Deciphering gene regulatory circuitry governing cell fate changes

A dissertation submitted to obtain the academic degree

"Doctor of Natural Sciences"

at the Faculty of Biology of the

Johannes Gutenberg University, Mainz

by

Sanjeeb Kumar Sahu

born on 27th May 1987

Mainz, 2017

Abstract

During development, cellular differentiation is tightly controlled via integrating several regulatory layers such as signaling pathways as well as transcriptional and epigenetic mechanisms to guarantee precise spatio-temporal gene expression programs. In order to understand the gene regulatory circuitry governing cell fate changes, in this thesis I employed two established model systems that display defined cascades of phenotypic remodeling, i.e. Neurogenesis and Epithelial to Mesenchymal Transition (EMT). During brain development, the process of neurogenesis comprises a highly defined set of cell fate decisions that involves the transition of proliferative and multipotent neuroepithelial cells towards terminally differentiated post-mitotic neurons. A combinatorial analysis of genome-wide datasets during this process profiling the transcriptome, chromatin accessibility and the epigenome profiles, including H3K27ac, the mark for active enhancers and promoters, reveals the importance and highly dynamic nature of distal gene regulation during neurogenesis. We further show that terminally differentiated neurons also undergo remodeling of the distal regulatory landscape to ensure proper transcriptional output upon exposure to stimuli that induce neuronal activity. Interestingly, further such epigenetic and transcriptional response during neuronal activation conferring a transient loss of neuronal identity, gain of cellular plasticity and induction of pro-survival genes. Within another project we functionally characterized the radial glia cell specific transcription factor Tox3, and unravel its essential function during development of neocortex as it directly binds to the promoter of *Nestin* thereby ensuring its timely induction during embryonic neurogenesis.

The differentiation of static epithelial cells into motile mesenchymal cells, a process known as EMT, is integral in development and wound healing and it contributes pathologically to fibrosis and cancer progression. While studying this process, we uncovered a kinetically distinct role of the JNK signaling, which is not required for initiation, but progression of EMT. Furthermore, we identified FBXO32 as a key regulator of EMT that functions via ubiquitination of transcriptional corepressor protein CTBP1 to modulate its cellular localization. Such tight regulation of CtBP1 levels during EMT is required as epigenetic remodeling and transcriptional induction of CTBP1 target genes create a suitable microenvironment for EMT progression. We further identified the C2H2-zinc finger containing protein ZNF827, to be a critical player during EMT that modulates alternative splicing of numerous genes during EMT. ZNF827 mediates such stage specific transcript diversity by directly targeting these genomic loci, modulating their epigenetic landscape to alter RNA Pol II kinetics and also facilitates the recruitment of core splicing components to nascent RNA. Overall, this thesis delineates diverse gene regulatory mechanisms which cells utilize to orchestrate cell fate changes during development and disease.

Deutsche Zusammenfassung

Zelluläre Spezifizierungen während der Embryonalentwicklung bedürfen engmaschiger Kontrolle in welcher eine Vielzahl regulatorischer Ebenen integriert werden müssen. Eine derartige Kontrolle involviert eine spezifische Integration von äußeren Einflüssen auf die jeweilige Zelle im Kontext des sich entwickelnden Organismus auf der Ebene von Signalwegen sowie transkriptioneller und epigenetischer Regulation, um einen präzises raum-zeitliches Expressionsprogram zu gewährleisten. Zur Erlangung eines umfassenderen Verständnisses der genregulatorischen Netzwerke, die derartige Zellschicksalspezifikationen gewährleisten, habe ich mich in dieser Arbeit zwei etablierten Modelsysteme bedient, die jeweils definierte Kaskaden phänotypischer Spezifikationen durchlaufen: Neurogenes und Epitheliale Mesenchymale Transistion (EMT). Während der Gehirnentwicklung durchlaufen Zellen in der Neurogenese eine hochspezifische, fein-regulierte Abfolge von Zellschicksalsentscheidungen die den Übergang von proliferierenden und multipotenten neuroepithelialen Zellen zu terminal differenzierten postmitotischen Neuronen beinhaltet. Durch eine kombinatorische Analyse genomweiter Datensätze während dieses Prozesses, die die Profilierung des Transkriptoms, der Genom-Zugänglichkeit und epigenetischer Profile (H3K27ac als Markierung für aktive Enhancer und Promotoren) beinhaltete, konnten wir die Wichtigkeit und hohe Dynamik distaler, regulatorischer Elemente während der Neurogenese belegen. Weiterhin zeigen wir das auch terminal differenzierte Nervenzellen noch eine Modulation ihrer regulatorischen Regionen aufweisen, um beispielsweise eine präzise transkriptionelle Antwort auf Stimulation der neuronalen Aktivität zu gewährleisten. Interessanterweise vermittelt eine derartige epigenetische und transkriptionelle Antwort während neuronaler Aktivierung einen transienten Verlust der neuronalen Identität, eine erhöhte zelluläre Plastizität sowie Induktion von überlebens-fördernden Genen. In einem weiteren Projekt charakterisierten wir den für radiale Gliazellen spezifischen Trankriptionsfaktor Tox3 und belegten eine essentielle Funktion dieses Faktors während der Entwicklung des Neocortex, da dieser Faktor direkt an den Promoter von Nestin bindet und daher dessen zeitgenaue Induktion während der embryonalen Neurogenese gewährleistet.

Die Differenzierung von epithelialen Zellen zu beweglichen mesenchymalen Zellen involviert einen Prozess, der als EMT bekannt ist und während der embryonalen Entwicklung und der Wundheilung abläuft sowie zu pathologischen Erscheinungen wie Fibrose und Krebsmetastasierung beiträgt. Während der Untersuchung dieses Prozesses haben wir die Kinetik des JNK Signalweges detailliert aufschlüsseln können, welcher nicht essentiell ist für die Initiierung aber das Fortschreiten von EMT. Außerdem haben wir FBXO32 als entscheidenden Regulator von EMT identifiziert, ein Faktor, der den transkriptionellen Ko-Repressor CTBP1 ubiquitiniert und dadurch seine zelluläre Lokalisierung verändert. Eine derart gezielte

Regulation der CtBP1 Leveln ist notwendig für EMT, da epigenetische Modulationen und die transkriptionelle Induktion von CTBP1 Zielgenen eine Mikroumgebung etablieren, die für das Fortschreiten des EMT förderlich ist. Weiterhin haben wir das C2H2-Zinkfinger Protein ZNF827 untersucht und herausgefunden, dass es eine essentielle Rolle als Regulator des alternativen Spleißens zahlreicher Gene während EMT einnimmt. ZNF827 vermittelt derartige zelltypspezifische Transkriptdiversität über direkte Bindung an die jeweiligen genomischen Loci und Modulation ihrer epigenetischen Umgebung, um die RNA Pol II Kinetiken entsprechen anzupassen und die Rekrutierung von Spleiß-Komponenten zu erleichtern. Insgesamt legt diese Arbeit diverse genregulatorische Mechanismen dar, welche von Zellen herangezogen werden um Zellschicksalsveränderungen während der embryonalen Entwicklung sowie progredienten Krankheitsverläufen zu koordinieren.

Table of Contents

1. General Introduction

The cell is a complex network of interlinked genetic and biochemical processes and needs to ensure hereditary of molecules and efficient coordination of many cellular processes that occur inside a membranous capsule [1]. Prokaryotic cells lack a clear physical separation between the genetic material and the cytoplasm. While many prokaryotes possess internal sub-organization that may even include membranous structures, eukaryotic cells have optimized such structures and display clear distinction within multiple organelles such as nucleus, ribosomes, endoplasmic reticulum, Golgi apparatus, vacuoles, lysosomes, mitochondria, and additionally chloroplasts in plants [2]. The complicated internal membrane compartments of modern eukaryotes are able to respond to changing environmental conditions and new requirements of the cell in a highly coordinated but adaptable fashion to allow functional flexibility [3-5]. The transition from prokaryotic to eukaryotic cells occurred over 1.5 billion years ago and is considered one of the most important alterations to cellular structure during all of evolution [6]. Such advancement allowed subsequent evolution of complex, multicellular organisms such as humans that consist of more than 200 different cell types [7]. The organizational subunits of cells make up tissues that are able to form tissues that organize into organs and subsequently build organ systems whose integrated activity is often coordinated by the endocrine and nervous systems [4,8,9]. The tightly regulated and complex interconnection of such organ systems that perform different functions allow humans and other complex multicellular organisms to carry out complex processes to survive and evolve.

1.1. Genetic material

Even though in multicellular organisms the cell types and their functions are highly diverse, each cell in one individual contains nearly the same genetic information (genome). This information is stored in the sequence of the DNA, which represents the heritable material encoding the composition and architectural plan of an organism [10-12]. However in the exceptional case of RNA viruses, RNA instead of DNA acts heritable genetic material. DNA is composed of four nucleotides containing the nucleobases

adenine (A), cytosine (C), guanine (G) and thymine (T), which together with deoxyribose sugar and a phosphate group build the functional polymer [13]. This chain forms a double helix structure by twisting around with an inverted, complementary strand of nucleotides, whereby the nucleobases project towards inside and form hydrogen bonds (A with T, C with G) [13]. The structural arrangement into two strands of DNA enables the replication of DNA in a semiconservative manner for duplication of the genome before cell division or act as a template for transcription into functional RNAs [13,14].

The central dogma of molecular biology aims to explain the flow of genetic information within a biological system. It was first stated by Francis Crick in 1958. The classic view states that "the coded genetic information hard-wired into DNA is transcribed into individual transportable cassettes, composed of messenger Ribonucleic acid (mRNA); each mRNA cassette contains the program for synthesis of a particular protein. Our genome consists of hundreds of genes referred to as hereditary information that is usually encoded on a defined piece of DNA. A functional gene contains the basic information for the production of a biologically active RNA by a process called as transcription, where a complimentary copy of the coding strand of DNA is produced in form of RNA. An essential function of the RNA is the conversion of genetic information into proteins (translation) via mRNA that acts as an information carrier. RNA also plays a critical role in protein biosynthesis; snRNA and snoRNA, the catalytic components of the ribosomes, are involved in the maturation of mRNA, rRNA, and tRNA transports the building blocks for the proteins. During translation, a partial process of protein biosynthesis, the amino acid sequence of proteins is assembled from the information stored in the mRNA molecule. Proteins are made up of hundreds or thousands of smaller subunits the so called amino acids, which are attached to one another in long chains. There are 20 different types of amino acids that can be combined in huge combinatorial variety to build functional proteins. The sequence of amino acids determines each protein's unique 3-dimensional structure and its specific function. Beside proteins, noncoding RNAs (ncRNAs) play a very

critical role during various biological process including gene regulation [15]. The activity state of each gene or its expression, can be regulated differently in individual cells and is essential for the phenotypic output.

1.2. Epigenetics: Overview and Concepts

Genetics conceptually, describes genes and gene function that is inherited to the next generation via the offspring. While Epigenetics encompasses heritable changes in gene expression (active vs. inactive genes) that does not involve changes in the underlying DNA sequence [16]. DNA methylation, posttranslational modifications (PTMs) on histone tails, chromatin remodeling, histone variants and noncoding RNAs are epigenetic modifications because they do not change the underlining DNA sequences but can be transmitted by both sister strands from mother to daughter cells during DNA replication and, thus, can be inherited through cell division [17-19] (Figure 1).

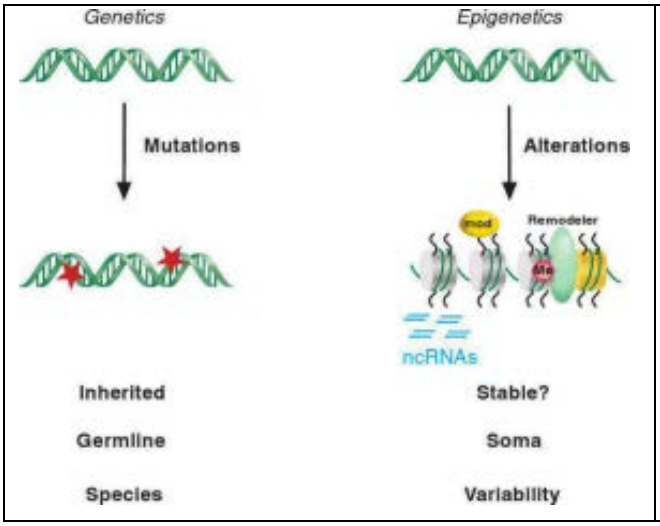


Figure 1. Genetics versus epigenetics. Genetics: Mutations (red stars) of the DNA template (green helix) are heritable as well somatically as through the germline. Epigenetics: Variations in chromatin structure that modulate the usage of the genome. Histone variant composition: yellow nucleosomes, DNA methylation: Me. (Adapted from Epigenetics, 2nd Edition, by Allis)

Since its discovery around 35 years ago, DNA methylation at cytosine residues is recognized as one of the prime epigenetic mechanisms to correlate with gene repression [20-22]. This modification converts cytosines by the addition of a methyl group into 5-methylcytosine (5mC) in the DNA template. Such modifications mainly occur at CpG dinucleotides in mammals but are observed with less frequency in non-CpG context especially in the neuronal system likely to increase neuronal diversity. The distribution of DNA

methylation within the genome shows enrichment at noncoding regions such as centromeric heterochromatin and interspersed repetitive elements. DNA methyltransferases (DNMTs) confer DNA methylation and catalyze either *de novo* methylation (DNMT3A, DNMT3B) or ensure the maintenance (DNMT1) of hemimethylated DNA following replication [22,23]. In order to demethylate the DNA template 5mC can be oxidized by the TET (ten eleven translocation) enzymes to generate 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) [22]. The combined action of thymine-DNA glycosylase (TDG) and the base excision repair (BER) machinery on 5caC sites finally results in unmodified cytosine. There are several readers for 5mC (e.g., MeCP2, MBDs), but a specific reader for 5hmC remains unknown [24].

To fit the very large genome (for e.g. human containing around 3.2×10^9 base pairs (bp) in 23 chromosome pairs) inside the nucleus of a cell that consist of 5 µm length, the DNA is compacted with the help of specialized highly basic proteins called histones [25]. A wedge-shaped histone octamer consists of two copies of the four core histones: H2A, H2B, H3 and H4 surrounded by approximately 1.7 turns of 146 bp DNA form the core unit called as nucleosome, that are organized in the so-called "beads-on-a-string" structure [26-29] (Figure 2A-B). The length of the linker DNA between the nucleosomes can vary around 10-100 bp [28,30].

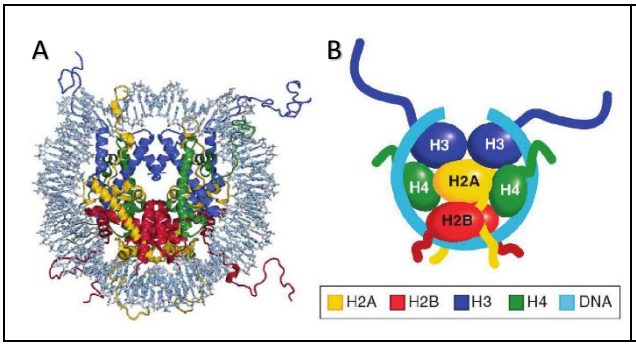


Figure 2. Nucleosome structure. (A) The 2.8-Å structure of a nucleosome. **(B)** Schematic representation of histone organization within the core octamer around which the DNA (light blue) is wrapped. Unstructured amino-terminal histone tails extrude from the nucleosome core with globular domains of the eight-histones. (adapted from [27])

Based on the amino acid sequence, histone proteins are highly conserved from yeast to humans, supporting the general view that these proteins, likely serve critical functions. The amino acid residues in the tails and some in the more structured globular core domains hold important clues to nucleosomal

and, hence, chromatin variability, as they are subject to extensive post-translational modifications (PTMs). Acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP-ribosylation, biotinylation, crotonylation and proline isomerization of core histones are among the most common PTMs [31] (Figure 2A). Modifications on histones are reversible and epigenetic enzymes and regulatory proteins confer their modification (writer), recognition it (reader), or removal (erasers) and were long proposed to correlate with positive and negative changes in transcriptional activity (i.e. DNA accessibility or inaccessibility to the transcription machinery) [31-38] (Figure 2B).

Among those histone acetyltransferase (HATs) acetylate specific lysine residues in histone substrates and are consequently erased by the action of histone deacetylases (HDACs) [32,33]. The histone kinase family of enzymes phosphorylates specific serine, threonine, or tyrosine residues while specific phosphatases remove such phosphorylation marks. Two general classes of methylating enzymes have been described: the histone lysine methyltransferases (KMTs) that act on lysines in mono-, di-, or tri-methylated states, and the protein arginine methyltransferases which modify arginines [34,35]. Arginine methylation is indirectly reversed by the action of protein arginine deiminases, which convert methyl-arginine into a citrulline residue [36]. Lysine methylation reverting enzymes, comprise two family, one with the amino oxidases represented by lysine-specific histone demethylase (KDM1A and KDM1B) that use FAD and oxygen as cofactors for demethylation and exclusively target mono- and dimethylation of H3K4 and H3K9 [37,38]. While the second family has the JMJC domain, that mediates demethylation using 2-oxoglutarate and iron as cofactors [39].

1.3. Chromatin mediated gene regulation

The accurate and robust regulation of gene expression is key during cell fate changes that accompany biological processes such as organismal development. Such precise regulation is controlled by various histone modifications which do not function in isolation and often require the presence or absence of other modifications and their potential information is transmit through their recognition by other factors.

Specific epigenetic marks on histone tails and DNA modification can correlate with patterns of chromatin structure that may provide active or inactive signatures [40]. Figure 3 illustrates two examples of established hierarchies of histone modifications that seem distinguish active and transcribed chromatin regions from heterochromatic domains that display a repressive transcriptional state. Such diverse combinations of "active" marks can function in a concerted fashion to simultaneously counteract "repressive" modifications and vice versa.

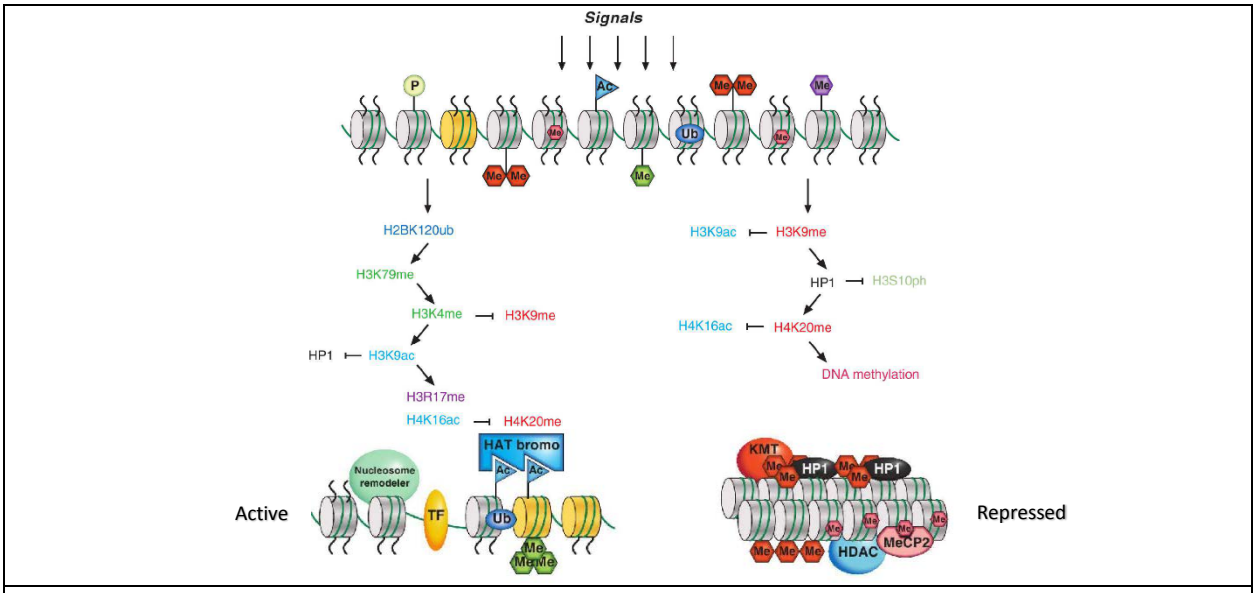


Figure 3. Chromatin mediated gene regulation. The transition of a naïve chromatin template to active euchromatin (left) or the establishment of repressive heterochromatin (right) involves a series of coordinated chromatin modifications. histone variants (yellow, namely, H3.3), histone acetylation (Ac), methylation (H3K4me2/3:green Me, K27me3: red Me, DNA methylation: Pink and small red Me) (This image was adapted from Epigenetics, 2nd Edition, by Allis)

Furthermore a class of *cis*-regulatory elements so called enhancers, canonically defined as short (100–1,000 bp) noncoding DNA sequences, play a central role in orchestrating spatiotemporally precise gene-expression programs during development and act independent of their relative distance, location, or orientation to their target promoter [41] (Figure 4). While enhancers share many features similar with promoters [40] they are able to activate transcription over long genomic distances what makes them unique. Once enhancers are brought into the spatial proximity of the promoter sequence by the formation of DNA loop with the help other regulatory factors they convey transcriptional enhancement of the associated

gene. Enhancers have a high ratio of H3K4me1 to H3K4me3, while the reverse is true at promoters. Both active enhancers and promoters are characterized by DNA hypomethylation, with the methylation status of enhancers being more dynamic and cell-type specific. This feature allows a gene to be regulated by multiple distal enhancers with different spatiotemporal activities, facilitating enormous combinatorial complexity of gene expression repertoires using a relatively limited set of genes. In addition to their central role in developmental gene regulation, enhancers are fertile targets for evolutionary change, as they are both cell-type specific and commonly exist in groups of redundant elements (facilitating accumulation of genetic variation by buffering the risk) [42,43].

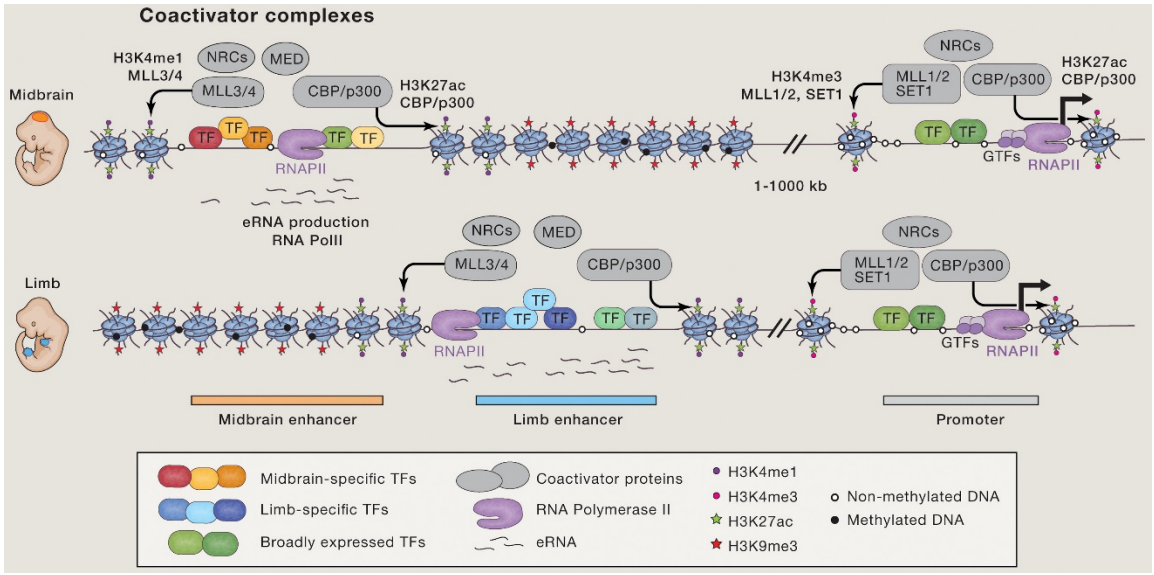


Figure 4. Chromatin mediated gene regulation. Presence of cell-type-specific and broadly expressed TFs, RNA Polymerase II (RNAPII) and associated enhancer RNAs (eRNAs), coactivator complexes such as Mediator (Med), nucleosome remodeling complexes (NRCs), histone acetyltransferases (CBP/p300), and methyltransferases (MLL3/4) at tissue-specific enhancer elements is schematically depicted for a midbrain (top) or limb (bottom) specific enhancer. (Adapted from [43])

1.4. Higher-order chromatin organization

Chromatin, the DNA-nucleosome polymer, is a dynamic molecule existing in many configurations. Chromatin has been classified as either euchromatic (a decondensed version with histone hyper-acetylation and mostly transcriptionally active) or heterochromatic (highly compacted and silenced). Heterochromatin exists either as permanently silent constitutive heterochromatin (pericentric and

subtelomeric regions, represented by the enrichment of H3K9me3 mark), or as facultative heterochromatin, in which genes can be derepressed during specific cell cycle or developmental stages. The simplest chromatin structure is a roughly 11-nm polymer that represents a largely unfolded configuration of chromatin wherein DNA is periodically wrapped around repeating units of nucleosomes (Figure 5). A repressive higher-order chromatin structure (30 nm) is achieved partially through the recruitment of linker histone H1 [44]. Nucleosomes may be irregularly packed and folded into versatile higher-order conformations (300–700-nm to most condensed 1.5-µm) that occur along the length of the genome, changing more subtle during interphase and dramatically during mitotic chromatin stages of the cell cycle [45,46] (Figure 5).

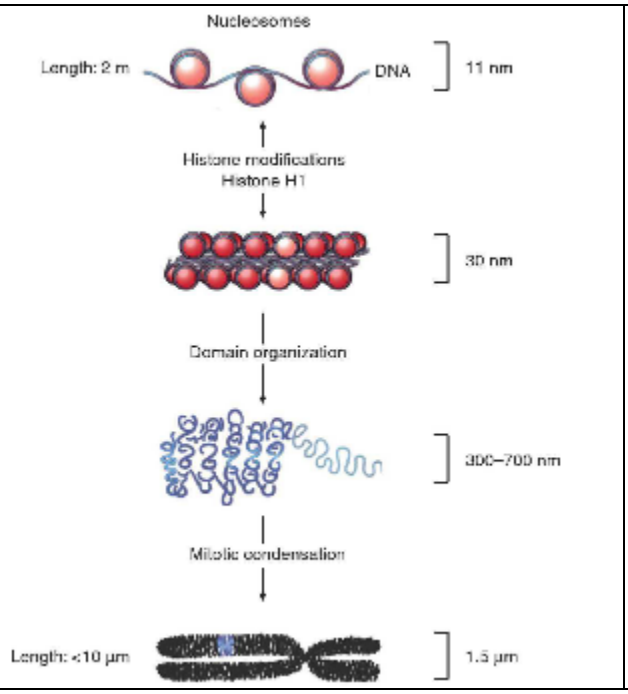


Figure 5. Higher-order chromatin structure. Model of chromatin condensation needed to achieve the packaging observed in metaphase chromosomes. (This image was adapted from Epigenetics, 2nd Edition, by Allis)

Many genes and enhancers originally discovered in multicellular organism have now been detected in single-celled outgroups suggesting that the evolution of gene regulation coincides with gene multiplications [9,47-49]. Cells must continually adapt to changing conditions by altering their gene expression patterns. One of the central effectors involves transcriptional regulatory interactions. The recent development of high-throughput methods and computational approaches allowes to survey these

complex molecular interactions by modeling them as networks [50-52]. In order to understand the gene regulatory circuitry governing cell fate changes, we employed two established model systems that display defined cascades of phenotypic remodeling - Neurogenesis and Epithelial to Mesenchymal Transition (EMT). The development of the mammalian brain is one of the most complex developmental processes which requires the precise regulation of several gene networks at different time points. On the other hand the differentiation of static epithelial cells into motile mesenchymal cells, a process known as EMT, is integral in development and wound healing and it contributes pathologically to fibrosis and cancer progression.

Neurogenesis

2.1.1. Neurogenesis

Neurogenesis is the process by which the various cell types of nervous system are produced from pluripotent stem cells. The brain is the most complex organ in our body and considered as the core component of human identity. Furthermore with regard to the diversity of cell types, the cyto-architecture and neural circuitry the complexity is greatest in the neocortex, the area of higher cognitive functions. Although certain fundamental aspects of neocortex development and structure are conserved among the various mammalian species, there are enormous differences with regard to morphology, folding and size of the neocortex [9,53-55]. Cells of the nervous system, neurons, are the most diverse type of specialized cells in animals and also morphologically most distinct [43]. It is fascinating, how cellular differentiation during nervous system development is coordinated, by a series of transitions starting from embryonic stem cells of the inner cell mass (Figure 6). We call those processes, in which the embryonic stem cells differentiate into mature neurons as "Neurogenesis". This process is mainly occurring during embryonic development, but it partially persists in adult stages in restricted stem cell niches such as in the dentate gyrus in the mammalian brain. New neurons born in the dentate gyrus integrate into the existing network and have been suggested to play a role in memory processing. Neurons do not exist in isolation but are connected via synapse structures which are formed during development but continuously changing with

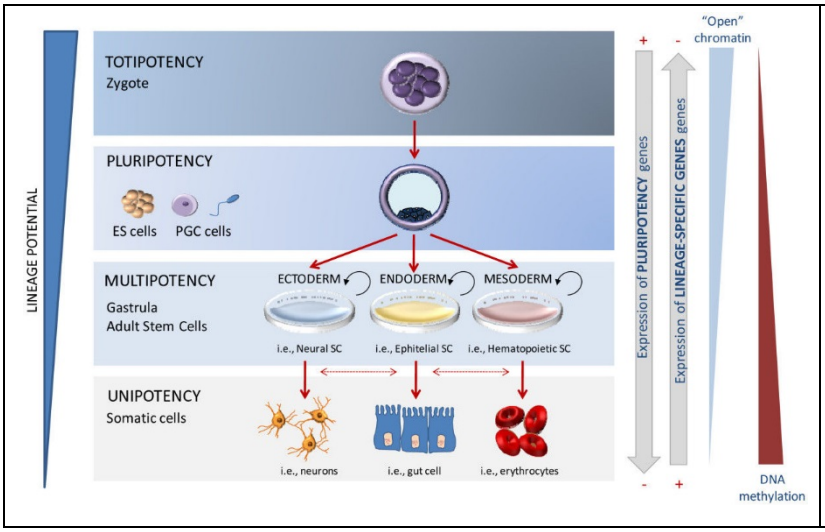


Figure 6. Lineage restriction of human developmental potency. Totipotent cells at the morula stage have the ability to self-renew and differentiate into all of the cell types of an organism, including extraembryonic tissues. (The image adopted from [56])

plasticity: they are exquisite communication devices, and their activity depends on the dynamic interplay between their intrinsic properties and inputs from other neurons. The activity of neural circuits interfaces animals with their environments and produces behavior.

2.1.2. Signaling pathways and transcriptional networks during early neurogenesis

Advancement in stem cell research allowed the development of an *in vitro* model system that recapitulates *in vivo* differentiation. Here e.g. "neural stem cells" (NSCs), which are a self-renewing cells can differentiated into neurons, oligodendrocytes and astrocytes by stimulation of a number of signaling pathways, such as Sonic Hedgehog, BMP, Wnt/β-Catenin, Notch, and FGF [57-64]. During nervous system development these cells reside in the ventricular zone of embryonic neural tube or in the subventricular zone in the developing cerebral cortex [57-63] (Figure 7A). In the embryo as well as in NSC cultures, the self-renewal ability of these cells is regulated by numerous transcription factors, including Hes, Sox, Bmi1, Tlx, and Gli proteins [65-68] (Figure 7B).

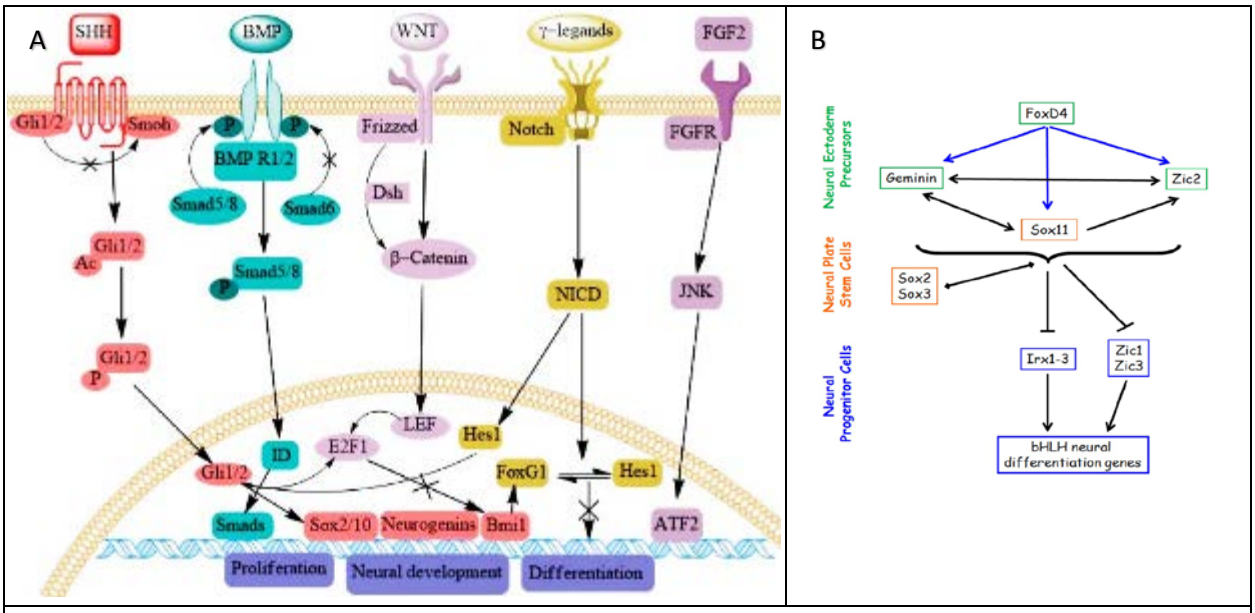


Figure 7. A schematic representation of signaling pathways and transcription factor involved in neural stem cell maintenance and differentiation. (A) Five major signaling pathways including SHH, BMP, WNT, Notch ligands and FGF2, together crosstalk during neural stem cell maintenance differentiation. **(B)** A gene regulatory network of neural transcription factors that regulate the earliest steps in vertebrate neural development. (All images were adopted from [57])

2.1.3. Progenitor specification and diversity during corticogenesis

Early in development, the telencephalic wall is composed of undifferentiated neuroepithelial cells. As these progenitors proliferate and expand in number, some begin to differentiate into radial glia, establishing the ventricular zone (VZ). Radial glia, in turn, give rise to additional progenitor classes, including outer radial glia and intermediate progenitors, which together form the subventricular zone (SVZ). Those progenitor cells further differentiate into neurons and migrate towards cortical plate (CP). Each of these progenitor populations has distinct morphological properties and follows a specific pattern of cell divisions. Radial glia span the entire thickness of the developing cortex, from the ventricular (apical) to the pial (basal) surface, and provide a scaffold along which newly born neurons migrate towards upper cortical layers. Such progenitors primarily divide asymmetrically to self-renew and at the same time also give rise to outer radial glia, intermediate progenitors or neurons. Outer radial glial cells are also unipolar but can be distinguished from radial glia by their lack of an apical process. They were first characterized in the outer SVZ of the developing human cortex and, until recently, were thought to be present only in gyrencephalic animals. However, a small population also exists in the SVZ of rodents, undergoing asymmetrical divisions to self-renew and generate neurons. Intermediate progenitors have a multipolar morphology and, unlike radial glia, are not anchored to either the apical or basal surface. They act primarily as transit-amplifying cells, undergoing limited proliferative divisions, and more often divide symmetrically to produce two neurons. A fourth class of progenitors, the short neural precursors, reside in the VZ but have a basal process that does not reach the pial surface and they might resemble radial glia that are in the process of becoming intermediate progenitors [69] (Figure 8A). Neocortical progenitors begin to produce excitatory projection neurons around embryonic day 10.5 (E10.5) in mice. The earliest-born neurons migrate away from the ventricular surface to segregate from progenitors and form the preplate. Later-born neurons migrate into the preplate, splitting it into the marginal zone and subplate, and establishing the cortical plate between these two regions. Throughout the rest of corticogenesis, newly

born neurons migrate into the cortical plate, organizing themselves in an 'inside-out' fashion, such that early-born neurons populate deeper neocortical layers (layer VI, then layer V), and late-born neurons migrate on top of earlier born neurons to progressively populate more superficial layers (layer IV, then layer II/III) (Figure 8A-B).

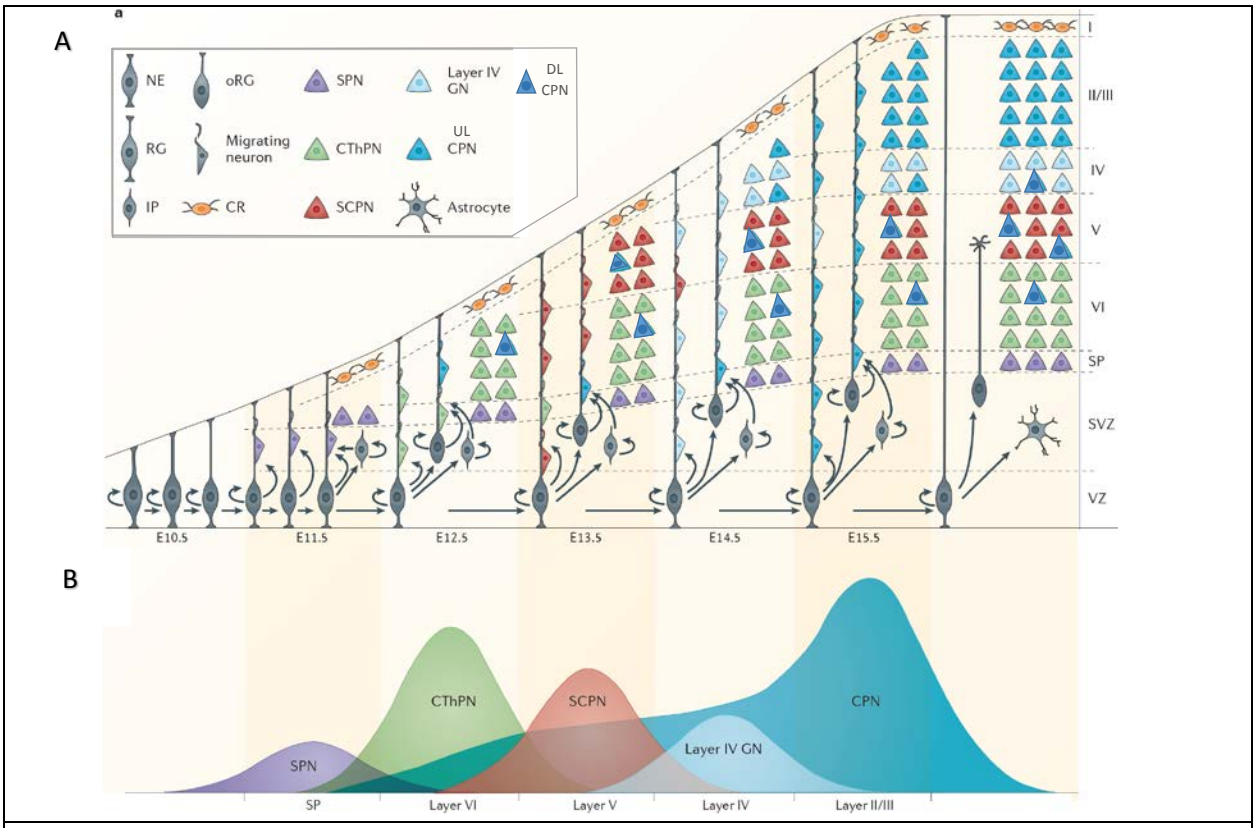


Figure 8. Embryonic cortical neurogenesis. (A) Neocortical projection neurons are generated in an 'inside-out' fashion by diverse progenitor types in the VZ and SVZ. This schematic depicts the sequential generation of neocortical projection neuron subtypes and their migration to appropriate layers over the course of mouse embryonic development. **(B)** Distinct projection neuron subtypes are born in sequential waves over the course of neurogenesis. (Adopted from [69])

Different layers of the cortical plate are marked by a distinct set of neurons and this positional information of neurons is already established by regulatory mechanisms active in neuronal progenitors. This information established in progenitors must be transmitted to neurons in a layer specific manner a process executed by a specific set of transcription factors regulating the neuronal identity and function. In the past decade a number of studies have revealed transcription factors involved in post-mitotic control

and sub-specification of neurons. Newly formed post-mitotic neurons usually express a number of transcription factors that get restricted to specific sub-type population with time. Ctip2 and Tbr1/Fog2 are co-expressed initially between E12.5 and E14.5 but later are expressed in distinct populations of sub-cerebral projection neurons (SCPN) and corticothalamic projection neurons (CThPN). The transcription factor Fezf2 for example is crucial for sub-type specification of SCPN and CThPN. Fezf2 is expressed at very high levels by SCPN while in CThPN it is expressed at very low levels and this differential expression results in formation of SCPN [70] (Figure 9).

