown to be much more positively (n= 7 and n= 24, respectively) than negatively (n= 3 and n= 2, respectively) correlated with genes repressed by ZEB1 in NCH421k cells (data not shown). The two other data sets revealed to be highly heterogeneous.

Given the unexpected differences of results obtained with the different data sets, we next characterized how the expression of each ZEB gene compares across GBM samples. According to all data sets, ZEB1 is overexpressed in GBM samples as compared to non-

tumour samples. By contrast, ZEB2 expression was found to be reduced in GBM samples from Gravendeel and REMBRANDT data sets, as compared to non-tumour samples, while being overexpressed in TCGA. Interestingly, when comparing ZEB1 and ZEB2 expression across subtypes, the two TFs show distinct profiles, suggesting they are differently expressed in GBM tumours. Again, differences were observed between data sets, with Gravendeel and TCGA showing similar results, but different from REMBRANDT (Figure 3.3).
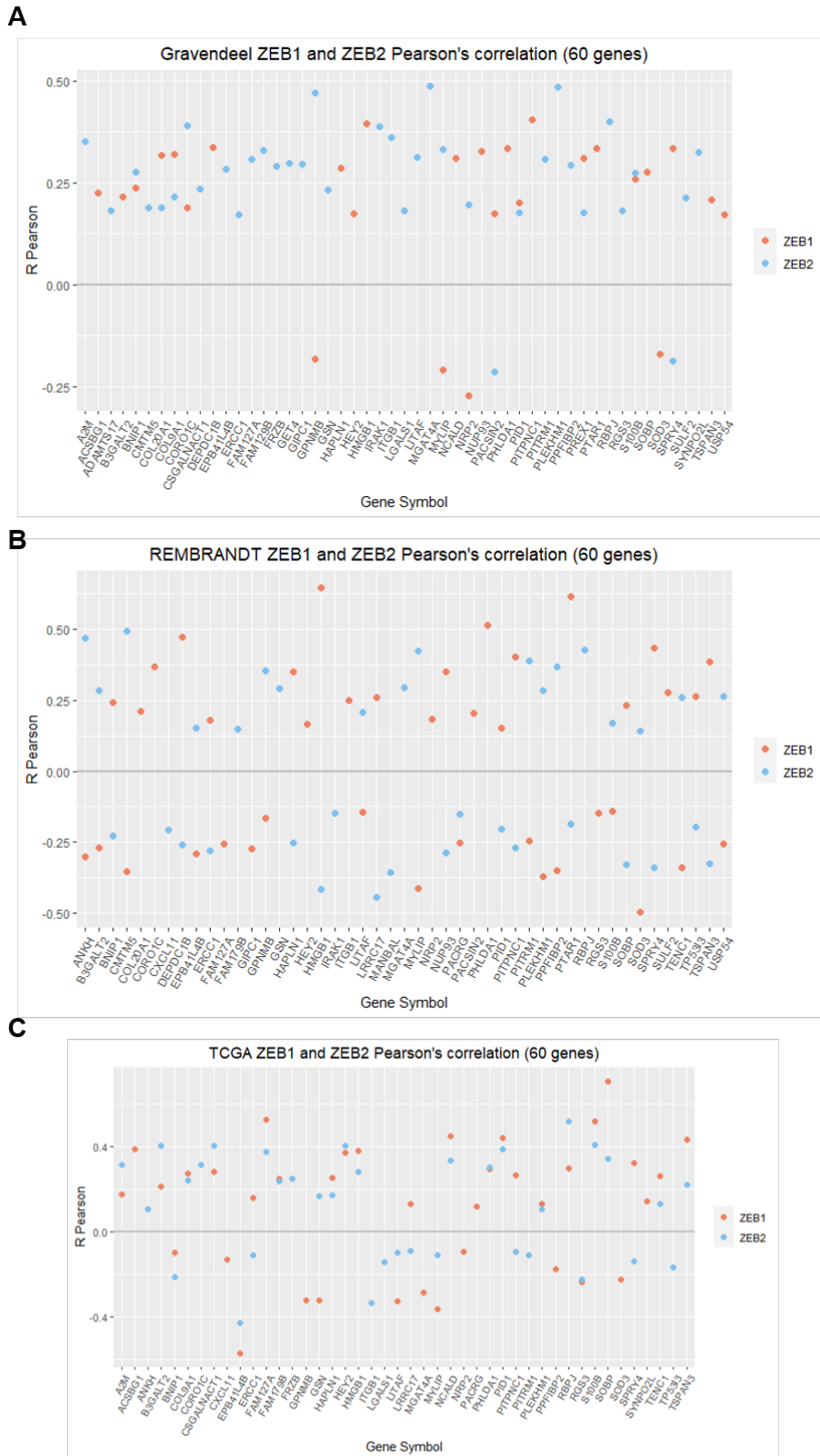
Figure 3.2 | ZEB1 targets downregulated upon ZEB1 knockdown correlated with ZEB1 and ZEB2 in primary GBM samples

Pearson's correlation of ZEB1 and ZEB2 expression levels with ZEB1 downregulated targets upon ZEB1 knockdown in primary GBM samples in **(A)** Gravendeel, **(B)** REMBRANDT, and **(C)** TCGA data sets. All correlations presented obtained p-value < 0.05.
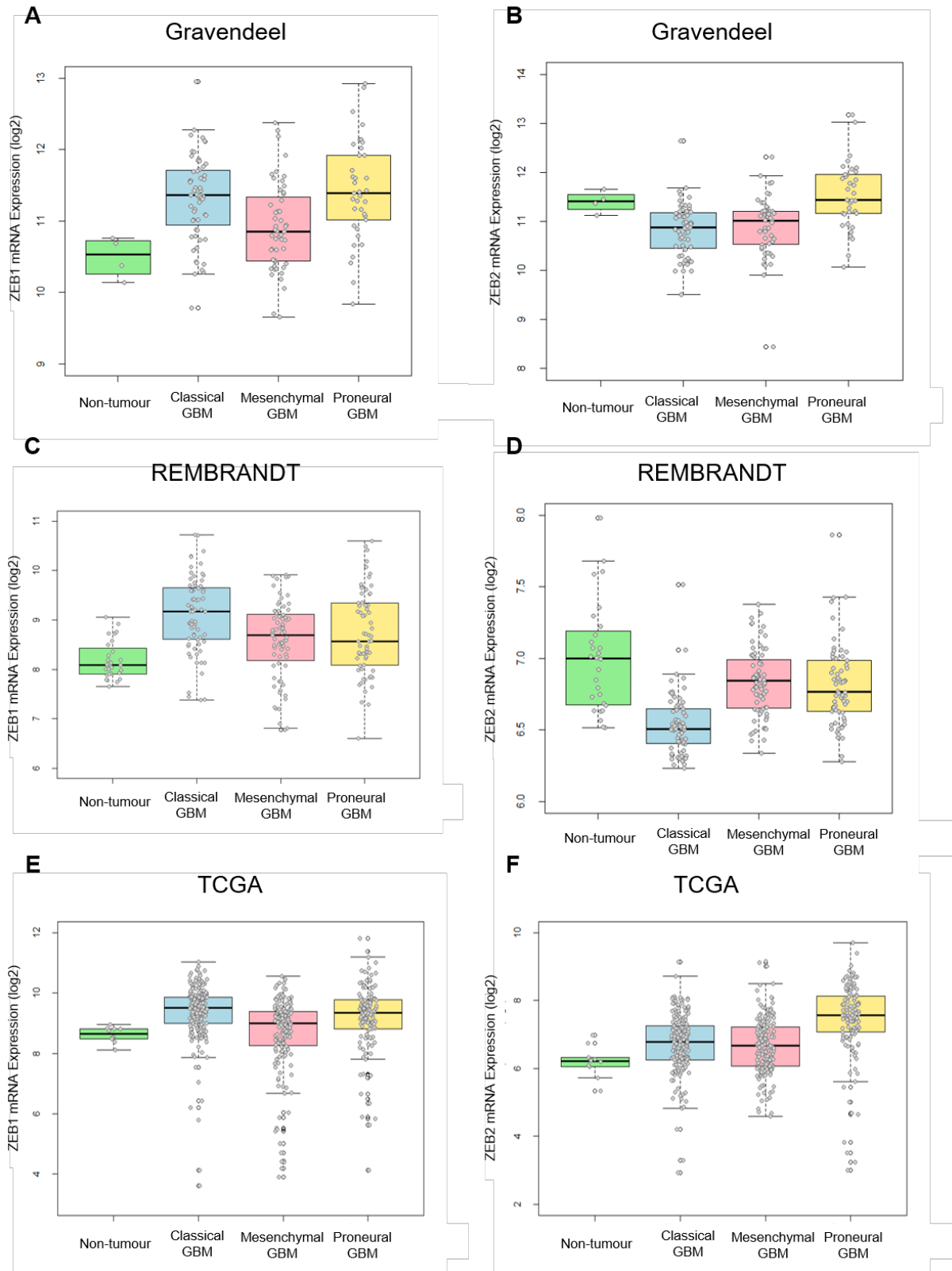
Figure 3.3 | Expression levels of ZEB1/2 genes in transcriptomics data from primary GBM tumours from publicly available data sets