In the absence of Ctip2, layer V neurons show drastic defects in axon formation and it is proposed that Ctip2 acts downstream of Fezf2 as in absence of Ctip2 SCPN can properly migrate to layer V, while Fezf2 is required to positioning of neurons to layer V. Although Fezf2 mediates up-regulation of Ctip2 is crucial of SCPN formation, several other transcription factors, such as Satb2, Sox5 or Couptf1, act via repression of Ctip2 to regulate development of several sub-types of cortical neurons. Tbr1, another neuronal transcription factor, acts as an antagonist to Fezf2/Ctip2 and Tbr1 expressing neurons project towards the thalamus. Previous studies have shown that Tbr1 acts via repression of Fezf2 and consequently Ctip2, by directly binding to regulatory regions controlling the expression of Fezf2. Sox5 binds to an enhancer element of Fezf2 and represses its expression in the forebrain and as a result neurons lacking Sox5 express very high levels of Ctip2. Such reciprocal inhibition of transcription factors specifying different sub-type populations have indicated that neuronal diversity is created by inhibition of alternate differentiation programs to form a particular sub-type rather than active enhancement of fate decisions (Figure 9). Generation of new neurons persists in the dentate gyrus in the mammalian brain in the adult life. New neurons born in the dentate gyrus integrate into the existing network and have been suggested to play a role in memory processing[70].

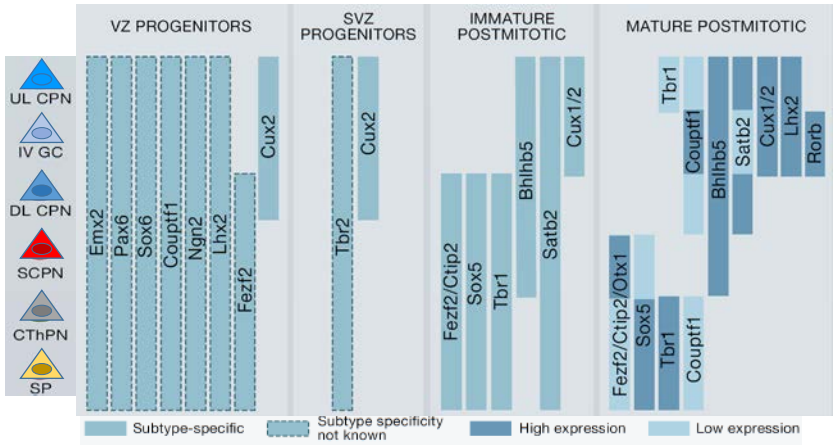


Figure 9. Transcriptional control of corticogenesis. (A) Combinatorial transcription factor programs progressively delineate projection neuron identity. (Image was adopted from [70].)

2.1.4. Neuronal activity

Each neuron consists of a cell body, an axon and dendrites. The axon looks like a long slender thread without branches while dendrites appear like the branches of a tree. Each neuron communicates within the circuitry via synapses that permit a neuron to pass an electrical signal like action potential or chemical signal of neurotransmitters to another neuron. It is widely accepted that the synapse plays a key role in the formation of memory. As neurotransmitters activate receptors across the synaptic cleft, this connection between two neurons is strengthened when both neurons are active at the same time. The strength of two connected neural pathways is thought to result in the storage of information, resulting in memory. The neuronal signaling via N-methyl-D-aspartate (NMDA) [71] receptors plays a critical role in the development of the central nervous system (CNS) and in adult neuroplasticity, learning, and memory that require associated transcriptional changes to mediate physiological responses [72,73]. The NMDA receptor has been extensively characterized via electrophysiological assays, genetic and chemical manipulations. NMDA receptors require the binding of glycine and glutamate in combination with the release of voltage-dependent magnesium blockage (Figure 10A). In addition to its role during CNS development, neuronal activity via NMDA receptors also imparts neuronal plasticity, memory formation and learning[74-78]. Evidences suggest that NMDA receptor activation leads to strengthening of synapses through long-term

potentiation [79] and to the weakening of synapses through long-term depression (LTD) [74,75,80-82]. Recent findings have revealed early transcriptome responses following neuronal activity including via NMDA receptors that further underlie many aspects of neuronal development[74,75,83-88] (Figure 10B-C). Malfunction of NMDA receptors have also been implicated in the etiology of several neurological disorders[89-92]. Excessive stimulation of the NMDA receptor results in disorders involving acute insult to the brain with deprivation of blood supply (e.g., ischemic stroke, traumatic brain injury)[93,94]. Moreover, neurotoxicity induced via hyper-stimulation of the NMDA receptor also facilitates slow-progressing neurodegenerative diseases (e.g., Huntington, Alzheimer's, Parkinson's and amyotrophic lateral sclerosis) [95-102] and disorders arising from the sensitization of neurons (e.g., epilepsy, neuropathic pain)[103,104].

In response to environmental stimuli, synaptic activity and neurotransmitters, NMDA receptors (NMDAR) open, allowing the influx of Ca2+ across the plasma membrane. Calmodulin (CaM) transduces the local elevation of calcium into changes in the neuronal gene expression pattern through the activation of CREB in the nucleus via the Ras–MAPK and CAMK kinase pathways. CREB transcriptional activity is positively regulated by its phosphorylation at serine-133, which allows the recruitment of the transcriptional coactivator CBP allowing the activation of immediate early genes, such as c-fos. Simultaneously, calcineurin (CaN) leads to TORC and NFAT dephosphorylation allowing nuclear translocation and regulation of their target genes. Together, these changes in gene expression affect many aspects of nervous system development including dendritic morphogenesis, neuronal survival, and synapse development as well as the adaptive responses that underlie learning and memory in the mature nervous system[105](Figure 10D).

Among those many aspects of nervous system development mediated by NMDAR activity-dependent modulation of synapses is critical for brain development and cognitive functions in the adult [105]. Learning and memory are basic functions of the brain essential for human evolution. It is well accepted that during learning and memory formation the dynamic establishment of new active synaptic connections is crucial.

Persistent synaptic activation leads to molecular events that include elevated release of neurotransmitters along with increased expression of receptors on the

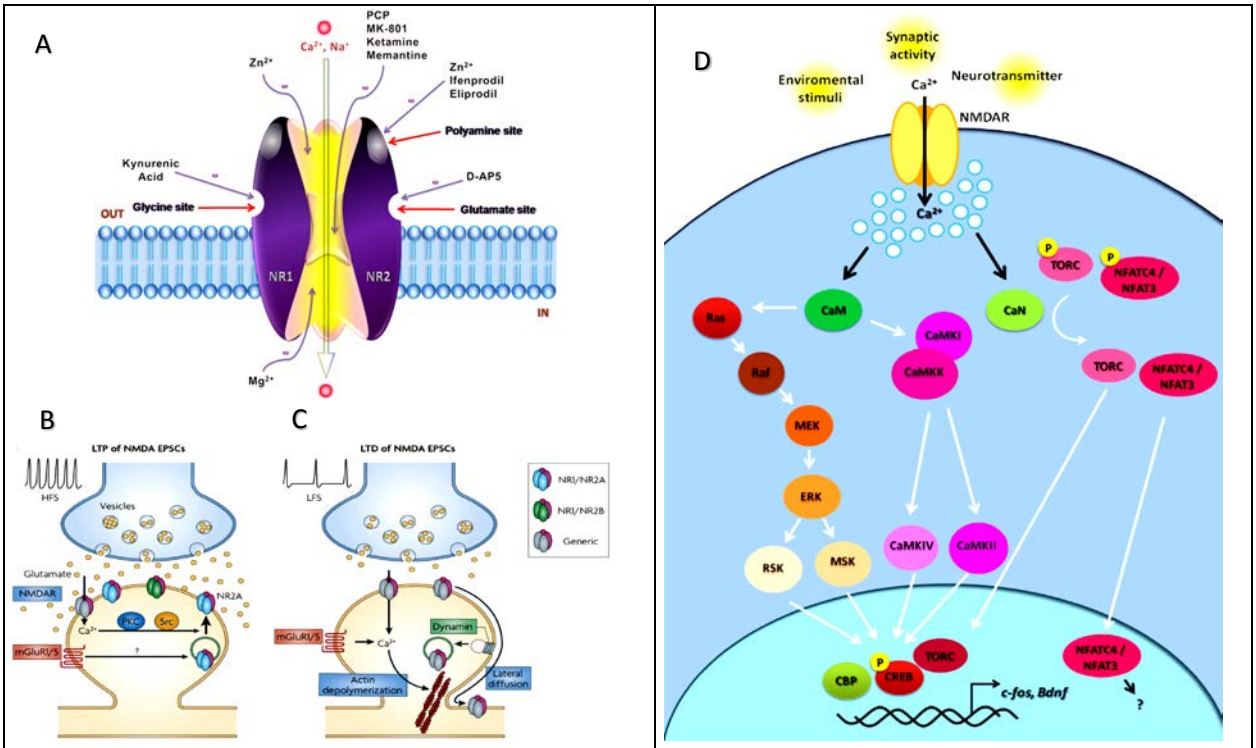


Figure 10. NMDAR structure and function. (A) Model showing binding sites for agonists and antagonists to NMDAR (N-methyl-d-aspartate receptors). The extracellular portions of NR1 and NR2 subunits consist of two domains, the modulatory domain and the agonist binding domain. Glycine and D-serine are agonists for NR1 subunits, whereas glutamate and NMDA act on the NR2 subunit agonistic site. Kynurenic acid is an endogenous ligand for NR1. Zn2+ at higher concentrations can also act as an NMDAR pore blocker. D-AP5 is an antagonist for the agonistic site of NR2 subunit. Mg2+, MK-801, ketamine, PCP, and memantine act as noncompetitive antagonists whose binding sites are within the ion channel pore region. (Adopted from [106]). **(B)** High-frequency stimulation and long-term potentiation (LTP) promote rapid insertion of NR2A-containing NMDARs and an increase in NMD field excitatory postsynaptic potentials at synapses through a protein kinase C and Src-dependent pathway. **(C)** Homosynaptic long-term depression of synaptic NMDA occurs by retrieval of NMDARs from synaptic sites. Low frequency stimulation can induce LTD in a manner that is dependent on metabotropic glutamate receptors that enhanced internalization of NMDARs. NMDAR-triggered LTD promotes actin depolymerization and lateral diffusion of NMDARs away from the synapse site. (Adopted from [107]). **(D)** Signaling pathways that lead to activity-mediated transcription. (Adopted from[105])

Post-synaptic neuron, thus creating a positive feedback that results in the activation of distinct signaling pathways that temporally and permanently alter specific patterns of gene expression. Importantly, the epigenetic changes that allow the establishment of long term genetic programs that control learning and

memory are not completely understood. Even less is known regarding the signaling events triggered by synaptic activity that regulate the establishment of such epigenetic marks. Importantly however, the effect of long-term excitation of the NDMAR on the gene expression program remains unknown. Furthermore, a systematic genome-wide analysis to delineate whether epigenetic reprogramming underlies the response to long-term NMDA activity has not been conducted. Towards systematic understanding of epigenetic regulation of neuronal development and activity, we performed two studies – dissecting the role of Tox3 in neurogenesis, and changes of epigenomic landscape during neurogenesis and neuronal activation - by combining neurobiology, developmental biology and computational biology tools.

Results

Chapter 1: **TOX3 regulates neural progenitor identity**

Chapter 2: **Dynamics & function of distal regulatory elements during neurogenesis & Neuroplasticity**

TOX3 regulates neural progenitor identity

Sanjeeb Kumar Sahu [a,1], Alina Fritz [b,1], Neha Tiwari [c,1], Zsuzsa Kovacs [b], Alireza Pouya [b], Verena Wüllner [b], Pablo Bora [a], Teresa Schacht [b], Jan Baumgart [d], Sophie Peron [c], Benedikt Berninger c, Vijay K. Tiwari [a,*], Axel Methner [b,*]

[a] Institute of Molecular Biology, (IMB), Mainz, Germany

[b] Focus Program Translational Neuroscience (FTN), Rhine Main Neuroscience Network (rmn2), Johannes Gutenberg University Medical Center Mainz, Department of Neurology, Langenbeckstr. 1, D-55131 Mainz, Germany

[c] Institute of Physiological Chemistry, University Medical Center, Hanns-Dieter-Hüsch-Weg 19, D-55128 Mainz, Germany

[d] Institute of Microscopic Anatomy and Neurobiology, University Medical Center of the Johannes-Gutenberg University, Mainz, Germany

1 These authors contributed equally to this work.

* Corresponding authors: E-mail addresses: v.tiwari@imb-mainz.de, axel.methner@gmail.com

Keywords: Transcription Neuronal progenitors Notch, Proliferation, Neuronal differentiation

Biochim Biophys Acta. 2016 Jul;1859(7):833-40.

Abstract

The human genomic locus for the transcription factor TOX3 has been implicated in susceptibility to restless legs syndrome and breast cancer in genome-wide association studies, but the physiological role of TOX3 remains largely unknown. We found Tox3 to be predominantly expressed in the developing mouse brain with a peak at embryonic day E14 where it co-localizes with the neural stem and progenitor markers Nestin and Sox2 in radial glia of the ventricular zone and intermediate progenitors of the subventricular zone. Tox3 is also expressed in neural progenitor cells obtained from the ganglionic eminence of E15 mice that express Nestin, and it specifically binds the Nestin promoter in chromatin immunoprecipitation assays. In line with this, over-expression of Tox3 increased Nestin promoter activity, which was cooperatively enhanced by treatment with the stem cell self-renewal promoting Notch ligand Jagged and repressed by pharmacological inhibition of Notch signaling. Knock- down of *Tox3* in the subventricular zone of E12.5 mouse embryos by *in utero* electroporation of *Tox3* shRNA revealed a reduced Nestin expression and decreased proliferation at E14 and a reduced migration to the cortical plate in E16 embryos in electroporated cells. Together, these results argue for a role of Tox3 in the development of the nervous system.

Introduction

TOX3, also known as TNRC9 (trinucleotide repeat containing 9), is a neuronal transcription factor containing a nuclear localization signal and a high mobility group (HMG)-box domain followed by a C-terminal poly-glutamine stretch. It was first identified as a calcium- dependent neuronal transcription factor, which contributes to calcium-induced activation of c-fos expression[108], and is expressed downstream a cytoprotective cascade involving RhoA activation[109]. Rho family GTPases regulate the actin and adhesion dynamics that control cell migration [110]. Recently, Tox3 was found to be increased in neurons and oligodendrocytes of the injured spinal cord of IL-1-KO mice [111]. These mice have a reduced lesion volume resulting in a better functional recovery after spinal cord injury and this was in part attributed to Tox3 because its over-expression protected human oligodendrocytes from hypoxia [111]. A paralog of Tox3, Tox (thymocyte selection-associated HMG-box), plays a role in brain development [112]. Variants of the human genomic TOX3 locus have been implicated in susceptibility to breast cancer [113,114] and the neurological disease, restless legs syndrome (RLS) [115] in genome-wide association studies (GWAS). RLS is a complex genetic disorder with up to 50% heritability characterized by an urge to move the legs and uncomfortable sensations in the lower limbs.

The strongest RLS association signal was identified in an intron of the human gene *MEIS1*, a member of the TALE family of homeo-box transcription factors. Mouse Meis1 is involved in the specification of neural progenitor cells (NPCs) in the ganglionic eminence similar to Tox [112] and the risk allele reduces an enhancer activity in the Meis1 expression domain [116]. Heterozygous Meis1-deficient mice are hyperactive and thereby resemble the RLS phenotype [116]. Others detected a significant decrease in MEIS1 expression at the mRNA and protein level in brain tissues from RLS patients homozygous for the intronic RLS risk haplotype, compared with those homozygous for the non-risk haplotype [117]. Together, these data suggest that RLS is caused by a reduction of MEIS1 expression and possibly MEIS1 target gene expression during neuronal development. Interestingly, CREB1, which interacts with TOX3 [108,109], binds

the MEIS1 enhancer and the risk allele affects this binding. These results suggest that MEIS1 and TOX3 might both be involved in neuronal differentiation pathways which, when perturbed, increase the risk of RLS in adulthood. This prompted us to investigate the expression of Tox3 in neuronal development. We found that Tox3 is prominently expressed in the subventricular zone of the developing mouse brain; it increases transcription from the Nestin promoter and directly binds to it. Knock- down of Tox3 by *in utero* electroporation of small hairpin RNAs (shRNAs) led to reduced Nestin expression and proliferation of neural progenitor cells and a reduced cortical migration of newborn neurons.

1.Results

1.1. Tox3 is strongly expressed in the ventricular zone of the developing mouse brain

We first investigated the expression pattern of Tox3 during neurogenesis and asked whether it resembles the expression course of Meis1. To this end, we differentiated mouse embryonic stem cells to NPCs with the characteristics of radial glial cells and then further to post-mitotic glutamatergic neurons as described previously [118]. Interestingly, expression analysis of *Meis1* and *Tox3* abundance by quantitative RT-PCR demonstrated that both RLS susceptibility genes were expressed in an identical pattern with a peak at day1 of neurogenesis and a decline towards terminal neuronal differentiation (Fig 1A). To validate these findings *in vivo*, we then analyzed *Tox3* levels in mouse brain samples from embryonic days 11, 14 and 17 and postnatal day 30 using again quantitative RT-PCR verified by immunoblotting (Fig. 1B). Both analyses showed a clear increase in *Tox3* levels at E14 and downregulation in later stages of development. These data further matched publicly available *in situ* hybridization data from the Allen Brain Atlas; at the time point of peak expression, E14.5, *Tox3* is prominently expressed al- most exclusively in the nervous system, with the strongest expression in the ventricular zone (VZ) of the developing mouse brain (Fig. 1C). This

also matches the expression in the adult brain where *Tox3* is mainly expressed in the remnant of the VZ, the ventricular/subventricular zone which harbors the adult stem cell niche (Fig. 1D).

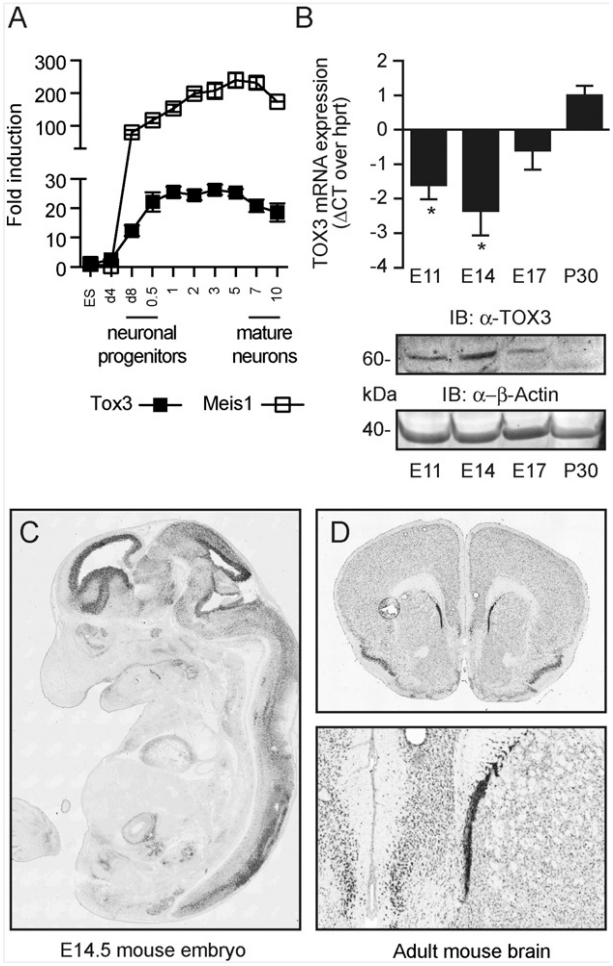


Fig. 1. Tox3 is expressed in the stem cell layers of the developing and adult mouse brain. (A) Similar course of regulation (fold over ES cells) of Tox3 and Meis1 during neuronal differentiation. Quantitative PCRs were performed to quantify the expression of Meis1 and Tox3 during various stages of neuronal differentiation comprising NPC (days d4 and d8) and mature neurons (0.5, 1, 2, 3, 5, 7 and 10 days). The data was normalized to the expression of the housekeeping gene CTCF. (B) Tox3 expression peaks at E14, as shown by quantitative RT-PCR and immunoblotting. mRNA expression was normalized to the expression of housekeeping hypoxanthine-phospho-ribosyl-transferase gene (Hprt) and is expressed as ΔCT values; protein expression was normalized to actin levels, size is indicated. In situ hybridization of a Tox3 antisense probe from the Allen Brain Atlas displaying expression of Tox3 in (C) the VZ of an E14.5 embryo and (D) the adult mouse brain. (Results in this figure partially contributed by me)

In the E14 mouse brain, Tox3 co-localized with other established markers of neural stem cells in the VZ like Sox2 and Nestin (an intermediate filament protein that is often used as a marker for NPCs) (Fig. 2). As Tox3 and Sox2 are both transcription factors and therefore both localize to the nucleus, we calculated the Pearson's co-localization coefficient which was highly positive with 0.65 (SD ± 0.06). MEIS1, in contrast, is not expressed in the VZ or subventricular zone, but in the mantle zone of the ganglionic eminence [116]. Together, these results support a role of Tox3 in stem cell development and self-renewal spatially distinct but temporally coincident with Meis1.

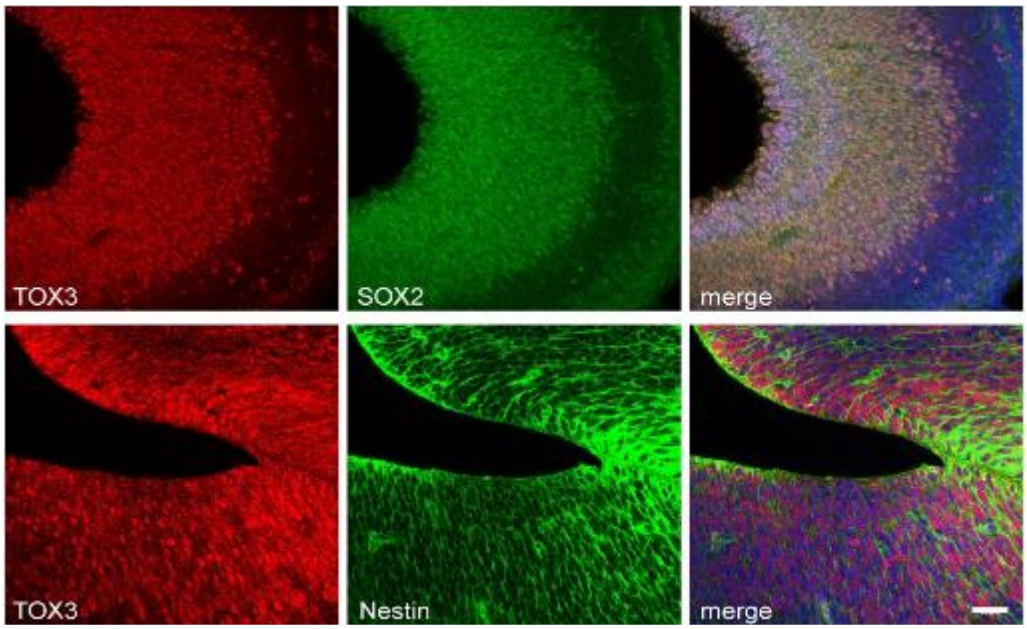


Fig. 2. Tox3 is expressed in Nestin and Sox2-positive cells in the E14 ventricular zone. Mouse embryo sections were stained via immunohistochemistry and visualized by confocal microscopy. The size bar corresponds to 25 μm.

1.2.Tox3 expression in vitro correlates with the expression of the stem cell marker Nestin

We then prepared primary NPCs from the ganglionic eminence from E14.5 mice. In these cells, Tox3 was strongly expressed right after plating in growth factor-deprived medium (Fig. 3A) and seemed to correlate with Nestin expression (Fig. 3B arrow heads). To validate this suspected correlation, we quantified the expression of nuclear Tox3 (defined as co-localization with DAPI) and the expression of Nestin in a region of interest defined by a circle around the nucleus using high-content imaging analysis of immunocyto-chemically stained cultures at day 1 after plating. This indeed verified that Nestin and Tox3 expressions positively correlate with each other ($r^2 = 0.23$, P b 0.0001, n = 6787 cells) (Fig. 3C). When we defined Nestin-positive and -negative cells using an arbitrary threshold, we observed that Nestin-positive cells had a significantly stronger Tox3 expression (Fig. 3D). Also, when we analyzed the transcript levels of *Tox3* and *Nestin* at various time points during the neuronal differentiation of embryonic stem cells, we detected that *Nestin* expression closely followed *Tox3* expression. While *Tox3* expression is induced immediately following commitment to the neural

progenitor state, the induction of *Nestin* is slightly delayed and occurs after *Tox3* expression (Fig. 3E). We therefore suspected that the transcription factor Tox3 positively influences the expression of Nestin.

1.3. Tox3 binds and activates the Nestin promoter

Tox3 can induce transcription from a variety of different promoters [108,109] and we therefore assumed that Tox3 also induces the Nestin promoter. To verify this assumption, we co-expressed *Tox3* with a construct consisting of the second intron of the Nestin gene upstream of firefly luciferase. We chose this promoter construct because its activation directs Nestin expression in NPCs [119,120]. Renilla luciferase driven from the constitutively active CMV promoter was cotransfected as a control. Interestingly, also Notch signaling, which is essential in maintaining a pool of undifferentiated stem cells [121], induces Nestin transcription from this promoter [122]. We therefore used the Notch ligand Jagged and the γ-secretase inhibitor N-N-(3,5-difluorophenacetyl)-L-alanyl-S-phenyl-glycine t-butyl ester (DAPT) as controls. We observed that Tox3 expression caused a more than five-fold induction of transcription from this promoter, which was almost tripled by Jagged, leading to a 15-fold induction of promoter activity, whereas DAPT had the opposite effect (Fig. 3F). To prove that Tox3 indeed interacts with the Nestin promoter, we then performed chromatin immunoprecipitation assays in NPCs using a Tox3-specific antibody. This revealed that Tox3 interacts with the Nestin promoter, but not with a control intergenic region or the Hoxd3 promoter (Fig. 3G). The C3 promoter, served as positive control as it was previously shown by us to be activated by Tox3 [109]. We conclude that Tox3 binds and activates the Nestin promoter and that this is potentiated by the activation of the Notch signaling pathway. It remains to be clarified whether this crosstalk between Notch and Tox3 signaling is additive or synergistic.

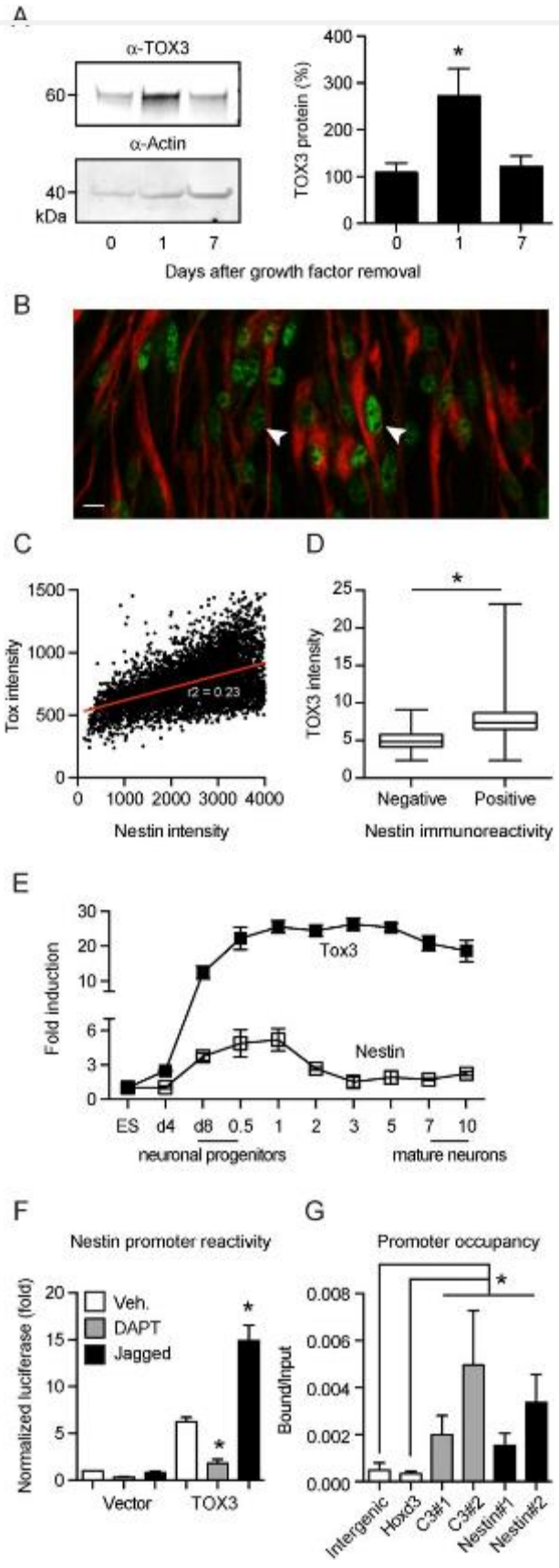


Fig. 3. Tox3 correlates with Nestin expression and binds and induces the Nestin promoter. (A) Tox3 is expressed in NPCs with a strong expression at day 1 after plating and growth factor deprivation. Protein expression was quantitated by immunoblotting and normalized to actin. (B) Nestin-positive cells express more Tox3. NPCs at day 1 after plating were stained with α-Tox3 and α-Nestin antibodies. The size bar corresponds to 10 μm. The arrow heads indicate cells with different Tox3 expression levels. (C) Tox3 expression correlates with Nestin expression. High-content imaging was used to measure Tox3 fluorescence in the nucleus (identified by DAPI) and Nestin fluorescence in a defined perimeter around the nucleus. Fluorescence is displayed in arbitrary units, the red line represents the linear regression of Tox3 and Nestin expression. (D) Nestin-negative cells were defined by the expression values of the secondary antibody only and Tox3 expression quantitated in the nucleus. The results are depicted as box and whisker plot with the box extending from the 25th quartile to the 75th quartile and the whiskers representing the minimum and maximum values of n = 6647 Nestin+ and 140 Nestin− cells. The asterisk indicates P b 0.0001 as determined by an unpaired student's t test. (E) Nestin expression follows Tox3 expression during neuronal differentiation. Quantitative PCRs were performed to quantify the expression of Nes and Tox3 during various stages of neuronal differentiation comprising NPC (days d4 and d8) and mature neurons (0.5, 1, 2, 3, 5, 7 and 10 days). The data was normalized to the expression of the housekeeping gene CTCF.

(F) Tox3 overexpression induces transcriptional activity of the Nestin promoter, which is enhanced by treatment with the Notch ligand jagged and repressed by the Notch inhibitor DAPT. Promoter activation was performed by dual luciferase assays in Neuro2a cells transfected with empty vector or Tox3 overexpressing vector and a construct expressing Renilla luciferase. Light emission was quantitated 24 h later and normalized to Renilla activity. Values are provided in fold luminescence versus empty vector. Bar graphs of all experiments are presented as the mean ± SEM. of at least three independent experiments. An asterisk indicates P b 0.05 as determined by one-way ANOVA. (G) Tox3 directly binds the Nestin promoter in NPCs. Sonicated chromatin from NPCs at day 8 was immunoprecipitated using a Tox3-specific antibody and the enrichment of Tox3 at the Nestin promoter (Nes#1 and Nes#2) analyzed by quantitative PCR. An intergenic region and the Hoxd3 locus were used as negative control, and two different regions of the C3 promoter (C3#1 and C3#2) served as positive control. The data is shown as the ratio of precipitated DNA (bound) to the total input DNA and normalized to the intergenic region. The error bars represent the standard deviation from three different biological samples. Asterisks indicate significant P values measured by the Mann Whitney U test. (Results in this figure partially contributed by me)

1.4. TOX3 knockdown reduces Nestin expression and cellular proliferation in E14 mouse embryos

We then aimed to study the effect of Tox3 knockdown, on Nestin expression in vivo and first identified small hairpin (sh)-RNAs that strongly reduced Tox3 expression in a transfectable cell line with stem cell-like properties [123] (Fig. 4A). The two most potent shRNAs were combined and driven from the H1 promoter and in-utero electroporated into E12.5 mouse embryos. A shRNA directed against luciferase (shLuc) served as negative control. These constructs express GFP from a separate constitutively active promoter and successfully transfected cells can be identified by GFP expression. When we harvested the embryos for characterization at E14.5, we indeed observed a significant downregulation of Tox3 in GFP+ shTox3-expressing cells as compared to GFP+ shLuc-expressing cells, proving the feasibility of this approach (Fig. 4B, GFP+ cells with reduced Tox3 expression are marked with arrowheads).

We then quantitated Nestin expression in GFP-positive cells of E14 embryos because of the above-described observations and indeed observed a significantly reduced Nestin expression in knockdown cells, further substantiating the assumption that Nestin expression is controlled, at least in part, by Tox3 (Fig. 4C). Nestin is a marker of a rapidly cycling subtype of NPCs characterized by their proliferative capacity

[124]. We therefore also quantified the proliferative activity of transfected cells by bromodeoxyuridine (BrdU) incorporation. BrdU is a synthetic nucleoside that is a structural analog of thymidine, which is incorporated into the newly synthesized DNA of replicating cells. In our experiments, BrdU was given on day 1 after electroporation. We observed a significant reduction of BrdU-positive cells in Tox3 knock-down cells. These results suggest that Tox3 induces transcription from the Nestin promoter during brain development and is involved in proliferation of neural progenitors (Fig. 4D).

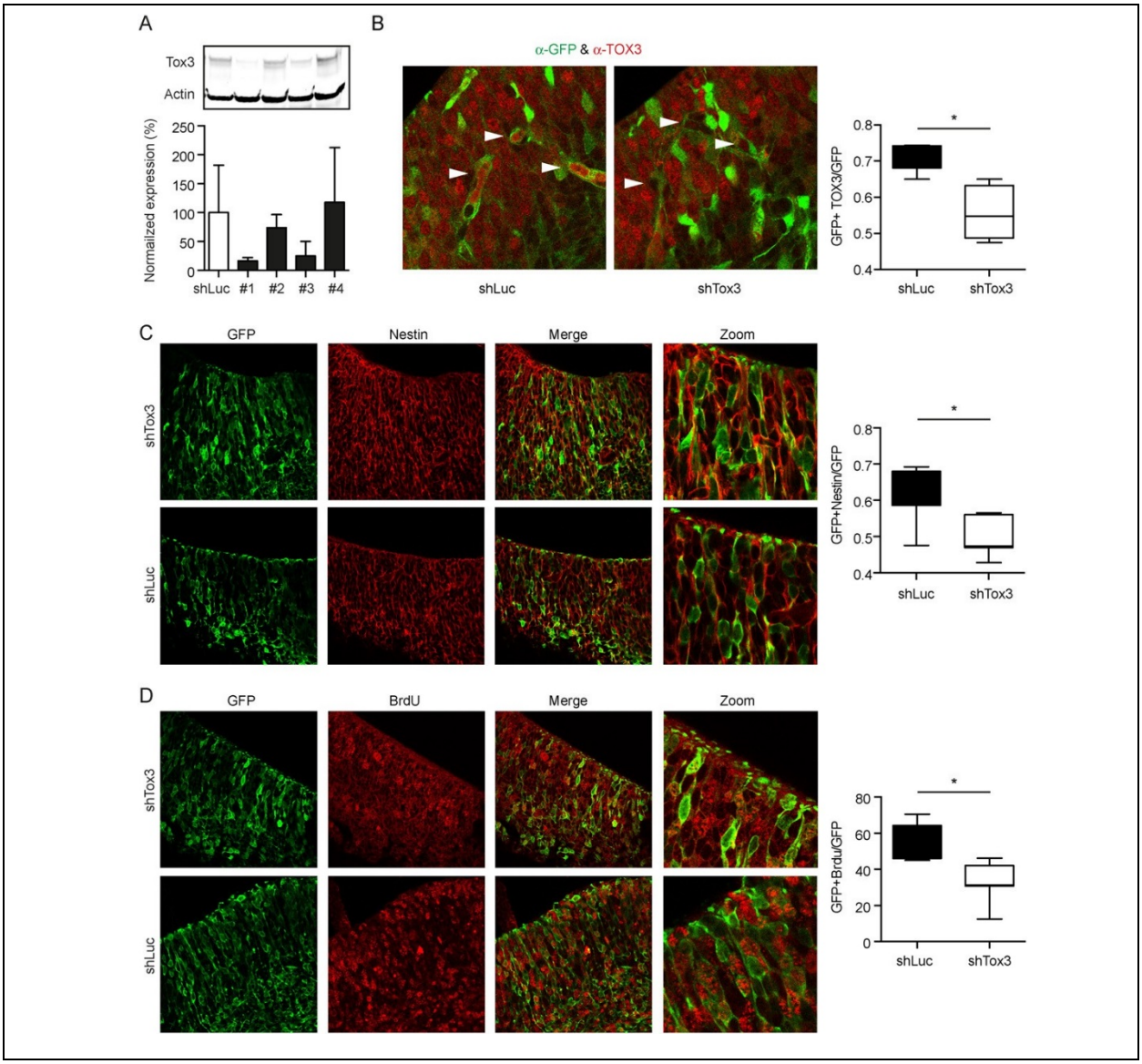


Fig. 4. Tox3 knockdown reduces Nestin expression and proliferation of neural progenitors in vivo. (A) A set of four different shRNAs against Tox3 (#1–4) and an shRNA against luciferase (shLuc) were transfected into glioblastoma-derived stem like cells and the Tox3 abundance quantitated by immunoblotting 48 h

later. Actin on the same blot served as loading control. The bar graphs show the mean ± SEM of three independent biological replicates (B) In utero electroporation of shRNAs against Tox3 significantly downregulates Tox3 in vivo. Mouse E12.5 embryos were electroporated in utero with a plasmid driving shRNAs #1 and #3 from the H1 RNA promoter and GFP from the constitutive active chicken β-actin promoter. All GFP+ cells therefore express the Tox3 or the luciferase control shRNA. Two days after electroporation, the embryos were sacrificed, and the Tox3 expression quantified by immunohistochemistry and confocal microscopy. (C–D) Tox3 knockdown reduces (C) Nestin expression and (D) BrdU incorporation in GFP+ cells quantified using confocal microscopy. BrdU was given 24 h after electroporation. The results in B–D are depicted as box and whisker plots with the box extending from the 25th quartile to the 75th quartile and the whiskers representing the minimum and maximum values of n = 7 embryos. An asterisk indicates P b 0.05 as determined by student's t test.

1.5. Tox3 is involved in cortical migration

We then attempted to investigate Tox3 function during cortical neurogenesis. We again performed *in utero electroporation* at embryonic stage E12.5 and harvested the embryos at E16.5 for further analysis. In the control shRNA electroporated brains, we observed GFP-positive cells in the cortex, some throughout various layers of the brain and very few still in the ventricular zone, reflecting the cortical

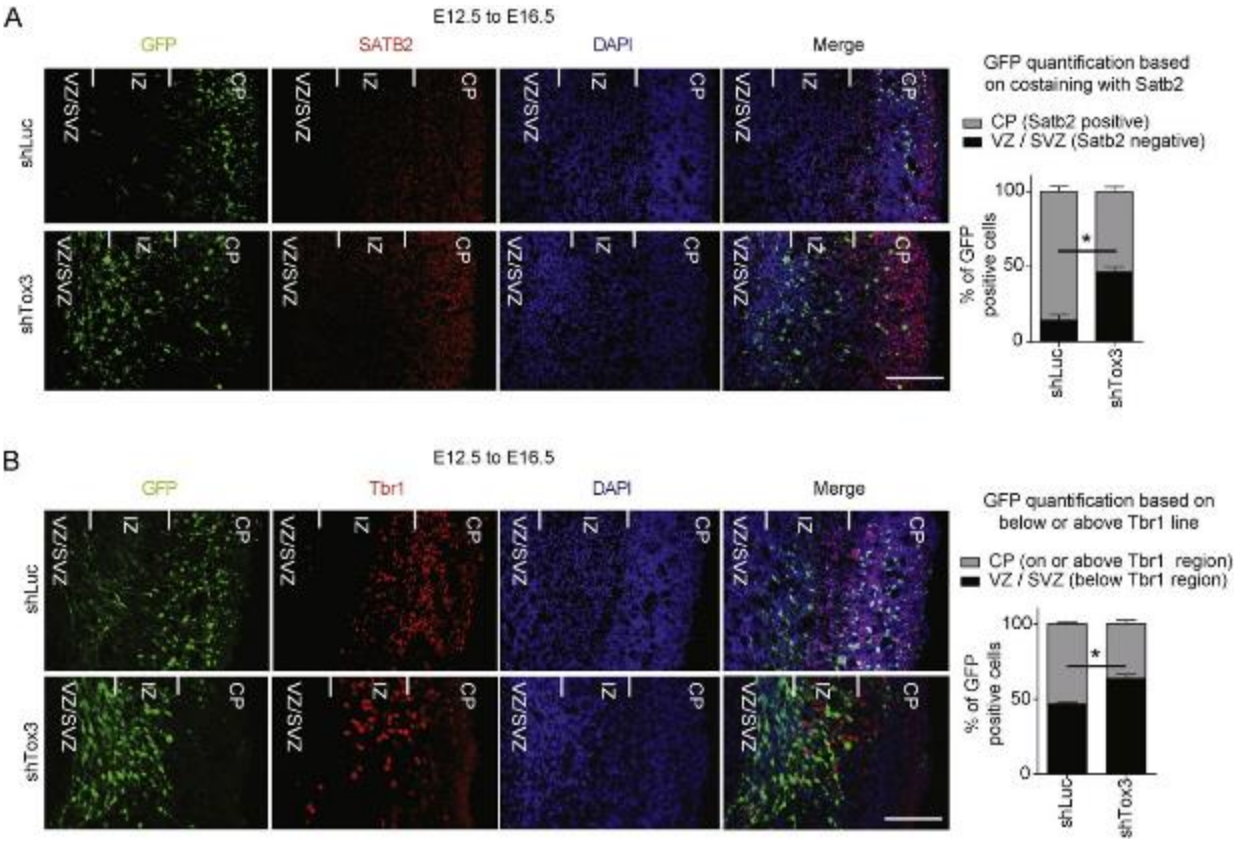


Fig. 5. Tox3 knockdown reduces cortical migration of neural progenitors. Mouse E12.5 embryos were electroporated in utero with shTox3 and shRNA against luciferase as negative control. Four days later, the embryos were sacrificed and fixed brains stained with anti-GFP and Satb2 or Tbr1 antibodies. The spatial localization of GFP+ cells in regard to (A) Satb2 and (B) Tbr1 positive cells was quantified by confocal microscopy and the individual three experiments depicted. (Results in this figure contributed by me)

migration of newborn neurons. Tox3 shRNA electroporated brains, in contrast, showed a clear phenotype where the GFP-positive cells were still largely confined to the ventricular zone. This was quantified by counting GFP-positive cells that were simultaneously positive for Satb2, a post-mitotic determinant and marker for upper layer neurons in the neocortex [125]. This demonstrated a significantly reduced number of double- positive cells in Tox3 knockdown cells compared with controls (Fig. 5A). A similar effect was observed for Tbr1, an established deep layer neuronal marker (Fig. 5B). Together, these findings suggest that Tox3 is critical for neurogenesis during embryonic development.

Discussion

Our data demonstrate that the transcription factor Tox3 is expressed in the ventricular zone of the developing mouse brain and is necessary for the proliferation of neural stem cells. Apparently, TOX3 is also expressed by neuroepithelial cells and early radial glia derived from human ES cells [126]. Lack of Tox3 reduces the proliferative potency of NPCs which most probably causes the reduced amount of post-mitotic neurons in the upper layer of the neocortex at later embryonic states observed in this study (Fig. 5). Interestingly, knockdown of Tox3 resembles the phenotype observed after in utero electroporation of the founding member of the TOX family, Tox, which also causes reduced proliferation and a reduced amount of neurons in the cortical plate [112]. Tox over-expression can induce or repress transcription of various genes involved in neuronal differentiation [112] and it would be interesting to know whether the effect of Tox overexpression is mediated via repression of Tox3.

We further demonstrate that Tox3 binds and activates the Nestin promoter and possibly other promoters turned on in Nestin-positive NPCs. The HMG-box class of transcription factors in general can be di- vided into subfamilies that recognize DNA in either a sequence dependent or sequence-independent manner. Tox again was first thought to bind to DNA in a sequence-independent manner based on a bioinformatic analysis [127] but was recently shown by deep sequencing to bind a specific DNA sequence [112]. This still needs to be clarified for Tox3.

We observed that Tox3 expression facilitated Nestin promoter activation by the Notch ligand Jagged. Jagged is essential in maintaining a pool of undifferentiated stem cells [120]. Inhibition of Notch signaling in neural stem cells with DAPT also delays the G1/S phase transition [128] and therefore reduces the proliferative activity of these cells similar to knockdown of Tox3. Notch signaling is activated via proteolytic cleavage of Notch mediated by the γ-secretase and release of the Notch intracellular domain, which then enters the nucleus where it associates with the DNA-binding protein CSL (also known as RBP-Jκ, CBF1, Suppressor of Hairless, and Lag-1) and recruits Mastermind-like transcriptional coactivators to

form a transcriptional activation complex [129]. HMG-box transcription factors can enable the simultaneous binding of other transcriptional regulators to the DNA by modifying chromatin structure through bending and unwinding the DNA through an interaction of the HMG box with the minor groove [130]. It is thus possible that Tox3 facilitates binding of this complex to the Nestin promoter without a specific binding motif for Tox3.

We previously showed that Tox3 can enhance transcriptional activation from different promoters and that this is dependent on the pre- dominance of other nuclear proteins such as phosphorylated CREB or CITED1 and possibly other factors, e.g., CSL within the transcriptionally active complex. Alternatively, Tox3 itself binds to an as of yet unknown DNA recognition site in all of these promoters. The transcription factor CREB enhances transcription from the intronic MEIS1 enhancer which increases susceptibility to RLS [115] and, when cotransfected with Tox3, from the Bcl-2 promoter [109]. Co-transfection of the transcriptionally less active S133A mutant [131], in contrast, abolished this effect of Tox3 on the Bcl-2 promoter. Lack of phosphorylation of CREB at S133 attenuates CREB function by inhibiting binding of CBP to CREB after phosphorylation by Ca2+/calmodulin-dependent protein kinases [132]. It is therefore conceivable that Tox3 facilitates Meis1 expression by enhancing the transcriptional activity of phosphorylated CREB at the Meis1 promoter. However, the spatially distinct expression of Meis1 during the peak expression of Tox3 speaks against a major crosstalk of these two transcriptionally active proteins implicated in RLS.

Author contributions

JB, VT and AM designed the study; SKS, AF, NT, ZK, PB and TS performed the experiments and analyzed the data. BB provided critical suggestions and helped write the manuscript. VT and AM wrote the manuscript.

Conflicts of interest

The authors have no conflicts of interest.

Transparency document

The Transparency document associated with this article can be found in the online version.

Acknowledgment

We would like to thank all members of the Tiwari and Methner lab for their cooperation and critical feedback during the progress of the project. Support from the core facilities of the Institute of Molecular Bi- ology (IMB), Mainz, is gratefully acknowledged; especially the microscopy facility. TS is a recipient of Susan G. Komen Breast Cancer Research Foundation grants BCTR0503620 and FAS0703860. This work is part of the MD thesis of AF. The research performed at the laboratory of VT was supported by the Wilhelm Sander Stiftung 2012.009.1, EpiGeneSys RISE1 program, Marie Curie CIG 322210 and Deutsche Forschungsgemeinschaft (DFG) Grant TI 799/1-1. The research per- formed in the laboratory of AM was supported by Wilhelm Sander- Stiftung (2009.054.1) and DFG CRC 1080 A10.

Materials and methods

1.1. Cells, proteins, shRNA, and plasmids

ES cells derived from blastocysts (3.5 PC) of mixed 129-C57Bl/6 background (called 159.2) were cultured and differentiated as previously described [118]. Neuro2a cells were maintained in DMEM (PAA laboratories) supplemented with 10% FCS, 100 IU/ml penicillin and 100 µg/ml streptomycin. Primary NPCs were derived from the ganglionic eminence of C57/Bl6J mice in the embryonic state E14 and were cultured in neurosphere medium consisting of neurobasal medium (Gibco) supplemented with 10% B27 supplement, 1% Glutamax (all Life Technologies), FGF-2 and EGF (both PeproTech 20 ng/ml) and 1% penicillin/streptomycin (Gibco). Glioblastoma-derived stem cells were cultured additionally with heparin (32 IE/ml Ratiopharm) as previously described [123]. Spheres were mechanically dissociated and filtered. The Tox3 construct and the shRNAs have been previously described [109]. The two most efficient shRNAs (#1 and #3) were cloned after each other downstream of the H1 promoter into pCGLH which also expresses GFP from the chicken β-actin promoter. The Nestin promoter luciferase construct was a kind gift from Leping Cheng from the University of Shanghai, China [133].

1.2. Quantitative RT-PCR

Total RNA isolation, cDNA synthesis and quantitative RT-PCR were performed according to the manufacturer's (Qiagen) protocol. Primer sequences (name, forward, reverse) were: Nes 5'gcctggatctggaagtcaacaga3', 5'ccctcagatgcaactctgcctta3'; Tox3 5'aagatggcgtgcttcatagc3', 5'gatcctggcggtactgtgac3'; Pax6 5'gccaccagactcacctgacacc3', 5'ctcaccgcccttggttaaagtc3'; Meis1 5'gcatccactcgttcaggaggaac3', 5'cactgttgtccaagccatcacct3'; and Ctcf 5'cacacacacaggtactcgtcctca3', 5'ccactggtcacaaaggccatatc3'.

1.3. Immunocytochemistry/immunohistochemistry

NPCs and tissue were fixed for 15 min in 2% paraformaldehyde (PFA) at room temperature, blocked and permeabilized with 0.2% FCS and 0.2% BSA in PBS-0.01% TritonX-100 for 30 min on ice. Antibodies (α-

Tox3 (Sigma HPA040376; 1:200), Nestin (Millipore MAB353; 1:100), and Sox2 (Santa Cruz, 1:200)) were incubated overnight before washing with PBS and incubating the secondary antibody (CY3-conju- gated, 1:500) for 2 h on ice. Nuclei were counterstained with 100 ng/ml DAPI for 15 min at room temperature, washed and embedded in Vectashield™ (Biozol) and visualized by microscopy (Olympus BX51).

1.4. Immunoblotting

Total cell lysates and eluates were separated on 8–16% polyacrylamide gels (Thermo Scientific), transferred onto nitrocellulose mem- branes (Invitrogen) and blocked in 3% non-fat dry milk for 1 h at room temperature before overnight incubation with a primary rabbit polyclonal α-Tox3 antibody (Sigma; 1:1000), rabbit monoclonal α–β- actin antibody (Millipore; 1:5000), and mouse monoclonal α-GAPDH antibody (Cell Signaling; 1:5000), followed by IgG (Fc) infrared fluorescence-conjugated (Licor, 1:30000) secondary antibody. The membranes were scanned for infrared fluorescence at 680 nm or 800 nm using the Odyssey system (Licor).

1.5. Luciferase assays

High-purity plasmids were prepared using Nucleobond AX 500 columns (Machery & Nagel). Neuro2a cells were transiently transfected in a 48-well plate at 80–90% confluence with the Nestin luciferase reporter plasmid [133], a Renilla control plasmid [109] and the indicated expression constructs using Lipofectamine 2000 (Invitrogen) or Attractene (Qiagen). The cells were then washed and lysed in 200 µl of passive lysis buffer 24 h after transfection (Promocell), the lysates centrifuged at 12,000 × g for 1 min, and 20 µl of the supernatants transferred to a white 96-well microtiter plate. A total of 100 µl of luciferase assay buffer (Promocell) was directly injected into each well before measurement. Luminescence was measured by a Genios Pro microplate reader (Tecan) and integrated for 10,000 ms. Normalization was conducted by dual luciferase assays, in which 20 µl of lysate were transferred to a second white microtiter plate and 40 µl of Renilla assay-enhancer solution added to the wells. Coelenterazine (Promocell) in the appropriate assay

buffer was injected into each well, and luminescence was measured as described above. Recombinant rat Jagged1 protein (R&D Systems; 1 µg/ml) or the γ-secretase inhibitor DAPT (Sigma; 1 µM) was added after 24 h and left for 2 h before measurement.

1.6. Chromatin immunoprecipitation assay

ChIP experiments were performed as previously described [134]. In brief, cross-linked chromatin was sonicated to achieve an average fragment size of 500 bp. Starting with 70 µg of chromatin and 5 µg of Tox3 antibodies, 1 µl of ChIP material and 1 µl of input material were used for quantitative real-time PCR using specific primers. Primers covering an intergenic region were used as control. The efficiencies of the PCR amplifications were normalized to those of the PCR products of the intergenic regions. Primer sequences (name, forward, reverse) were: Intergenic region: 5'atgcccctcagctatcacac3', 5'ggacagacatctgccaaggt3'; Hoxd3 promoter: 5'tgtcacaaatggatgtttgctatgg3', 5'ccgaaatgcctcacatttctcaa3'; C3 promoter #1: 5'cccagcaactgaggaggtag3', 5'tgattggaggtcagcacaac3'; C3 promoter #2: 5'aaccaagtacaacagggacca3', 5' aacaagccttggggtgagta3'; Nestin promoter #1: 5'ctctgctgagctgggatgat3', 5'ggagcttttaaggggggagaa3'; and Nestin promoter #2: 5'ggactagagcaggctgagggtca3', 5'cctgaggaccaggagtctccatc3'.

1.7. In utero electroporation

In utero electroporation experiments were carried out essentially as described [135]. All experimental procedures were conducted in accordance with European, national, and institutional guidelines for animal care. Timed-pregnant (Theiler stages 20 = E12.5) C57BL/6 mice were anesthetized with isoflurane (2.5% via mask, Forene®, Abbod) and analgesized subcutaneously with carprofen (4 mg/kg bodyweight, Rimadyl, Pfizer). After opening the abdominal cavity the embryos were carefully exposed and 1 µl colored solution containing 3–4 µg of the pCGLH plasmid expressing chained shRNAs against TOX3 or against luciferase and GFP injected into one of the lateral ventricles. With specialized platinum electrodes (Nepagene CUY

650P) the appropriate voltage was applied (50 ms, interval 950 ms, 5 pulses). Then, the uterine horns were replaced after electroporation in the abdominal cavity.

1.8. Cortical migration immunofluorescence assay

Isolated E16.5 embryonic brains were fixed immediately for 24 h in 4% PFA in PBS at 4 °C. Brains were then cryoprotected in 10% sucrose for 2 h then in 30% sucrose (in PBS) overnight, embedded in Tissue-Tek, stored at − 20 °C and cryosectioned 12 µm. Sections on coverslips were preblocked with 2%BSA, 0.5% Triton (in PBS) for 1 h. Primary antibodies (Satb2, 1:500, Abcam (ab34735), Tbr1, 1:500, Abcam (ab31940) and GFP (B-2), 1:100 Santa Cruz (sc-9996)) were applied in blocking solution overnight at 4 °C. Fluorescent secondary antibodies were applied according to the manufacturer's protocol (life technologies). The coverslips were counterstained with Hoechst, mounted with immomount and imaged with a confocal laser-scanning microscope Leica (SP5). Data were processed with ImageJ software.

1.9. Statistical analysis

The data are summarized as the mean ± S.E.M., and statistical significance was assessed using two-tailed t-tests or analysis of variance (ANOVA), with Tukey's or Dunnett's multiple comparison test, as indicated.

Dynamics and Function of Distal Regulatory Elements during Neurogenesis and Neuroplasticity

Sudhir Thakurela[#], Sanjeeb Kumar Sahu[#], Angela Garding, Vijay K. Tiwari[*]

Institute of Molecular Biology (IMB), Mainz, Germany

*Correspondence should be addressed to v.tiwari@imb-mainz.de.

[#] These authors contributed equally to this work.

Running title: Dynamics of distal regulatory landscape during neurogenesis and neuronal activity

Key words: Gene regulation, epigenetic mechanisms, transcription factors, chromatin accessibility, neurogenesis, neuronal activity

Genome Res. 2015 Sep;25(9):1309-24.

ABSTRACT

Gene regulation in mammals involves a complex interplay between promoters and distal regulatory elements that function in concert to drive precise spatio-temporal gene expression programs. However, the dynamics of distal gene regulatory landscape and its function in the transcriptional reprogramming that underlies neurogenesis and neuronal activity remain largely unknown. Here, we performed a combinatorial analysis of genome-wide datasets for chromatin accessibility (FAIRE-seq) and the enhancer mark H3K27ac that reveal the highly dynamic nature of distal gene regulation during neurogenesis, which gets progressively restricted to distinct genomic regions as neurons acquire a post-mitotic, terminally differentiated state. We further find that the distal accessible and active regions serve as target sites for distinct transcription factors that function in a stage-specific manner to contribute to the transcriptional program underlying neuronal commitment and maturation. Mature neurons respond to a sustained activity of NMDA receptors by epigenetic reprogramming at a large number of distal regulatory regions as well as dramatic reorganization of super-enhancers. Such massive remodeling of distal regulatory landscape in turn results in a transcriptome that confers a transient loss of neuronal identity and gain of cellular plasticity. Furthermore, NMDA receptor activity also induces many novel pro-survival genes that function in neuroprotective pathways. Taken together, these findings reveal the dynamics of the distal regulatory landscape during neurogenesis and uncover novel regulatory elements that function in concert with epigenetic mechanisms and transcription factors to generate the transcriptome underlying neuronal development and activity.

INTRODUCTION

The nervous system is the most complex mammalian organ. How this complexity is generated during development remains very poorly understood. The primordium of the central nervous system (CNS), the neural plate, is composed of a single sheet of neuroepithelial (NE) cells that undergo rapid symmetric divisions to enable planar expansion of the neural plate and generation of the neural tube. At mid-gestation, between embryonic day 9 (E9) and E10 in mice, the first neurons of the CNS are born, heralding an important transition in the development of neural progenitor cells. In developing brain radial glial (RG) cells in the ventricular zone (VZ) undergo asymmetric division, and daughter cells migrate toward the cortical plate (CP), passing through the subventricular zone (SVZ). Our understanding of the transcriptional control of neurogenesis in the cerebral cortex by sequence-specific transcription factors has increased but despite exciting developments [136-139], very little is known about the dynamics of chromatin accessibility and distal epigenetic gene regulation during embryonic neurogenesis and how stage-specific transcription factors utilize differential regulatory regions in driving the stage-specific transcriptional program.

The neuronal signaling via N-methyl-D-aspartate (NMDA) receptors plays a critical role in the development of the CNS and in adult neuroplasticity, learning, and memory [72,73]. NMDA receptors have also been implicated in the etiology of several neurological disorders [89-92]. Excessive stimulation of the NMDA receptor results in disorders involving acute insult to the brain with deprivation of blood supply (e.g., ischemic stroke, traumatic brain injury) [93,94]. Moreover, neurotoxicity induced via hyper-stimulation of the NMDA receptor also facilitates slow-progressing neurodegenerative diseases (e.g., Huntington, Alzheimer's, Parkinson's and amyotrophic lateral sclerosis) [95-102] and disorders arising from the sensitization of neurons (e.g., epilepsy, neuropathic pain) [103,104]. The NMDA receptor has been extensively characterized via electrophysiological assays, genetic and chemical manipulations. NMDA receptors require the binding of glycine and glutamate in combination with the release of voltage-dependent magnesium blockage. In addition to its role during CNS development, neuronal activity via NMDA

receptors also imparts neuronal plasticity, memory formation and learning that require associated transcriptional changes to mediate physiological responses [74-78]. Evidences suggest that NMDA receptor activation leads to strengthening of synapses through long-term potentiation (LTP) and to the weakening of synapses through long-term depression (LTD)[74,75,80-82]. Recent findings have revealed early transcriptome responses following neuronal activity including via NMDA receptors that further underlie many aspects of neuronal development [74,75,83-88]. Importantly however, the effect of long-term excitation of the NDMA receptor on the gene expression program remains unknown. Furthermore, a systematic genome-wide analysis to delineate whether epigenetic reprogramming underlies the response to long-term NMDA activity has not been conducted.

In eukaryotes, nucleosomes provide a basic layer of transcription repression by reducing access to DNA. A widely emerging concept is that nucleosome positioning and occupancy are organized by the combinatorial action of transcription factors, epigenetic regulators and DNA sequence to regulate DNA accessibility [140]. In this regard, open chromatin profiling has been widely used to identify regulatory elements that predict cell type-specific functional behaviors [141]. Active regulatory elements such as promoters and enhancers are accessible and are marked by H3K27ac, thereby allowing to study the activity state of both proximal and distal regulatory elements [142-144]. Recently, the term 'super-enhancer' was used to describe groups of putative enhancers in close genomic proximity which were further shown to drive expression of genes that define cell identity [145-147]. To understand cell-fate determinants at the level of transcriptional regulation, it is crucial to identify a reliable set of regulatory elements that actively contribute to regulation of gene expression during processes of cell fate commitment. One of the most effective means of discovering these regulatory elements is through the identification of nucleosome-depleted regions (“open chromatin”). The FAIRE assay has emerged as a simple, unbiased but robust and high-throughput method to identify such functional regulatory regions in a broad range of organisms and cell types [148-151]. Here we performed a combinatorial analysis of chromatin accessibility using FAIRE assay

and the enhancer mark H3K27ac to elucidate the dynamics and function of distal gene regulatory landscape during neurogenesis and in response to neuronal activity.

RESULTS

FAIRE-seq sensitively identifies accessible chromatin during neurogenesis

To investigate changes in chromatin accessibility during neurogenesis, we adapted a highly refined system that uses embryonic stem cells (ESC) to generate >95% pure neuronal progenitor (NP) cells ("radial glial" like) and subsequently terminally differentiated pyramidal neurons (TN) [152]. Previous studies, including our own work, have revealed highly synchronous and reproducible changes in the epigenome and transcriptome during neuronal differentiation in this system that were also in good agreement with mouse primary cortical neurons [134,153-157] (Fig. 1A, B). We performed the FAIRE assay at three distinct time points during neuronal differentiation: NP, TN day 1 (TND1) (immediate/early neurons, marking onset of neurogenesis), and TN day 10 (TND10) (late neurons, representing terminally differentiated postmitotic neurons) and subjected the derived material to high-throughput sequencing (FAIRE-seq) (Supplemental Fig. S1A-E, Supplemental Table S1). Exploration of UCSC browser tracks revealed the expected patterns of accessible (*Pax6*) and inaccessible (*Pou5f1*) regions in our FAIRE-seq data (Fig. 1C). Further analysis demonstrated that neuronal differentiation accompanies a progressive reduction in the number of both total and stage-specific (unique) accessible sites (Fig. 1D and Supplemental Fig. S1F). This indicates that post-mitotic differentiated cells have a more compact and defined accessible chromatin state which may reflect their restricted developmental potential. Genomic distribution of FAIRE sites revealed that a large number of peaks occurred at intergenic regions followed by promoters in NP and TND1, whereas in TND10, promoters exhibited slightly higher accessibility than intergenic regions (Supplemental Fig. S1G, H). Exons in all stages exhibited much less openness, followed by introns (Supplemental Fig. S1G). Interestingly, we also observed that fully differentiated neurons have significantly higher peak widths and promoters exhibit higher accessibility than other genomic locations at all stages of neuronal

differentiation (Fig. 1E, F). An examination of the relationship between promoter accessibility and transcription revealed

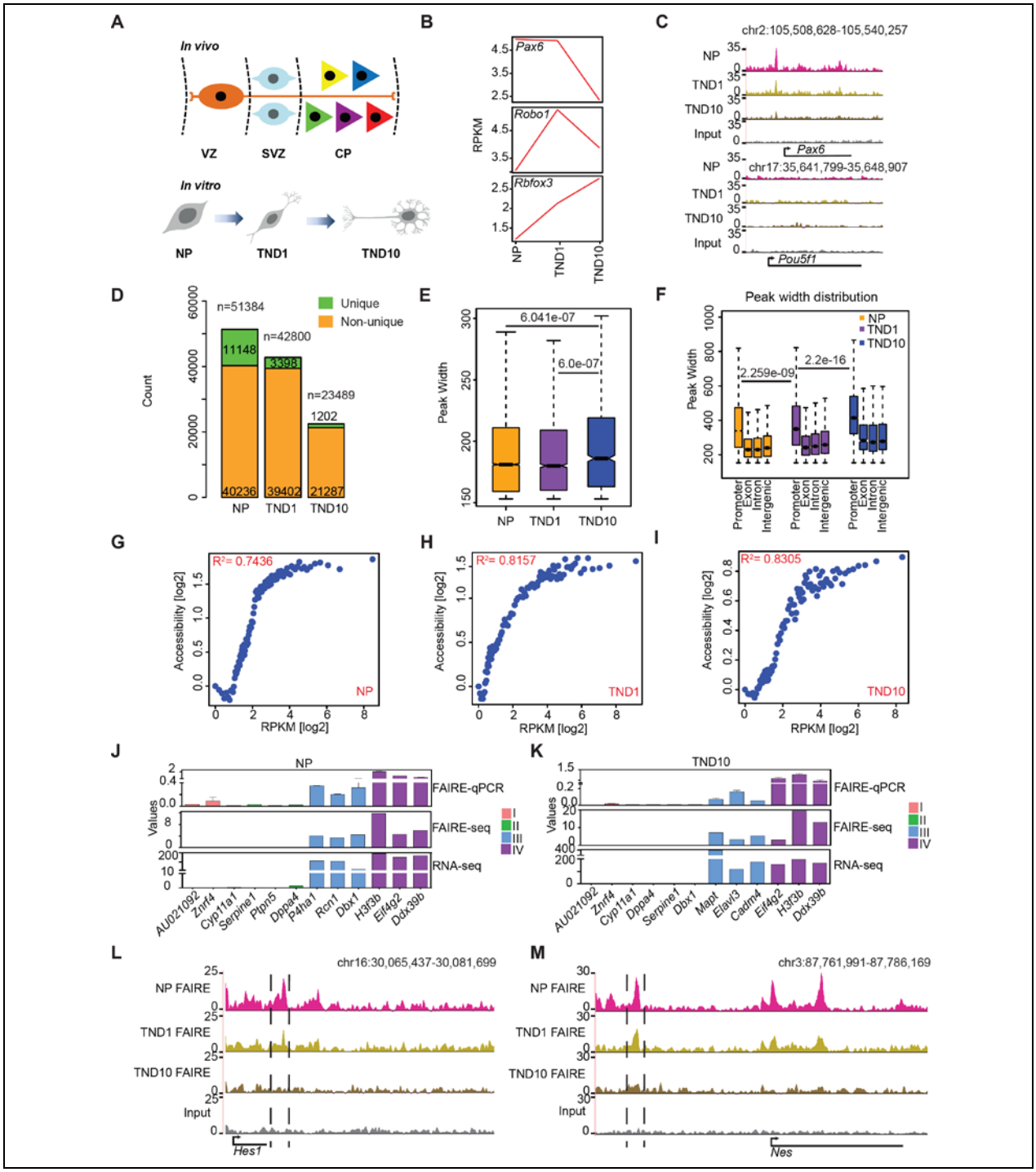


Figure 1. FAIRE-seq sensitively identifies proximal and distal accessible regions during neuronal differentiation. (A) Schematic representation of the in vitro differentiation system compared to respective in vivo stages. (B) Line plot showing the expression of established marker genes for different neuronal stages. (C) Genome browser tracks indicating the presence of an open region at Pax6 promoter (gene active in NP) and the absence of any active region in the Pou5f1 gene locus (active in ES cells only). (D) Stacked bar plot of the total and unique FAIRE peak count. (E) Box plot showing distribution of peak widths of unique NP, TND1 and TND10 peaks. P-values are calculated using Wilcoxon test. (F) Peak width distribution in different genomic regions. P-values are calculated using Wilcoxon test (G-I) Scatter plot comparing gene expression and promoter accessibility. Genes were binned into percentiles, and the mean expression and mean promoter accessibility of each bin were plotted. The x-axis shows gene expression in reads per kilobase of transcript per million mapped reads (RPKM); the y-axis represents normalized FAIRE enrichment at promoters. (J-K) FAIRE-qPCR validation of promoter accessibility and gene expression. Genes are categorized into 4 classes: (I) Not expressed in any of the three stages, (II) Not expressed in the analyzed stage but expressed in at least one other stage, (III) Expression specific to the analyzed stage, and (IV) Expressed in all the stages. Values on y-axis represent FAIRE/Input normalized to Hspa8 (FAIRE qPCR), normalized enrichment score (FAIRE-seq) and RPKM for normalized expression (RNA-seq). (L) Genome browser track showing the presence of a distal regulatory element near *Hes1* (a NP specific gene). (M) Same as in L for the nestin gene (*Nes*). (Results in this figure partially contributed by me)

a very high positive correlation between promoter openness and gene expression at all three stages (Fig. 1G-I). Surprisingly, such correlation increased following acquisition of the neuronal state as compared to neuronal progenitors (Fig. 1G-I). We further validated a number of genes for their promoter accessibility using FAIRE-qPCR which fully corroborated with the observations based on genome-wide FAIRE-seq and gene expression data (Fig. 1J, K). As recent data suggest that distal regulatory elements are critical players in orchestrating the stage-specific gene expression profile underlying cellular identity during differentiation, we next focused on uncovering distal regulatory elements and their contribution to the transcriptional reprogramming underlying neurogenesis [87,88]. A closer look at UCSC browser tracks revealed that FAIRE-seq is also highly sensitive and specific for the identification of potential distal regulatory regions (Fig. 1L, M). Overall, these results suggest that FAIRE-seq can sensitively identify both proximal and distal accessible regions during neurogenesis.

Distal open regions marked with H3K27ac define cell-type specific transcriptional program

The acetylation of lysine 27 at histone 3 (H3K27ac) has been used as a mark to distinguish active versus inactive distal regulatory elements (hereafter referred to as "potential distal regulatory regions" when marked by H3K27ac) that function to primarily regulate the expression of proximal genes (Ref). To identify such regulatory elements, we generated H3K27ac ChIP-seq profiles for the three stages of neurogenesis, and extensive quality controls confirmed the high quality of the derived data (Supplemental Fig. S2A-E, Supplemental Table S2). Computational analysis revealed a large number of H3K27ac-enriched regions in each cell type (Supplemental Fig. S2F). Further investigation demonstrated that although the total number of H3K27ac sites remained similar, fully differentiated neurons gained more unique H3K27ac sites compared with NP and TND1 (Supplemental Fig. S2F, G). Furthermore, these unique peaks in neurons displayed significantly higher peak widths (Supplemental Fig. S2H). We also observed a gradual loss of H3K27ac enrichment at promoters (Supplemental Fig. S2I). We speculated that accessible H3K27ac regions that are targeted by regulatory factors could function as sites of high regulatory activity. Interestingly, we observed that the majority of FAIRE-identified open regions harbored H3K27ac within 1kb, for all three cell types but the fraction of such overlap was higher in differentiated neurons than dividing neuronal progenitors. (Fig. 2A, Supplemental Fig. S2J). Such occurrence of FAIRE positive sites with H3K27ac was also observed at a large number of stage-specific accessible regions (Fig. 2B and Supplemental Fig. S2K). Furthermore, reverse overlap of the H3K27ac sites with FAIRE revealed that many H3K27ac sites were not open and, consequently, may not be functionally active (Supplemental Fig. S2L-O). The genomic distribution of all double-positive sites indicated that a majority were located at promoters, least in exons followed by introns (Fig. 2C, Supplemental Fig. 2P). After promoters, intergenic regions contained the highest number of such sites, likely representing potential distal regulatory elements (Fig. 2C).

As H3K27ac also marks active promoters, we classified all accessible sites enriched in H3K27ac into two classes: promoters and non-promoters. Interestingly, sites that are H3K27ac-unique and also accessible were almost exclusively located in non-promoter regions, arguing that the unique transcription profiles of different cellular stages during neuronal differentiation may be largely determined by distal regulatory elements (Fig. 2D). Moreover, the gene sets located in proximity to these H3K27ac-unique and accessible sites were mostly non-overlapping (Fig. 2E). Expression analysis of these nearest genes demonstrated that approximately 40-60% of genes were up-regulated compared with the adjacent state (Fig. 2F, Supplemental Fig. S2Q). The genes that do not exhibit induced behavior in association with such distal elements may be wrongly assigned due to computational limitations and that the enhancer-nearest TSS pairing only holds for up to 40% of enhancers [158-160]. The correlation of unique distal H3K27ac-accessible sites with nearest gene expression was much higher in TND1 and TND10 than in NPs, further suggesting that distal gene regulation becomes more defined as cells reach a terminally differentiated state. We

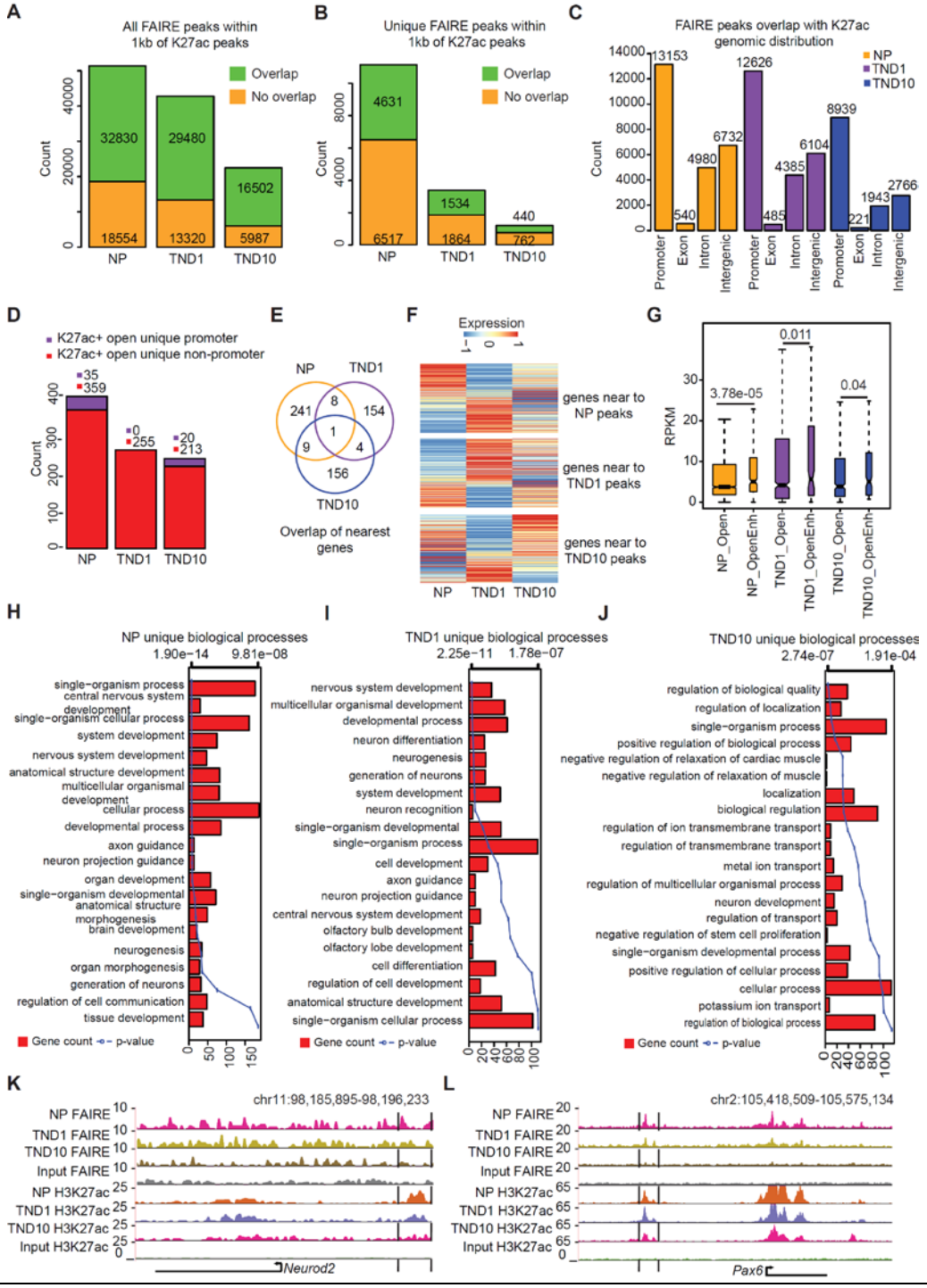


Figure 2. Distal regulatory elements define the cell type-specific transcriptome during neuronal development. (A) Stacked bar plot showing fraction of total FAIRE peaks overlapping with H3K27ac sites in NP, TND1 and TND10. Peaks were considered to be overlapping if they were within a distance of 1kb. **(B)** Same as in (A) but for unique peaks. **(C)** Bar plot showing the distribution of FAIRE and H3K27ac overlapping peaks in different genomic regions in NP, TND1 and TND10 stages. **(D)** Stack bar plot showing promoter and non-promoter peaks that are open and uniquely acquire H3K27ac in NP, TND1 and TND10.

(E) Venn diagram representing the overlap of genes near to the sites identified in (D). **(F)** Heat map showing expression of genes nearest to the peaks that uniquely gain H2K27ac in NP, TND1 or TND10. **(G)** Boxplot showing differences in expression for genes that are near to open H3K27ac sites (OpenEnh) and only open sites (Open). P-values are calculated using Wilcoxon test. **(H-J)** Bar plots showing enrichment of biological processes for genes near to NP, TND1 and TND10 open and unique H3K27ac positive sites. The bars reflect the number of genes in each category, and the lines represent the multiple tested corrected p-values, displayed as alternate x-axis, of the corresponding GO terms. **(K-L)** Browser tracks for *Neurod2* and *Pax6* showing potential distal regulatory regions. (Results in this figure partially contributed by me)

further observed that the genes that have nearby open and H3K27ac-positive regions exhibit significantly higher expression than genes that only have accessible regions without H3K27ac, suggesting that the presence of H3K27ac at these open sites has an enhancing effect on transcription (Fig. 2G), as previously suggested [143,144]. GO enrichment analysis showed that these genes are involved in neuronal development for NP (e.g. *Pax6* and *Neurod2*) and TND1 (e.g. *Robo1*), whereas TND10 (e.g. *Nlgn1*) genes were involved in more specialized neuronal functions such as ion transport, localization, regulation of transmembrane transport etc. (Fig. 2H-L). Furthermore, a comparison of TND10 unique H3K27ac sites with similar data from other cell-types confirmed that these sites are neuron-specific and not a general consequence of the differentiation process (Supplemental Fig. S2R). These findings suggest that distal accessible regions marked with H3K27ac contribute to the stage-specific transcriptional program during neurogenesis.

Distal regulatory elements recruit distinct transcription factors to regulate gene expression during neurogenesis

To gain understanding into the action of these distal regulatory elements, we next investigated whether they are enriched for binding sites of regulatory factors. Motif prediction analysis of these open and H3K27ac-positive sites revealed motifs for many known factors, a majority of which were identified in a stage-specific manner (Fig. 3A-C). Furthermore, these transcription factors were expressed in the predicted stage (Fig. 3A-C, line graph). Importantly, a number of these factors have been previously implicated in neuronal development (Supplemental Table S3). To test the targeting of predicted transcription factors at these distal sites, we chose two factors for each stage for which genome-wide

binding datasets were publicly available for matching stages of neuronal differentiation (SOX2 and SOX3 for NP, LHX3 and ISL1 for TND1 and CTCF and RFX1 for TND10) and were shown to be critical for the respective stages [161-167] (Fig. 3D-F). Comparative analysis of these ChIP-seq datasets with open and H3K27ac-positive sites revealed a high enrichment for these factors at distal regions in the respective stages (Fig. 3G-L). To provide ultimate validation of such transcription factor occupancy within our neuronal differentiation system, we performed ChIP assay for one transcription factor for each stage of neurogenesis (SOX2 for NP, ISL1 for TND1 and CTCF for TND10). These results show that indeed these three factors, viz. SOX2, ISL1 and CTCF, are significantly enriched at the identified distal regulatory elements in neuronal progenitors, early neurons and mature neurons respectively (Fig 3 M-O).. These findings also argue that the function of distal open sites involves a crosstalk of epigenetic mechanisms and critical transcription factors in defining the cell-type specific transcriptional profile during neuronal differentiation.

NMDA receptor activation results in massive transcriptional reprogramming in neurons

Our observations suggested a highly dynamic nature of epigenome and transcriptome during neurogenesis that becomes progressively more defined as neurons acquire a terminally differentiated, postmitotic identity. We were next curious to investigate whether the acquired epigenetic and transcriptional profile of neurons could be modulated in response to neuronal function such as neuronal activity. Neuronal activity is mediated via calcium-permeable receptors such as NMDA receptors (NMDARs). To study effects of a prolonged NMDAR stimulation, we treated differentiated neurons with the established agonist NDMA for six hours. We confirmed that such treatment did not to lead any cytotoxicity or apoptotic response (Supplemental Fig. S3A-C). Furthermore, while the NMDA treatment led to induction of classical plasticity genes in a manner similar to glutamate, this was effectively blocked by the NMDAR antagonist D-2-amino-5-phosphonovaleric acid (DAPV), suggesting the specificity of the

response (Supplemental Fig. S3D). Moreover, the established neuronal activity genes showed a dose-dependent modulation in both ESC-derived neurons and cortical neurons (Supplemental Fig. S3E, F).