Box plots representing levels of ZEB1 (left) and ZEB2 (right) transcripts in non-tumour and primary GBM tumour samples categorized in subtypes from the **(A,B)** Gravendeel, **(C,D)** REMBRANDT, and **(E,F)** TCGA datasets.

Next, we investigated any possible correlation between ZEB genes themselves. Even though their expression was not found to be correlated in Gravendeel data set (**Error! Reference source not found.**A,D), REMBRANDT (**Error! Reference source not found.**B,E) and TCGA (**Error! Reference source not found.**C,F) showed opposite correlation directions, and this trend was consistent among the different GBM subtypes.

In conclusion, using this type of correlative studies to compare the activities of ZEB TFs and validate the mechanistic insights provided by the transcriptional assays may be difficult, due to the confounding effect resulting from the fact that expression of both ZEB genes seem to be contradictory, highly oppositely correlated in two of the three data sets examined. The basis for the striking differences observed across data sets, namely REMBRANDT and TCGA, is currently not understood (see Discussion).

Figure 3.4 | Correlative expression levels between ZEB1 and ZEB2 in all GBM or subtype-specific GBM samples

(A-C) Scatter plots showing correlative expression levels between ZEB1 and ZEB2 genes in all primary GBM samples from Gravendeel, REMBRANDT, and TCGA datasets, respectively. R is Pearson's correlation coefficient. **(D-F)** Correlative expression levels between ZEB1 and ZEB2 genes in primary GBM samples categorized in different subtypes from Gravendeel, REMBRANDT, and TCGA datasets, respectively.

40

A

Gravendeel

B

REMBRANDT

C

TCGA

D

ZEB1 and ZEB2 correlation
(Gravendeel)

Subtype	Pearson's r	p-value
Proneural	-0,23425	0,162879
Classical	0,05267	0,72513
Mesenchymal	0,210255	0,134628

E

ZEB1 and ZEB2 correlation
(REMBRANDT)

Subtype	Pearson's r	p-value
Proneural	-0.2697161	0.02851844
Classical	-0.4697961	3.574561e-05
Mesenchymal	-0.1875143	0.1173713

F

ZEB1 and ZEB2 correlation
(TCGA)

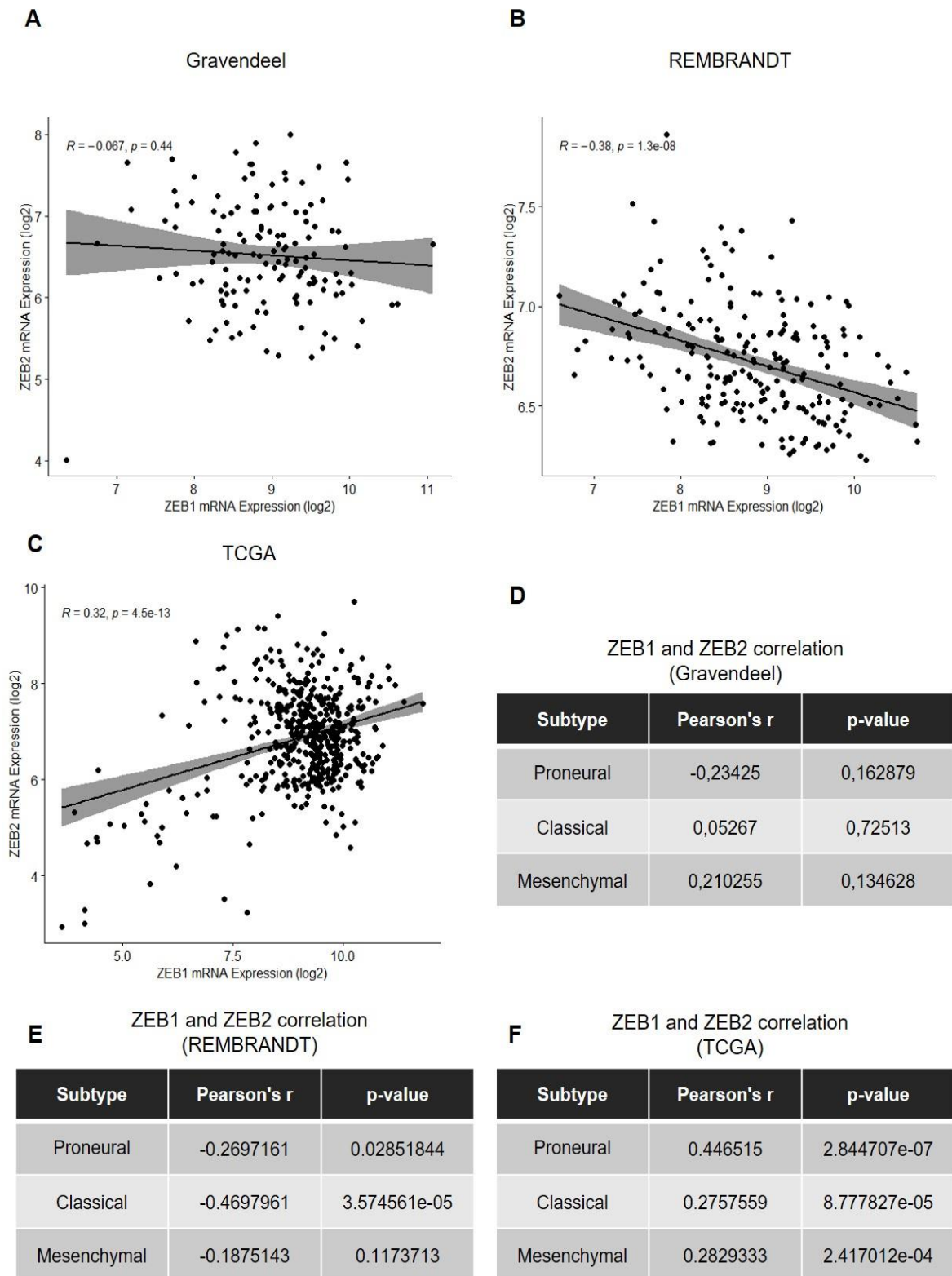| Subtype	Pearson's r	p-value
Proneural	0.446515	2.844707e-07
Classical	0.2757559	8.777827e-05
Mesenchymal	0.2829333	2.417012e-04

Figure 3.4

CHAPTER FOUR

ESTABLISHING A CHROMATIN IMMUNOPRECIPITATION PROTOCOL FOR ZEB2

1. Materials and methods

1.1. Cell culture

1.1.1. NCH421K cells

NCH421K cells (Campos et al., 2010) were cultured in DMEM-F12 GlutaMAX medium (Gibco) supplemented with 1 × N-2 supplement (Gibco), 0.05 × B-27 supplement (Gibco), Penicillin-Streptomycin (100 U/mL, Gibco), epidermal growth factor (EGF, 10 ng/mL, Peprotech) and basic fibroblast growth factor (bFGF, 10 ng/mL, Peprotech) in T-flasks, plates or well plates (Corning) pre-coated with sterile filtered Poly-L-Lysine (Sigma-Aldrich) and Laminin (1 µg/mL, Sigma-Aldrich) for approximately 45 and 60 min, respectively. When indicated, cells were grown in the presence of 5 µM of the Wnt agonist CHIR99021 (Sigma-Aldrich) for 24h.