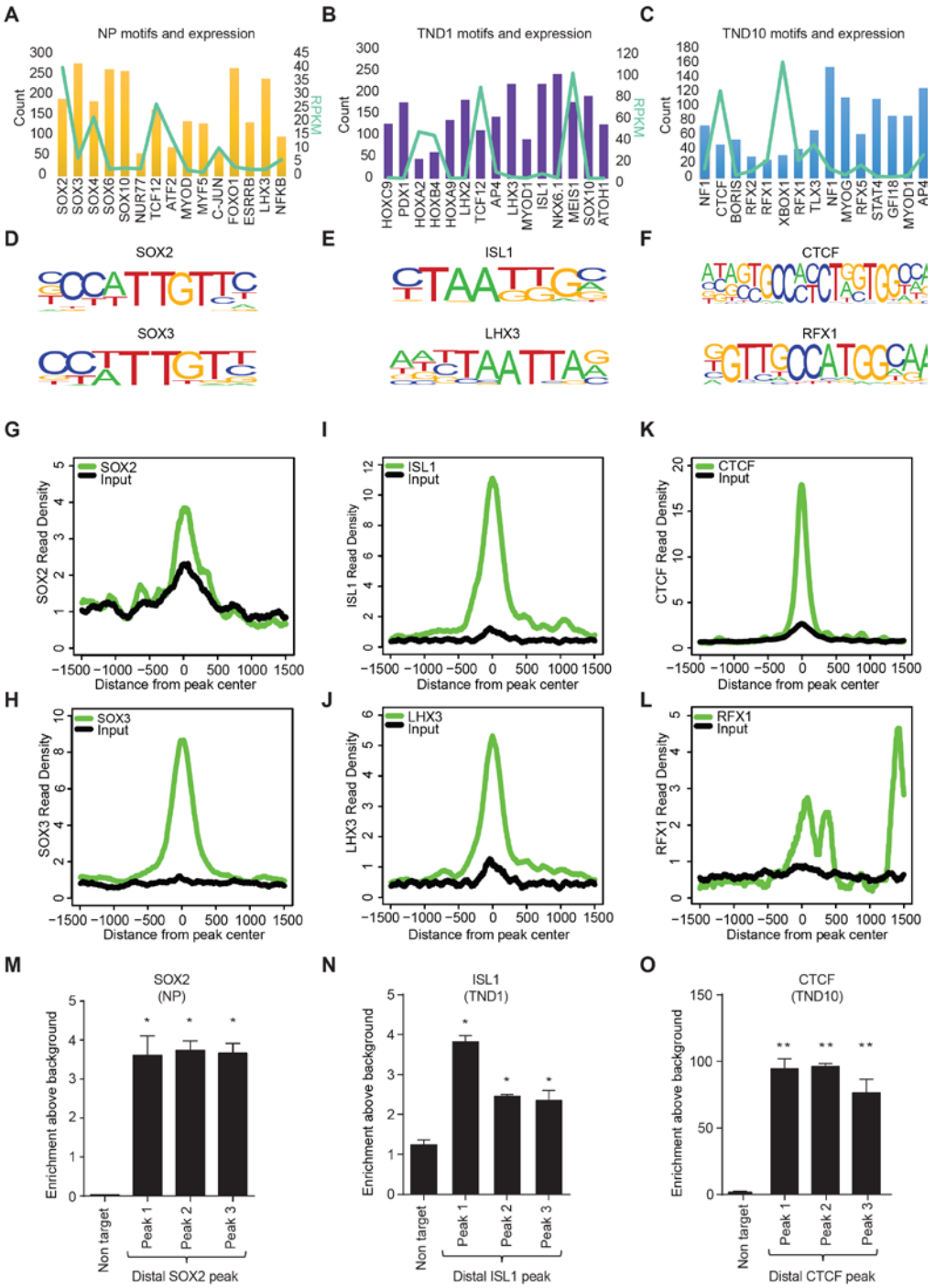


Figure 3. Distinct transcription factors are targeted to distal regulatory elements in a stage-specific manner during neurogenesis. (A-C) Bar and overlapping line plot showing identified motifs and their counts in NP, TND1 and TND10 at uniquely gained H3K27ac non-promoter sites (bars, main y-axis) and there expression (Line, alternate y-axis). **(D-F)** Motif sequence of the selected transcription factors from

the three stages. **(G-L)** Plots showing enrichment of the corresponding transcription factors around the center of the uniquely gained open H3K27ac peaks in NP (SOX2 & SOX3; G, H), TND1 (LHX3 & ISL1; I, J) and TND10 (CTCF & RFX1; K, L). The x-axis shows the distance from the peak center; the y-axis represents normalized enrichment of the corresponding transcription factor. **(M-O)** ChIP-qPCR validation of selected distal regions for their occupancy using specific antibody, SOX2 in NP (M), ISL1 in TND1 (N) and CTCF in TND10 (O). Average enrichments from independent assays are plotted on the y-axis as ratio of precipitated DNA (bound) to total input DNA and then further normalized to an intergenic region. Error bars represent SEM of independent biological replicates. * $p<0.05$, ** $p<0.01$, *** $p< 0.001$, Student's t-test. (Results in this figure partially contributed by me)

We next profiled the transcriptome (RNA-seq) of NMDA-treated neurons to illuminate the gene expression changes induced by the prolonged activation of NMDA receptors (Supplemental Table S4). Differential gene expression analysis comparing the control and NMDA-treated neurons revealed massive transcriptional reprogramming (Fig. 4A). Furthermore, we observed that many established neuronal activity and neuro-plasticity genes were also up-regulated (Fig. 4B-C). Surprisingly, this further accompanied down-regulation of many neuronal genes, particularly those related to synaptic transmission (e.g. *Syt2* and *Glra1*) (Fig. 4D). GO analysis of down-regulated genes revealed a dominant enrichment of genes involved in synaptic transmission and neuronal development (Fig. 4E-F). In contrast, up-regulated genes were enriched for processes related to extracellular matrix reorganization, protein localization and translation (Fig. 4G-H). Mouse phenotype enrichment analysis demonstrated that the defects in the down-regulated genes were almost exclusively associated with phenotypes such as abnormal learning/memory/conditioning and behavior, whereas those upregulated were dominantly linked to defects in the development of non-neuronal tissues (Supplemental Fig. S3G, H). The importance of long intergenic non-coding RNAs (lincRNAs) in neuronal development has been increasingly appreciated [168]. Our analysis further revealed that NMDA-driven neuronal activity also involves modulation of large number of lincRNAs (Supplemental Fig. S3I). Furthermore, the expression of the genes nearest to these lincRNAs was also altered under these conditions, suggesting a potential function of these lincRNAs in gene regulation in *cis* (Supplemental Fig. S3J-L). The genes potentially down-regulated by

lincRNAs were enriched for GO terms related to synaptic transmission and neuronal activity while up-regulated genes were dominantly enriched for many non-neuronal genes (Supplemental Fig. S3M-N).

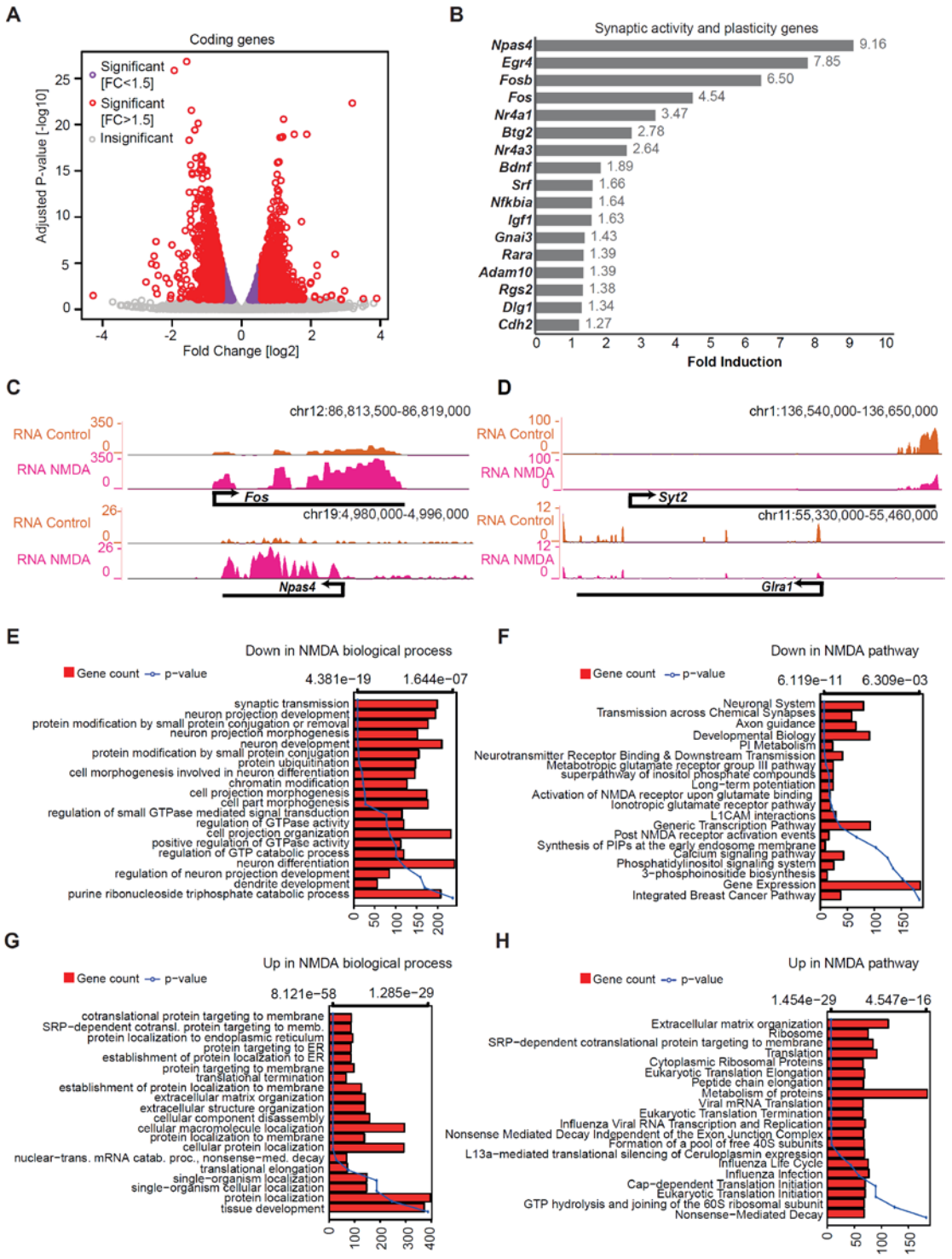


Figure 4. NMDA receptor signaling causes dramatic transcriptional changes. (A) Volcano plot showing up- and down-regulated genes upon NDMA treatment in TND10 neurons. The x-axis represents the fold change between control and NDMA in log2; the y-axis shows the adjusted p-value in -log10. **(B)** Bar plot showing fold indcution of few selected known synaptic activity and neuronal plasticity related genes. X-axis show gain of expression fold change upon NMDA stimulation in neurons. **(C-D)** Browser tracks showing gain (C) and loss (D) of expression for key activity or neuronal-related genes. (E-H) Bar plots showing enrichment of biological processes and pathways in downregulated (E-F) and upregulated (G-H) genes upon NDMA stimulation. The bars reflect the number of genes in each category; the lines represent the multiple tested corrected p-value, displayed as alternate x-axis, of the corresponding GO term. (Results in this figure partially contributed by me)

Induction of NMDA receptor pathway results in neuroplasticity and a neuroprotective response

Interestingly, a number of NMDA-induced genes have previously been shown to be important in NMDA-driven neuronal survival (e.g., *Btg2, Bdnf, Atf1*) [74,75,169,170]. In addition, this list also contained genes that, based on the literature, could play a role in cell proliferation but have not been implicated in cell survival, especially in the context of neurons (Fig. 5A). Such NMDA-induced expression of these genes was also observed in E16.5 cortical neurons (Fig. 5B). To establish the function of these genes in cell survival, we decided to assess cell death following their siRNA mediated knockdown in cortical neurons. Interestingly, the depletion of a majority of these genes showed a very strong effect on the viability of cortical neurons (Fig. 5C) which was further extensively validated in mouse epithelial cells (Fig. 5D; Supplemental Fig. S4A-C).

To gain further biological insights in the global gene expression changes induced by NMDA stimulation, we performed a Principal Component Analysis (PCA) based comparison of the DMSO and NMDA transcriptomes with the transcriptomes of several embryonic tissues from the three germ layers (ectoderm: ventricular zone (VZ), subventricular zone (SVZ) and cortical plate (CP) from mouse cortex; endoderm: lung and pancreas; mesoderm: heart and MEF). This analysis demonstrated that although control neurons were very similar to CP, NMDA treatment of neurons resulted in a transcriptome that was closer to cells from VZ and SVZ but not CP, further confirming a reversal toward the progenitor state (Fig. 5E). Further analysis showed that genes down-regulated upon NMDA treatment were generally

highly expressed in control neurons and CP and exhibited least expression in other tissues (Fig. 5F). By contrast, up-regulated genes were expressed at much higher levels in either VZ/SVZ or in other tissues compared with CP (Fig. 5G). Intrigued by these patterns, a closer gene-wise analysis of up- and down-regulated protein-coding genes further revealed that many established progenitor markers (e.g. *Pax6, Nes, Sox2, Notch1, Dll1* and *Hes1*) were significantly induced while known neuronal genes (e.g. *Ncam1, Syn1, Rbfox3, Dscam, Nlgn* and *Epha2*) were significantly down-regulated (Supplemental Fig. 5A, B). Such NMDA-induced upregulation of neuronal progenitor genes was further independently validated both in ESC-derived neurons and cortical neurons (Fig. 5H-I).

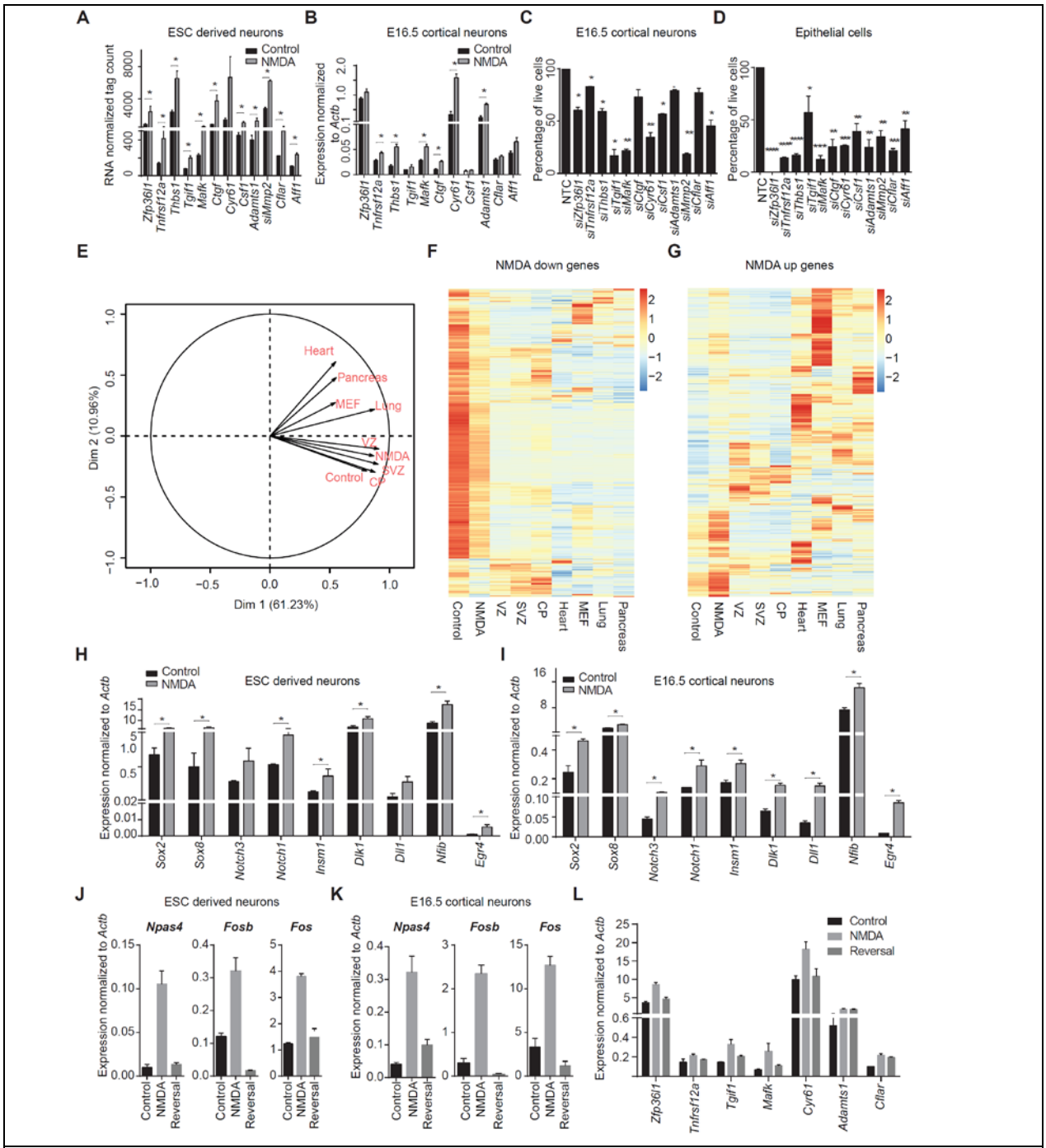


Figure 5. NMDA receptor stimulation causes induction of non-neuronal and pro-survival genes. **(A)** Expression values from RNA-seq (normalized tag counts) of selected potential pro-survival candidate genes that are induced upon NMDA. Error bars represent SEM of independent biological replicates. **(B)** mRNA levels for selected pro-survival genes (as in A) in NMDA induced cortical neurons were measured by qRT-PCR relative to *Actb* and plotted on the y-axis. Error bars represent SEM of independent biological replicates. **(C)** Bar plot showing percentage of live cells upon siRNA mediated knock down of pro-survival candidates as compared to non-targeting control (NTC) in cortical neurons. Error bars represent SEM of independent biological replicates. **(D)** Similar analysis as in (C) but knock down performed in epithelial

cells. Error bars represent SEM of independent biological replicates **(E)** Principal Component Analysis (PCA) plot of control- and NMDA-treated day-10 neurons with VZ (Ventricular Zone), SVZ (Sub-Ventricular Zone), CP (Cortical Plate), lung, pancreas, heart and MEF transcriptomes. **(F)** Heat map showing the expression of genes down-regulated by NMDA treatment in different tissues from 3 germ layers (ectoderm: VZ, SVZ and CP, endoderm: pancreas and lung, mesoderm: heart and MEF). **(G)** Same as in (F) but for up-regulated genes. **(H-I)** Bar plots showing mRNA levels for known neuronal progenitor genes in NMDA induced ESC-derived neurons (H) and cortical neurons (I) measured by qRT-PCR relative to *Actb* and plotted on the y-axis. Error bars represent SEM of independent biological replicates. **(J-K)** Bar plots showing mRNA levels of key synaptic activity genes upon NMDA treatment and following NMDA withdrawal both in ESC-derived neurons (J) and cortical neurons (K) measured by qRT-PCR relative to *Actb* and plotted on the y-axis. Error bars represent SEM of independent biological replicates. **(L)** Similar analysis as in (J) but for pro-survival genes. . * $p<0.05$, ** $p<0.01$, *** $p< 0.001$, Student's t-test. (Results in this figure partially contributed by me)

Since NMDA-dependent neuronal activity has been implicated to cause neuro-plasticity [171], we next probed whether our experimental setup recapitulates similar phenomenon. To investigate this, we performed NMDA withdrawal experiment on ESC-derived neurons as well as cortical neurons and measured the expression of classical neuronal plasticity genes. This analysis indeed showed that while NMDA stimulation results in a strong upregulation of neuronal plasticity genes (*Npas4, Fosb and Fos*), they were completely reverted back after NMDA withdrawal (Fig 5. J, K). Furthermore, similar behavior was observed for many neuroprotective genes (Fig. 5L). Altogether, these observations indeed establish that NMDA stimulation involves induction of neuroplasticity and neuro-protective response. These findings further imply that under prolonged neuronal activity via NMDA receptors, neurons transiently lose their identity and exhibit greater developmental potential, which may be essential to allow neuronal plasticity.

NMDA-induced neural activity results in a massive epigenetic reprogramming of distal regulatory elements

We had previously observed that in differentiated neurons, stage-specific distal regulatory regions marked with H3K27ac regulate the expression of genes involved in ion transport, localization, and the regulation of transmembrane transport- the functions that are known to be involved in neuronal activity (Fig. 2D and 2J). Furthermore, the gain of stage-specific active regulatory regions was found to occur

exclusively at distal sites, which in turn defines cell type-unique transcriptional programs during neuronal development (Fig. 2D). Although researchers have touched upon the distal regulation of basal neuronal activity, there are no reports studying how distal gene regulation is involved in transcriptional responses following prolonged neuronal activity via NMDA receptors [87,88]. To examine NMDA-dependent transcriptional changes we performed a ChIP assay using H3K27ac-specific antibody with control and NMDA-treated neurons and performed next-generation sequencing (ChIP-seq) (Supplemental Table S5). Our analysis revealed that NMDA treatment led to a substantial loss of H3K27ac peaks (n=16349) and a gain of lesser new peaks (n=4850) (Fig. 6A). Using stringent enrichment criteria, we further shortlisted genomic sites that strongly lost or gained H3K27ac enrichment (n=6571, down-regulated; n=1047, up-regulated) (Fig. 6B). Although decrease in enrichment was significant among all peaks, a dramatic reduction was observed at unique peaks (Fig. 6C). Deeper analysis did not reveal major changes in the percentage of H3K27ac sites distributed across promoters, exons, introns and intergenic regions following NMDA treatment (Fig. 6D, E). Interestingly, we observed that, upon NMDA stimulation, introns and intergenic regions exhibited most prominent reprogramming of H3K27ac mark (Fig. 6F-G). The nearest genes to these sites included known neuronal and synaptic-activity modulated genes (e.g., *Dscam* and *Epha10* which lose H3K27ac and expression upon NMDA treatment, and *Npas4* and *Gpm6a*, which gain H3K27ac and transcription upon NMDA treatment) (Fig. 6H-I). Comparison of epigenome and transcriptome dynamics between neuronal activity versus neurogenesis (NP to TND10) revealed that decreased H3K27ac peaks or downregulated genes during neuronal activity had more overlap with increased H3K27ac peaks or upregulated genes during neurogenesis and vice-versa (Supplemental Fig. 6A-C). GO enrichment analysis of these overlapping set of genes further supported our earlier observations that NMDA-stimulated neurons transiently lose their identity (Supplemental Table S6). We also found that a large fraction of these activity-modulated H3K27ac positive distal regulatory regions were also active *in vivo* during murine brain development [172] (Supplemental Fig. S6D, E).

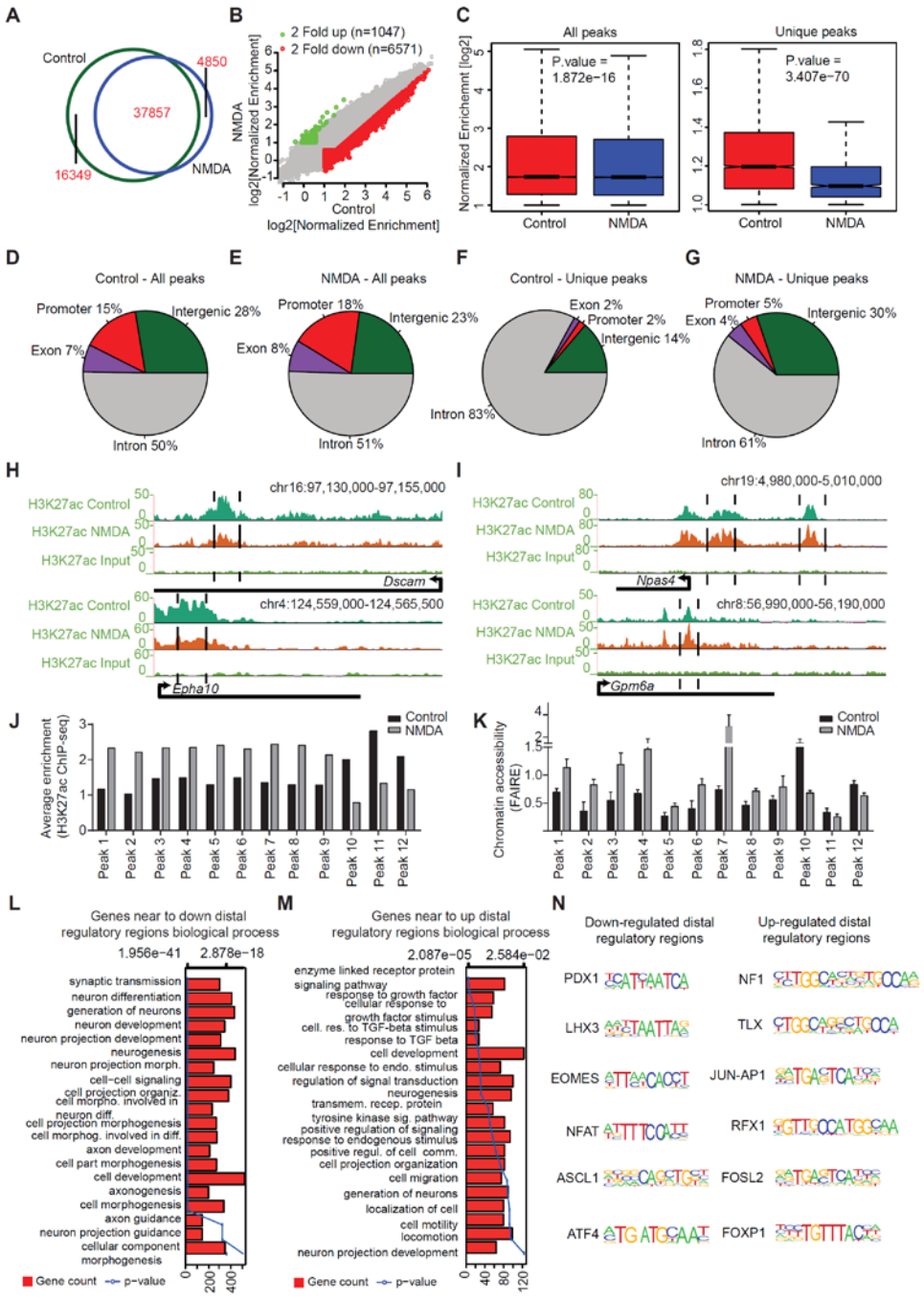


Figure 6. Prolonged NMDA activity results in a massive modulation of distal regulatory landscape.

(A) Venn diagram showing overlap of H3K27ac peaks identified in control- and NDMA-treated day-10 neurons. **(B)** Scatter plot showing the gain and loss of H3K27ac sites. Peaks marked in green are either only present in NMDA or at least 2-fold up-regulated compared with control; peaks marked in red represent the opposite. The x- and y-axes represent the normalized enrichment of H3K27ac in control and NDMA treated neurons. **(C)** Box plots showing the loss of H3K27ac upon NMDA treatment compared with control at all and unique peaks. P-values are calculated using Wilcoxon test. **(D-G)** Pie charts showing the distribution of control or NMDA peaks in different genomic regions for all (D, E) and unique (F, G) peaks. **(H)** Browser plots showing loss of H3K27ac at potential distal regulatory regions of neuronal

genes (*Dscam* and *Epha10*). **(I)** Same as in (H) but showing gain of H3K27ac at potential distal regulatory regions of activity-related genes. **(J)** Bar plot showing normalized H3K27ac enrichment of selected distal peaks in control and NMDA treated neurons as derived from the ChIP-seq dataset. **(K)** FAIRE assay was performed in cells treated with NMDA and qRT-PCRs were performed for the same distal regulatory regions as in (J). Average enrichments from independent assays are plotted on the y-axis as ratio of precipitated DNA (bound) to total input DNA and then further normalized to an intergenic region. Error bars represent SEM of independent biological replicates. * p<0.05, ** p<0.01, *** *p*< 0.001, Student's t-test. **(L-M)** Bar plots showing the enrichment of biological processes for genes near down-regulated (L) and up-regulated (M) distal regulatory regions. The bars reflect the number of genes in each category; the lines represent the multiple tested corrected p-value, displayed on alternate x-axis, of the corresponding GO term. **(N).** Representative motifs identified in down- and up-regulated distal regulatory regions. (Results in this figure partially contributed by me)

In line with our previous observations that the distal accessible chromatin sites are largely enriched with H3K27ac in neurons (Fig. 2A), we found that all tested distal regions that gained H3K27ac following NMDA stimulation also gained chromatin accessibility and *vice versa* (Fig. 6J-K). A GO analysis of genes in proximity to distal regulatory sites modulated via NMDA treatment showed that

the genes nearest to those loosing enrichment were dominantly enriched for synaptic transmission and neurogenesis (Fig. 6L) while genes in closest proximity to sites gaining H3K27ac were involved in signaling pathways and cell motility (Fig. 6M). Furthermore, a motif enrichment analysis of these regions revealed their enrichment for binding sites for a number of transcription factors that are known to be critical for neurogenesis (Fig. 6N and Supplemental Fig. 6F, G). Overall, these findings establish that prolonged NMDA activity involves dramatic epigenetic reprogramming at distal regulatory elements.

NMDA-mediated epigenetic reprogramming of distal regulatory elements functions in gene regulation

We next attempted to determine whether distal regions that exhibit changes in H3K27ac in response to NMDA activity correlate with expression changes in the nearest genes. To address this question, we created bins of genes based on the distance from these distal regions and plotted their expression levels under control and NMDA-treatment conditions. In line with previous observations, nearby genes displayed a greater decrease or increase in expression when distal regions exhibited a loss or gain of

H3K27ac respectively (Fig. 7A-D; Supplemental Fig. S7A, B). Importantly, these genes included a number of genes that were differentially expressed by at least 1.5-fold upon NMDA treatment (n=538 genes, down-regulated; n=173, up-regulated) (Fig. 7E-G). These down-regulated genes were again enriched for synaptic transmission and ion transport, whereas up-regulated genes were enriched in extracellular matrix reorganization (Supplemental Fig. S7C, D). We further independently validated that chromatin changes at distal regulatory elements accompany changes in the expression levels of associated genes both in ESC-derived neurons and cortical neurons derived from E16.5 brain (Fig. 7H-K, Supplemental Fig. S7E).

We next attempted to assess the downstream epigenetic machinery involved in altering chromatin landscape and gene expression in response to NMDA stimulation. To address this, we employed C646, a specific inhibitor of the histone acetyltransferase (HAT) EP300 (also known as p300)/CREBBP (also known as CBP) [173] and analyzed distal regulatory elements for certain genes that showed increase in H3K27ac upon NMDA treatment following treatment with NMDA in the presence and absence of C646. These results reveal that for NMDA-induced distal H3K27ac peaks, the histone acetyltransferase (HAT) EP300/CREBBP activity is required for a gain of H3K27ac enrichment in response to NMDA treatment (Fig 7L). In addition, such depletion of H3K27ac enrichment further accompanied loss of chromatin accessibility at all measured distal regulatory elements (Fig. 7M). Furthermore, this also accompanied downregulation in transcription of genes nearest to these distal sites (Fig. 7N). Taken together, these results suggest that NMDA pathway employs distinct epigenetic players such as histone acetyltransferase (HAT) EP300/CREBBP at distal regulatory elements to govern epigenetic state and transcriptional dynamics following NMDA receptor activation.

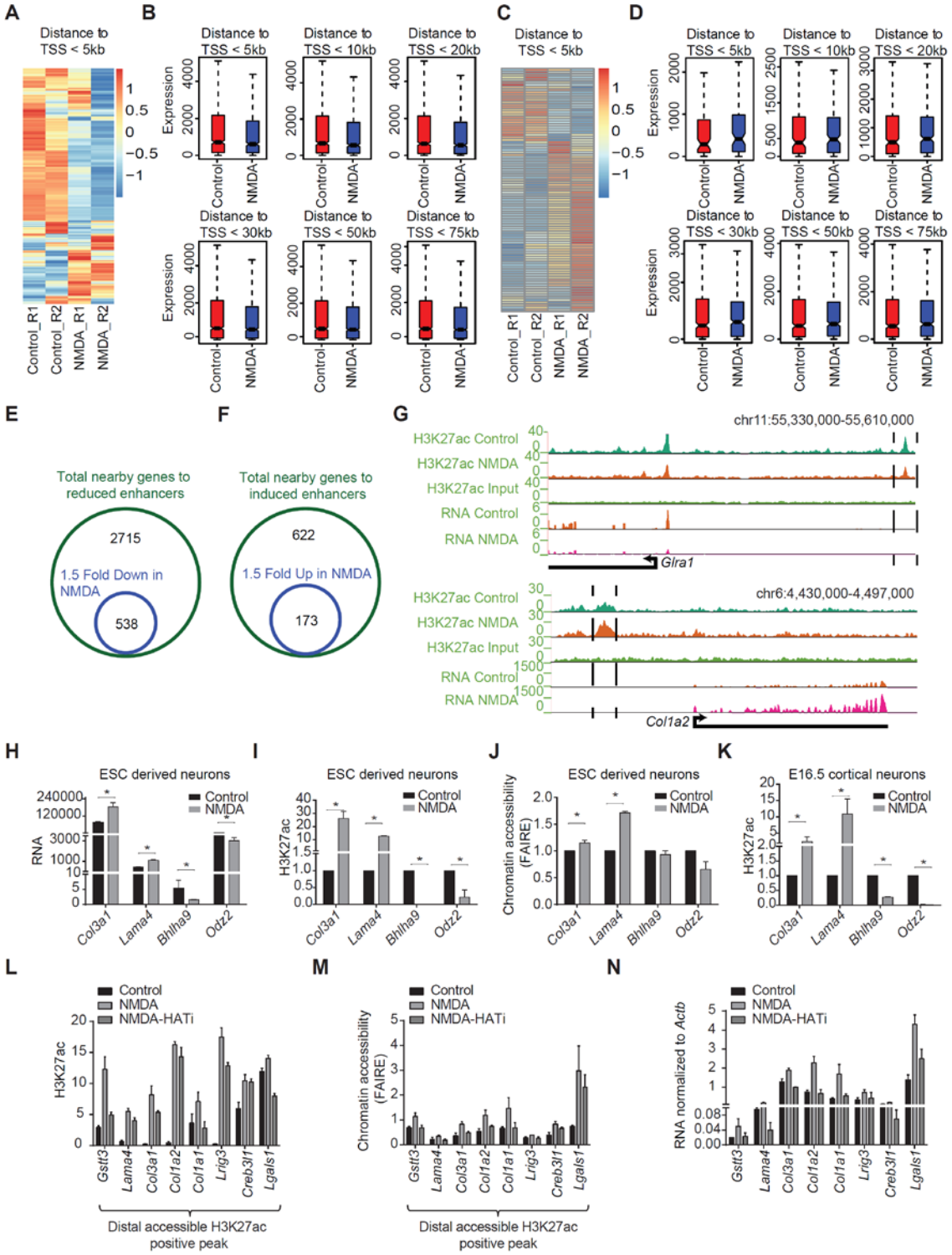


Figure 7. Neuronal activity-dependent transcriptome changes are driven via distal regulatory elements. **(A)** Heat map showing the expression of genes (within 5 kb window) near to down-regulated H3K27ac positive distal regulatory regions in the control and NMDA treatments. **(B)** Box plots showing differences in the expression of genes near down-regulated H3K27ac positive distal regulatory regions after the genes were binned based on their distance (5kb, 10kb, 20kb, 30kb, 50kb and 75kb) from these sites. P-values for the differences between control and NMDA conditions for distances 5kb (n=911), 10kb (n=1076), 20kb

(n=1367), 30kb (n=1621), 50kb (n=1970) and 75kb (n=2252) are 1.5e-54, 5.7e-56, 3.4 e-51, 1.4e-55, 1.5e-54 and 4.4e-56 respectively. P-values are calculated using Wilcoxon test. **(C)** Same as in (A) but for up-regulated H3K27ac positive distal regulatory regions. **(D)** Same as in (B) for up-regulated H3K27ac positive distal regulatory regions. P-values for the differences between control and NMDA conditions for distances 5kb (n=97), 10kb (n=171), 20kb (310), 30kb (406), 50kb (528) and 75kb (n=608) are 0.0005, 0.05, 0.04, 0.13, 0.56 and 0.91 respectively. **(E-F)** Venn diagram showing overlap of genes near to down-regulated distal regulatory regions (green circle) with genes at least 1.5 fold down-regulated upon NMDA treatment (blue circle) (E) and up-regulated distal regulatory regions (green circle) with genes at least 1.5 fold up-regulated (blue circle) (F). **(G)** Browser plots showing concomitant loss (*Glra1*) or gain (*Col1a2*) of distal H3K27ac and expression. **(H)** Normalized tag counts from RNA-seq data of selected genes that show changes in H3K27ac at their distal regions upon NMDA treatment. **(I)** ChIP-qPCR validation of NMDA-dependent H3K27ac changes for the distal regions for genes shown in (H). **(J)** FAIRE-qPCRs to asses changes in chromatin accessibility at the same regions shown in (I) upon NMDA treatment. **(K)** ChIP-qPCR validation of NMDA-dependent H3K27ac changes for the distal regions of genes shown in (H) in mouse cortical neurons treated with NMDA. **(L-N)** Bar plots showing analysis of H3K27ac (by ChIP-qPCRs) (L), chromatin accessibility (by FAIRE-qPCRs) (M) of selected distal regions and expression of nearest genes (by RT-qPCRs) (N) following NMDA stimulation in the presence and absence HAT inhibitor (C646). All error bars represent SEM of independent biological replicates. * $p<0.05$, ** $p<0.01$, *** $p< 0.001$, Student t-test. (Results in this figure partially contributed by me)

NMDA-induced neural activity involves a dramatic reorganization of super-enhancers

Recent studies have demonstrated that super-enhancers are large clusters of transcriptional enhancers that drive the expression of genes defining cell identity [145,146]. Prompted by our observations that NMDA-mediated gene expression responses involve epigenetic reprogramming of distal regulatory elements, we next investigated whether the NMDA response involves the reorganization of super-enhancers. Surprisingly, despite the massive loss of classical H3K27ac marked distal regulatory regions, we observed a substantial increase in the number of super-enhancers upon NMDA treatment (Fig. 8A). Furthermore, when we assigned these super-enhancers to nearby genes, we observed that few genes lost super-enhancers (n=113), whereas a larger number of genes gained super-enhancers (n=295) following NMDA treatment (Fig. 8B). Among the genes in proximity to control- or NMDA-specific super-enhancers were a number of genes that are known to be crucial for neuronal identity (e.g. *Ncam1, Mapt, Rbfox3*, control super-enhancers) or neuronal activity (*Fos, Per1* and *Ephb2,* NMDA super-enhancers), respectively (Supplemental Table S7). Comparing the loss and gain of super-enhancers with the transcription changes

following NMDA treatment demonstrated that the substantial number of control- or NMDA-specific super-enhancers were associated with higher expression levels of the associated genes in the respective states (Fig. 8C, D). Considering the expression of nearby genes in the three layers of the embryonic cortex and representative tissues from all three lineages, we observed that the majority of genes associated with super-enhancers that were present in control-treated neurons but lost after NMDA treatment were almost exclusively expressed in the cortical plate (CP), where mature neurons reside (Fig. 8E).

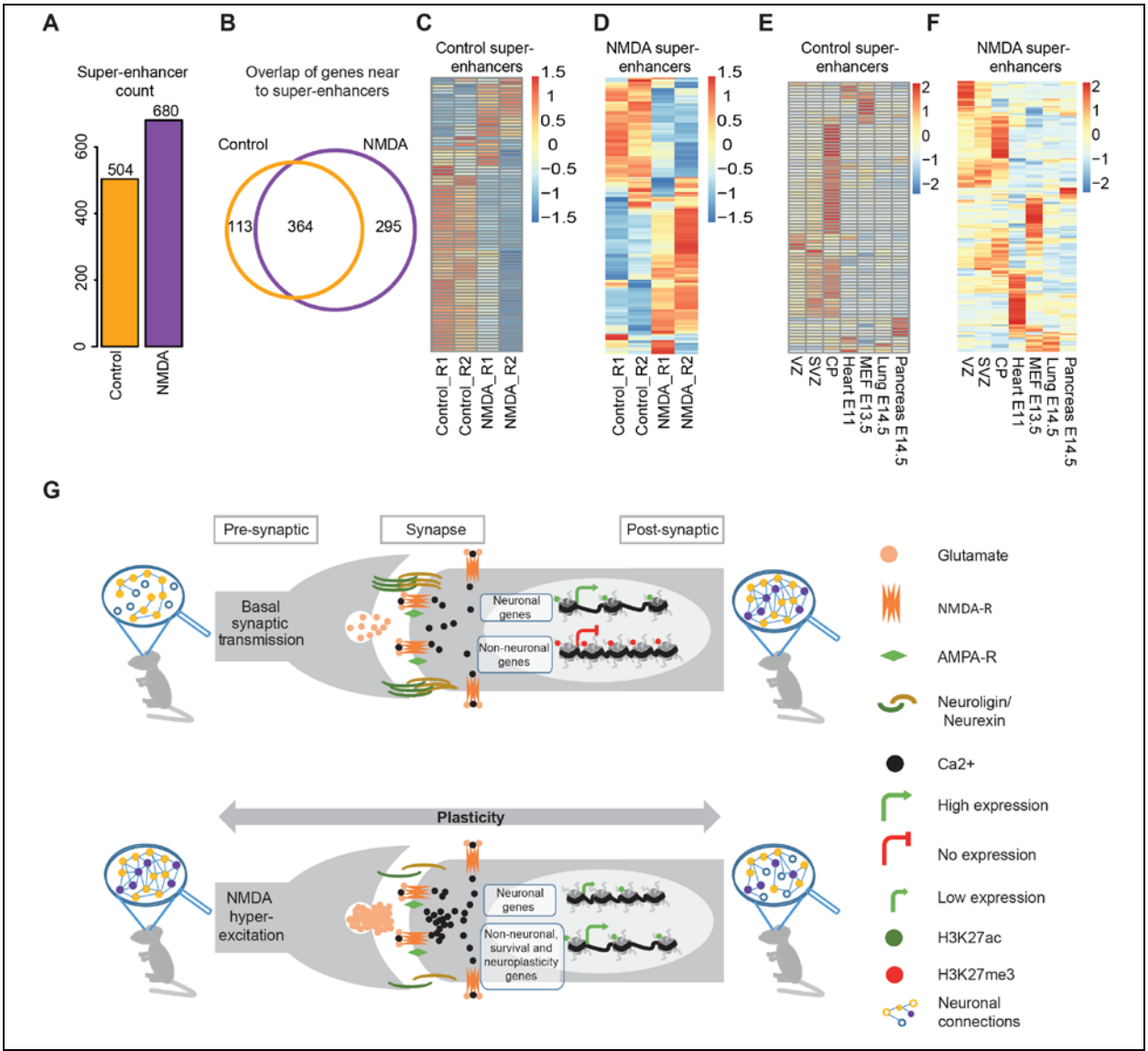


Figure 8. Neuronal activity results in a global reorganization of super-enhancers at key responsive genes. (A) Bar plot showing the number of super-enhancers in the control and NMDA treated neurons. **(B)** Venn diagram showing the overlap of genes near control and NMDA super-enhancer. **(C-D)**

Expression of genes near to control (C) and NDMA (D) non-promoter super-enhancers in control- and NDMA-treated neurons (Enlarged heat maps along with gene names are provided as Supplemental Figure S9). **(E-F)** Heat map showing expression of same genes as in (C) and (D) in tissues from three germ layers. **(G)** Schematic representation of a model showing epigenome and transcriptome changes in response to sustained NMDA receptor activity. Neurons exhibiting basal synaptic activity express neuronal genes that are marked by an active chromatin state while non-neuronal genes are kept silent in heterochromatin. A prolonged NMDA activity results in a dramatic epigenetic reprogramming including at many H3K27ac positive distal regulatory elements and super-enhancers resulting in the downregulation of neuronal genes (loss of H3K27ac) and activation of neuronal plasticity, neuroprotective and several non-neuronal genes (gain of H3K27ac). (Results in this figure partially contributed by me)

Interestingly, while many genes associated with super-enhancers acquired exclusively following NMDA treatment were expressed in the cortex, a major fraction of these genes were also expressed in other tissues (Fig. 8F). Accordingly, super-enhancers identified in control neurons were enriched with genes related to neurogenesis (e.g. *Ncam1, Mapt, Rbfox3*) (Supplemental Fig. 8A), whereas genes associated with NMDA-induced super-enhancers were related to general transcription or metabolic regulation and included many established neuronal activity up-regulated genes (e.g. *Fos, Per1* and *Ephb2*) (Supplemental Fig. 8B; Supplemental Table S7). These findings suggest that prolonged NMDA activity causes a massive reorganization of super-enhancers, which then mediate changes in the expression of critical genes.

DISCUSSION

Differentiation is characterized by sequential changes in the cellular state specified by transcriptional reprogramming; that is believed to be generated by a combinatorial action of epigenetic mechanisms and transcription factors. Despite progress, we lack a comprehensive understanding of how specific genes are precisely turned on and off during neuronal development and how regulatory sequences contribute to this process. Recent evidence has indicated a crucial role for distal regulatory elements in defining a cell type-specific gene expression program during development. Using FAIRE-seq assay here we have revealed genome-wide remodeling of chromatin accessibility during successive stages of neurogenesis. These data reveal that the H3K27ac marked distal regulatory regions are uniquely gained in each stage of differentiation and function to drive the expression of genes that define cell identity. We further demonstrated that, in addition to chromatin competence with respect to the presence of the H3K27ac marker, these distal elements recruit distinct transcription factors that are known to be critical for the respective stages of neuronal differentiation. These findings further highlight a complex crosstalk among epigenetic mechanisms, DNA sequence and transcription factors in determining the potential of distal regulatory elements that in turn define the transcriptome underlying cellular identity. Furthermore, such combinatorial analysis of chromatin accessibility (FAIRE-seq) and H3K27ac (ChIP-seq) has led to the first high-sensitivity map of regulatory elements for progressive stages of neuronal differentiation.

The identification of open chromatin regions using FAIRE-seq revealed that accessible chromatin is rapidly remodeled during neurogenesis. We observed that dividing neuronal progenitors have more open chromatin than post-mitotic neurons, reflecting genome compaction upon differentiation. This result also suggests that a large number of accessible sites from differentiated neurons may pre-exist in neuronal progenitors, arguing for 'epigenetic priming'-like phenomenon that later gains transcriptional competence upon the induction of neurogenesis. This suggestion was further supported by our observations that promoter accessibility correlates more strongly with gene expression in post-mitotic

neurons than dividing neuronal progenitors. Alternatively, a certain fraction of the openness detected in dividing neuronal progenitor cells may be contributed by DNA replication while in post-mitotic neurons the open chromatin state may be exclusively related to transcriptional activity. These observations were further extended to distal regulatory elements because the fraction of open distal regions enriched with the active mark H3K27ac was much higher in post-mitotic cells than dividing neuronal progenitors. Overall, these observations suggest that neurogenesis accompanies establishment of a more structured landscape of regulatory elements and chromatin accessibility that function primarily in gene regulation in post-mitotic, terminally differentiated neurons.

Enhancers are distal regulatory elements that function to activate gene expression and are primary drivers of tissue-specific gene expression programs. Using our high-resolution map of chromatin accessibility and H3K27ac, which marks active enhancers, we illustrated the dynamic changes in usage of regulatory elements during neurogenesis. Interestingly, genes in close proximity to stage-specific non-promoter accessible H3K27ac sites were non-overlapping between the consecutive stages of neuronal differentiation (NP, TND1 and TND10) and gained expression in the stage associated with H3K27ac. A GO enrichment analysis for these genes revealed involvement in biological processes and pathways that reflects the cellular identity of each of the three cell types. These findings further argue for a critical importance of distal gene regulation in defining cell-fates during neurogenesis.

We were next tempted to ask whether the acquired epigenetic and transcriptional state in neurons is subject to modulation during neuronal function such as neuronal activity. The activity-dependent plasticity of vertebrate neurons allows the brain to respond to external stimuli. Neuronal activity-dependent LTP and LTD impart activity-driven neuronal plasticity. Basal neuronal activity or LTP results in the formation of synapses and thus creates memories, but after prolonged neuronal activity, LTD is required to rewire old synapses to allow the acquisition of new learning. This study unravels many novel aspects of neuronal activity induced by NMDA receptors and demonstrates that prolonged NMDA activity

results in significant epigenetic remodeling of the distal regulatory elements to mediate the required gene expression responses.

Here we provide the first comprehensive map of all coding and noncoding genes that are transcriptionally modulated in response to sustained NMDA receptor activity. The genes downregulated following NMDA receptor stimulation included many neuron-specific genes while those induced contained extracellular matrix genes and many other genes expressed in non-neuronal lineages in addition to the classical neuronal activity genes (Fig. 8G). This result is also in agreement with a previous study demonstrating that the NDMA response leads to a major re-organization of the extracellular matrix [74]. Interestingly, such NMDA receptor activity driven transcriptome resembled closer to neuronal progenitors than to neurons and showed induction of classical neuronal progenitor genes (e.g. *Pax6, Sox2*) and downregulation of hallmark neuronal genes (e.g. *Mapt, Rbfox3*). Importantly, transcription of neuronal plasticity genes was completely reversed soon after the NMDA withdrawal, suggesting a true neuronal plasticity in response to neuronal activity. These findings suggested that sustained activity of NMDA receptor results in a transient loss of neuronal identity and we propose that this may be essential to allow cellular and chromatin plasticity. Future studies should aim to investigate the relevance of this phenomenon in the *in vivo* contexts of such neuronal activity. Furthermore, given generation of such 'plastic' cellular state in response to neuronal activity, it would be interesting to assess whether this condition predisposes neurons for reprogramming to other cell types.

Furthermore, this study also discovered for the first time a number of novel genes that were transcriptionally induced by NMDA and function in neuronal survival (Fig. 8G). In addition, NMDA-induced genes also included several genes established to be important in NMDA-driven neuronal survival (e.g. *Btg2, Bdnf, Atf1*) [74,75,169,170]. Such induction of pro-survival genes upon neuronal activity may reflect a neuroprotective response (Fig. 8G) and envisage the possible protective effects of neuronal activity against neurodegeneration. It would be interesting to explore in future whether these novel pro-survival

genes could offer therapeutic avenues against neurodegenerative diseases such as Parkinson's, Alzheimer's or Huntington's disease.

Current literature suggests that basal neuronal activity results in a gain of H3K27ac at distal sites and assists memory formation [74,87,139,169,170,174,175]. Therefore, any LTD event that results in a selective loss of synaptic strength may also involve the erasure of memory stored on chromatin during basal activity. We observed that prolonged NMDA activity results in a massive loss of potential distal regulatory regions, whereas only a limited set of potential distal regulatory regions were gained. The genes in close proximity to these lost distal regulatory regions were dominantly enriched in neuronal genes such as those involved in neuronal transmission and neuronal development (Fig. 8G). Furthermore, such NMDA-induced lost and gained distal H3K27ac sites showed opposite dynamics during neurogenesis (upon transition from NP to TND10), further suggesting that a sustained neuronal activity also generates an epigenome that is closer to the neuronal progenitor state and reflects a partial de-differentiation.

In contrast to these classical H3K27ac marked distal regulatory regions that were mostly lost upon NMDA stimulation, such treatment led to a massive reorganization of super-enhancers that were both lost and gained. Furthermore, genes gaining super-enhancers following NMDA treatment included key neuronal activity genes while genes that lose super-enhancers contained many neuronal genes (Fig. 8G). These observations also imply that NMDA receptor signaling also functions to reprogram critical regulatory regions such as super-enhancers, which have been shown to drive expression of genes that define cell identity [145-147] and thus may have a profound effect on the cell fate. We also observed that the activation of NMDA pathway results in the recruitment of histone acetyltransferase (HAT) EP300/CREBBP at distal regulatory elements that then acetylates Histone H3 at Lysine 27 to result in the observed chromatin accessibility of these distal sites as well as transcriptional activation of the nearest genes.

Taken together, our findings reveal the dynamics of distal regulatory elements and a vast trove of potential distal regulatory regions and super-enhancers that function in determining transcriptome

underlying stage-specific cellular identity during neurogenesis and in response to NMDA-induced neuronal activity. It is very exciting to uncover how the distal regulatory landscape is actively remodeled during neuronal development as well as upon activation of the NMDA receptor pathway to mediate neuroplasticity and neuroprotective responses (Fig. 8G). Future work should aim to unravel the mechanistic details of the collaborative partnership between transcription factors and epigenetic mechanisms at distal regulatory regions and how this partnership contributes to the gene expression program that underlies neuronal development and activity.

METHODS

Cell culture

Wild-type embryonic stem cells derived from blastocysts (3.5 PC) from a mixed 129-C57BL/6 background (called 159.2) were cultured and differentiated as previously described [152]. A subclone of NMuMG cells has been described previously and was grown in DMEM supplemented with 10% FBS, 2 mM L-Glutamine and 1X Non-essential amino acids [176]. All cells were grown in 7% CO2 at 37 degrees.

Cortical neuron culture

Cortical neurons were cultured as described previously [87]. Please refer to supplemental methods for more details.

FAIRE assay

FAIRE-assay was performed as described previously [148]. Please refer to supplemental methods for more details.

ChIP assay

ChIP experiments were performed as described previously [154], starting with 60 µg of chromatin and 5 µg of anti-H3K27ac (ab4729; Abcam) antibodies. ChIP-real time PCR was performed using SYBR Green chemistry (ABI). Primers used for ChIP-qPCRs are listed in Supplemental Table S8.

Neuronal activation

Neurons were washed three times with BSS (balanced salt solution: 25 mM Tris, 120 mM NaCl, 15 mM glucose, 5.4 mM KCl, 1.8 mM $CaCl_2$, 0.8 mM $MgCl_2$, pH 7). After treating the neurons with 200 µM NMDA (Sigma) or control only for 6 h, cells were processed for RT-qPCR, ChIP and FAIRE analysis or washed three times with BSS and fresh neuronal medium added for other downstreal experiment. For recovery cells were incubated 72 hours in neuronal medium without NMDA.

Inhibition of the histone acetyltransferase (HAT) EP300/CREBBP

Neurons were pre-treated with the histone acetyltransferase (HAT) EP300/CREBBP inhibitor (10 µM, C646, Sigma) for 30 minutes and then NMDA treatment was performed as described above but in the presence of C646.

Next Generation Sequencing data analysis

Details of genomics data analysis including RNA-seq, ChIP-seq and FAIRE-seq are provided as supplemental methods.

Data Access

FAIRE-seq, ChIP-seq and RNA-seq data generated in this study data have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE65713.

ACKNOWLEDGEMENTS

We thank members of the Tiwari lab for support and critical feedback during the project. Support by the Core Facilities of the Institute of Molecular Biology (IMB), Mainz, is gratefully acknowledged, particularly the Genomics Core Facility and the Bioinformatics Core Facility. We thank Prof. Jason Lieb (The University of Chicago) for providing protocols and critical comments during the study. Research in the laboratory of V.K.T. is supported by the Deutsche Forschungsgemeinschaft (DFG) Grant TI 799/1-1, Marie Curie CIG 322210, EpiGeneSys RISE1 program and Wilhelm Sander Stiftung 2012.009.2.

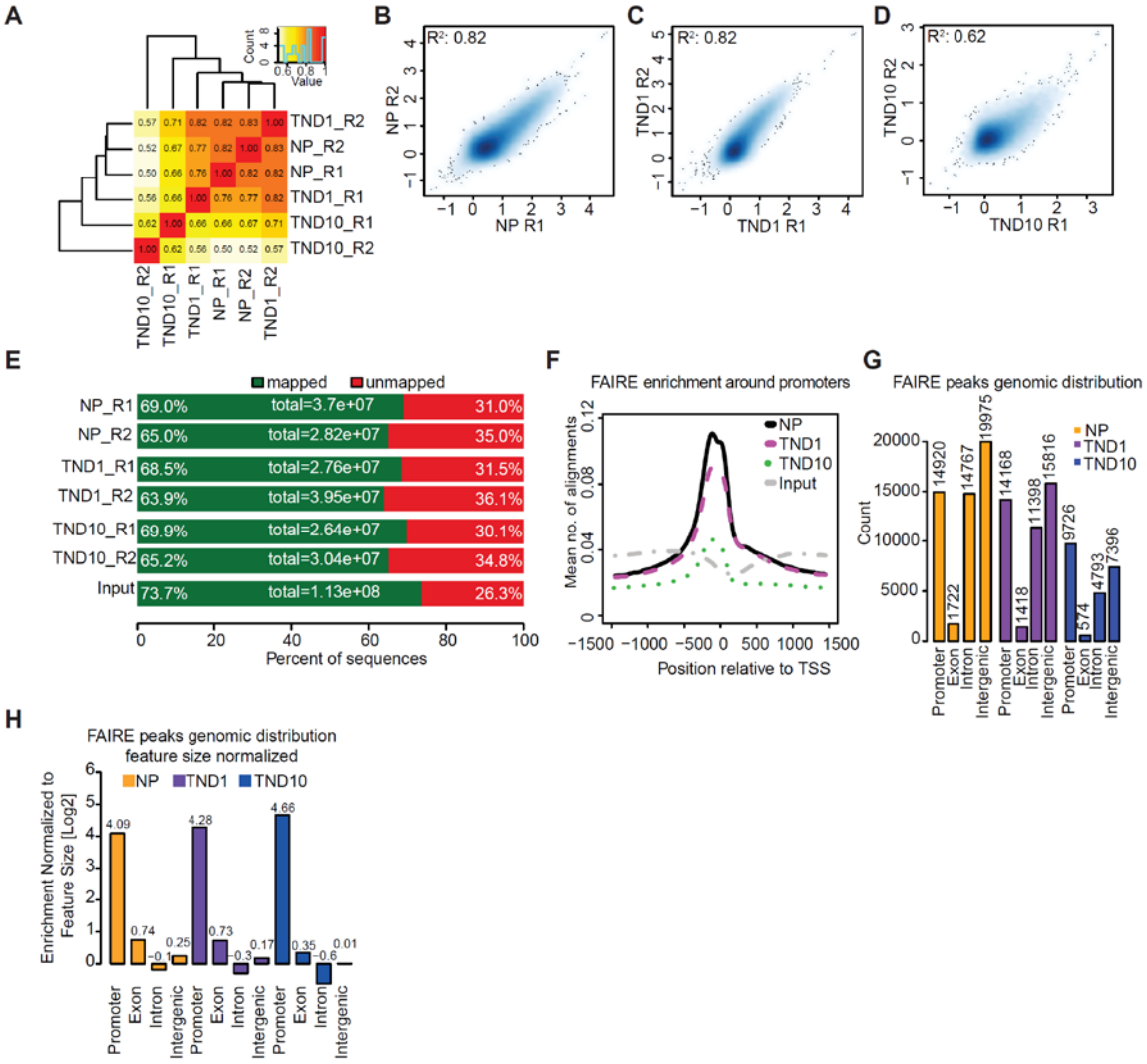
AUTHOR CONTRIBUTIONS

S.T. initiated and conceptually designed the study, developed analytical ideas, analyzed data and wrote the manuscript. S.K.S. performed experiments, analyzed data and wrote the manuscript. A.G. performed experiments. V.T. designed the study, analyzed data and wrote the manuscript.

CONFLICT OF INTEREST

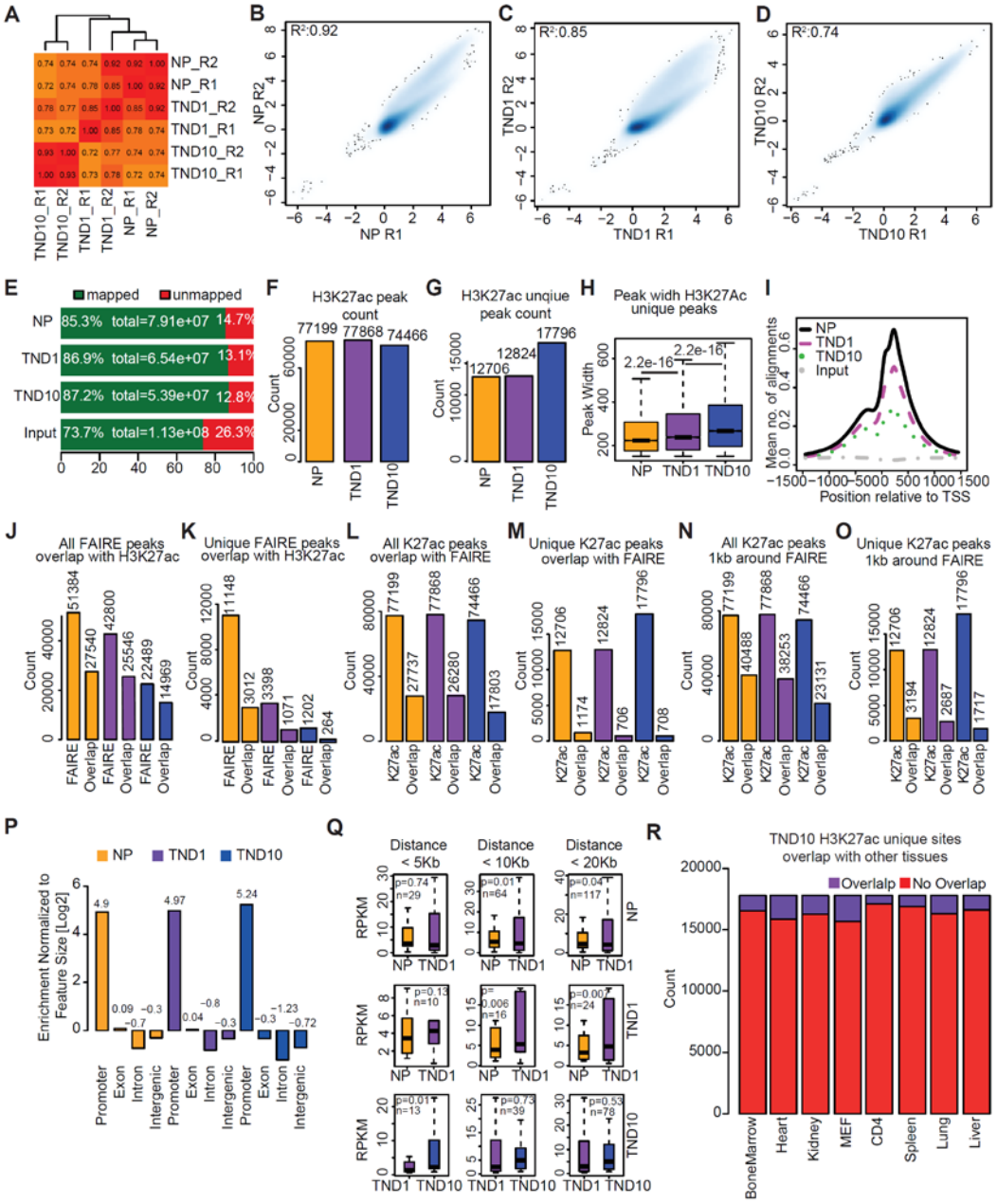
The authors declare no competing financial interests.

SUPPLEMENTAL FIGURES



Supplemental Figure S1.

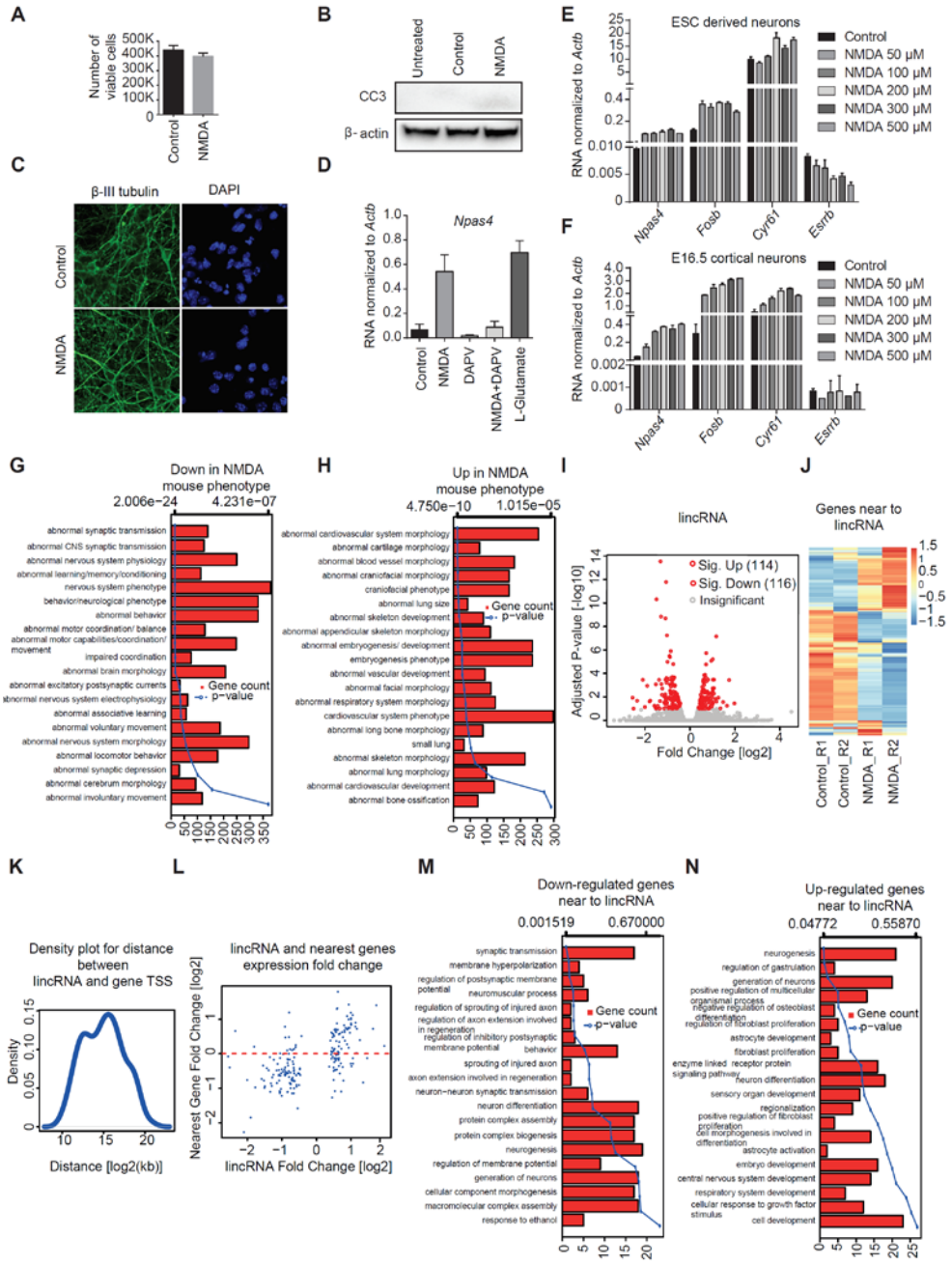
(A) Heat map showing correlation of FAIRE-seq replicates. **(B-D)** Scatter plot showing replicate correlation and enrichment distribution of all identified peaks for NP (B), TND1 (C) and TND10 (D). **(E)** Bar plot showing mapping statistics of reads obtained for each FAIRE-seq sample. X-axis represents percentage of alignment. (F) A profile plot showing normalized FAIRE enrichment around TSS of all genes in different stages of neurogenesis. X-axis shows distance from TSS, while y-axis represents library size normalized enrichment. **(G)** Bar plot showing distribution of FAIRE peaks in different genomic locations in NP, TND1 and TND10. **(H)** Bar plot showing feature size normalized distribution of FAIRE peaks in different genomic locations in NP, TND1 and TND10.



Supplemental Figure S2.

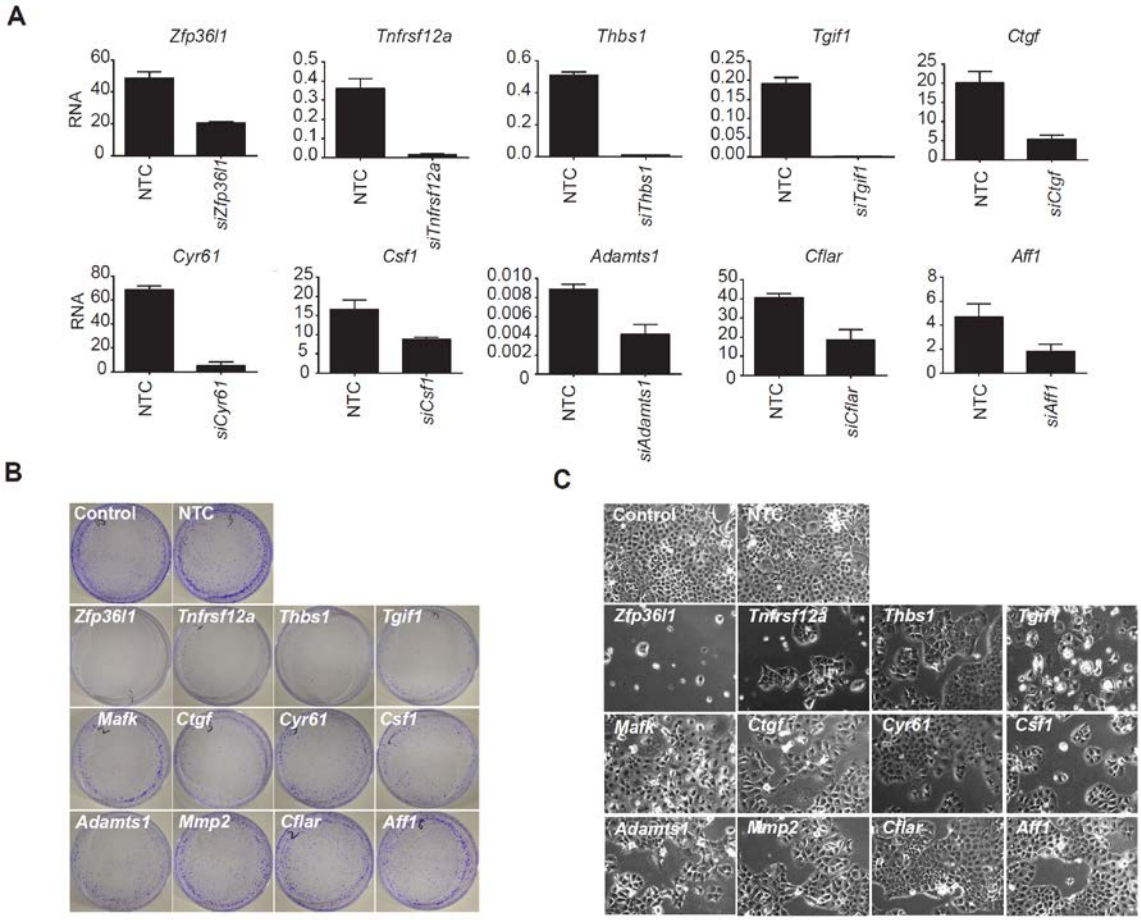
(A) Heat map showing the correlation of H3K27ac replicates. **(B-D)** Scatter plot showing the replicate correlation and enrichment distribution of all identified H3K27ac peaks for NP (B), TND1 (C) and TND10 (D). **(E)** Bar plot showing the mapping statistics of H3K27ac ChIP-seq reads obtained for each sample. The x-axis represents the percentage of alignment. **(F)** Bar plot showing the total number of enriched peaks identified in NP, TND1 and TND10. **(G)** Bar plot showing the number of unique enriched peaks identified in NP, TND1 and TND10. **(H)** Box plot showing the distribution of peak width in NP, TND1 and TND10. P-values are calculated using Wilcoxon test. **(I)** A profile plot showing normalized H3K27ac enrichment around the TSS of all genes. The x-axis shows the distance from the TSS; the y-axis represents the library size-normalized enrichment at different stages of neurogenesis. **(J)** Bar plot showing the overlap of all FAIRE peaks with H3K27ac peaks. A peak was considered to be overlapping if the coordinates of FAIRE peaks showed at least 20% overlap with H3K27ac peak. **(K)** Bar plot showing the overlap of unique FAIRE peaks with H3K27ac peaks using the same approach as in J. **(L)** Bar plot showing the overlap of all H3K27ac

peaks with FAIRE peaks using 20% overlapping criteria. **(M)** Bar plot showing the overlap of unique H3K27ac peaks with FAIRE peaks using same criteria as in (L). **(N)** Bar plot showing the number of all H3K27ac peaks that lie within 1 kb from FAIRE peaks. **(O)** Bar plot showing the number of unique H3K27ac peaks that lie within 1 kb from FAIRE peaks. **(P)** Bar plot showing feature size normalized distribution of FAIRE and H3K27ac overlapping peaks in different genomic locations in NP, TND1 and TND10. **(Q)** Box plot showing the expression of adjacent stages for the genes closest to NP, TND1 and TND10 unique H3K27ac sites. Genes were divided into overlapping bins based on the distance of the distal regulatory elements from the TSS. Expression is reported as reads per kilobase of transcript per million mapped reads (RPKM). **(R)** Stacked bar plot showing the overlap of all uniquely identified H3K27ac sites in TND10 with H3K27ac sites from other tissues.



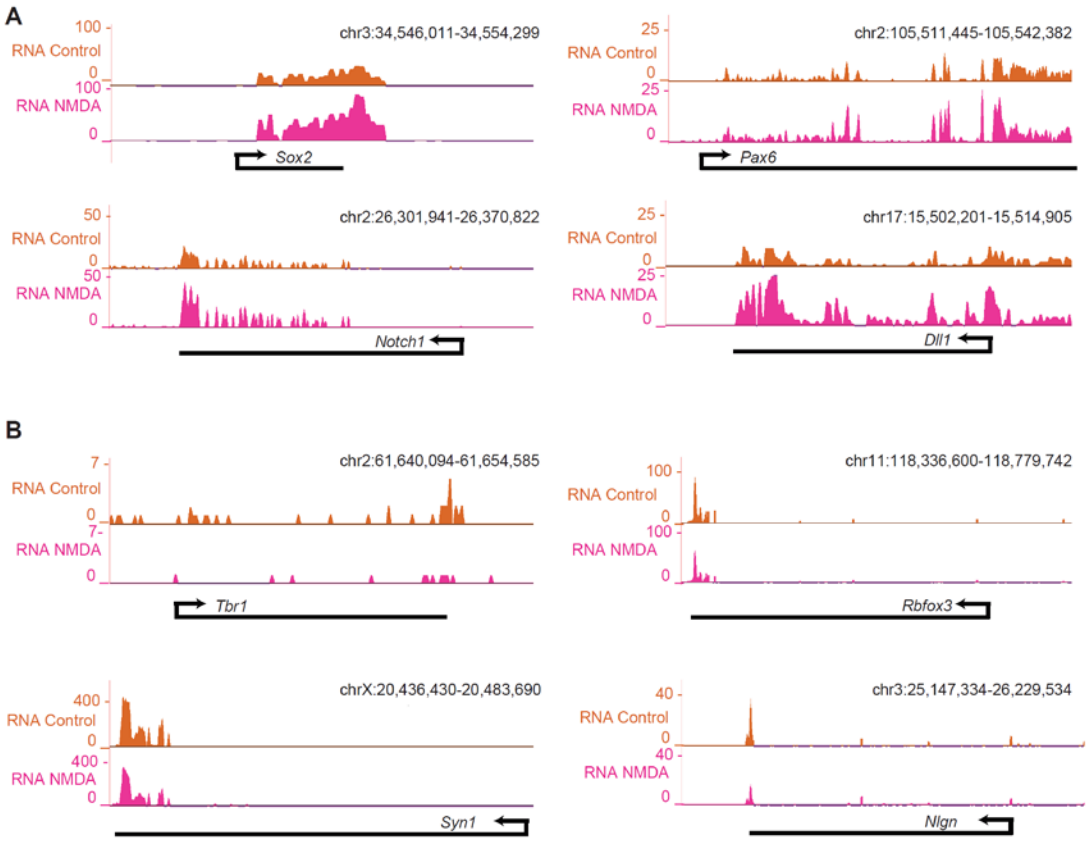
Supplemental Figure S3. (A) Cell viability assay showing total number of viable cells in control and NMDA treated neurons. **(B)** Representative immunoblots (n=2) showing caspase3 in NMDA induced neuron. β-actin is shown as a loading control. **(C)** Immunofluorescence microscopy for β III Tubulin staining in control and NMDA treated neurons confirming no symptoms of neuronal degeneration. Images were taken at 63X magnification **(D)** mRNA levels for classical activity modulated gene *Npas4* in neuron treated with NMDA or NMDA in presence of NMDAR antagonist DAPV or L-glutamate were measured by qRT-PCRs relative to *Actb* plotted on the y-axis. Error bars represent SEM from independent biological replicates. **(E-F)** mRNA expression kinetics of neuronal activity modulated genes upon various concentrations of NMDA were measured by qRT-PCRs relative to *Actb* and plotted on the y-axis for ES-derived neurons (E) and cortical neurons (F). **(G-H)** Bar plot showing the enrichment of known mouse phenotypes in genes

down-regulated (G) and upregulated (H) upon NMDA treatment. The bars reflect the number of genes in each category; the lines represent the multiple tested corrected p-value of the corresponding GO term. **(I)** Volcano plot showing up- and down-regulated lincRNAs upon NDMA treatment in TND10 neurons. The x-axis represents the fold change between control and NDMA in log2; the y-axis shows the adjusted p-value in -log10. **(J)** Heat map showing the expression of the genes nearest to differentially expressed lincRNAs. **(K)** Density plot showing distance (in kb, log2) of lincRNA from nearest gene TSS. **(L)** Scatter plot displaying changes in expression of lincRNA and corresponding nearest genes. **(M-N)** Bar plots showing enrichment of biological processes in down-(M) and up-regulated (N) genes near down- or up-regulated lincRNAs upon NDMA treatment respectivelly. The bars reflect the number of genes in each category; the lines represent the multiple tested corrected p-value of the corresponding GO term. (Results in this figure partially contributed by me)

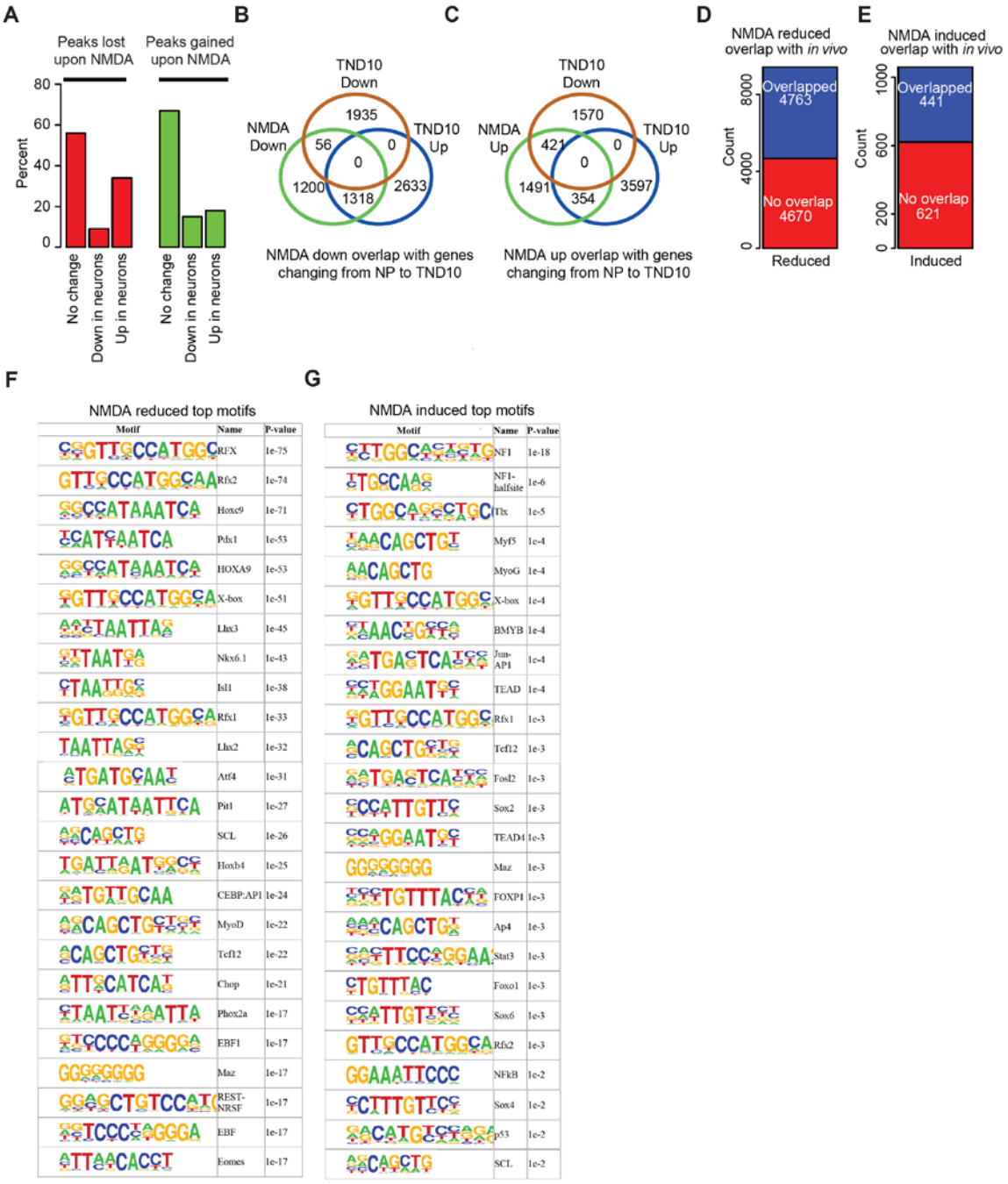


Supplemental Figure S4.

(A) mRNA levels for knockdown efficiency of target gene in epithelial cells pre-depleted for two days with specific siRNA against pro-survival genes were measured by qRT-PCR relative to *Actb* and plotted on the y-axis. Error bars represent SEM from independent biological replicates. **(B)** Representative images of colony forming assay performed on epithelial cell after 7 days of siRNA-mediated depletion of pro-survival factors compared to non-targeting control (NTC). **(C)** Representative Bright field images of the cells described in (B). (Results in this figure contributed by me)



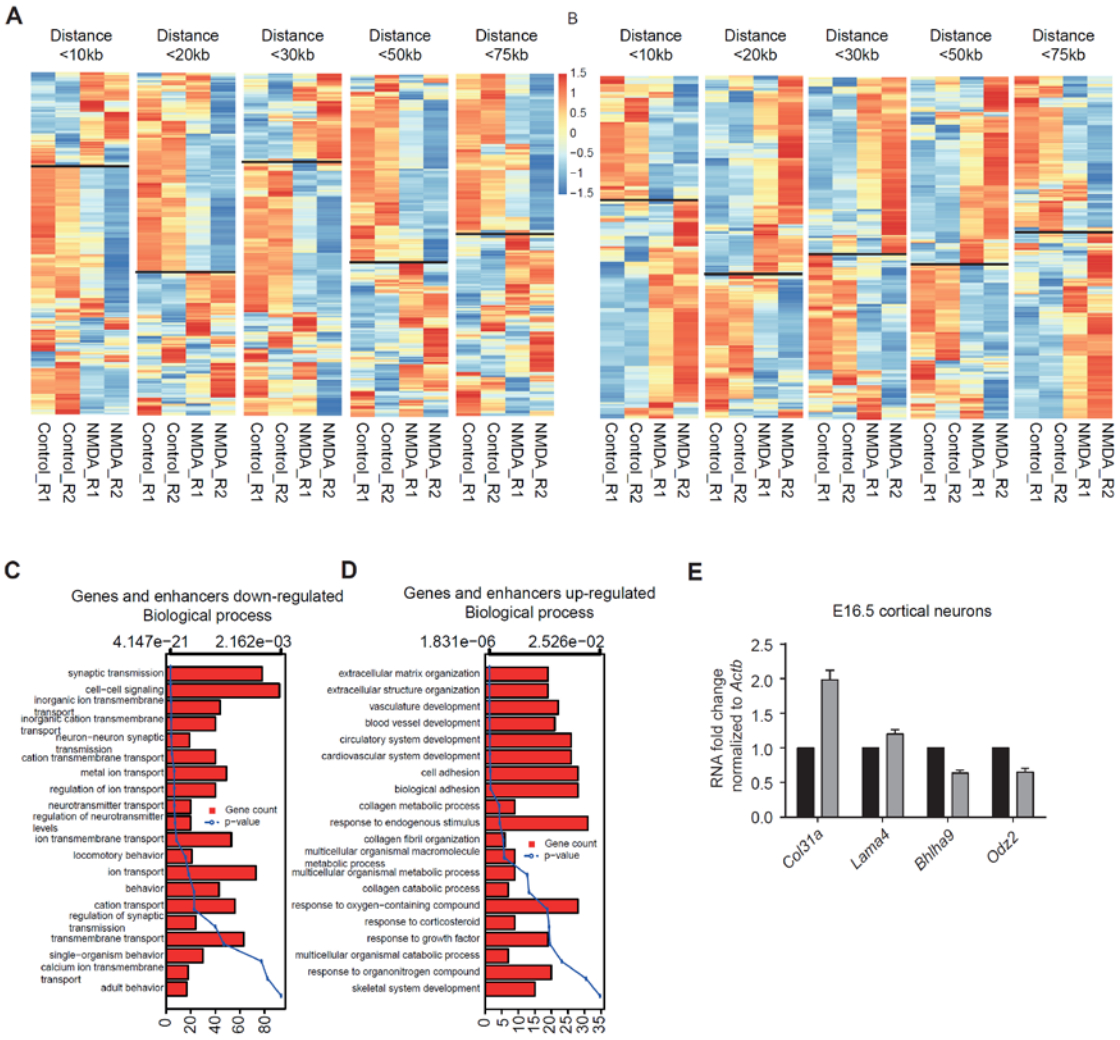
Supplemental Figure S5.
(A) Representative browser tracks of few progenitor genes showing increase in expression upon NMDA.
(B) Same as in (A), but showing down-regulation of selected neuronal genes.