1.2. Preparation of protein lysates from P19 and NCH421k cells

Transfection and preparation of extracts from P19 cells was carried out as described in Chapter 2. Protein lysates from NCH421k cells were obtained and quantified as for P19 cells, from NCH421k cells growing in normal culture conditions in 6 well plates.

1.3. Western Blot

Western blot analysis was performed as previously described in Chapter 2, using primary and secondary antibodies listed in Table 2.4 and Table 2.5.

1.4. Chromatin isolation from NCH421K cells

Three T175 cm^2 flasks containing approximately 25 million NCH421k cells each were used per experiment. Firstly, cells were fixed by replacing the existing media with 2 mM of a freshly made 0.5 M disuccinimidyl-glutarate (DSG, Sigma-Aldrich) solution in PBS (Sigma-Aldrich). After incubation with occasional agitation for 45 min at RT, the PBS-DSG fixative solution was replaced by 1 % of formaldehyde (Sigma-Aldrich) in PBS and incubated for 10 min at RT. To quench the reaction, 125 mM of glycine pH 5.2 in PBS was added to each flask, followed by an incubation of 5 min at RT under soft agitation. Cells were washed once with PBS and then scrapped into 8 mL of PBS containing 1 × protease inhibitors cocktail (Roche) and 0.75 % of bovine serum albumin (BSA, Gibco). Cells were transferred into two 15 mL conical falcon tubes and centrifuged for 7 minutes at 1,200 rpm and 4 ºC. Cell pellets were resuspended in 5 times the volume of SDS lysis buffer (1 % SDS, 10 mM EDTA, 50 mM Tris pH 8.0, 1 × complete protease inhibitors cocktail (Roche)), transferred to non-sticky RNase-

free tubes (Ambion) and incubated for 20 minutes at 4 ºC on a rocking platform. Chromatin was then sheared by sonication using the Bioruptor Plus (Diagenode) sonicator at high power settings (200 watts (W)) for 30 min in 30 seconds ON/OFF cycles at 4 ºC. After a 10 min centrifugation at 4 ºC and 15,000 rpm to remove any precipitated material, the DNA concentration was measured in NanoDrop 1000 Spectrophotometer (Thermo Fisher). Chromatin preparations were aliquoted, snap-frozen in liquid nitrogen and stored at -80 ºC. To verify the efficiency of the sonication, one aliquot of chromatin was subjected to reverse crosslink by incubating ON at 65 ºC, followed by proteinase K (0.1 mg/mL, Roche) digestion during 2 h at 42 ºC. DNA was purified by phenol-chloroform extraction and the efficiency of chromatin sonication was assessed by running 1 and 2 µg of DNA in a 2 % agarose gel.

1.5. Chromatin immunoprecipitation (ChIP)

At the beginning of the experiment, the total amount of Protein G Dynabeads (Invitrogen) required for this protocol was blocked by washing 4 × with freshly made IP buffer (0.2 M of HEPES pH 8.0 (Sigma-Aldrich), 2 M of NaCl (Sigma-Aldrich) and 0.02 M of EDTA (Sigma-Aldrich), 0.1 % of sodium deoxycholate (Na-DOC, Sigma-Aldrich), 1 % of Triton X-100 (Sigma-Aldrich), 1 mg/mL of BSA (Promega), and 1 × of complete protease inhibitor (Roche)).

Chromatin was pre-cleared by adding 50 µL of the pre-blocked beads to the whole amount of chromatin used per experiment (diluted in twice the volume of IP buffer), rocking, for 2 h at 4 ºC, before setting up IP reactions. At this point, 50 µL of IP reaction was saved from the mock IP reaction (corresponding to 5 % of input chromatin). Each IP reaction contained at least 50 µg of chromatin (diluted in 100 µL of lysis buffer), in a total volume of 1ml in IP buffer. Antibodies used were 4 µl of anti-ZEB1 (HPA027524, Sigma-Aldrich) and 3 µl or 6 µl of anti-ZEB2 (ab223688, ABCAM). One control tube without antibody (mock) was used. All tubes stayed rocking ON at 4 ºC.

On the next day, 50 µL of pre-blocked beads were added to each IP sample, following by incubation for 2 h at 4 ºC. Beads were captured with a magnet and washed 5 × with freshly made washing buffer (0.5 M of HEPES pH 7.6, 1 mM of EDTA, 1 % of NP-40, 0.7 % of Na-DOC, and 0.5 M of LiCl), followed by a single wash with Tris-EDTA (TE). All washes were done for 4 minutes at 4 ºC, rocking in between. Beads were eluted with 500 µL of freshly made elution buffer (10 mM of Tris pH 8.0 and 1 % of SDS). Input chromatin was diluted with 450 µL of elution buffer and treated in parallel with IP samples starting from this point. Samples were incubated for 10 minutes at 65 ºC, and beads captured with a magnet. Supernatant was recovered and digested upon addition of 5 µL of proteinase K (10 mg/mL, Roche) and 11 µL of 5 M NaCl, by incubation at 42 ºC for 2 h. Tubes were then transferred to 65 ºC and incubated ON to reverse the crosslinks.

46

On the following day, samples were extracted twice with an equal volume of phenol:chloroform (equilibrated with TE, pH 8.0) and once with phenol:chloroform:isoamyl alcohol (25:24:1, Sigma Aldrich). Next, DNA precipitation was performed by adding 2 µL of glycogen (40 µg, Sigma-Aldrich), 50 µL of 3 M of NaAc pH 5.2 (Sigma-Aldrich), and 0.9 mL of isopropanol, following by a 20 min incubation at - 20 ºC. After a 15,000 rpm centrifugation at 4 ºC, for 15 min, the supernatant was discarded and the pellets were washed with 75 % ethanol, left to dry for approximately 2 min and resuspended in 120 µL of water (Sigma-Aldrich).

1.6. Quantitative PCR

Quantitative PCR (qPCR) was performed in 384 well plates, with three technical replicates. Reactions were performed in 10 µL volume and contained 3 µL of immunoprecipitated chromatin, 5 µL of SYBR Green (Invitrogen), 0.5 µL of forward primer (0.25 µM), 0.5 µL of reverse primer (0.25 µM), and 1.5 µL of water (Sigma-Aldrich), using CFX384 Touch Real-Time System (Bio-Rad). The software was set to perform 40 cycles of 3 min and 20 seconds at 95 ºC followed by 30 seconds at 60 ºC; lastly, ending with an increment of 0.5 ºC every 10 seconds from 55 to 95 ºC. Data were analysed using Bio-Rad CFX Maestro 1.0 software.