Supplemental Figure S6.

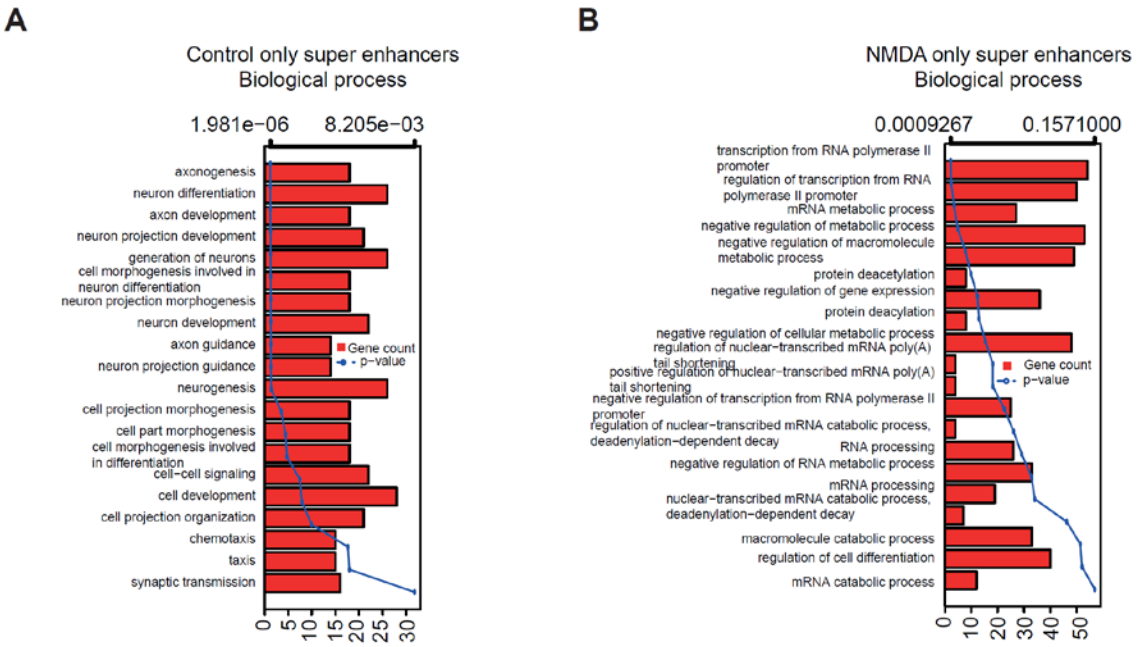
(A) Comparison of NMDA modulated H3K27ac peaks with H3K27ac peaks changing between NP and TND10. Peaks that are either losing or gaining two fold enrichment upon NMDA treatment were compared with two fold changing H3K27ac peaks between NP and TND10. A peak was considered to be overlapping, if it was found to be within 1kb distance of peaks modulated during neurogenesis (NP to TND10). **(B)** Venn diagram showing overlap of significantly down-regulated genes upon NMDA with significantly down or up-regulated genes between NP and TND10. **(C)** Same as in (B) but for genes significantly up-regulated upon NMDA. **(D-E)** Stacked bar plot showing the overlap of NMDA-induced (D) or -reduced (E) peaks with H3K27ac peaks from several embryonic and adult stages taken from E11.5, E14.5, E.17.5, P0, P7, P21, P56

cortex [172]. **(F-G)** List of top 25 transcription factors motifs identified in NMDA-reduced (F) and NMDA-induced (G) peaks.

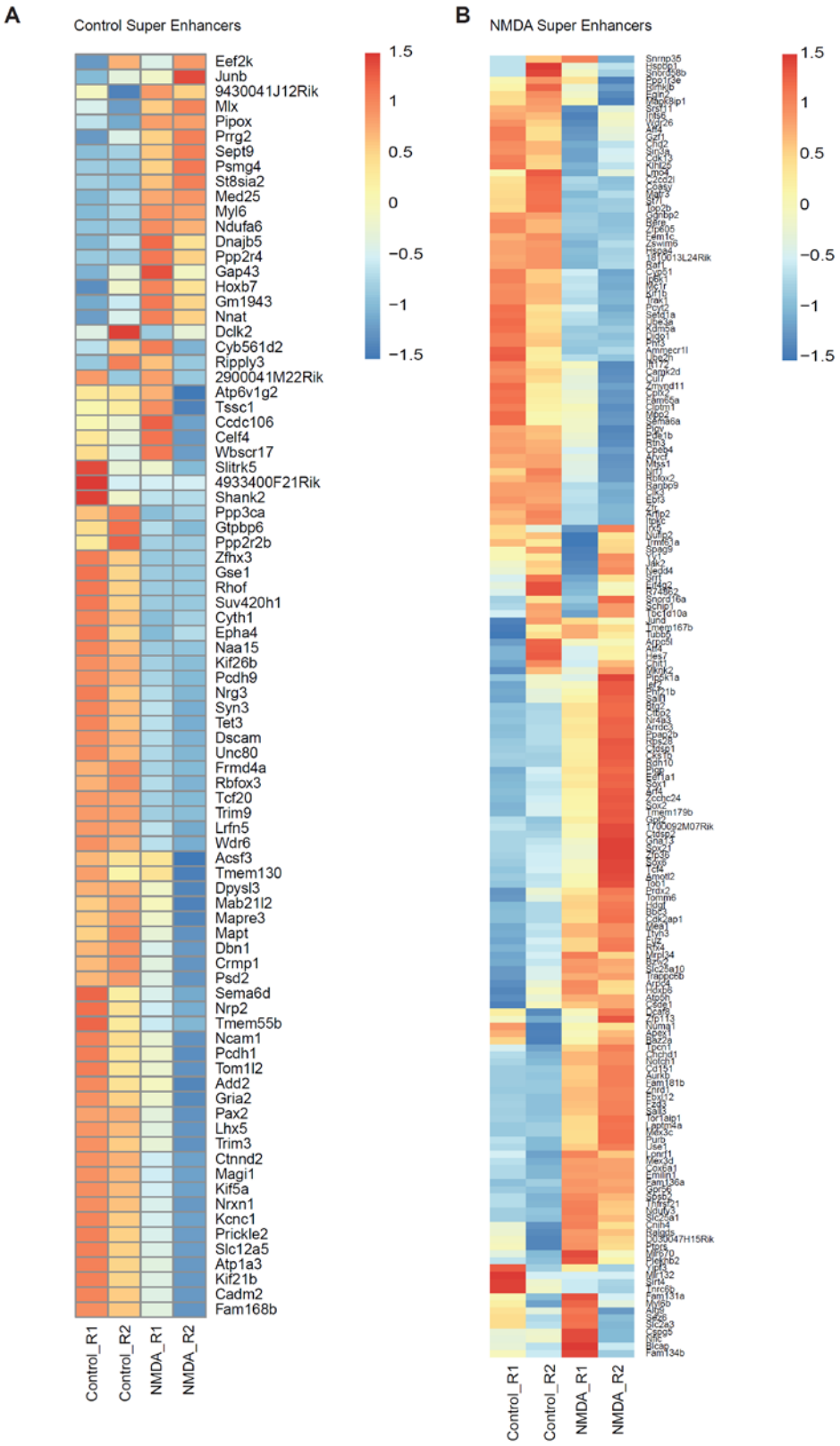


Supplemental Figure S7.

(A-B) Heat map showing the expression of genes, binned based on distance (10kb, 20kb, 30kb, 50kb and 75kb) from NMDA-reduced (A) or NMDA-induced distal peaks (B) in control- and NMDA-treated day 10 neurons. **(C-D)** Bar plots showing the enrichment of biological processes in genes that are near to down- and up-regulated distal regulatory regions and their expression reproducibly changed by at least 1.5-fold. The bars reflect the number of genes in each category; the lines represent the multiple tested corrected p-value of the corresponding GO term. ***(E)*** mRNA levels for genes shown in Fig. 7H upon NMDA mediated activation in mouse cortical neurons isolated from E16.5 mouse brain measured by qRT-PCR relative to *Actb* and plotted on the y-axis. Error bars represent SEM from independent biological replicates. (Results in this figure partially contributed by me)



Supplemental Figure S8.
(A) Enrichment of biological process in genes near to non-promoter control only super-enhancers of day 10 neurons. Bars reflect count of gene in each category while line represent multiple tested corrected p-value, displayed as alternate x-axis, of corresponding GO term. **(B)** Same as in (A) but for NMDA unique super-enhancers. (Results in this figure partially contributed by me)



Supplemental Figure S9.

(A-B) Heat maps showing expression of genes near to control (A) and NDMA (B) non-promoter super-enhancers in control and NDMA-treated neurons.

LEGENDS FOR SUPPLEMENTAL TABLES

Supplemental Table S1. Table showing enrichment for all FAIRE peaks in NP, TND1 and TND10.

Supplemental Table S2. Table showing enrichment for all H3K27ac peaks in NP, TND1 and TND10.

Supplemental Table S3. References for studies that have shown a function of transcription factors predicted to target open H3K27ac-positive sites in neuronal development.

Supplemental Table S4. Normalized expression values all genes upon NMDA treatment.

Supplemental Table S5. Table for enrichment of all H3K27ac sites in control and NMDA treated neurons.

Supplemental Table S6. Gene Ontology enrichment table for the genes shown in Supplemental figure S6B and C.

Supplemental Table S7. List of activity and neuronal genes close to super-enhancers.

Supplemental Table S8. List of all primers used in this study.

SUPPLEMENTAL METHODS

Cortical Culture

Cortical neurons were cultured as described previously [87]. Embryonic mouse cortex (E16.5) was dissected and tissue was dissociated with Trypsin for 10 minutes. After neutralizing Trypsin with 10% FBS in DMEM/F12 cell suspension was obtained by passing the samples through thin glass Pasteur pipette and then through cell strainer. All experiments were performed in cold conditions and cells were kept in cold until plating. Plates were pre-coated overnight with 20 µg/ml poly-D-Lysine (Sigma) and 4 µg/ml laminin (Sigma) in water. After overnight coating plates were washed three times with water and once with Neurobasal media. Neurons were plated at a density of 25,000/cm^2 in Neurobasal media containing 2% B27 supplement (Invitrogen). Fresh media was added after two hours (and day 3 and 6) and neurons were kept in 7% CO2 at 37 degrees. Experiments were performed on day 7 neuron.

FAIRE assay

FAIRE-assay was performed as described previously [148]. Briefly, neuronal progenitors, day-1 neurons and differentiated neurons (day 10) were crosslinked in medium containing 1% formaldehyde for 10 min at room temperature. Cells were collected by scraping in cold PBS, resuspended in buffer L1 (50 mM HEPES/KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, pH 8.0, 10% glycerol, 5% NP-40, 0.25% Triton X-100) and incubated for 10 min at 4 °C. The cells were then centrifuged for 5 min at 4 °C at 1,300 g. The pellet was

resuspended in buffer L2 (200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0) and incubated for 10 min at room temperature, followed by centrifugation for 5 min at 4 °C at 1,300 g. The pellet was then resuspended in buffer L3 (1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, 100 mM NaCl, 0.1% Na-deoxycholate, 0.17 mM N-lauroyl sarcosine) containing protease inhibitors and incubated at 4 °C for 3 h. Samples were then sonicated using Bioruptor plus (Diagenode), and cellular debris was cleared by spinning at 14,000 g for 10 min at 4 °C. DNA was isolated from 60 µg of total chromatin by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), vortexing and spinning at 12,000 g for 5 min at room temperature. The aqueous phase was isolated, and a second round of phenol:chloroform:isoamyl alcohol purification was performed. Following collection of the aqueous phase, chloroform:isoamyl alcohol (24:1) was added, vortexed and spun at 12,000 g for 5 min at room temperature. The aqueous phase was then collected, and the DNA was ethanol precipitated and recovered in 40 µl of TE buffer. This recovered material was then re-extracted with phenol followed by chloroform:isoamyl alcohol (24:1) and subsequently ethanol precipitated. Next, the DNA was recovered in 40 µl TE buffer. To obtain the input DNA control, 10% input was treated with RNase A (0.2 mg/ml) for 30 min at 37 °C and then with Proteinase K (50 mg/ml) for 2.5 h at 55 °C. The crosslinking was reversed at 65 °C overnight with gentle shaking. The DNA was purified by phenol-chloroform extraction followed by ethanol precipitation and recovered in 40 µl of TE buffer. Real-time PCR using FAIRE or input DNA was performed using SYBR Green chemistry (ABI). Primers used for FAIRE-qPCRs are listed in Supplemental Table S8.

RT-PCR

RNA was extracted either using trizol or the SurePrep TrueTotal RNA Purification Kit (Fisher Scientific) and subsequently used for cDNA synthesis by applying the First Strand cDNA Synthesis Kit (Fermentas). Gene expression levels were measured using SYBR Green chemistry (ABI). Primers used for RT-qPCRs are listed in Supplemental Table S8.

Cell survival assay

Cells were pre-depleted with siRNA (pool of 4 siRNA targeting same gene, SMARTpool, Dharmacon) for two days and 40,000 cells per 6cm were platted and kept for 6 day. Fresh medium along with siRNA was added each alternate day. Cells were either counted or bright field images were taken or fixed with 4% PFA. For visualization of fixed colonies, cells were stained with solution containing crystal violet and 10% ethanol for 10 minutes and the plates were washed in water till the excessive dye washed away. Pictures of the plates were taken for down-stream analysis.

RNA-seq data analysis

RNA-Seq data was generated using Illumina sequencing. Reads were aligned to mouse genome (mm9) using TopHat [177] (version 2.0.9) with default options. Expression was quantified after library size normalization using DESeq [178] and in Reads Per Kilobase of transcript per Million mapped reads (RPKM) using cufflink (version 2.1.1)[179]. Differential expression analysis was performed using the DESeq package with a FDR rate of 0.1 [178].

FAIRE-seq analysis

The FAIRE libraries processed for Illumina sequencing were prepared according to Illumina's instructions (Catalogue number IP-102-1001). Sequencing reads were aligned with the reference mouse genome (mm9) using bowtie (version 0.12.9) with a maximum of 2 mismatches in the seed region [180]. Aligned reads from all samples and replicates were merged before peak calling. Input FAIRE samples from all stages were pooled to create a single input file. Potential peaks were called using MACS [181] (version 2.0.10.2013071) without providing an input file and with default parameters. These potential peaks were then used to calculate their raw enrichment using the aligned bam files for each merged sample with respect to input using the QuasR package [182]. Normalized enrichment based on library size and input was calculated, and only peaks that displayed at least 1.5-fold enrichment above input were selected for further analysis. Normalized enrichment was calculated using the following formula:

$$Enrichment = log2((\frac{ns}{Ns} * \min(ns, nb) + p)/(\frac{nb}{Nb} * \min(ns, nb) + \mathrm{p}))$$

where ns is the total number of reads aligning in the sample for the genomic feature; Ns is the total number of aligned reads in the sample; nb is the total number of reads aligned in the input/background for the genomic feature; Nb is the total number of aligned reads in the input/background; and p stands for pseudocount (correction for low alignment reads), which was set to 16. The FAIRE-seq data derived during this study are available under GEO accession number GSE65713.

ChIP-seq analysis

The ChIP libraries processed for Illumina sequencing were prepared using Illumina's standard instructions. ChIP-seq analysis of H3K27ac was performed in a manner similar to that described above for FAIRE-seq.

Wig files

Library size-normalized wig files were created for all the samples using the QuasR package [182].

lincRNA

We collected lincRNA from NonCode (www.noncode.org)[183], ENSEMBL (http://www.ensembl.org) [184], UCSC (http://genome.ucsc.edu) and other published resources. Differential expression analysis was performed using the DESeq package with a FDR rate of 0.1.

Genomic distribution of peaks

The gene annotation file for mouse (mm9) from UCSC was used, and the genomic coordinates of the promoter, exon, intron and intergenic regions were generated in a hierarchical manner. Promoters were defined as 1000 bp upstream and downstream of the TSS; exon coordinates from the annotation file were used to define exons; and the remaining regions between the TSS and TTS were called introns. The rest of the genome was annotated as intergenic.

Peak overlap calculation

Overlapping peaks were called using BEDTools with at least 20% overlap. An alternative approach also used to identifay nearby peaks between FAIRE and H3K27ac peaks, where any peaks within 1 kb of the base peak were considered.

Motif prediction

Transcription factor motif prediction for stage-specific intergenic accessible regions was performed using Homer [185].

Super-enhancer identification

Super-enhancers were identified using the Homer package [185]. Peaks that were within 6 kb were stitched together, and an enhancer score was calculated.

GO enrichment analysis

GO enrichment analysis were performed using the ToppGene package [186]. Only the top 20 enriched terms from the GO analysis were plotted. All p-values plotted are corrected for multiple testing.

Epithelial to Mesenchymal Transition

3.1.1. Epithelial mesenchymal transition

In the early 1980s Elizabeth Hay, coined the term epithelial to mesenchymal transfer to describe the phenotypic changes in the primitive streak of chick embryos, where epithelial cells can downregulate epithelial characteristics and acquire mesenchymal features [187]. Epithelial cells remain in close contact with neighbouring cells and maintain an apical-basal axis of polarity by sequential arrangement of adherent junctions, desmosomes, and tight junctions [188]. Epithelial to mesenchymal transition (EMT) describes the process, which involves phenotypic remodelling by cytoskeletal reorganization, loss of cell polarity and cell-cell junction and gain of migratory phenotype[187,189] (Figure 11A). Depending on the tissue and signalling contexts, epithelial cells may lose only some characteristics or may show some epithelial and mesenchymal properties; this can be considered as partial EMT. Conversely mesenchymal–epithelial transition (MET) is a reversible biological process of EMT that involves the transition from motile, multipolar or spindle-shaped mesenchymal cells to planar arrays of polarized epithelial cells (Figure 11A).

EMT can be classified into three broad types depending on the physiological tissue context. Type-1 EMT occurs during embryogenesis and organ development, type-2 EMT is important for tissue regeneration as well as organ fibrosis, and type-3 EMT is associated with cancer progression and cancer stem cell properties (Figure 11B). EMT is essential during various developmental processes, such as embryo implantation, embryonic layer formation, gastrulation, neural crest migration while in adults, EMT is known to be only activated to ensure tissue regeneration and wound healing [188,190,191]. However, aberrant activation of EMT program is associated with disease phenotypes such as organ fibrosis [192-194] as well as tumour progression and metastasis [195-197] (Figure 11B). Additionally, cells that have undergone EMT acquire resistance to senescence and apoptosis. The transition of epithelial cells into mesenchymal cells, in development or in pathological conditions, follows a common and conserved programme with hallmarks of gene expression during EMT. EMT and MET have been closely linked to ‘stemness’ in development and cancer. Pluripotent embryonic stem (ES) cells in the inner mass of the blastocyst have

epithelial characteristics. During gastrulation, pluripotent epithelial epiblast cells ingress to form the primary mesoderm through EMT. EMT thus represents an initial differentiation event in the generation of the three germ layers from pluripotent cells. Conversely, the reprogramming of fibroblasts into induced pluripotent stem (iPS) cells requires MET i.e. the transition from a mesenchymal phenotype to an epithelial phenotype[198].

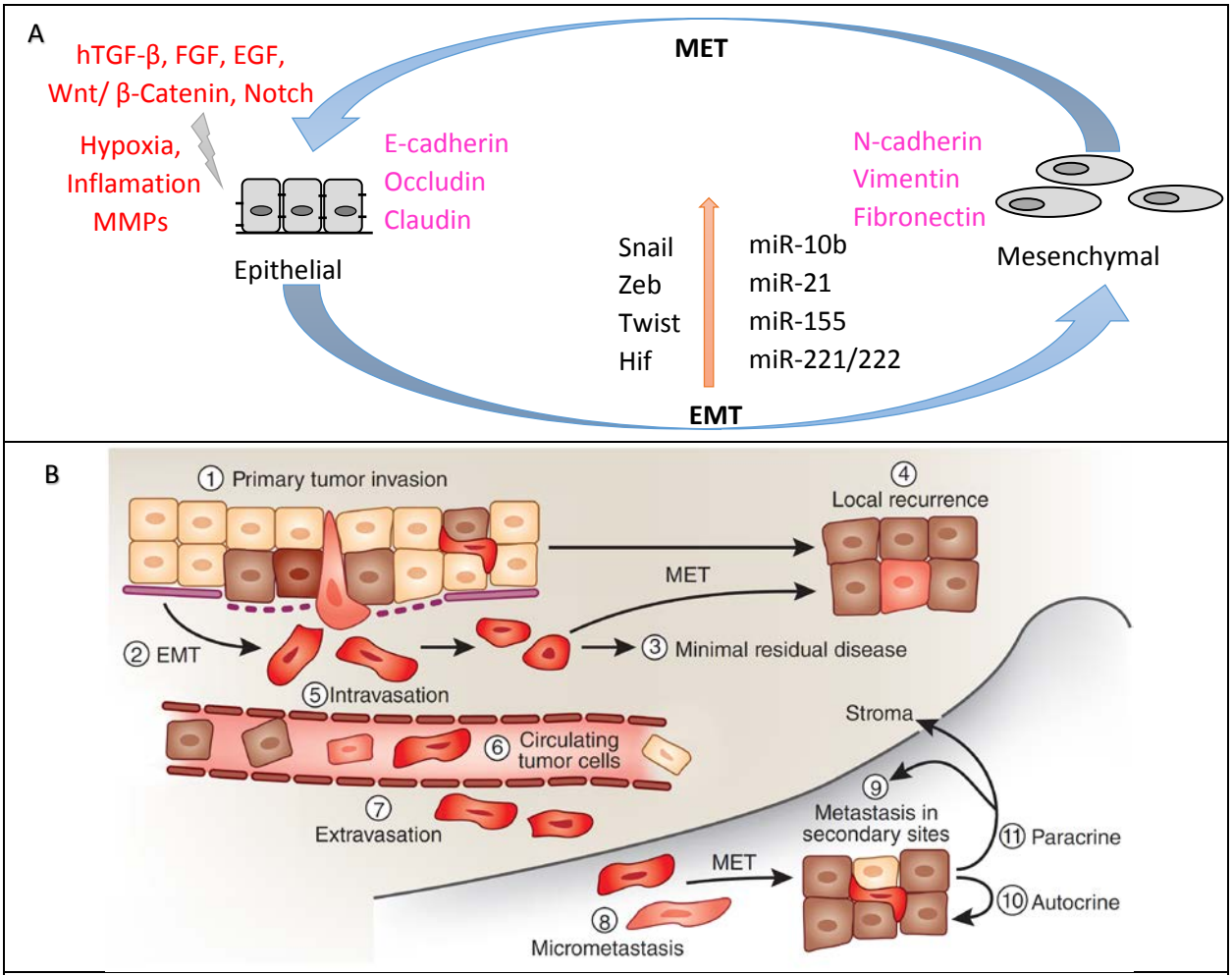


Figure 11. EMT and its roles during carcinoma progression. (A) Schematic representation of EMT and MET**. (B)** Schematic representation of tumour metastasis. Carcinoma cells usually have a primarily epithelial phenotype but contain small numbers of mesenchymal-like cancer stem cells, and/or they may otherwise undergo EMT induced by genetic, autocrine or paracrine factors. (Adopted from [199]).

3.1.2. Signalling pathways involved during EMT

A number of signalling pathways such as TGF-β, FGF, EGF, HGF, Wnt/beta-catenin and Notch are known to induce EMT *in vitro* and *in vivo* [189,196,197,200] (Figure 12A). TGF-β is among the most established and potent

inducers of EMT and considered as prototypic stimulus[191,201]. Such stimulation induces the canonical pathway in which Smad2 and Smad3 become C-terminally phosphorylated and form an active trimer with Smad4, which in turn translocates to the nucleus to bind target gene promoters for transcriptional regulation[202,203] resulting in down regulation of epithelial and cell cycle genes and up regulation of mesenchymal markers [204-206] (Figure 12B). TGF-β also induces non-canonical signalling via a number of MAP kinases like Erk, P38 and JNK [207-209] that are also crucial for proper EMT (Figure 2B).

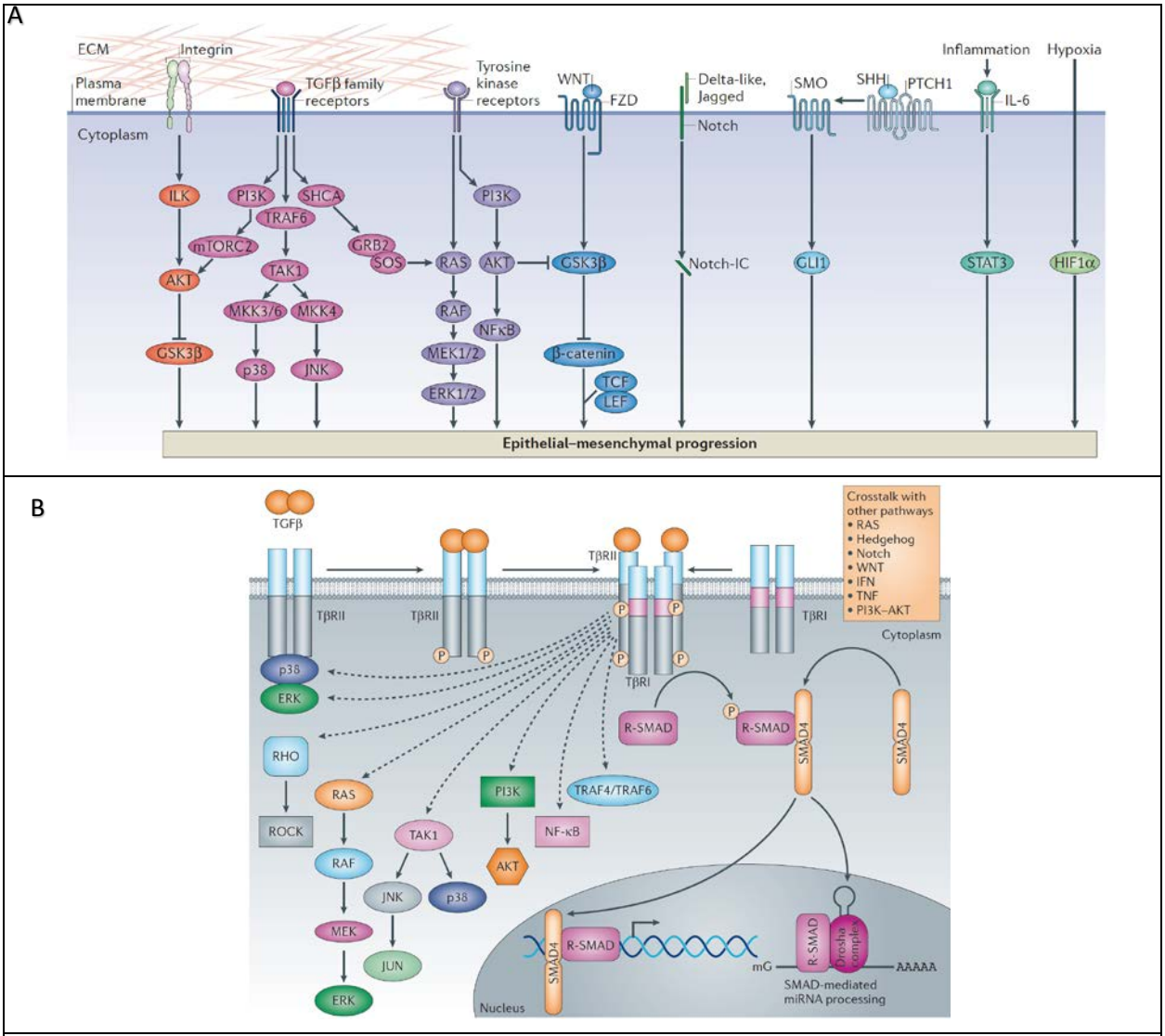


Figure 12. Signalling pathways involved in EMT. (A) Various signalling pathways regulating EMT. **(B)** Schematic representation of non-canonical TGFβ signalling and crosstalk with other signalling pathways. (Adopted from [198,210])

3.1.3. Cytoskeletal changes and acquisition of motility during EMT

Epithelial cells are connected through subapical tight junctions, adherent junctions and desmosomes at lateral surfaces, and gap junctions at lateral surfaces. Upon the initiation of EMT, these junctions are deconstructed by delocalization and degradation of involved proteins. The dissolution of tight junctions is accompanied by decreased claudin and occludin expression, and the diffusion of zonula occludens 1 (ZO1/TJP1) from the sites of cell–cell contacts. During the destabilization of adherent junctions, epithelial cadherin (E-cadherin) is cleaved at the plasma membrane and subsequently degraded. Consequently, β-catenin can no longer interact with E-cadherin and becomes free for degradation or its action in transcriptional regulation. p120 catenin (catenin δ1) also accumulates in the nucleus and participates in transcription [198]. EMT initiation also disrupts desmosomes and the integrity of gap junctions. As EMT progresses, the expression of junction proteins is transcriptionally repressed and the cell organizes their cortical actin cytoskeleton into one that enables dynamic cell elongation and directional motility[198]. New actin-rich membrane projections include sheet-like membrane protrusions called lamellipodia and spike-like extensions called filopodia at the edge of lamellipodia that in turn facilitate cellular movement. Actin-rich invadopodia exert a proteolytic function in ECM degradation, thus facilitating cell invasion. Overall there is an increased cell contractility and actin stress fibre formation. Various matrix metalloproteinases (MMPs) assist in degradation and modification of the extracellular matrix (ECM) as well as cell-ECM and cell-cell contacts and promote detachment of epithelial cells from the surrounding tissue [211]. A hallmark of EMT is the downregulation of E-cadherin, activation of genes N-cadherin, the protein products of which promote mesenchymal adhesion and neural cell adhesion molecule (NCAM) [198] (Figure 13). The partitioning-defective (PAR) complexes and Crumbs complexes localize apically in association with tight junctions and define the apical compartment; Scribble complexes define the basolateral compartment. Consequently, the dissolution of epithelial junctions during EMT confers a loss of apical–basal polarity [198] (Figure 13).

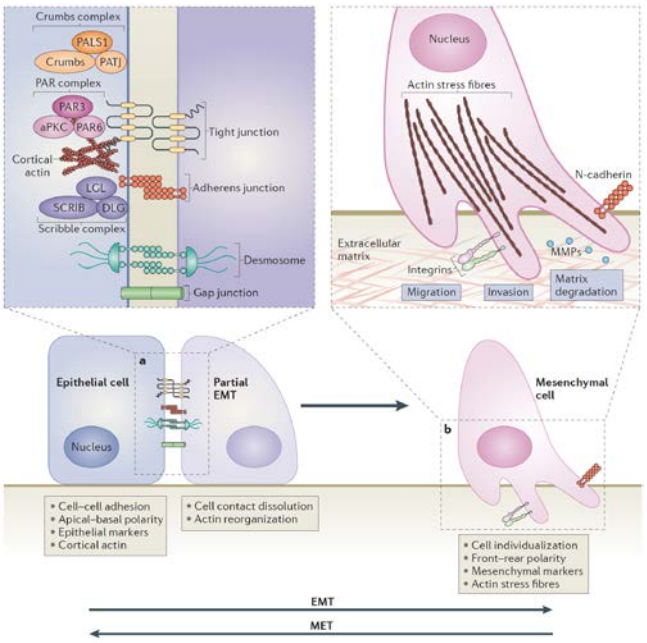


Figure 13. Cellular events during EMT. (a) The first steps of epithelial–mesenchymal transition (EMT) are the disassembly of epithelial cell–cell contacts, that is, tight junctions, adherent junctions, desmosomes and gap junctions and the loss of cell polarity through the disruption of the Crumbs, partitioning defective and Scribble polarity complexes. Next, in (b) the epithelial actin architecture reorganizes, and cells acquire motility and invasive capacities by forming lamellipodia, filopodia and invadopodia, and by expressing matrix metalloproteinases (MMPs) that can degrade extracellular matrix (ECM) proteins. (Adopted from [198])

3.1.4. Transcription factors and epigenetic machinery driving EMT

The changes in gene expression that contribute to the repression of the epithelial phenotype and activation of the mesenchymal properties involve master regulators, including SNAIL, TWIST and zinc-finger E-box-binding (ZEB) transcription factors**.** These factors often control the expression of each other and functionally cooperate at target genes. SNAIL proteins repress epithelial genes by binding to E-box DNA sequences through their carboxy-terminal zinc-finger domains and recruits Polycomb repressive complex 2 (PRC2) that coordinates histone modifications, specifically methylation of histone H3 Lys 27 (H3K27me3). SNAIL activity itself is regulated by post-translational modifications such as phosphorylation at various serine residues and ubiquitination to modulate its activity, stability and nuclear localization (Figure 14A).

TWIST1 and TWIST2, together with E12, E47, and inhibitor of differentiation (ID) proteins are belonging to the Basic helix–loop–helix (bHLH) transcription factors that function as master regulators of EMT progression. TWIST1 and TWIST2 form homodimers, as well as heterodimers, with E12 or E47 to regulate E-box DNA binding and transcriptional regulation. TWIST recruits the methyltransferase SETD8, which mediates H4K20 monomethylation, a histone mark that is associated generally with transcriptional

repression. Mechanical stress and hypoxia-inducible factor 1α (HIF1α) induce TWIST expression under hypoxic conditions (Figure 14B).

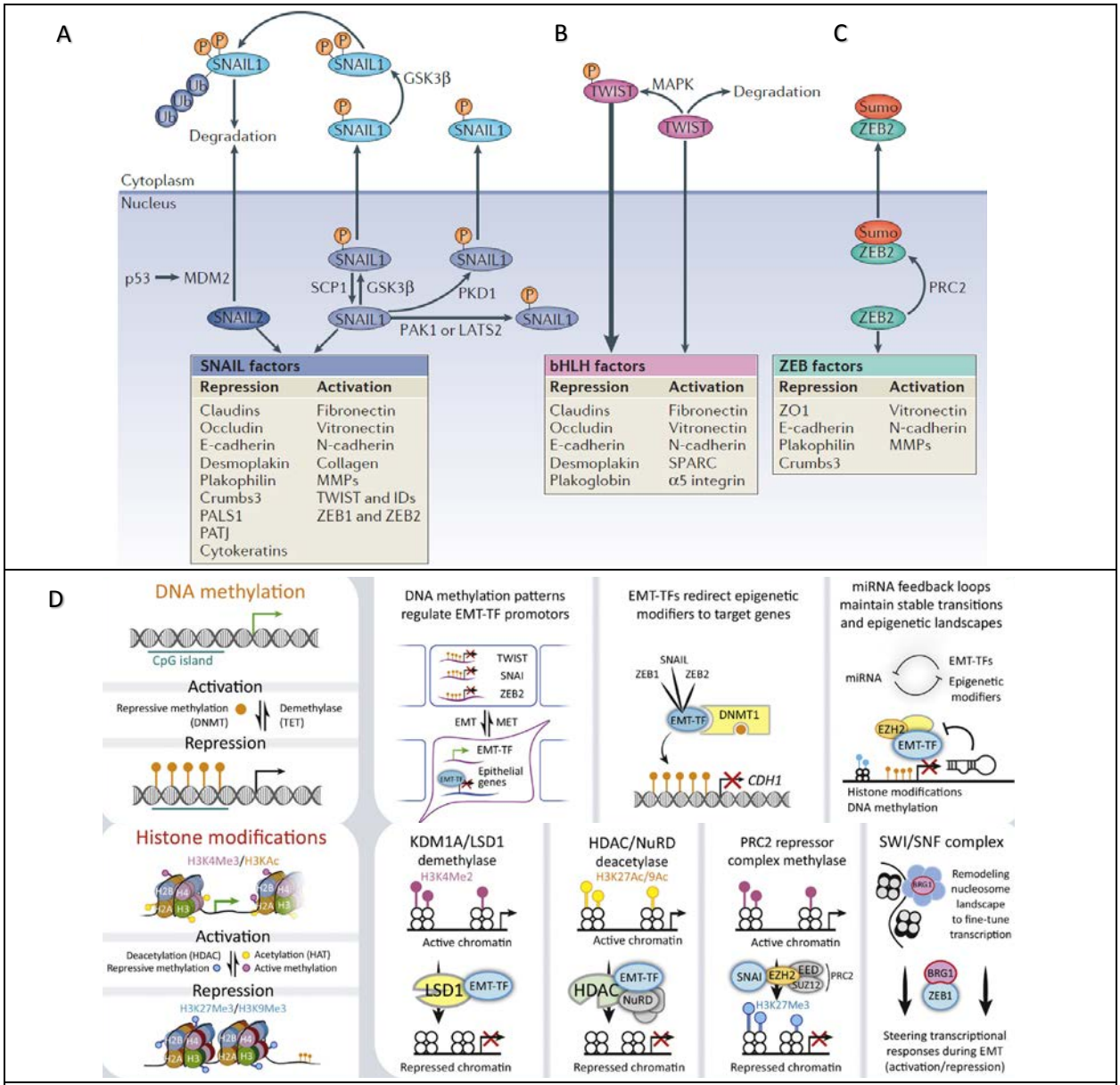


Figure 14. Roles and regulation of major EMT transcription factors and epigenetic machinery. (A-C) Epithelial–mesenchymal transition (EMT) is driven by SNAIL, zinc-finger E-box-binding (ZEB) and basic helix–loop–helix (bHLH) transcription factors that repress epithelial marker genes and activate genes associated with the mesenchymal phenotype. Post-translational modifications regulate their activities, subcellular localization and stability (Adopted from [198]). **(D)** The most relevant enzymatic systems known to modify DNA and/or histones are depicted. EMT Transcription Factors (TFs) are able to recruit the methyltransferase DNMT1, resulting in DNA methylation and gene repression. A large number of

histone-modification enzymes have been identified in complex with EMT-TFs, affecting chromatin compaction and locus accessibility during gene regulation (Adopted from [212]).

ZEB-mediated transcriptional repression often involves the recruitment of a corepressor C-terminal-binding protein (CTBP) or sometime independently by recruiting the Switch/sucrose-non-fermentable (SWI/SNF) p300/CBP-associated factor (PCAF/KAT2B) and p300, which switches its activity from a transcriptional repressor to a transcriptional activator (Figure 14C). Along with known EMT TFs several of the forkhead box (FOX) transcription factors, GATA family, and SRY box (SOX) transcription factors in coordination with epigenetic modulators play critical role in mesenchymal transition [212] (Figure 14D).

3.1.5. Post-transcriptional regulation of EMT

In addition to the direct effects of EMT transcription factors and epigenetic regulators on gene expression, changes at the post-transcriptional level via alternative splicing have been discovered to regulate EMT progression. EMT is characterized by extensive changes in splicing of a multitude of mRNAs, that allows generation of different protein isoforms in mesenchymal cells as compared to epithelial counterparts e.g. p120 catenin, cluster of differentiation 44 (CD44), FGFR2 display such fate-specific splicing differences. Many changes in splicing result from the rapid downregulation of epithelial splicing regulatory protein 1 (ESRP1) and ESRP2 during EMT. Furthermore increased expression, during EMT, of RNA binding protein FOX1 homologue 2 (RBFOX2) and Ser-Arg-rich splicing factor 1 (SRSF1) convey alternative splicing events in epithelial vs. mesenchymal cells. The differential splicing of nascent RNAs into mRNAs that generate proteins with structural and functional differences have been shown to regulate the activities of key proteins in EMT [198,213] (Figure 15).

Non-coding miRNAs that selectively bind mRNAs, thus inhibiting their translation or promoting their degradation, also regulate the epithelial phenotype as well as EMT. For example, miR-29b and miR-30a repress SNAIL1 and miR-1 and miR-200b can repress SNAIL2 expression. Members of the miR-200 family and miR-205 repress the translation of *ZEB1* and *ZEB2* mRNAs and a double-negative feedback controls

ZEB and miR-200 expression. The expression of high mobility group A2 (HMGA2), which activates SNAIL1 and TWIST expression, is downregulated by the miRNA let-7 and miR-365. In addition miRNAs also target genes that are essential to define epithelial versus mesenchymal phenotype, such as E-cadherin, N-cadherin, TJP1 and Nectin1 (Figure 15).

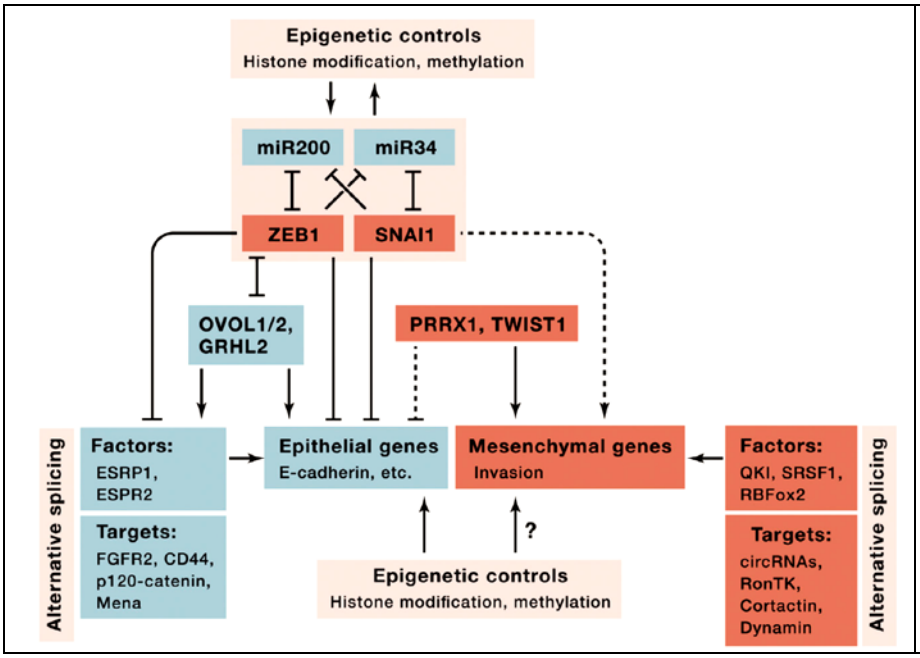


Figure 15. The Core Regulatory Machinery of EMT. (Adopted from [214])

However, the contribution of signaling pathway to global transcriptional reprogramming underlying cell-fate changes during EMT remains elusive. Furthermore, epigenetic mechanisms and repertoire of other proteins including transcription factors that function downstream signaling during EMT are not comprehensively known. To overcome those limitation, I investigated the role of JNK signaling during EMT and also discovered novel regulators including FBXO32 and ZNF827 during EMT using genome-wide analysis of transcriptome and epigenome during EMT time-course.