Table 4.1 | Primers used in ChIP-qPCR

Primer	Forward sequence	Reverse sequence
Axin2 (ORF I)	CATCCCATCCAACACAACCC	TTTGCACTACGTCCCTCCAA
Fbxw7 (ORF II)	ATTCACCCGTTTTCAAGTCC	CTAGGTCCCAACAAGCATCA
Pard6b	AGCCGAGCCCTTCTTCAG	CTCCTCAAAACCCCGCCTA
ZEB1 promoter	CAGAGCCCAGCACTATTCT	GCCGGAACCTTGTTGCTA
Prex1	CTCACACTCAGGCCTTTGTC	GAGTGTTTGTGGGGAAGTGTC
Nrp2	AAACTCCTACAGGCCAGGTC	GTCACATGAGGCATTCATCC

2. Results

With the final aim of generating a ZEB2 ChIP-seq sample from GBM CSCs, the expression of ZEB2 in NCH421k cells was first investigated by Western blot analysis. A sample of NCH421k cells was treated with the Wnt pathway agonist CHIR99021, since this was previously shown to decrease ZEB1 protein expression, with a concomitant increase of ZEB2 transcript (Pedro Rosmaninho, unpublished). As controls for the western blot, lysates of P19 cells transfected with either human ZEB1 or ZEB2 expression plasmids were analysed in parallel. Western blot analysis revealed the presence of a band recognized by the ZEB2 antibody in NCH421k cells, running at the expected size. Treatment with CHIR99021 resulted in decreased expression of ZEB1, but ZEB2 protein levels were not altered. Importantly, the ZEB2 antibody tested did not recognize ZEB1 protein generated in transfected P19 cells and recognized by a ZEB1-specific antibody.
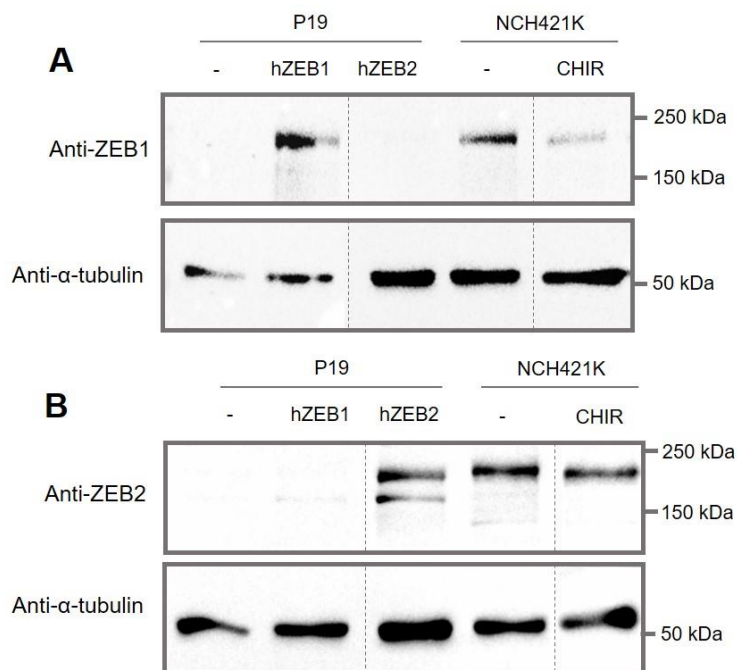
Figure 4.1 | Western blot analysis of ZEB proteins expression in NCH421K cells.

(A,B) Expression of ZEB1 and ZEB2, respectively, in P19 cells transfected with both ZEB factors expression constructs and NCH421K cells treated and not treated with CHIR99021, as indicated, assessed by Western blot analysis. α-tubulin is shown as loading control.

Since NCH421k cells were shown to express ZEB2 protein, chromatin from these cells (grown in the absence of CHIR99021) was extracted and sonicated (Figure 4.2), in preparation for a chromatin immunoprecipitation protocol. Agarose gel electrophoresis revealed proper chromatin sonication, with an accumulation of DNA running below 0.7 Kb.

48

Figure 4.2 | Assessment of sonication efficiency of NCH421k chromatin, by electrophoresis on a 2 % agarose gel

Chromatin from NCH421k cells was extracted and sonicated. Agarose gel electrophoresis was used to assess the sonication efficiency.
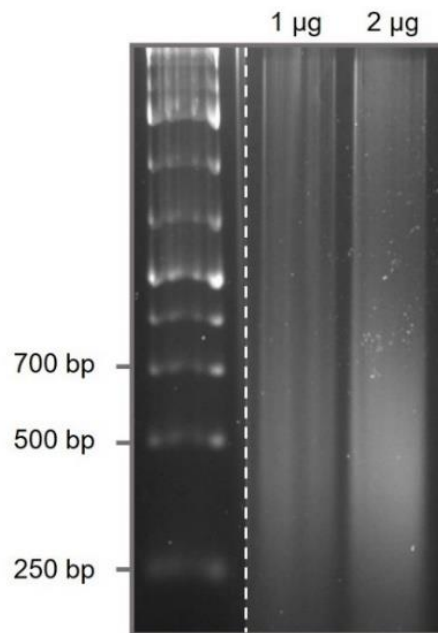
Next, extracted chromatin was used in a ChIP-qPCR protocol testing two quantities of the ZEB2 antibody, and assessing ZEB2 recruitment to a series of genomic regions. Previously characterized ZEB1 antibody was used as control. As expected, strong enrichment of proximal promoter regions of Pard6b and ZEB1 genes (spanning various E-box sequences) (Rosmaninho et al) as compared to a negative control region located within the open reading frame (ORF) of the Axn2 gene was observed, in presence of the ZEB1 antibody (Figure 4.2A). Importantly, strong enrichment was also observed with ZEB2 antibody, with no significant differences between the two tested conditions (3 µL and 6 µL) (Figure 4.2A). In a second ChIP-qPCR experiment, binding to regulatory regions of Prex1 and Nrp2 genes was also tested (Figure 4.2B). As expected, weaker but significant enrichment of Prex1 and Nrp2 regions was detected, in the condition with ZEB1 antibody. These regions were also significantly enriched in the presence of ZEB2 antibody, while no enrichment was observed at any region in the absence of antibody. Thus, results indicate the ZEB2 antibody works in ChIP assay. In addition, it reveals binding of ZEB2 to regulatory regions of Prex1 and Nrp2 genes, in line with previous transcriptional assays. (Figure 4.2B).

Next, the ChIP protocol was used in the preparation of a ZEB2 ChIP-seq sample. This required upscaling the protocol so that enough ChIPed material could be collected for sequencing (10 ng of DNA). With this aim, three ZEB2 samples were used in parallel, in the absence of a control (mock) reaction. Pellets were resuspended in 30 µL of water, with a small

fraction used for qPCR analysis. In order to save a maximum amount for sequencing, only two primers sets were tested. Reassuringly, a strong enrichment in Pard6b regulatory region was seen, as compared to the ORF negative region (Figure 4.4). Thus, the sample concentration was further measured using Qubit 2.0 Fluorometer (Invitrogen) and found to contain 18 ng of DNA.
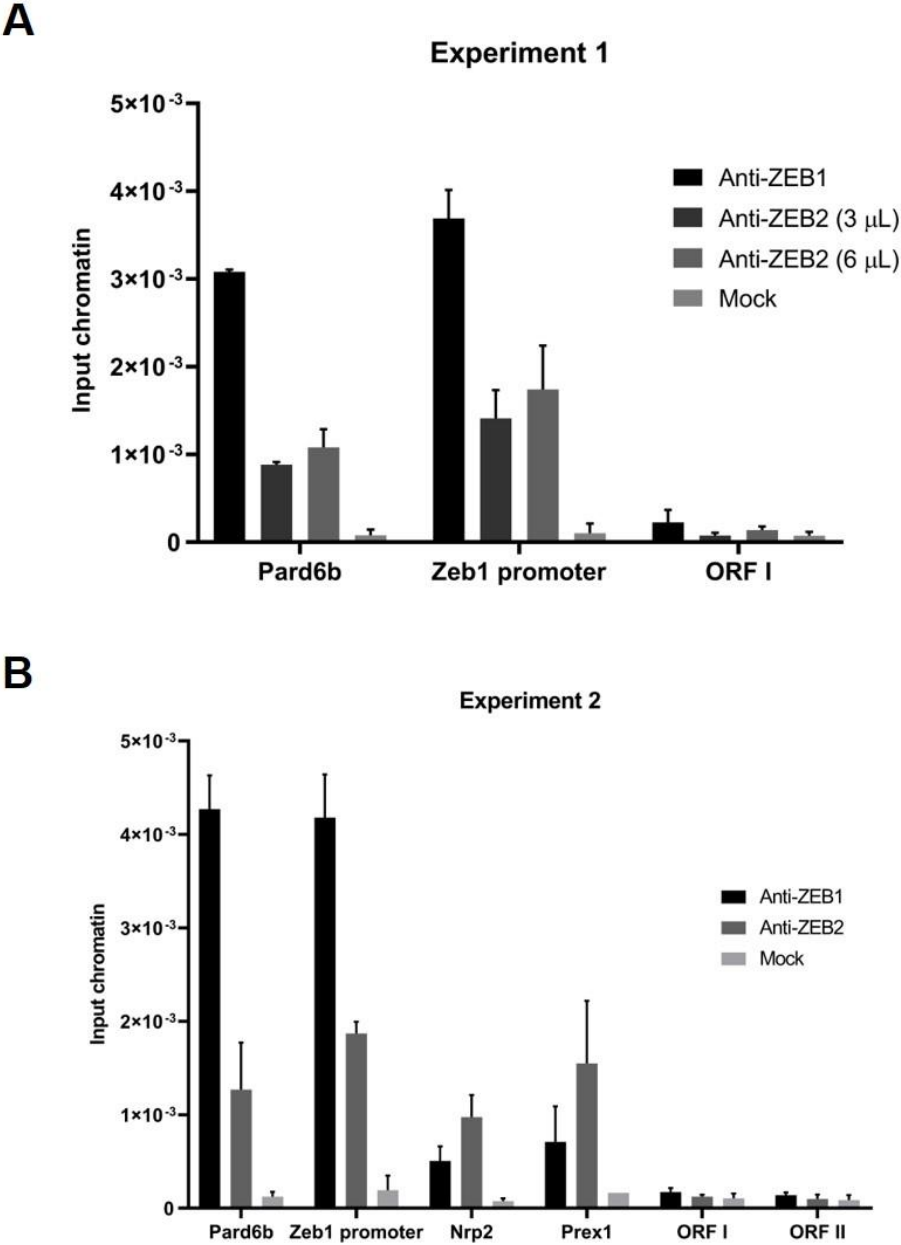
Figure 4.3 | Two ChIP-qPCR experiments in NCH421k cells, assessing ZEB1 and ZEB2 recruitment to previously characterized regulatory regions.

Genomic regions within the ORFs of Axin2 (ORF I) and Fbxw7 (ORF II) genes were used as negative control (non-bound) regions. Data are presented as mean ± SD of three technical replicates.
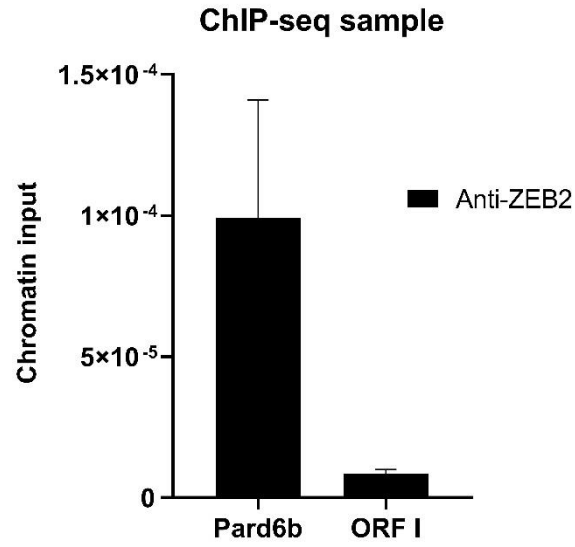
ChIP-seq sample

Figure 4.4 | Quality control qPCR experiment of a ChIP-seq sample prepared from ZEB2 in NCH21k cells, assessing ZEB2 recruitment to a Pard6b regulatory region in NCH421k cells. A genomic region within the ORF of Axin2 gene (ORF I) was used as negative control (non-bound) region. Data are presented as mean ± SD of three technical replicates.

CHAPTER FIVE

GENERAL DISCUSSION

Discussion

The focus of this project settled in the previous observations from our group that ZEB1 interacts with Lef1 to activate a large number of its target genes in GBM (e.g. Nrp2 and Prex1) (Rosmaninho *et al.*, 2018). By contrast to ZEB1, ZEB2 transcriptional mechanisms remain poorly understood. Thus, one of the main goals of this study was to understand how ZEB2 behaves in several transcriptional paradigms already studied for ZEB1. Using transcriptional assays, we first compared the activity of ZEB TFs in three paradigms where ZEB1 functions as an activator (Sánchez-Tilló *et al.*, 2015; Lehmann *et al.*, 2016; Rosmaninho *et al.*, 2018). Strikingly, our results demonstrate that ZEB2 behaves differently than ZEB1 in the gene activation paradigms tested. Instead of functioning in synergy with LEF1 or TEAD/YAP, ZEB2 shows repressive activity. Results with a third model, regulating the WNT signalling target genes LAMC2 and uPA, were not reproduced. The original experiments by Sánchez-Tilló and colleagues were performed in SW480, SW620, HCT116 and HT29 colorectal cancer cells. Because we used P19 cells in all our transcriptional assays, one very likely possibility for our negative results is the lack of an appropriate cell context. We were therefore unable to test ZEB2 in this gene regulation model.

There are some hypotheses for the molecular basis for the opposing activities of ZEB1 and ZEB2. ZEB proteins are quite conserved, and both TFs present a conserved domain essential for recruitment of CtBP, partially responsible for transcriptional repression. In contrast to ZEB2 however, ZEB1 presents two known activation domains – one located closer to the N-terminus, responsible for p300 and P/CAF co-activators recruitment, and another closer to C-terminus (Figure 1.3). Some regions close to N- and C- terminus are not well conserved between the two ZEB TFs. The lack of required activator domains may be a possibility why ZEB2 does not promote gene activation in the assays tested (Postigo *et al.*, 2003; Vandewalle, Van Roy and Berx, 2009). To continue addressing this point, namely in the context of previous work by Rosmaninho and colleagues, it will be important to clearly map the protein domains of ZEB1 important for activation of the Prex1 and Nrp2 genes in synergy with LEF1. Another possibility to explain the results obtained would be the inability for ZEB2 to be recruited to the regulatory regions associated with gene activation, for example by not being able to physically interact with other TFs involved. This does not seem to be the case in the Prex1 and Nrp2 regulatory regions, which are bound by ZEB2 in ChIP-qPCR, and where ZEB2 has a repressive activity in transcriptional assays.

Our results are similar to the only previous study that tested ZEB2 activity in a ZEB1 activation model. Postigo and colleagues have shown that ZEB1 activates gene transcription downstream TGF-β signalling, in a mechanism that requires direct interaction with SMAD

proteins. In case of ZEB2, a similar interaction with SMAD proteins results in gene repression instead, and this difference has been attributed to the absence of the p300/PCAF interaction domain in ZEB2 (Postigo *et al.*, 2003).

Furthermore, we used protein deletion mutants to investigate whether the presence of known activation domains of ZEB1 is crucial for the activation of its target genes. The results show that deletions of N-, C- or both terminus regions are sufficient to abrogate the combined Lef1 and ZEB1 highly efficient transcriptional activation of Nrp2 regulatory region. Although, one should be cautious because only one experiment was performed due to time constraints and protein levels of these various ZEB1 derivatives were not controlled. Protein deletions often compromise protein stability, resulting in processes (e.g. structural modifications, target for degradation) that lower the plasmid expression (Shortle and Sondek, 1995). Since protein levels with these deletions' plasmids were not determined, it would be interesting to perform a Western blot to discard a possible protein malfunction and consecutively lack of expression.

Even though studies are being done to understand the molecular mechanisms of ZEB proteins, we still lack a complete picture of how the expression patterns of the two TFs compare in some contexts, as it is the case in GBM. The fact that both proteins are expressed in the GSC model used (i.e. NCH421k cell line), raises the possibility that these could be co-expressed in some GBM cells. Therefore, one interesting experiment to perform in the future would be the co-expression of ZEB1 and ZEB2 in the activating paradigms investigated in this study, to assess to which extent can ZEB2 abolish ZEB1 activation competitively and/or if it is concentration-dependent. Future studies should address this using transcriptional assays.