Results:

Epithelial to Mesenchymal Transition

Chapter3: **JNK Signaling Employs Novel Transcription Factors to Specify Mesenchymal Fate**

Chapter4: **FBXO32 mediates transcriptional program critical to promote microenvironment underlying Epithelial-Mesenchymal Transition**

Chapter5: **ZNF827-dependent splicing dynamics governs epithelial to mesenchymal transition**

JNK Signaling Employs Novel Transcription Factors to Specify Mesenchymal Fate

Sanjeeb Kumar Sahu1, Angela Garding1, Neha Tiwari2, Sudhir Thakurela1, Joern Toedling1, Susanne Gebhard3, Felipe Ortega2, Nikolai Schmarowski4, Benedikt Berninger2, Robert Nitsch4, Marcus Schmidt3, Vijay K. Tiwari1,*

1. Institute of Molecular Biology (IMB), Mainz, Germany
2. Institute of Physiological Chemistry, University Medical Center, Johannes Gutenberg University, Mainz, Germany
3. Department of Obstetrics and Gynecology, Johannes Gutenberg University, Mainz, Germany
4. Institute for Microscopic Anatomy and Neurobiology, University Medical Center, Johannes Gutenberg University, Mainz, Germany

*Correspondence should be addressed to v.tiwari@imb-mainz.de.

Running title: Mapping JNK-dependent gene regulatory circuitry underlying EMT

EMBO J. 2015 Aug 13;34(16):2162-81.

ABSTRACT

The epithelial to mesenchymal transition (EMT) is a biological process in which cells lose cell-cell contacts and become motile. EMT is used during development, for example in triggering neural crest migration, and in cancer metastasis. Despite progress, the dynamics of JNK signaling, its role in genomewide transcriptional reprogramming and involved downstream effectors during EMT remain largely unknown. Here we show that JNK is not required for initiation, but progression of phenotypic changes associated with EMT. Such dependency resulted from JNK-driven transcriptional reprogramming of critical EMT genes and involved changes in their chromatin state. Furthermore, we identified eight novel JNK-induced transcription factors that were required for proper EMT. Three of these factors were also highly expressed in invasive cancer cells where they function in gene regulation to maintain mesenchymal identity. These factors were also induced during neuronal development and function in neuronal migration *in vivo*. These comprehensive findings uncovered a kinetically distinct role for JNK pathway in defining transcriptome that underlies mesenchymal identity and revealed novel transcription factors that mediate these responses during development and disease.

Introduction

Epithelial to mesenchymal transition (EMT) refers to the phenotypic remodeling that involves changes such as cytoskeletal reorganization and loss of cell-cell junctions allowing epithelial cells to evade from their original location by acquiring a motile, migratory, mesenchymal phenotype [187]. Such dramatic change in cell-fate is essential during key developmental processes such as embryo implantation, embryonic layer formation during gastrulation and neural crest migration [188,189,191,196,215-217]. While widely employed during embryonic development, in adults, EMT is known to be only activated to ensure tissue regeneration and wound healing [188,190,191]. However, aberrant activation of EMT program is associated with disease phenotypes such as organ fibrosis [192-194] as well as tumor progression and metastasis [195-197]. A number of signaling pathways such as TGF-β, FGF, EGF, HGF, Wnt/beta-catenin and Notch are known to induce EMT *in vitro* and *in vivo* [189,196,197,200].

TGF-β is among the most established drivers of EMT during development, cancer progression and fibrosis [191,201]. TGF-β is further considered as prototypic EMT inducer whereas other molecules display a more context-dependent action [218,219]. TGF-β stimulation induces the canonical pathway in which Smad2 and Smad3 become C-terminally phosphorylated and form an active trimer with Smad4, which in turn translocates to the nucleus to bind target gene promoters for transcriptional regulation [202,203]. This results in gene expression changes that are crucial for mediating early steps of reprogramming from epithelial to mesenchymal identity including downregulation of classical epithelial and cell cycle genes as well as upregulation of mesenchymal markers [204-206]. It is also known that besides Smad-dependent signaling (canonical) TGF-β also induces Smad-independent signaling (non-canonical) via a number of additional cascades including MAP kinases [207-209,220,221]. Chemical inhibition of these non-canonical signaling pathways including JNK or perturbation of their components alters EMT suggesting their important function during epithelial to mesenchymal transition [209,222-224]. However, the contribution of JNK pathway to global transcriptional reprogramming underlying cell-fate changes during EMT remains elusive.

Furthermore, epigenetic mechanisms and repertoire of transcription factors that function downstream of JNK signaling during EMT are not known.

Here we reveal the functional kinetics of JNK signaling during EMT in mammary epithelial cells and show that it is not required for onset but for the progression of phenotypic changes hallmark of EMT. We further reveal the underlying cause by showing that the transcriptional reprogramming of many critical EMT genes relies on JNK signaling as cells progress to a mesenchymal state. Such JNK-dependent gene regulation involves modulation of the epigenetic state as well as chromatin packaging at target gene promoters. We further identified eight previously uncharacterized transcription factors that rely on JNK signaling for their expression during mesenchymal progression. Depletion of these factors perturbs phenotypic properties of EMT and reverses EMT gene signature. We find that three of these factors are also crucial for maintaining the mesenchymal identity by regulating expression of a large number of metastasis relevant genes. Importantly, these factors show significantly higher expression in invasive tumors which exhibit high JNK and low Smad signaling. These transcription factors are similarly induced during neurogenesis in a JNK-dependent fashion and regulate neuronal EMT during development. This study goes beyond previous notions on the involvement of JNK signaling in EMT, detailing kinetically distinct modes of JNK-activity during the process and establishing its role in maintenance of the mesenchymal fate via regulation of critical genes. Importantly further, these comprehensive findings uncover JNK-dependent gene regulatory circuitry underlying cell-fate changes from epithelial to mesenchymal state, at the same time reveal novel transcription factors that participate in mediating these changes during development and disease.

RESULTS

JNK shows kinetically distinct phase of activation during EMT and progressively overtakes as a critical determinant of mesenchymal properties

To decipher gene regulatory mechanisms underlying EMT, we used untransformed normal murine mammary gland (NMuMG) epithelial cells exposed to TGF-β, which is an established model for EMT in cell culture [176,197]. Over a period of seven days of exposure to TGF-β, we found that NMuMG cells lost markers of cell-cell junctions, increased stress fiber formation, and increased the production of extracellular matrix and focal adhesion proteins (Supplementary Fig S1A), consistent with the acquisition of a complete mesenchymal identity. To determine if TGF-β-induced EMT requires *de novo* transcription and translation and cannot be established exclusively by signaling mediated modulation of the existing proteome, we treated NMuMG cells with inhibitors of transcription or translation shortly before exposure to TGF-β and observed that EMT was completely blocked under these conditions (Supplementary Fig S1B).

To understand specifically which genes changed expression during TGF-β induced EMT, we performed high coverage deep-sequencing of the transcriptome (RNA-seq) at several timepoints of TGF-β exposure (day 0, day 1, day 4 and day 7) spanning early, intermediate, and late stages of EMT. The time-resolved TGF-β-dependent transcriptomes revealed many known EMT-associated changes in gene expression, including the classical cadherin switch, lincRNAs such as Malat1, other genes enriched for gene ontology (GO) annotations hallmark of EMT [189] [225], as well as a number of genes that were previously unknown to be modulated during specific phases of EMT (Fig 1, A-C and Supplementary Fig S1, C-J). We further confirmed that these transcriptional changes during EMT were specifically induced by TGFβ and were not a result of culture conditions (Supplementary Fig S1K).

To gain insight into the upstream regulators of observed transcriptional remodeling, we investigated the dynamics of Smad-dependent and Smad-independent TGF-β signaling during EMT. We found that phosphorylation of Smad2 was induced as early as 30 minutes and then declined 6 hours after exposure

to TGF-β (Fig 1D-E). In contrast, JNK showed a very distinct mode of activation as compared to Smad signaling. A high level of active JNK was detected only after 24 hours and this was maintained throughout EMT, including the later stages where Smad signaling loses activity (Fig 1D-E). We also observed a similar pattern of phosphorylation of Smad2 and JNK in human epithelial (MCF7) and mesenchymal (BT549) breast cancer cell lines (Supplementary Fig S2A). To investigate whether such patterns are also observed in human cancers of different invasiveness, we analyzed these signaling pathways in various breast cancer samples. These findings show that a significant fraction of invasive tumors exhibit high p-JNK and low p-Smad levels whereas this pattern was not observed in non-invasive tumors (Fig 1F-G, Supplementary Fig S2B-C, Supplementary Table S1).

We next assessed the contribution of canonical (Smad-dependent) and non-canonical (Smad-independent) pathways to TGFβ-induced JNK activation. Towards this, we analyzed phospho-JNK levels 24 hours after TGFβ-stimulation following Smad4 depletion or blockage of Akt or PI3 Kinase signaling. Interestingly, siRNA-mediated knockdown of Smad4 led to a significant reduction in p-JNK levels (Supplementary Fig S2D) while chemical inhibition of the Akt or PI3 Kinase pathway led to a significant increase in JNK activity (Supplementary Fig S2E).

Intrigued by the different activation kinetics of Smad and JNK pathways, we next attempted to test their functional requirement in distinct stages of TGF-β-induced EMT. siRNA mediated depletion of Smad4 severely impaired the onset of EMT (Supplementary Fig S2F). In contrast, inhibition of JNK signaling in early stages did not affect the induction of EMT (Supplementary Fig S2G). However, JNK inhibition in later stages of EMT led to a drastic downregulation of crucial mesenchymal markers such as fibronectin and severe reduction in remodeling of the cytoskeleton from cortical actin to stress fibers as well as focal adhesion formation indicating a reversal towards the epithelial state (Fig 1H). Such reversion of EMT following JNK inhibition was also accompanied by changes in the expression levels of classical EMT markers (Fig 1I). These effects of inhibiting JNK signaling were also reflected in cellular properties such as

a significant reduction in their migration capacity (Fig 1J). The effectiveness of JNK inhibitor was validated by an observed decrease in the phosphorylation of downstream substrates during different time points of TGF-β induced EMT (Supplementary Fig S3A, B, G and H). These observations were further validated using siRNA-mediated simultaneous knockdown of JNK1 and JNK2 (JNK1/2) as well as another established JNK inhibitor (JNK-IN-8), both of which caused similar reversal of EMT-associated features (Supplementary Fig S3, C-F). Overall, these interesting data uncover a kinetically distinct mode of action by signaling cascades during EMT where Smad pathway initiates EMT while JNK acquires a dominant role in its progression.

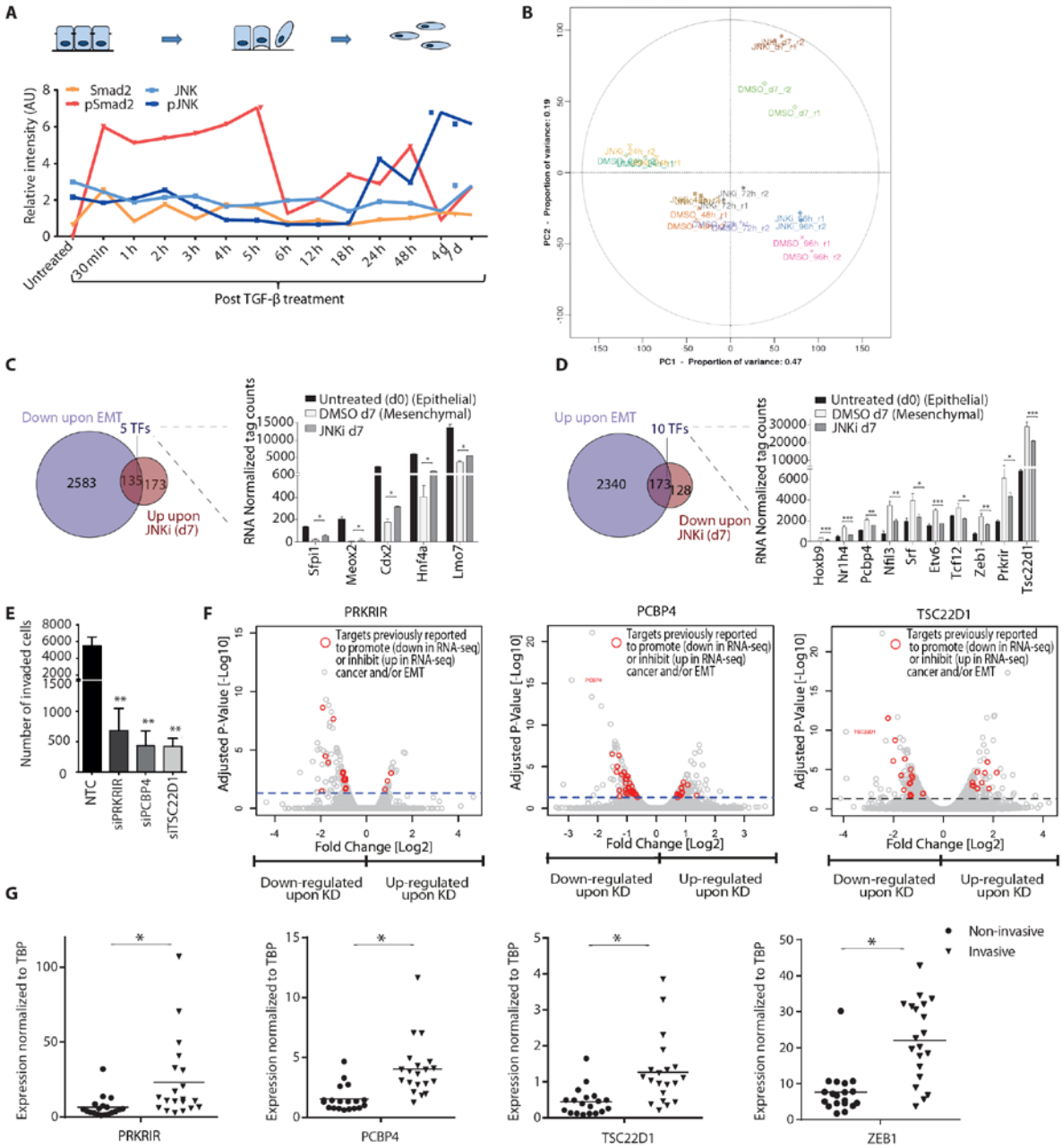


Figure 1. JNK signaling is activated during a distinct phase of EMT and becomes a critical determinant of mesenchymal identity. **(A)** Heatmap showing standardized expression (Z-scores) of genes differentially expressed upon TGF-β-induced EMT in NMuMG cells. **(B)** Venn diagram showing overlap of upregulated genes at individual time-points. **(C)** Same as in B for downregulated genes. **(D)** Representative immunoblots (n=3) showing phosphorylation state of Smad2 and JNK during TGF-β-induced EMT in NMuMG cells. Lamin B serves as loading control. **(E)** Quantification of one set of representative immunoblots (in figure D) and its graphical representation showing dynamic of Smad2 and JNK activation during TGF-β-induced EMT in NMuMG. **(F)** Representative immunohistochemistry image for pSmad2 and pJNK in non-invasive (n=34) and invasive (n=14) breast cancer samples. Scale bars for 10X and 40X images are 200 µm and 50 µm respectively. A relative score is assigned for pSMAD2 and pJNK staining in each clinical sample based on the IHC staining intensity (0 to 3 where 0 is lowest and 3 is highest). **(G)** Quantification of Immunohistochemical staining for p-JNK and p-Smad2 as in (F) in invasive tumor samples (n=34). The pie chart shows the percentage of tumors with high p-Smad2 and low p-JNK or high p-JNK and

low p-Smad2 or equal p-Smad2 and p-JNK. **(H)** Immunofluorescence microscopy for changes in the localization and expression levels of EMT markers in epithelial NMuMG cells (d0) and cells treated with TGF-β for seven days in the presence of JNK inhibitor (d7 JNKi) or control (d7 DMSO). Staining was performed with antibodies against Fibronectin 1, with Phalloidin to visualize the actin cytoskeleton and against Paxillin to detect focal adhesion plaques. Scale bar, 100 μm, 63X magnification. (I) mRNA levels for classical EMT markers in cells treated as in (F) were measured by qRT-PCRs relative to Ctcf and fold change as compare to untreated and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates. (J) Transwell migration assay for NMuMG cells treated as in G. Plotted Mean ± SEM of three biological replicates. * $p<0.05$, ** $p<0.01$, *** $p< 0.001$, Student's t-test. (Results in this figure partially contributed by me)

Mesenchymal cells rely on active JNK signaling to define epigenome and transcriptome underlying their identity

The unique pattern of JNK activation and its requirement in mesenchymal progression during EMT prompted us to investigate whether this function involves gene regulatory events. Towards this, we carried out genomewide transcriptome analysis (RNA-seq) at several time points during TGF-β induced EMT in continuous presence of the JNK inhibitor. We observed no changes in gene expression profile during early stages of EMT (Fig 2A), in agreement with the lack of phenotypic changes following JNK inhibition during early EMT (Supplementary Fig S2G). As cells progressed through advanced mesenchymal states, the effect of JNK inhibition on gene expression progressively became more prominent (Fig 2A). While first differentially expressed genes following JNK inhibition were visible at 72 hours of EMT (24 genes), major transcriptome changes were observed only at day 7 (611 genes) (Fig 2A, B and C). A Principle Component Analysis (PCA) on these datasets further confirmed the progressive dependency of EMT transcriptome on JNK signaling (Fig 2D). The genes upregulated upon JNK inhibition contained cell cycle regulators (Ccnd1, Ccnf and Ccng1) and epithelial genes (e.g. Fmn1, Brca1 and Ckap2) and were overall enriched for genes involved in defining epithelial identity (e.g. tight junction formation and cytoskeletal organization as well as cell cycle) (Supplementary Fig S4A and B). The downregulated genes included classical mesenchymal markers (e.g. Mcam, Fn1, Tnc) and GO term analysis further confirmed enrichment for genes defining mesenchymal properties (e.g. response to wounding, cytoskeletal reorganization and migration) (Supplementary Fig S4C and D). Analysis of lincRNA expression revealed that JNK inhibition also

affects expression of a selected set of lincRNAs only at day 7 of EMT (Supplementary Fig S4E and F). We further selected additional EMT relevant genes that were upregulated upon treatment with the JNK inhibitor (Supplementary Fig S4G) and analyzed their expression following addition of JNK inhibitor at various time-points during EMT. These genes were similarly affected irrespective of whether the JNK inhibitor was present from the very beginning or employed later following onset or progression of EMT (Supplementary Fig S4H). This was also reflected in analysis of EMT markers that showed similar reversal of EMT in all tested conditions (Supplementary Fig S4I).

Having established that the mesenchymal transcriptome becomes progressively dependent on JNK signaling during EMT, we next asked whether this involves interplay with epigenetic machinery at the promoters of target genes. Polycomb Repressive Complex 2 (PRC2) mark trimethylation of Lysine 27 of Histone H3 (H3K27me3) has been implicated to play a role in the transcriptional reprogramming underlying EMT [197]. Therefore, we performed ChIP assay using antibodies against the PRC2 mark H3K27me3 in epithelial and mesenchymal cells as well as mesenchymal cells treated with the JNK inhibitor. We next quantified H3K27me3 enrichments at a number of genes that were strongly downregulated upon EMT but were reverted back closer to epithelial cells following loss of JNK signaling. All tested genes showed a strong increase in H3K27me3 levels at their promoters during EMT, indicating that PcG complex is involved in their repression during EMT (Fig 2E). Strikingly, such deposition of H3K27me3 was noticeably reduced in the absence of JNK activity (Fig 2E). We further assessed whether such gain of PcG associated mark during EMT accompanies changes in chromatin accessibility at these genes and thereby leading to a transcriptionally silenced state. Using FAIRE assay, we indeed find that these promoters lose chromatin accessibility during EMT. Interestingly further, JNK inhibition restores a more open chromatin state at many of these genes (Fig 2F). Such chromatin changes were not observed at control loci (Supplementary Fig S4J). Together, these novel findings reveal that JNK signaling becomes

progressively critical in gene regulation underlying EMT and involves epigenetic modulation at target gene promoters.

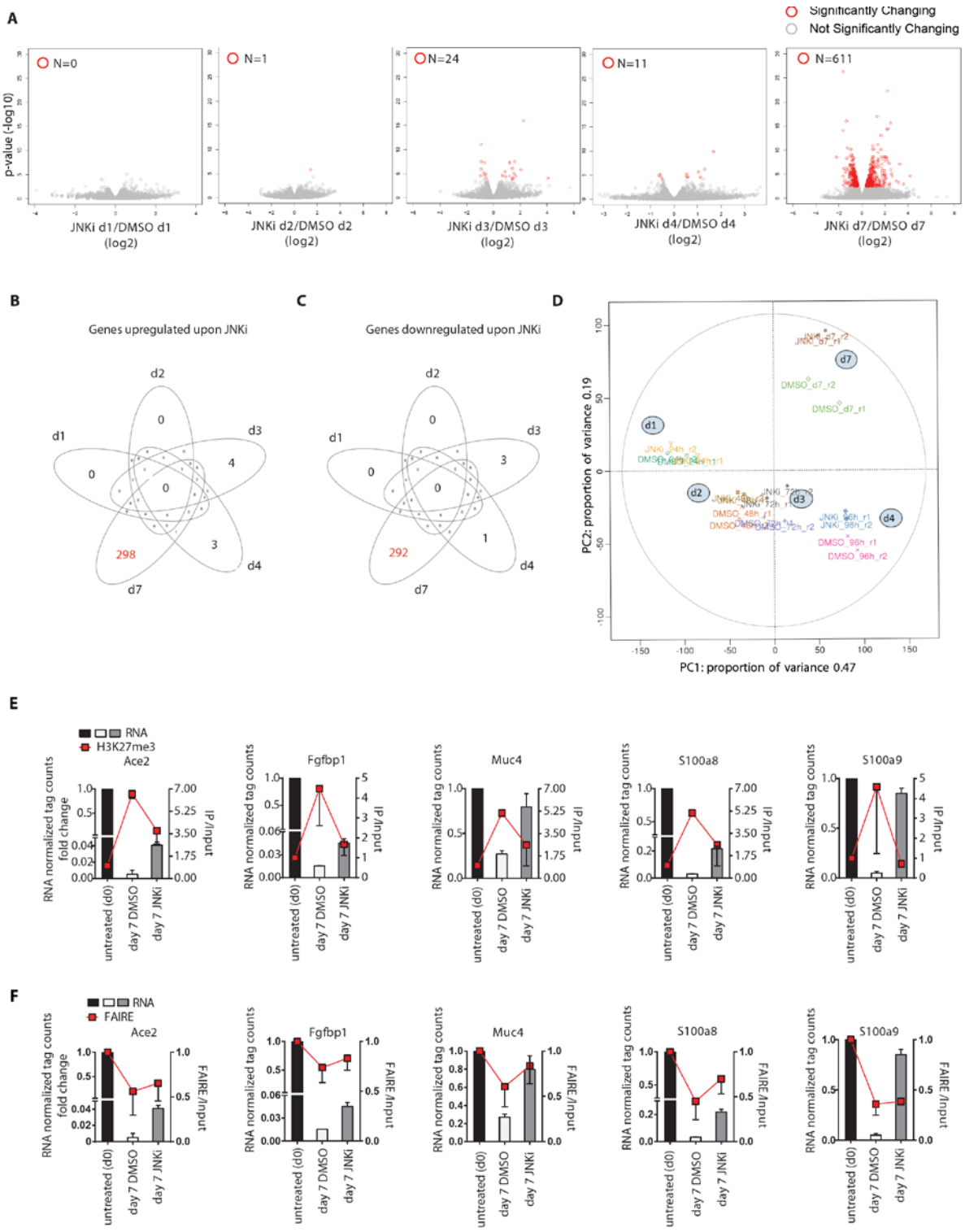


Figure 2. JNK signaling defines epigenome and transcriptome to specify mesenchymal fate. **(A)** Volcano plots showing significantly differentially expressed genes upon treatment with the JNK inhibitor SP600125

(JNKi) as compared to DMSO at day 1, 2, 3, 4 and 7 during TGF-β-induced EMT in NMuMG cells. X-axis indicates fold change in log2 scale while y-axis denotes -log 10 of the P-value. Differentially expressed genes are depicted with red circles. Blue horizontal line represents FDR adjusted Pvalue of 0.1. **(B)** Venn diagram showing the overlap of upregulated genes in JNKi versus DMSO treated cells. **(C)** Same as in (B) for genes downregulated upon JNKi. **(D)** A principal component analysis (PCA) using RNA-Seq data was performed comparing two replicates (r1, r2) each of DMSO and JNKi treated samples at different time points during TGF-β-induced EMT (d1, d2, d3, d4, d7). The first two principal components are presented to display the variation between samples. The distances on the x- and y-axis explain 47% and 19% of variation in the data respectively. Biological duplicates (r1 and r2) are depicted in the same color. **(E)** ChIP assay using H3K27me3-specific antibody in epithelial cells (day 0)) or NMuMG cells treated for seven days with TGF-β (mesenchymal) and DMSO or JNKi. qRT-PCRs were performed for indicated gene promoters and enrichments are plotted on the right y-axis as ratio of precipitated DNA (bound) to total input DNA. On the left y-axis RNA levels of analyzed genes derived from RNA-seq data (fold change to untreated normalized tag counts) are plotted. Mean and SEM is plotted from two independent biological replicates. **(F)** FAIRE assay was performed in cells treated as in (D) and qRT-PCRs were performed for the same gene promoters and plotted as in (E). (Results in this figure partially contributed by me)

JNK-induced novel transcription factors are critically required for EMT

To uncover downstream components of JNK-dependent gene regulatory network that could be critically involved in mesenchymal progression, we further explored the identity of genes that are misregulated in the absence of JNK signaling. Towards this, we compared JNK inhibitor affected genes (Fig 2B and C) with the transcriptome changes acquired during normal EMT (Fig 1B and C). A comparative analysis revealed that 44% (n=135) of genes that are upregulated in response to JNK inhibition were repressed during normal EMT (Fig 3A). These genes show enrichment for GO terms associated with an epithelial cell-fate such as cell-cycle (Supplementary Fig S5A). Furthermore, overlap of genes downregulated upon JNK inhibition with genes that are induced during EMT revealed 57% (n=173) of genes in common (Fig 3B). These genes included hallmark mesenchymal genes (e.g. Fn1, Cdh2, Mcam, Zeb1 and Tnc) and showed GO term enrichments such as actin cytoskeletal reorganization, cell junction and extracellular matrix organization (Supplementary Fig S5B). These findings further support our previous findings that the blockage of JNK signaling is able to significantly reverse the gene expression signature acquired during EMT. Having observed that a large number of genes rely on active JNK signaling for proper transcriptional dynamics during EMT, we next assessed whether JNK pathway regulates expression of any gene regulatory proteins during EMT that are then involved in mediating JNK-dependent transcriptional

program. Computational analysis searching for transcription factors in these overlap lists revealed five and ten specific transcription factors that were up and downregulated respectively following JNK inhibition (Fig. 3A and B). These patterns were further validated by independent expression analyses (Supplementary Fig S5C and D).

Out of five transcription factors (Sfpi1, Meox2, Cdx2, Hnf4a, and Lmo7) that are downregulated during EMT but transcriptionally induced following inhibition of JNK signaling, all are known to inhibit EMT and tumor progression as well as promote epithelial properties [226-232]. These data thus uncover JNK signaling as a novel upstream regulator functioning in suppression of these critical epithelial determinant genes during EMT. Among the ten transcription factors that are upregulated during EMT but become downregulated upon JNK inhibition we identified Zeb1 and Srf that have an established role in tumor progression [189,200,233,234]. Importantly however, remaining eight factors have been either only partly (Tcf12, Etv6, Hoxb9) or never (Nr1h4. Pcbp4, Prkrir, Tsc22d1, Nfil3) investigated in the context of EMT. We, therefore, decided to explore the functional role of these eight factors in EMT progression. Towards this, we employed siRNA mediated depletion of these factors during EMT and assessed its impact on cellular and molecular changes that accompany this process. Strikingly, knockdown of these factors perturbed morphological features hallmark of EMT as well as expression pattern of key EMT genes (Fig 3C, S6A-B (for completely novel factors); Supplementary Fig S6C-E, (for partly implicated factors)). These changes include drastic reduction in crucial mesenchymal markers such as Fibronectin, overall increase in expression and relocalization of ZO-1 and E-cadherin to the membrane, reduction in remodeling of the cytoskeleton from cortical actin to stress fibers as well as focal adhesion formation, in addition to the increased cell to cell contact. These observations were further validated using independent siRNAs (Supplementary Fig S7A-D).

We next attempted to identify downstream components in the JNK pathway through which it may regulate expression of these genes. Towards this we performed ChIP assay using antibodies against

various transcription factors that are established JNK effectors in different contexts viz. ATF4, ATF2, EGR1, c-JUN, NF-YA and ELK1 at day 7 of TGFβ-induced EMT in the presence and absence of JNK inhibitor and analyzed the promoter of these genes by quantitative PCRs. While most of these transcription factors did not show any enrichment, ATF2 was found to be binding at these gene promoters in mesenchymal cells and this occupancy was significantly diminished following JNK inhibition (Fig S8A). Furthermore, such reduced binding of ATF2 at target genes following JNK inhibition was accompanied by a diminished nuclear localization of ATF2 (Fig S8B).

The transcriptional regulation of these factors by JNK signaling was further validated by treatment with an alternate JNK inhibitor, JNK-IN-8 and siRNA mediated knockdown of JNK1/2 during TGF-β induced EMT that also showed downregulation of these genes following loss of JNK activity or protein (Supplementary Fig S8C-D). Encouraged by our findings of a functional role of these novel factors in EMT, we next induced MET by TGF-β withdrawal in mammary mesenchymal cells that have originally undergone EMT in response to TGF-β (Supplementary Fig S8E). Interestingly, such induction of MET led to a loss of JNK signaling (Supplementary Fig S8F) and transcriptional repression of all novel transcription factors (Supplementary Fig S8G). These findings establish JNK signaling in regulation of previously known important regulators of EMT, at the same time discover a new set of JNK-induced transcription factors that function as critical regulators of mesenchymal progression.

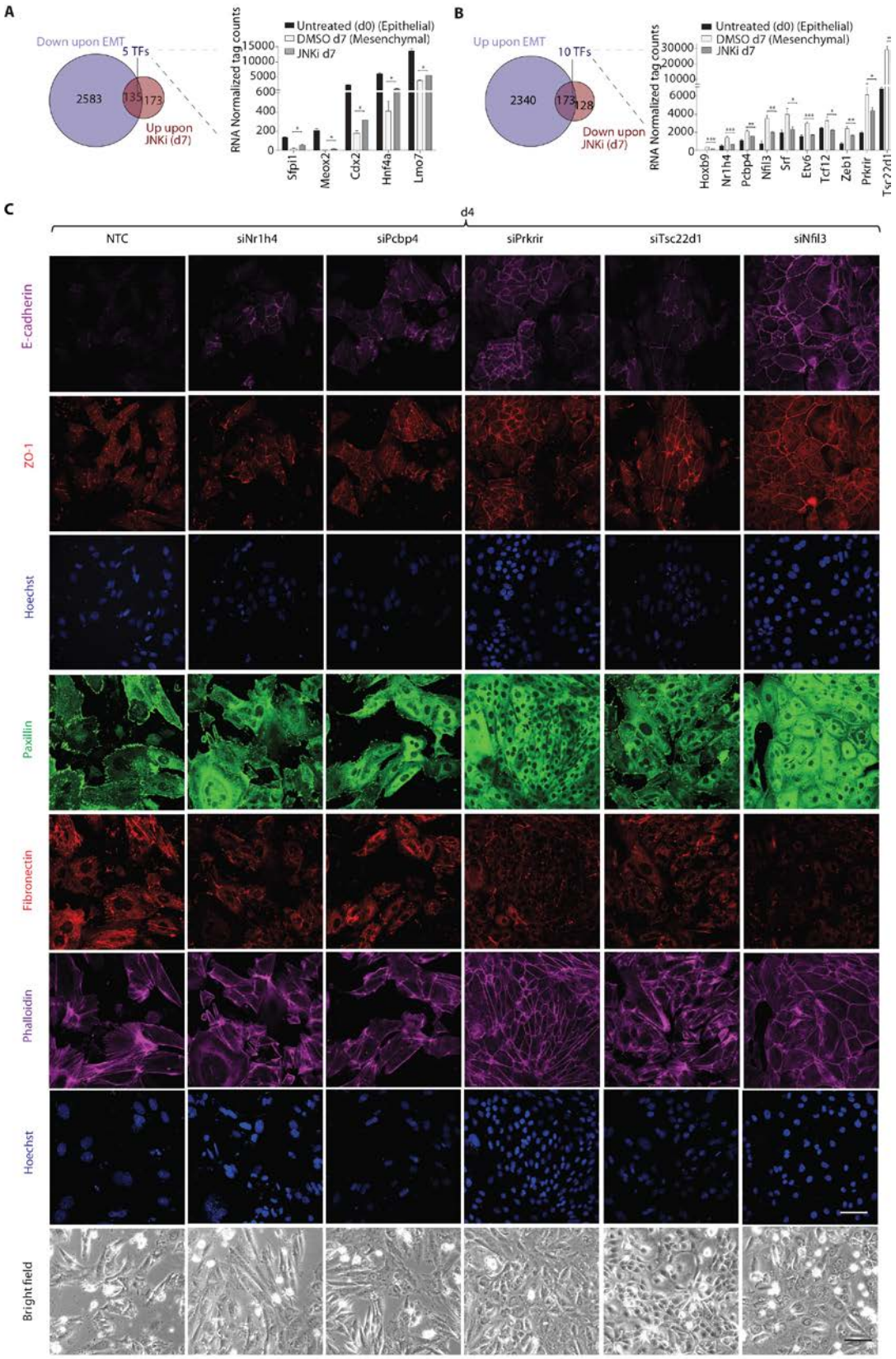


Figure 3. JNK-induced transcription factors are critically required for EMT**. (A)** Venn diagram showing overlap of genes upregulated upon JNKi treatment during EMT and downregulated during normal EMT at

day 7 that contains 5 transcription factors. Right panel shows RNA levels (normalized tag counts) for these factors. **(B)** Same overlap as in (A) for genes downregulated upon JNKi treatment during EMT and upregulated during normal EMT. Right panel displays RNA levels as in (A) for 10 transcription factors identified in this overlap. **(C)** Representative bright field and immunofluorescence images for localization and expression levels of EMT marker proteins after four days (d4) of siRNA-mediated depletion of factors compared to non-targeting control (NTC). Staining was performed with antibodies against epithelial markers E-cadherin and ZO1, mesenchymal markers Fibronectin 1, with Phalloidin to visualize the actin cytoskeleton and against Paxillin to detect focal adhesion plaques. Scale bar, 100 µm, 40X magnification. Plotted Mean ± SEM of three biological replicates. * $p<0.05$, ** $p<0.01$, *** $p< 0.001$, Student's *t*-test. (Results in this figure contributed by me)

Depletion of JNK-induced novel transcription factors during TGF-β induced EMT confers proliferative properties

One of the hallmarks of the epithelial state is their high proliferative capacity. We therefore attempted to ask whether the novel factors induced by JNK during EMT promote mesenchymal fate by reducing the rate of cell division. To address this in detail, we performed time-lapse video microscopy and single cell tracking of epithelial cells and during TGF-β-induced EMT following depletion of three of the selected factors, Tsc22d1, Prkrir and Pcbp4. Analysis of several experiments revealed a marked decrease in the number of cell divisions within the time window of tracking during TGF-β-mediated EMT as compared to the epithelial cells (Fig 4A-B, Supplementary Fig S9). Interestingly however, depletion of any of these three novel factors during EMT was sufficient to significantly rescue the effect of TGF-β on cell division and led to a noticeable increase in cell proliferation (Fig 4A-B, Supplementary Fig S9). This analysis further showed that the knockdown cells acquire morphology that closely resembles epithelial cells (data not shown). Overall, these experiments tracking cell-fate changes of single cells during EMT further showed how JNK-induced novel transcription factors contribute to various aspects of the mesenchymal fate, including regulation of cell division.

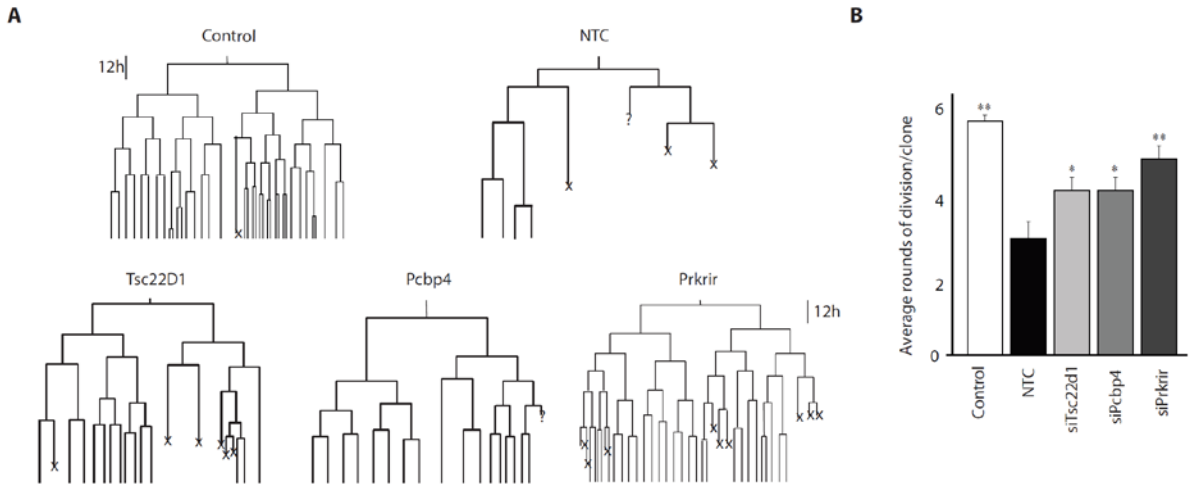


Figure 4. Depletion of Prkrir, Pcbp4 and Tsc22d1 rescues TGF-β-induced reduction in the rate of cell division. **(A)** Lineage trees tracked by live imaging under control, Non treated control (NTC) and siRNA-mediated depletion of factors (X= cell death/cells could not be tracked till the end). **(B)** Average number of rounds of division per clone in NTC compared with control and siRNA-mediated depletion of factors (ANOVA, Tukeys post-test, *=p<0.05,**=p<0.01). (Results in this figure partially contributed by me)

PRKRIR, PCBP4 and TSC22D1 are crucial for maintenance of the metastatic state

To extend our observations to the human contexts of metastasis, we next inhibited JNK signaling in an aggressive, human mesenchymal breast cancer cell line (MDA-MB-231). Such treatment led to a massive loss in their wound healing capacity, suggesting that JNK signaling is critical in maintenance of the metastatic state in these cells (Fig 5A). Furthermore, this also accompanied upregulation of epithelial markers and downregulation of mesenchymal markers, suggesting differentiation towards an epithelial state (Fig 5B). Interestingly, blockage of TGF-β receptor in these cells did not inhibit JNK signaling, but rather led to an increase in phospho-JNK levels that also accompanied a slight increase in their wound healing potential (Supplementary Fig S10A-B). These results further suggest that during tumor progression mesenchymal breast cancer cells may evolve TGF-β-independent mechanisms to sustain high JNK activity that allows maintenance of the mesenchymal fate independent of the signal (viz. TGF-β) that originally induced EMT.

Driven by these observations, we were next interested in investigating whether JNK-induced novel transcription factors also have a role in maintenance of the mesenchymal state, in addition to their

observed role in its progression. Interestingly, all newly identified transcription factors show much higher expression levels in the aggressive, mesenchymal breast cancer cell line (MDA-MB-231) compared to the isogenic non-invasive epithelial cell line (MDA-MB-361) (Fig 5C, S10C) (except NR1h4, which is not expressed in these cells). Prompted by these findings, we next assessed whether these factors are involved in maintaining the mesenchymal properties and thus the metastatic potential. Three of these factors (HOXB9, TCF12 and ETV6) have been partly implicated in promoting metastasis and we therefore chose the remaining four novel transcription factors (PRKRIR, NFIL3, PCBP4, TSC22D1) for further functional analysis. siRNA-mediated knockdown of these factors in the MDA-MB-231 cells showed a noticeable reduction in their wound healing potential (Fig 5D; Supplementary Fig S10D-F). Importantly further, such depletion of these factors also led to a significant reduction in the invasive capacity of these cells, a property that reflects a critical loss in their metastatic potential (Fig 5E). These effects following knockdown of JNK-induced factors were not linked to alternations in cell-viability or cell-cycle progression (Supplementary Fig S10G-J).