The previous study from Rosmaninho and colleagues performed a correlational analysis between ZEB1, and the 60 genes found to be activated by ZEB1 in NCH421k cells, using transcriptomics data from large cohorts of GBM patients. One important goal of this study was to expand these correlational studies to ZEB2. In addition to the two transcriptomics data sets used previously (Gravendeel and TCGA), studies with ZEB2 also used the REMBRANDT data set. Importantly, all three data sets used a similar Affymetrix array platform. We found 34 (out of 60) genes to be positively correlated with ZEB2 (with 11 genes correlated with both ZEB1 and ZEB2). This was unexpected, assuming most of the identified 60 genes are regulated in NCH421k cells by a similar mechanism to that described for Prex1 and Nrp2. Given the repressive activity of ZEB2 observed in the Prex1 and Nrp2 regulatory regions, one may thus expect this to result in a negative correlation (if any) between ZEB2 and those candidate targets. In conclusion, our observations point at the importance of performing these types of correlation studies with adequate negative controls.

Another surprising observation was the large difference in results observed across GBM data sets. The differences between REMBRANDT and TCGA were striking, with many activated ZEB1 targets in NCH421k cells being oppositely correlated with each ZEB factor in REMBRANDT but instead correlated in the equal direction in TCGA. Cross-regulatory interactions are very common amongst members of the same family of TFs. One possibility is that the observed correlations involving ZEB2 result mostly from its direct regulation by ZEB1. In support of this idea, ZEB1 is found to directly bind and repress the promoter of the ZEB2 gene in NCH421k cells (Rosmaninho *et al.*, 2018). However, this still does not explain the opposite contrasting results obtained across data sets. Correlational studies were also performed between Zeb genes and the 42 targets repressed by ZEB1 in NCH421k cells. Once again, results were very different across data sets, and largely inconclusive. Unexpectedly, *in silico* results were found to be highly inconsistent across data sets, which could be explained by some hypotheses. The method used to determine gene expression is not expected to be the reason, because data sets were chosen based on the use of similar platforms (DNA arrays). GBM has shown intra- and inter-tumoral heterogeneity (Phillips *et al.*, 2006; Darmanis *et al.*, 2017). The way the sample is collected is therefore crucial for the analysis, because depending on the region resected there may be differences on the molecular profile as well. This is also important, given the high number of infiltrating immune cells found in GBM tumours, and which have a different expression profile from cancer cells, affecting the tumour transcriptomic characterization.

It is important to note that these transcriptomic data are obtained using a whole-tumour, not single-cell approach, something to consider when considering gene regulatory mechanisms. A solution to this limitation would be to use single-cell RNA sequencing or spatial transcriptomics. This would allow us to correlate TFs with target genes within (or close to) single-cell resolution.

Lastly, the cell line used (NCH421k) may only be representative of a small fraction of GBM tumours, due to the high heterogeneity seen in GBM. The same way GBM tumours have subtypes depending on the gene expression profile, GSC lines can also be associated with a specific GBM subtype (Phillips *et al.*, 2006). Since NCH421k cells have a proneural gene expression profile, this cell line is representative of only one of the various GBM subtypes. A way to overcome this issue would be using, in parallel, other cell models that represent the other GBM subtypes.

DNA-protein interactions can be studied through a variety of assays. Chromatin immunoprecipitation (ChIP) allows these interactions to be studied in an *in vivo* context. ChIP is widely used when studying the regulation of many important cellular functions, including gene transcription (Das *et al.*, 2004). One important application is to study binding of TFs to

their genomic sites. ChIP can be combined with qPCR for testing interactions with candidate genomic regions (ChIP-qPCR), or with Next Generation Sequencing (NGS), to map protein/DNA interactions on a genome-wide level (ChIP-seq). A ChIP protocol for ZEB1 has been previously established by Rosmaninho and colleagues and used in ChIP-seq in the context of NCH421k cells. Their study revealed important information on how ZEB1 is recruited to target genes, and how this impacts gene transcription.

We successfully tested a ZEB2 antibody in a ChIP protocol, using chromatin extracted from NCH421k cells for ZEB2. The fact that NCH421k cells express both ZEB proteins in normal growing conditions, allows the activity of the two TFs to be compared in the same cellular context. As expected, ZEB1 direct binding events were validated with ZEB2 (i.e. Pard6b and Zeb1 promoter). In addition, binding of ZEB2 to the regulatory regions of Nrp2 and Prex1 genes was also confirmed. This was concordant with results from transcriptional assays from this study, confirming that repression of LEF1 in these regulatory regions is associated with direct binding of ZEB2.

Using the ZEB2 ChIP protocol, a ChIP-seq sample was prepared from NCH421k cells and sent for sequencing. This will allow the binding profiles of both ZEB TFs to be compared. Based on the results from ChIP-qPCR, one may expect these to be largely overlapping. However, in many cases, common binding events are not expected to result in identical regulatory events. In future studies, it will be important to combine ChIP-seq data of ZEB2 with gene knockdown experiments followed by gene expression profiling. This will indicate how binding events associate with gene activation and repression. Moreover, these studies should also be extended to CSC lines representative of the various GBM subtypes.

CHAPTER SIX

REFERENCES

References

Berghe, V. Van Den *et al.* (2013) 'Directed Migration of Cortical Interneurons Depends on the Cell-Autonomous Action of Sip1'. doi: 10.1016/j.neuron.2012.11.009.

Billin, A. N., Thirlwell, H. and Ayer, D. E. (2000) 'β-Catenin–Histone Deacetylase Interactions Regulate the Transition of LEF1 from a Transcriptional Repressor to an Activator', *Molecular and Cellular Biology*, 20(18), pp. 6882–6890. doi: 10.1128/mcb.20.18.6882-6890.2000.

Bressler, J. *et al.* (2011) 'P19 embryonic carcinoma cell line: A model to study gene-environment interactions', in Aschner, M., Sunol, C., and Bal-Price, A. (eds) *Cell Culture Techniques*. Neuromethods, pp. 223–240. doi: https://doi.org/10.1007/978-1-61779-077-5_10.

Castro, D. S. *et al.* (2006) 'Proneural bHLH and Brn Proteins Coregulate a Neurogenic Program through Cooperative Binding to a Conserved DNA Motif', *Developmental Cell*, 11(6), pp. 831–844. doi: 10.1016/j.devcel.2006.10.006.

Dang, L. *et al.* (2009) 'Cancer-associated IDH1 mutations produce 2-hydroxyglutarate', *Nature*. Nature Publishing Group, 462(7274), pp. 739–744. doi: 10.1038/nature08617.

Darmanis, S. *et al.* (2017) 'Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma', *Cell Reports*. ElsevierCompany., 21(5), pp. 1399–1410. doi: 10.1016/j.celrep.2017.10.030.

Das, P. M. *et al.* (2004) 'Chromatin immunoprecipitation assay', 37(6), pp. 1–9. Available at: papers2://publication/uuid/C108069F-D98D-41C5-B9E8-097F05AC3866.

Deng, Z. *et al.* (2016) 'Epigenetic regulation of IQGAP2 promotes ovarian cancer progression via activating Wnt/catenin signaling', *International Journal of Oncology*, 48(1), pp. 153–160. doi: 10.3892/ijo.2015.3228.

Deorah, S. *et al.* (2006) 'Trends in brain cancer incidence and survival in the United States: Surveillance, Epidemiology, and End Results Program, 1973 to 2001', *Neurosurgical Focus*, 20(4), pp. 1–7.

Depner, C. *et al.* (2016) 'EphrinB2 repression through ZEB2 mediates tumour invasion and anti-angiogenic resistance', *Nature Communications*, 7. doi: 10.1038/ncomms12329.

Derynck, R. and Zhang, Y. E. (2003) 'Smad-dependent and Smad-independent pathways in TGF-b family signalling', *Nature*, 4, pp. 577–584. doi: 10.1038/nature02006.

Eastman, Q. and Grosschedl, R. (1999) 'Regulation of LEF-1/TCF transcription factors by Wnt and other signals', *Current Opinion in Cell Biology*, 11(2), pp. 233–240. doi:

10.1016/S0955-0674(99)80031-3.

Euskirchen, P. *et al.* (2017) 'Cellular heterogeneity contributes to subtype-specific expression of ZEB1 in human glioblastoma', *PLoS ONE*, 12(9), pp. 1–15. doi: 10.1371/journal.pone.0185376.

Fernandes, C. *et al.* (2017) 'Current Standards of Care in Glioblastoma Therapy', in Vleeschouwer, S. De (ed.) *Glioblastoma*. 1st edn. Brisbane, Australia: Codon Publications. doi: 10.15586/codon.glioblastoma.2017.ch11.

Frosina, G. (2009) 'DNA repair and resistance of gliomas to chemotherapy and radiotherapy', *Molecular Cancer Research*, 7(7), pp. 989–999. doi: 10.1158/1541-7786.MCR-09-0030.

Gilbert, M. R. *et al.* (2014) 'A randomized trial of bevacizumab for newly diagnosed glioblastoma', *New England Journal of Medicine*, 370(8), pp. 699–708. doi: 10.1056/NEJMoa1308573.

Gravendeel, L. A. M. *et al.* (2009) 'Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology', *Cancer Research*, 69(23), pp. 9065–9072. doi: 10.1158/0008-5472.CAN-09-2307.

Gusev, Y. *et al.* (2018) 'Data descriptor: The REMBRANDT study, a large collection of genomic data from brain cancer patients', *Scientific Data*, 5, pp. 1–9. doi: 10.1038/sdata.2018.158.

H. J. Scherer, M. D. (1940) 'Cerebral Astrocytomas and Their Derivatives', *Am J Cancer*, 40, pp. 159–98. doi: 10.7723/antiochreview.72.3.0546.

Hänzelmann, S., Castelo, R. and Guinney, J. (2013) 'GSVA: Gene set variation analysis for microarray and RNA-Seq data', *BMC Bioinformatics*, 14, pp. 1–21. doi: 10.1186/1471-2105-14-7.

Hay, E. D. (1995) 'An overview of epithelio-mesenchymal transformation', *Acta Anatomica*, pp. 8–20. doi: 10.1159/000147748.

Holland, J. D. *et al.* (2013) 'Wnt signaling in stem and cancer stem cells', *Current Opinion in Cell Biology*. Elsevier Ltd, 25(2), pp. 254–264. doi: 10.1016/j.ceb.2013.01.004.

Horbelt, D., Denkis, A. and Knaus, P. (2012) 'A portrait of Transforming Growth Factor β superfamily signalling: Background matters', *International Journal of Biochemistry and Cell Biology*. Elsevier Ltd, 44(3), pp. 469–474. doi: 10.1016/j.biocel.2011.12.013.

Jensen, S. S. *et al.* (2016) 'Establishment and characterization of a tumor stem cell-based glioblastoma invasion model', *PLoS ONE*, 11(7). doi: 10.1371/journal.pone.0159746.

Johnson, D. R., Leeper, H. E. and Uhm, J. H. (2013) 'Glioblastoma survival in the United States improved after Food and Drug Administration approval of bevacizumab: A population-based analysis', *Cancer*, 119(19), pp. 3489–3495. doi: 10.1002/cncr.28259.

Kahlert, U. D. *et al.* (2015) 'ZEB1 Promotes Invasion in Human Fetal Neural Stem Cells and Hypoxic Glioma Neurospheres', *Brain Pathology*, 25(6), pp. 724–732. doi: 10.1111/bpa.12240.

Kalluri, Raghu, Neilson, E. G. and Kalluri, R (2003) 'Epithelial-mesenchymal transition and its implications for fibrosis Find the latest version : Epithelial-mesenchymal trasition and its implications for fibrosis', *J Clin Invest*, 112(12), pp. 1776–1784. doi: 10.1172/JCI200320530.For.

Kassambara, A. (2020) 'ggpubr: "ggplot2" Based Publication Ready Plots'. Available at: https://cran.r-project.org/package=ggpubr.

Kleihues, P. and Ohgaki, H. (1999) 'Primary and secondary glioblastomas : From concept to clinical diagnosis 1', *Neuro-Oncology*, pp. 44–51.

Kolligs, F. T. *et al.* (1999) 'Neoplastic Transformation of RK3E by Mutant β-Catenin Requires Deregulation of Tcf/Lef Transcription but Not Activation of c-myc Expression', *Molecular and Cellular Biology*, 19(8), pp. 5696–5706. doi: 10.1128/mcb.19.8.5696.

Lamouille, S., Xu, J. and Derynck, R. (2014) 'Molecular mechanisms of epithelial-mesenchymal transition', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 15(3), pp. 178–196. doi: 10.1038/nrm3758.

Lee, J. *et al.* (2006) 'Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines', *Cancer Cell*, 9(5), pp. 391–403. doi: 10.1016/j.ccr.2006.03.030.

Lee, J. M. *et al.* (2006) 'The epithelial-mesenchymal transition: New insights in signaling, development, and disease', *Journal of Cell Biology*, 172(7), pp. 973–981. doi: 10.1083/jcb.200601018.

Lehmann, W. *et al.* (2016) 'ZEB1 turns into a transcriptional activator by interacting with YAP1 in aggressive cancer types', *Nature Communications*. Nature Publishing Group, 7, pp. 1–15. doi: 10.1038/ncomms10498.

Levitt, P., Eagleson, K. L. and Powell, E. M. (2004) 'Regulation of neocortical interneuron development and the implications for neurodevelopmental disorders', 27(7). doi: 10.1016/j.tins.2004.05.008.

Louis, D. N. *et al.* (2007) 'The 2007 WHO classification of tumours of the central

nervous system', *Acta Neuropathologica*, 114(2), pp. 97–109. doi: 10.1007/s00401-007-0243-4.

Lu, C. *et al.* (2012) 'IDH mutation impairs histone demethylation and results in a block to cell differentiation', *Nature*. Nature Publishing Group, 483(7390), pp. 474–478. doi: 10.1038/nature10860.

Malta, T. M. *et al.* (2018) 'Glioma CpG island methylator phenotype (G-CIMP): Biological and clinical implications', *Neuro-Oncology*, 20(5), pp. 608–620. doi: 10.1093/neuonc/nox183.

Mansouri, A., Karamchandani, J. and Das, S. (2017) *GLIOBLASTOMA*. 1st edn. Edited by S. De Vleeschouwer. Brisbane, Australia.

Margadant, C. and Sonnenberg, A. (2010) 'Integrin–TGF -β crosstalk in fibrosis, cancer and wound healing', 11(2), pp. 97–105. doi: 10.1038/embor.2009.276.

McBurney, M. W. (1993) 'P19 embryonal carcinoma cells', *International Journal of Developmental Biology*, 37(1), pp. 135–140. doi: 10.1387/ijdb.8507558.

Niehrs, C. (2012) 'The complex world of WNT receptor signalling', *Nature Publishing Group*. Nature Publishing Group, 13(12), pp. 767–779. doi: 10.1038/nrm3470.

Ohgaki, H. and Kleihues, P. (2005) 'Population-Based Studies on Incidence , Survival Rates , and Genetic Alterations in Astrocytic and Oligodendroglial Gliomas', *J Neuropathol Exp Neurol*, 64(6), pp. 479–489.

Ohgaki, H. and Kleihues, P. (2013) 'The definition of primary and secondary glioblastoma', *Clinical Cancer Research*, 19(4), pp. 764–772. doi: 10.1158/1078-0432.CCR-12-3002.

Omuro, A. and DeAngelis, L. M. (2013) 'Glioblastoma and other malignant gliomas: A clinical review', *JAMA - Journal of the American Medical Association*, 310(17), pp. 1842–1850. doi: 10.1001/jama.2013.280319.

Ostrom, Q. T. *et al.* (2014) 'The epidemiology of glioma in adults: A state of the science review', *Neuro-Oncology*, 16(7), pp. 896–913. doi: 10.1093/neuonc/nou087.

Ostrom, Q. T. *et al.* (2015) 'CBTRUS statistical Report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012', *Neuro-Oncology*, 17, pp. iv1–iv62. doi: 10.1093/neuonc/nov189.

Parsons, D. W. *et al.* (2008) 'An Integrated Genomic Analysis of Human Glioblastoma Multiforme', *Science*, 321(5897). doi: 10.1126/science.1164382.An.

64

Peinado, H., Olmeda, D. and Cano, A. (2007) 'Snail, ZEB and bHLH factors in tumour progression: An alliance against the epithelial phenotype?', *Nature Reviews Cancer*, 7(6), pp. 415–428. doi: 10.1038/nrc2131.

Phillips, H. S. *et al.* (2006) 'Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis', *Cancer Cell*, 9(3), pp. 157–173. doi: 10.1016/j.ccr.2006.02.019.

Postigo, A. A. (2003) 'Opposing functions of ZEB proteins in the regulation of the TGFβ/BMP signaling pathway', *EMBO Journal*, 22(10), pp. 2443–2452. doi: 10.1093/emboj/cdg225.

Postigo, A. A. *et al.* (2003) 'Regulation of Smad signaling through a differential recruitment of coactivators and corepressors by ZEB proteins', *EMBO Journal*, 22(10), pp. 2453–2462. doi: 10.1093/emboj/cdg226.

Qi, S. *et al.* (2012) 'ZEB2 mediates multiple pathways regulating cell proliferation, migration, invasion, and apoptosis in glioma', *PLoS ONE*, 7(6), pp. 1–12. doi: 10.1371/journal.pone.0038842.

Reardon, D. A. *et al.* (2012) 'Bevacizumab continuation beyond initial bevacizumab progression among recurrent glioblastoma patients', *British Journal of Cancer*. Nature Publishing Group, 107(9), pp. 1481–1487. doi: 10.1038/bjc.2012.415.

Rosmaninho, P. *et al.* (2018) 'Zeb1 potentiates genome-wide gene transcription with Lef1 to promote glioblastoma cell invasion', *The EMBO Journal*, 37(15), pp. 1–21. doi: 10.15252/embj.201797115.

RStudio Team (2020) 'RStudio: Integrated Development for R'. Boston, MA: RStudio, PBC. Available at: http://www.rstudio.com/.

Sánchez-Tilló, E. *et al.* (2015) 'ZEB1 and TCF4 reciprocally modulate their transcriptional activities to regulate Wnt target gene expression', *Oncogene*, 34(46), pp. 5760–5770. doi: 10.1038/onc.2015.352.

Santiago, L. *et al.* (2017) 'Wnt signaling pathway protein LEF1 in cancer, as a biomarker for prognosis and a target for treatment', *American Journal of Cancer Research*, 7(6), pp. 1389–1406.

Schmierer, B. and Hill, C. S. (2007) 'TGFβ-SMAD signal transduction: Molecular specificity and functional flexibility', *Nature Reviews Molecular Cell Biology*, 8(12), pp. 970–982. doi: 10.1038/nrm2297.

Seymour, T., Nowak, A. and Kakulas, F. (2015) 'Targeting aggressive cancer stem cells

in glioblastoma', *Frontiers in Oncology*, 5(JUL), pp. 1–9. doi: 10.3389/fonc.2015.00159.

Shortle, D. and Sondek, J. (1995) 'The emerging role of insertions and deletions in protein engineering', *Current Opinion in Biotechnology*, 6(4), pp. 387–393. doi: 10.1016/0958-1669(95)80067-0.

Siebzehnrubl, F. A. *et al.* (2013) 'The ZEB1 pathway links glioblastoma initiation, invasion and chemoresistance', *EMBO Molecular Medicine*, 5(8), pp. 1196–1212. doi: 10.1002/emmm.201302827.

Singh, D. K. *et al.* (2017) 'Oncogenes Activate an Autonomous Transcriptional Regulatory Circuit That Drives Glioblastoma', *Cell Reports*. ElsevierCompany., 18(4), pp. 961–976. doi: 10.1016/j.celrep.2016.12.064.

Singh, S. *et al.* (2016) 'Zeb1 controls neuron differentiation and germinal zone exit by a mesenchymal-epithelial-like transition', *eLife*, 5(MAY2016), pp. 1–31. doi: 10.7554/eLife.12717.

Turcan, S. *et al.* (2012) 'IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype', *Nature*. Nature Publishing Group, 483(7390), pp. 479–483. doi: 10.1038/nature10866.

Vandewalle, C., Van Roy, F. and Berx, G. (2009) 'The role of the ZEB family of transcription factors in development and disease', *Cellular and Molecular Life Sciences*, 66(5), pp. 773–787. doi: 10.1007/s00018-008-8465-8.

Verhaak, R. G. W. *et al.* (2010) 'Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1', *Cancer Cell*, 17(1), pp. 98–110. doi: 10.1016/j.ccr.2009.12.020.

Vermeulen, L. *et al.* (2010) 'Wnt activity defines colon cancer stem cells and is regulated by the microenvironment', *Nature Cell Biology*. Nature Publishing Group, 12(5), pp. 468–476. doi: 10.1038/ncb2048.

Verschueren, K. *et al.* (1999) 'SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes', *Journal of Biological Chemistry*, 274(29), pp. 20489–20498. doi: 10.1074/jbc.274.29.20489.

Vicovac, L. and Aplin, J. D. (1996) 'Epithelial-Mesenchymal Transition during Trophnoblast Differentiation', *Acta Anatomica*, 156, pp. 202–216.

Wang, H. *et al.* (2019) 'ZEB1 Represses Neural Differentiation and Cooperates with CTBP2 to Dynamically Regulate Cell Migration during Neocortex Development', *Cell Reports*. ElsevierCompany., 27(8), pp. 2335-2353.e6. doi: 10.1016/j.celrep.2019.04.081.

Watanabe, K. *et al.* (1996) 'Overexpression of the EGF receptor and p53 mutations are mutually exclusive in the evolution of primary and secondary glioblastomas', *Brain Pathology*, 6(3), pp. 217–223. doi: 10.1111/j.1750-3639.1996.tb00848.x.

Wesseling, P. and Capper, D. (2018) 'WHO 2016 Classification of gliomas', *Neuropathology and Applied Neurobiology*, 44(2), pp. 139–150. doi: 10.1111/nan.12432.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Available at: https://ggplot2.tidyverse.org.

Wu, Z. Q. *et al.* (2012) 'Canonical Wnt signaling regulates Slug activity and links epithelial-mesenchymal transition with epigenetic Breast Cancer 1, Early Onset (BRCA1) repression', *Proceedings of the National Academy of Sciences of the United States of America*, 109(41), pp. 16654–16659. doi: 10.1073/pnas.1205822109.

Xie, Q., Mittal, S. and Berens, M. E. (2014) 'Targeting adaptive glioblastoma: An overview of proliferation and invasion', *Neuro-Oncology*, 16(12), pp. 1575–1584. doi: 10.1093/neuonc/nou147.

Xu, W. *et al.* (2011) 'Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of a -Ketoglutarate-Dependent Dioxygenases', *Cancer Cell*. Elsevier Inc., 19(1), pp. 17–30. doi: 10.1016/j.ccr.2010.12.014.

Zeisberg, M. and Neilson, E. G. (2009) 'Biomarkers for epithelial-mesenchymal transitions', *The Journal of Clinical Investigation*, 119(6), pp. 1429–1437. doi: 10.1172/JCI36183.protected.

Zhang, C. *et al.* (2017) 'TGF- β 2 initiates autophagy via Smad and non-Smad pathway to promote glioma cells ' invasion'. Journal of Experimental & Clinical Cancer Research, pp. 1–15. doi: 10.1186/s13046-017-0628-8.

Zhang, J., Tian, X.-J. and Xing, J. (2016) 'Signal Transduction Pathways of EMT Induced by TGF-β, SHH, and WNT and Their Crosstalks', *Journal of Clinical Medicine*, 5(4), p. 41. doi: 10.3390/jcm5040041.

Zhang, L. *et al.* (2016) 'SHP-2-upregulated ZEB1 is important for PDGFRα-driven glioma epithelial-mesenchymal transition and invasion in mice and humans', *Oncogene*, 35(43), pp. 5641–5652. doi: 10.1038/onc.2016.100.

Zhou, B. P. *et al.* (2004) 'Dual regulation of Snail by GSK-3 β -mediated phosphorylation in control of epithelial – mesenchymal transition', 6(10). doi: 10.1038/ncb1173.