We next asked whether the restoration of these transcription factors is able to counteract the effects of JNK inhibition in these cells. While control breast cancer cells treated with the JNK inhibitor showed gain of epithelial markers and loss of mesenchymal markers, overexpression of TSC22D1 and PCBP4 was able to significantly rescue cells from these JNK inhibitor induced changes (Fig 5F). These observations suggest that these factors indeed function downstream of JNK activity in breast cancer cells in maintenance of the mesenchymal state. Overall, these findings establish that the identified novel transcription factors are not only important for progression but also in maintenance of the mesenchymal state.

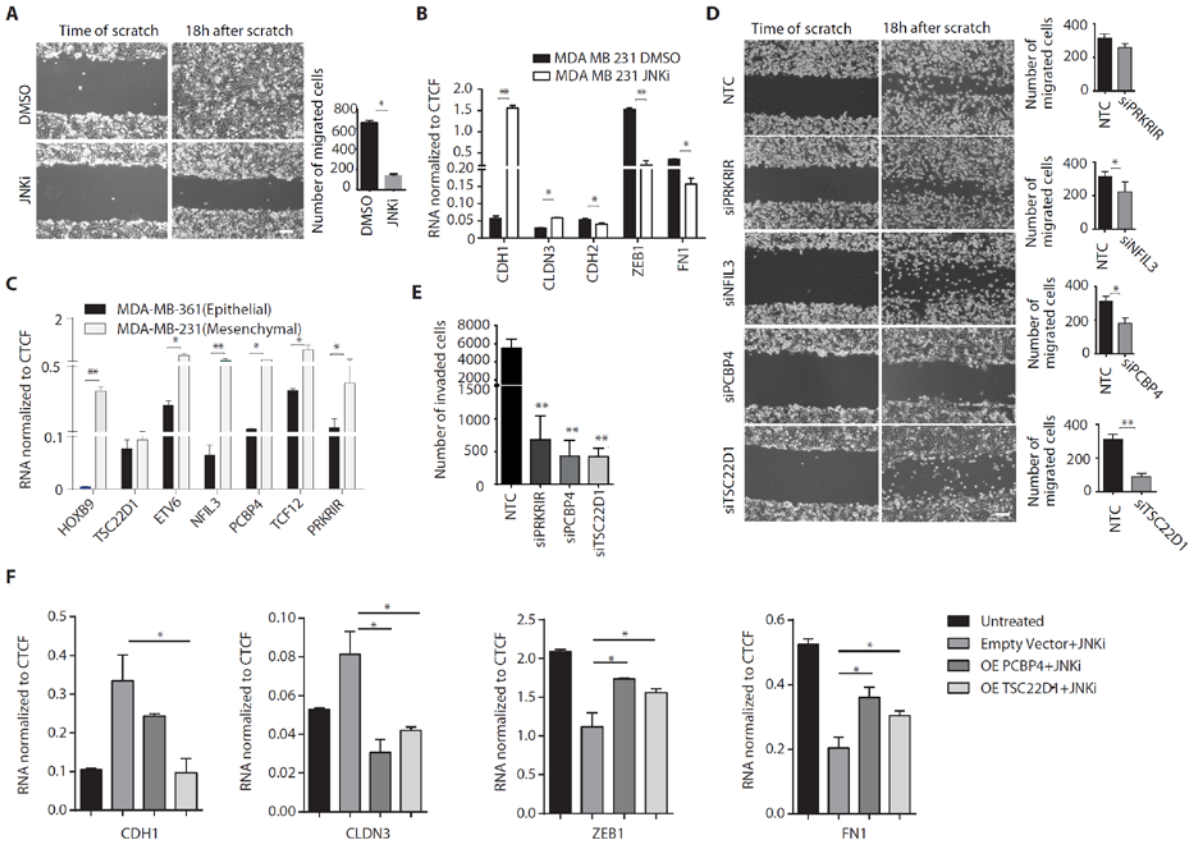


Figure 5. PRKRIR, PCBP4 and TSC22D1 are crucial for maintenance of the mesenchymal fate. **(A)** Scratch assay in MDA-MB-231 cells treated for four days with DMSO or JNKi. Scale bar, 200 µm; 20 X magnification. Mean and SEM is plotted from three independent biological replicates. **(B)** mRNA levels for key EMT marker genes in MDA-MB-231 treated with DMSO or JNKi for 4 days were measured by qRT-PCR relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates **(C)** mRNA levels for transcription factors in isogenic epithelial (MDA-MB-361) and mesenchymal (MDA-MB-231) breast cancer cells were measured by qRT-PCR relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates. **(D)** Scratch assay in MDA-MB-231 cells transfected with NTC or siRNA against PRKRIR, NFIL3, PCBP4 and TSC22D1 for 4 days as in A. Mean and SEM is plotted from three independent biological replicates as bar plot on right . **(E)** Invasion assay in MDA-MB-231 cells treated as in (D). Mean and SEM is plotted from three independent biological replicates. **(F)** mRNA levels of key EMT marker genes in MDA-MB-231 and MDA-MB-231 transfected with either empty vector or vector overexpressing PCBP4 or TSC22D1 for 48 hours and then treated with the JNK inhibitor (SP600125) for 24 hours were measured by qRT-PCRs relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates. Plotted Mean ± SEM of three biological replicates. * p<0.05, ** p<0.01, *** *p*< 0.001, Student's *t*-test. (Results in this figure contributed by me)

PRKRIR, PCBP4 and TSC22D1 regulate distinct set of genes in breast cancer cells

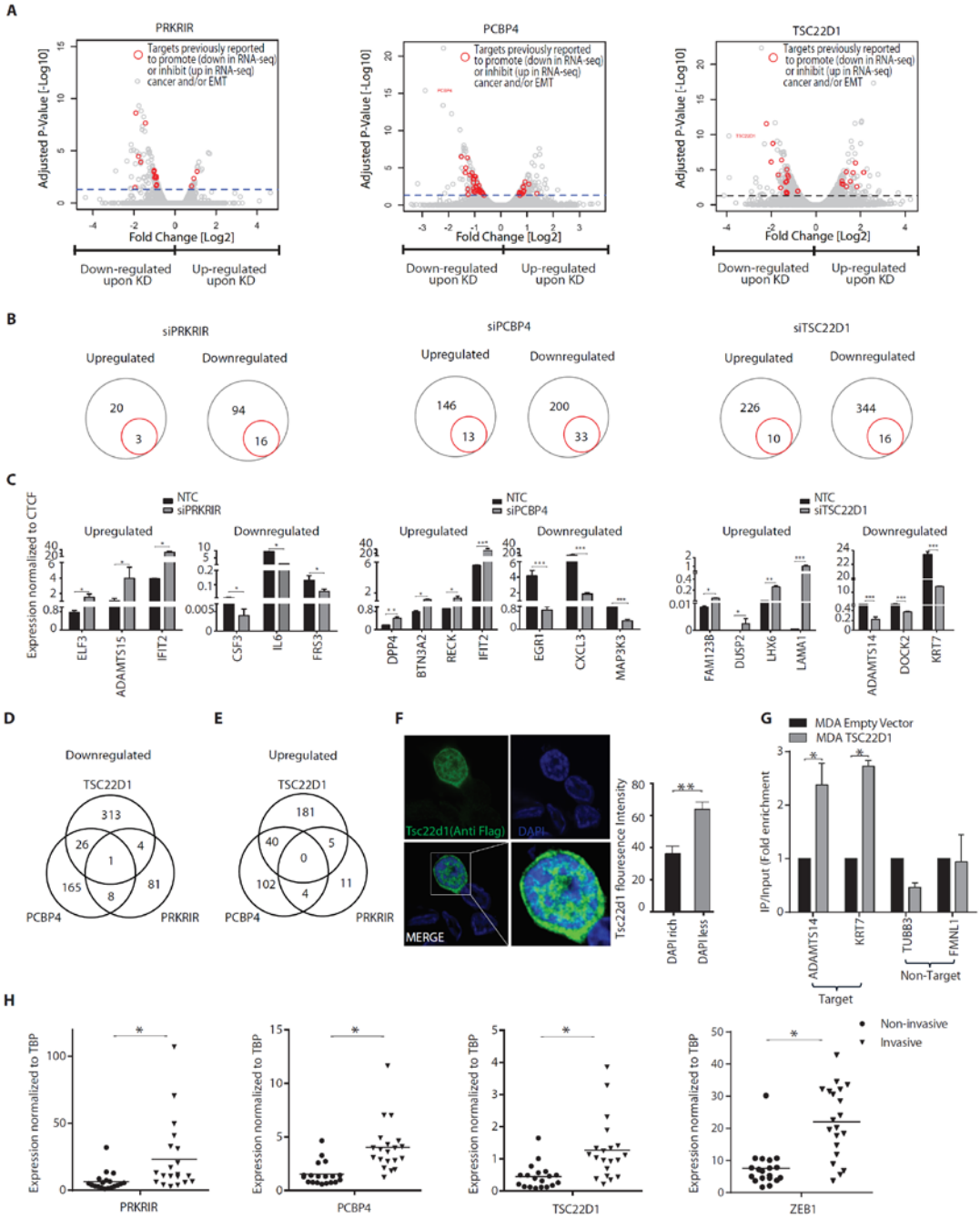
To further uncover genes under regulation of these transcription factors, we performed genomewide transcriptome profiling (RNA-seq) following their siRNA-mediated depletion in these cells. Strikingly,

computational analysis revealed a large number of genes that were significantly misregulated following knockdown of PRKRIR, PCBP4 and TSC22D1 (Fig 6A). To our surprise, with a stringent cut-off for significance, NFIL3 knockdown did not show any aberrantly expressed genes. Interestingly further, several genes that were downregulated following knockdown of these three novel factors have previously been shown to promote metastasis (Fig 6B; Supplementary Table S2). Similarly, a number of genes that were upregulated following depletion of these factors are known to be of anti-tumorigenic nature (Fig 6B; Supplementary Table S2). We further quantified expression of a number of these misregulated genes in independent knockdown experiments and the results fully validated the changes observed in the RNA-Seq assay (Fig 6C). Furthermore, an overlap between the differentially expressed genes in cells depleted of these three factors did not show extensive overlap, suggesting that these proteins function to regulate distinct set of genes to maintain mesenchymal properties (Fig 6D and E).

We next attempted to explore whether these transcription factors are directly targeting the regulatory elements of misregulated genes. To test this, we chose and characterized TSC22D1 for its distribution and DNA binding activity. Following transient expression, TSC22D1 showed an exclusive nuclear localization (Fig 6F). Interestingly, a deeper analysis revealed that TSC22D1 preferentially localizes with the DAPI-less regions in the nucleus that are known to mark euchromatin and transcriptionally active sites in the genome (Fig 6F). Given these patterns, we performed Chromatin Immunoprecipitation (ChIP) assay for TSC22D1 in MDA-MB-231 cells and analysis showed that it directly binds at the promoter of its target genes (Fig 6G). These TSC22D1 bound genes were discovered to be downregulated following TSC22D1 depletion in MDA-MB-231 cells (Fig 6C). These observations imply that TSC22D1 is targeted to active chromatin sites in the genome where it functions in transcriptional activation of critical EMT genes.

In order to extend these findings to clinic, we collected many non-invasive (n=20) and invasive tumor samples (n=20) and analyzed expression of the three novel transcription factors TSC22D1, PRKRIR, and PCBP4 as well as ZEB1, which is an established marker of highly invasive tumors. Interestingly, all three

factors were significantly higher expressed in invasive tumors as compared to noninvasive tumors (Fig 6H) that we have also shown to exhibit high phospho-JNK levels (Fig 1F-G). These observations clearly establish the clinical relevance of our findings and support our hypothesis that these factors are critical for the metastatic progression of breast cancer cells. Altogether, these findings imply that the JNK-induced novel transcription factors function in the transcriptional regulation of metastasis relevant genes in invasive



breast cancer cells.

Figure 6. PRKRIR, PCBP4 and TSC22D1 regulate expression of distinct set of genes in mesenchymal breast cancer cells. **(A)** Volcano plots showing significantly changing genes upon siRNA-mediated depletion of PRKRIR, PCBP4 and TSC22D1 as compared to NTC. The x-axis and y-axis show fold expression changes (log2 scale) and associated P-value (-log10 scale) respectively. The blue horizontal line marks the cut-off for genes considered significantly changing upon factor depletion. Red dots indicate genes previously implicated in a similar context (Supplementary Table 1). **(B)** Circle plots of genes highlighted in (A) (red) and all significantly changing genes (grey). **(C)** mRNA levels for representative genes highlighted in (A) for cells treated like in (Fig 5D) were measured by qRT-PCR relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates. **(D-E)** Venn diagrams showing the overlap of the genes differentially upregulated and down regulated following knockdown of the three factors mentioned in (A)**. (F)** Representative immunofluorescence images showing localization of Tsc22d1 in NMuMG cells following transient expression of Flag-HA tagged Tsc22d1 and detection using an anti-Flag antibody, Bar plot shows quantification of Tsc22d1 localization with respect to DAPI density within the nucleus. Y-axis shows the percentage of total fluorescence intensity for Tsc22d1 in DAPI dense and light regions. SEM is derived from independent biological replicates. **(G)** ChIP assay using anti-Flag antibody following expression of Flag-HA tagged TSC22D1 in MDA-MB-231 cells. Quantitative PCRs were performed for indicated gene promoters and enrichments are plotted on the y-axis as ratio of precipitated DNA (bound) to total input DNA and then further divided by the same obtained in the empty vector transfected cells. SEM is derived from independent biological replicates. **(H)** mRNA levels for PRKRIR, PCBP4, TSC22D1 in non-invasive (DCIS) (n=20) and invasive tumor (n=20) samples were measured by qRT PCRs relative to TBP and plotted on the y-axis. ZEB1 was used as an established positive control for invasive tumors. Plotted Mean ± SEM of three biological replicates. * p<0.05, ** p<0.01, *** *p*< 0.001, Student's *t*-test. (Results in this figure contributed by me)

Newly identified transcription factors are similarly upregulated during neurogenesis and are required for neuronal EMT

We next attempted to investigate whether the newly identified factors are also components of the core machinery involved in other contexts of EMT such as during development. The differentiation of embryonic stem cells was described to exhibit features of EMT in various studies [235-237]. Therefore we employed mouse embryonic stem cells (ESCs) that differentiate under well-defined conditions into Pax6-positive neural progenitors (NP), that go on to differentiate into post-mitotic glutamatergic neurons (TN) [118,152,238]. We find that the transition from embryonic stem cells to differentiated neurons indeed involves gene expression changes that are hallmark of EMT such as classical cadherin-switch (Fig 7A). Furthermore, this process also accompanied an activation of JNK signaling at later stages (Fig 7B) [153]. Interestingly further, concomitant with the induction of JNK activity, these newly identified transcription factors are significantly increased in their expression during neuronal differentiation (Fig 7C). Having observed a

correlation with JNK activity and induction of these novel transcription factors during neurogenesis, we investigated a previously published dataset comparing the transcriptome of neurons following a short-term exposure to the JNK inhibitor SP600125 [153]. Interestingly, the expression of all three transcription factors was significantly reduced following such inhibition of JNK signaling in neurons (Fig 7D).

To investigate whether the target genes of these transcription factors show any specific patterns during neuronal EMT, we chose Tsc22d1 for further validation. We selected a set of genes that were downregulated in TSC22D1-depleted MDA-MB-231 cells with the additional criteria that they have a known role in EMT and neurogenesis (Supplementary Fig S11). Interestingly, these genes also showed a significant increase in their expression during neuronal differentiation of embryonic stem cells (Fig 7E). We next attempted to further validate these observations *in vivo* by analyzing their expression in various cortical layers during mouse embryonic development. In support of our observations, Tsc22d1 as well as these target genes were similarly induced during neuronal development *in vivo* where radial glial cells (neuronal progenitors) generate neurons that then migrate to the cortical layer using EMT-like mechanisms (Fig 7F). These genes were also found to be significantly induced during EMT in mouse mammary epithelial cells (Fig 7G).

To functionally test the requirement of Tsc22d1 in neuronal migration, we attempted to knockdown Tsc22d1 *in vivo* during mouse brain development. Towards this, we performed *in utero* electroporation assay in mouse cortex at E12.5 with plasmids containing either an established non-target (control) shRNA or a validated shRNA against Tsc22d1 as described previously [239] and sacrificed the animals at E16.5 for sectioning and analysis (Fig 7H). Interestingly, such knockdown of Tsc22d1 led to a very strong reduction in the number of electroporated cells (GFP positive) in the upper neuronal layer and a noticeable retention in the lower layer, reflecting a defect in neuronal migration towards the cortical plate (CP) (Fig 7**I** and J). These observations establish that Tsc22d1 is indeed critical for neuronal EMT during mouse brain development. These findings further highlight that the EMT machinery involved in both development and

disease may involve common set of transcription factors that function via regulating a similar set of EMT genes and places JNK-induced transcription factors among these core EMT regulatory factors.

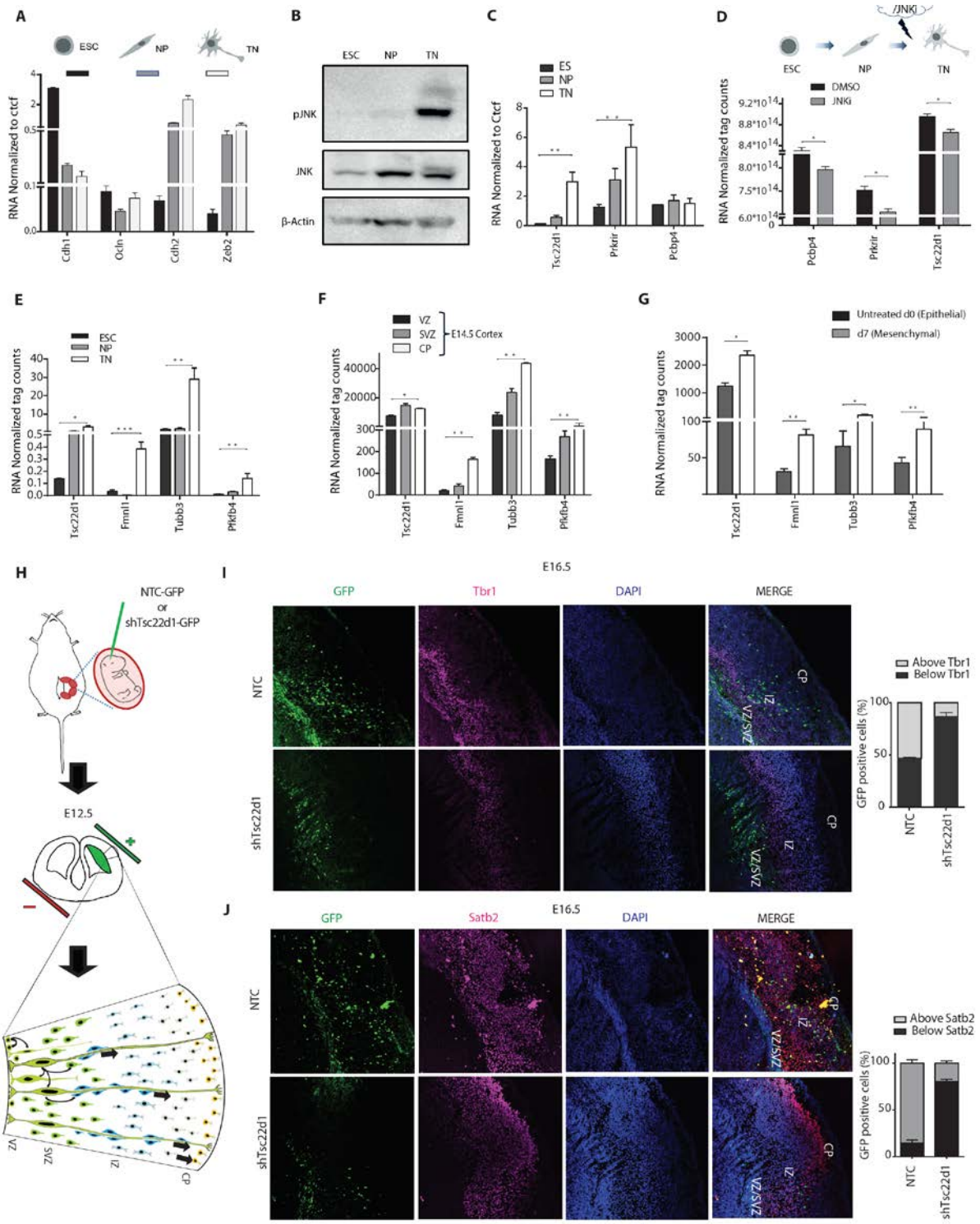


Figure 7. Newly identified transcription factors are similarly upregulated during neurogenesis and function in neuronal migration. **(A)** RT-qPCR analysis for key EMT markers in embryonic stem cells (ES), neuronal progenitors (NP) and terminally differentiated neurons (TN). mRNA levels were measured by RT-qPCR

relative to Ctcf and plotted on the y-axis. Error bars represent SEM from independent biological replicates. **(B)** Western blot analyzing total JNK and phospho-JNK (p-JNK) during neuronal differentiation of stem cells. β-actin serve as a loading control. **(C)** RNA levels of Pcbp4, Prkrir and Tsc22d1 during neuronal differentiation as shown in A, measured by RT-qPCR relative to Ctcf and plotted on the y-axis. Error bars represent SEM from independent biological replicates. **(D)** RNA levels of Pcbp4, Prkrir, Tsc22d1, represented by average counts derived from previously published microarray data of independent biological replicates from neurons treated with DMSO or JNK inhibitor SP600125 (JNKi) for 6 hours. **(E)** RNA levels of Tsc22d1 and three Tsc22d1 target genes during various stages of neuronal differentiation shown in (A), measured by RT-qPCR relative to Ctcf and plotted on the y-axis. Error bars represent SEM from independent biological replicates. **(F)** RNA levels of Tsc22d1 and the same three target genes is shown as average normalized tag counts derived from RNA-seq data from Ventricular Zone (VZ), Sub-Ventricular Zone (SVZ) and Cortical plate (CP) of E14.5 mouse Cortex. **(G)** RNA levels of genes shown in Figure (E), represented by average normalized tag counts derived from RNA-seq data of TGF-β-induced EMT in NMuMG cells and plotted on the y-axis. Error bars represent SEM from three biological replicates. **(H)** Graphical representation of *In Utero* Electroporation (IUE) performed at the embryonic stage 12.5 (E12.5) during murine development. E12.5 mouse embryos were electroporated with GFP-tagged constructs either with an established non-target control (NTC) shRNA or a validated shRNA against Tsc22d1. Four days later (E16.5), embryos were sacrificed and brains were fixed for immunofluorescence analysis with antibodies against GFP and Tbr1 or Satb2 (early neuronal markers). **(I)** A representative image from immunofluorescence analysis performed with anti-GFP and anti-Tbr1 antibody that shows retention of GFP positive cells below the Tbr1 layer in Tsc22d1 knockdown brain as compared to the control brain. The bar plot on the right side shows quantification of migrated GFP positive cells in control and Tsc22d1 depleted mouse brain with respect to the Tbr1 staining. y axis represent percentage of cells above or below the Tbr1 stain region. Error bars represent SEM from 3 independent biological replicates. **(J)** Similar analysis as in (I) but with respect to Satb2 staining. * $p<0.05$, ** $p<0.01$, *** $p< 0.001$, Student's *t*-test. (Results in this figure partially contributed by me)

DISCUSSION

Epithelial to mesenchymal transition (EMT) plays crucial roles in generating the body plan during embryonic development by contributing to the morphogenesis of multiple tissues and organs [188,190,191]. This process is further employed during wound healing and tissue regeneration in adults. However, its aberrant activation is known to cause organ fibrosis and promote carcinoma progression through a variety of mechanisms [188,189,191-194,196,197,200,215-217]. EMT confers cells with migratory and invasive properties, induces stem cell properties, prevents apoptosis and senescence and contributes to immunosuppression [188,191]. Thus, the acquisition of mesenchymal state accompanies gain in the capacity of cells to migrate to distant organs and maintain stemness, allowing their subsequent differentiation into multiple cell types during development and the initiation of metastasis during carcinogenesis. Our high resolution, high coverage deep-sequencing transcriptome data for various steps during progression from epithelial to mesenchymal state including later mesenchymal stages would be of vital importance to researchers studying EMT in these various contexts. These datasets will serve as a useful resource to not only study gene expression program during EMT but to also pursue functional investigation of novel coding and long noncoding RNAs that we discovered to be transcriptionally modulated during distinct phases of this phenotypic transition.

TGF-β-induced EMT has been shown to involve activation of a number of non-canonical pathways including MAP kinases such as ERK, p38 and JNK. However, the induction kinetics of these pathways as well as their crosstalk to canonical Smad signaling during EMT has not been explored in detail so far. Our observations show that while Smad signaling is required for the onset of EMT, JNK pathway is critical for the progression and maintenance of phenotypic and cellular changes associated with EMT. We further established the clinical relevance of these findings by showing that invasive tumors exhibit high JNK signaling and low Smad activity as compared to non-invasive tumors.

A number of previous, small-scale studies have suggested that JNK activation may contribute to the TGF-β-induced gene regulation including expression of Fibronectin and of other profibrotic genes [220,223,240,241]. However, these were single gene studies that neither uncovered all the genes regulated by JNK signaling nor investigated the differential requirement of JNK signaling during distinct phases of mesenchymal progression. Our genomewide expression profiling in combination with morphological analysis following JNK inhibition at various phases of EMT revealed that JNK signaling has no functional contribution in the onset of this process. Interestingly however, as cells progress through the mesenchymal state, JNK signaling gradually becomes a crucial transcriptional regulator of many genes that includes activation of established mesenchymal genes (e.g. Zeb1, Ncam, Mcam and Tnc1) and repression of known epithelial genes (e.g. Cdh1, Ccnd1, Ccnf and Ccng1) and therefore, is required to achieve morphological features typical of a progressed mesenchymal state. This study goes beyond previous notions on the involvement of JNK signaling during EMT, detailing kinetically distinct modes of JNK-activity during the process and establishing its role in genomewide transcriptome remodeling that drives mesenchymal progression.

It is increasingly being recognized that EMT involves active participation of epigenetic regulatory pathways [197,242,243]. However, there are no existing observations that link upstream signaling to the regulation of chromatin state in driving this process. We find that a number of genes that were downregulated upon EMT acquired PcG complex associated repressive mark H3K27me3 at their promoters in a JNK-signaling dependent fashion. Furthermore, JNK signaling also mediates a compact chromatin state of these gene promoters, facilitating their silencing during EMT. While the precise molecular mechanism underlying JNK signaling mediated regulation of chromatin states during EMT remains unclear, it is possible that JNK pathway modulates activity of certain epigenetic regulators or other factors involved in targeting of these proteins to distinct genomic sites. It also remains probable that JNK directly modifies chromatin as shown previously in the context of neurogenesis [153]. Nevertheless, these findings have uncovered a previously unknown function of JNK signaling in modulating the epigenetic state and chromatin packaging at target

gene promoters to mediate transcriptional reprogramming underlying EMT. These findings further set a new dimension how external signaling could modulate epigenome to drive changes in cell-fate.

The transcriptional reprogramming underlying transition from epithelial to mesenchymal fate is known to involve function of a number of critical transcription factors (e.g. ZEB1, ZEB2, TWIST, SNAIL1, SNAIL3, SOX4) [189,196,197,200]. However, these few transcription factors render unable to explain expression changes observed in several hundred genes during EMT and suggests that the entire circuitry of regulatory factors involved in driving this process is yet to be fully explored. Our study discovered eight novel transcription factors that were highly induced during EMT and were critically required for the phenotypic and molecular changes that accompany this process. Importantly further, active JNK signaling was required for the induced expression of these factors during EMT. Moreover, these factors were transcriptionally repressed upon MET that parallels loss of JNK signaling. These observations further suggest how JNK activity might participate in transcriptional reprogramming underlying EMT by regulating expression of critical transcription factors.

We also revealed that the newly identified factors not only function in acquisition of the mesenchymal fate but are also required for its maintenance. These factors were very highly expressed in the aggressive, human mesenchymal breast cancer cells and depletion of three of these factors (PRKRIR, PCBP4 and TSC22D1) in these cells led to a drastic loss in their wound healing and invasion capacity. In combination with genomewide transcriptional profiling, we find that these phenotypic effects of transcription factors is linked to their function in activation and repression of a large number of established metastasis promoting and tumor suppressor genes respectively. Importantly further, these datasets also uncovered a number of novel genes under transcriptional control of these factors but have never been studied in the context of tumor biology and warrants further investigation. It is further important to note that the set of genes regulated by these novel factors are largely non-overlapping, arguing that carcinogenic mechanisms may evolve a network of transcription factors that regulate distinct set of genes, but ultimately

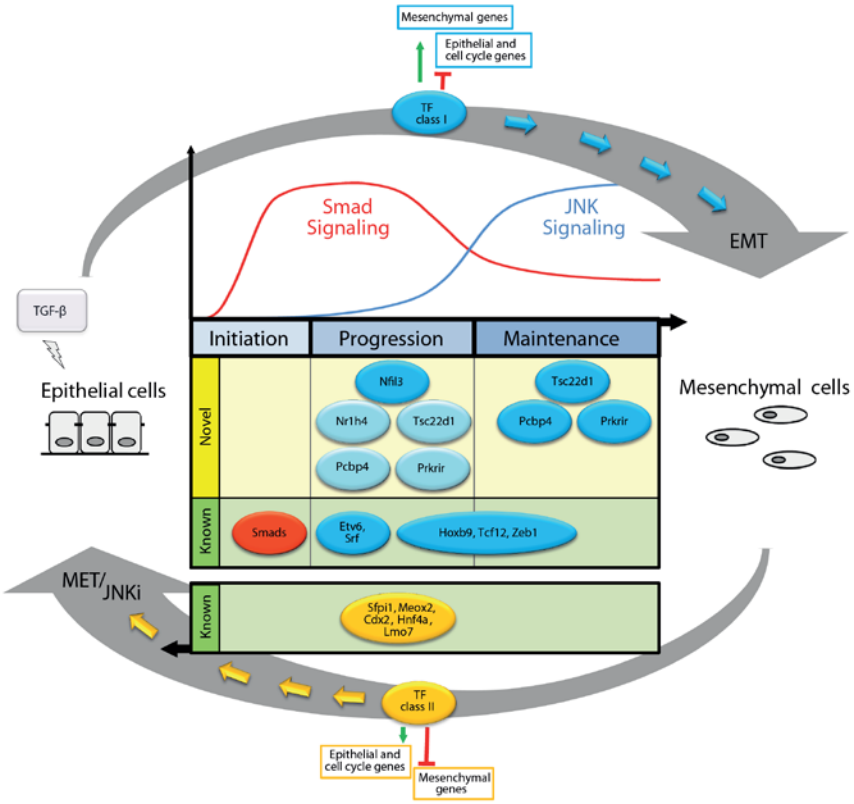
functionally culminate to cooperatively drive metastatic progression at the phenotypic level. We also explored whether these gene regulatory effects can be explained by a DNA binding function of these regulatory factors. Indeed we find that one such factor, Tsc22d1, not only localizes to the nucleus but is preferentially targeted to the euchromatic regions in the genome. Furthermore, Tsc22d1 directly binds at the upstream regulatory elements of critical EMT genes to induce their transcription. We further enhanced the translational impact of our findings by screening a large number of breast cancer patient samples and showing that the three JNK-induced novel transcription factors, PRKRIR, PCBP4 and TSC22D1, are significantly higher expressed in invasive tumors as compared to non-invasive tumors.

While abnormal EMT is involved in tumor progression and metastasis, this process is vitally involved in key developmental events such as embryonic layer formation during gastrulation and neural crest migration [188,215,216]. Motivated by our discovery of new transcription factors that play crucial role in EMT in the context of breast carcinogenesis, we also extended our investigation to understand whether they belong to the core EMT machinery that is equally employed in other contexts of EMT, such as during development. We find that neuronal differentiation, which shows expression changes hallmark of a classical EMT, involves activation of JNK pathway. Importantly further, all novel factors were transcriptionally induced during this process and this requires an active JNK signaling. Furthermore, *in utero* knockdown of one such factor, Tsc22d1, during brain development results in a defective neuronal migration. This is also in line with previous observations that JNK pathway is essential for neurogenesis [153,244,245]. Altogether, these observations strongly suggest that our newly identified transcription factors are components of the EMT core machinery that is utilized in both disease and developmental contexts.

Overall, these comprehensive findings extend beyond existing notions on the function of JNK activity during EMT, providing detailed kinetics of its phase of action during the process and unraveling the entire gene regulatory network downstream of this important signaling cascade through which it functions to determine mesenchymal fate. Moreover, these observations put forth new evidence how signaling

pathways such as JNK may directly modulate epigenetic state of critical genes to mediate transcriptional reprogramming underlying EMT, at the same time reveal a new repertoire of JNK-induced transcription factors that are critical in progression as well as maintenance of the mesenchymal fate (Fig 8).

(A)



(B)

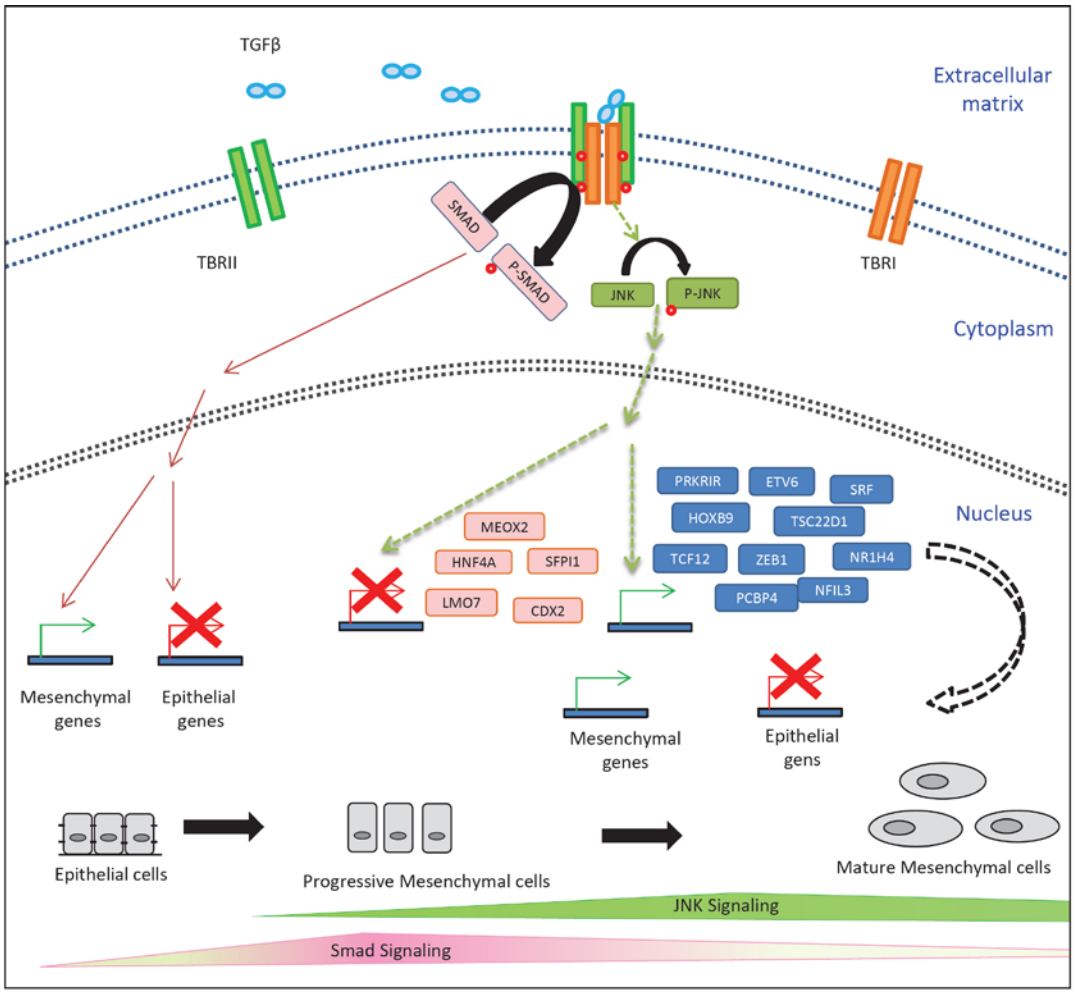


Figure 8. (A-B) Models showing distinct gene regulatory function of JNK signaling during EMT. Onset of EMT crucially relies on Smad-mediated [203,228] transcriptional changes and as cells advance through the mesenchymal fate, JNK pathway progressively becomes critical in gene regulation. We identified a set of transcription factors that are induced by JNK signaling and function in progression and/or maintenance of the mesenchymal state (class I, depicted in blue). Furthermore, JNK activity also represses another set of factors that promote an epithelial phenotype (class II, depicted in orange). Among these JNK regulated factors, in addition to the factors known to regulate EMT and carcinogenesis (Etv6 [246], Srf [247], Hoxb9 [248], Tcf12 [249], Zeb1 [250,251], Meox2 [228], Hnf4a [226], Sfpi1 [230,232], Lmo7 [231] and Cdx2 [252]), we also identified five novel transcription factors that are crucial for progression and/or maintenance of EMT (Nfil3, Nr1h4, Tsc22d1, Pcbp4 and Prkrir). Furthermore, MET involves loss of JNK signaling that facilitates reversion of the transcriptome changes acquired during EMT. The inhibition of JNK signaling results in cellular and molecular changes similar to MET.

MATERIALS AND METHODS

Cell culture

A subclone of NMuMG cells (NMuMG/E9, hereafter NMuMG) has been described previously [176] and was grown in DMEM supplemented with 10% FBS, 2 mM L-Glutamine and 1X Non-essential amino acids. HMEC cells were obtained from Lonza and cultured according to manufacturer's guidelines. Other cell lines were obtained from ATCC and cultured supplying the following media: N2a: DMEM, 10% FBS, 2.0 mM L-glutamine, 1.0mM sodium pyruvate, MDA-MB-231: DMEM, 10% FBS, MDA-MB-361: DMEM, 20% FBS, BT549: RPMI-1640, 10% FBS, 0.001 mg/ml bovine insulin, MCF7: DMEM, 10% FBS, 0.01 mg/ml bovine insulin. All cells were cultured at 37°C with 7% CO_2 in a humid incubator. For TGF-β timecourse experiments, NMuMG and HMEC cells were treated with 2 ng/ml TGF-β (rhTGF-β1 240-B, R&D systems) for indicated time points with TGF-β replenishment and medium change every 2 days. Neuronal differentiation in N2a cells was induced by adding 20 µM retinoic acid (Sigma, St. Louis, MO) in DMEM supplemented with 2% FBS.

siRNA mediated knockdown

For all siRNA mediated knockdown experiments, cells were seeded at same starting density and transfected with ON-TARGET plus SMARTpool siRNAs (i.e. a mixture of 4 siRNA provided as a single reagent) (Dharmacon) or single siRNAs (Dharmacon) wherever indicated every second day. For siRNA transfections, Lipofectamine RNAiMax (Invitrogen, 13778-150) was used according to the manufacturer's instructions. For experiments during TGF-β-induced EMT a four day pre-depletion was performed in epithelial cells before TGF-β-induction to ensure efficient knockdown prior to starting EMT. Sequences of siRNAs are provided in Table S3.

ChIP assay

NMuMG cells were cross-linked in medium containing 1% formaldehyde for 10 min at room temperature, neutralized with 0.125M glycine, scraped off and rinsed twice with 10 ml 1X PBS. Pellets were

resuspended in 3 mL of Buffer L1 (50mM Hepes KOH, pH 7.5, 140mM NaCl, 1mM EDTA pH 8.0, 10% glycerol, 5% NP-40, 0.25% Triton-X 100) and incubated for 10 min at 4°C. This was followed by centrifugation for 5 min at 4°C at 1300g. The pellet was then resuspended in 3 ml of Buffer L2 (200mM NaCl, 1mM EDTA pH 8.0, 0.5mM EGTA pH 8.0, 10mM Tris pH 8.0) and incubated for 10 min at room temperature, followed by centrifugation for 5 min at 4°C at 1300g. The pellet was then resuspended in 600 µl Buffer L3 (1mM EDTA pH 8.0, 0.5mM EGTA pH 8.0, 10mM Tris pH 8.0, 100mM NaCl, 0.1% Na-deoxycholate, 0.17mM N-Lauroyl sarcosine) containing protease inhibitors and incubated at 4°C for 3 hours following sonication using Bioruptor plus (Diagenode). 60 µg of chromatin was incubated overnight at 4 °C with 2 µg of the antibody targeting H3K27me3 (ab6002; Abcam). The mixture was then incubated for 3 h with 40 µl protein A-Sepharose beads preblocked with tRNA and BSA at 4 °C. Beads were washed twice with 1 ml buffer L3 and once with 1 ml DOC buffer (10 mM Tris (pH 8.0), 0.25 M LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA), and bound chromatin was eluted in 1% SDS/0.1 M $NaHCO_3$. This followed treatment with RNase A (0.2mg/ml) for 30 min at 37°C and then with proteinase K (50 µg/ml) for 2.5 h at 55°C. The crosslinking was reversed at 65°C overnight with gentle shaking. DNA was purified by phenol-chloroform extraction followed by ethanol precipitation and recovered in 40 µl TE buffer. Real-time PCR on this ChIP material was performed using SYBR Green (ABI) and ChIP (1:40) or input (1:100) DNA per PCR.

FAIRE assay

For FAIRE assay NMuMG cells were fixed, lysed and sonicated as described for ChIP. After clearing cellular debris by spinning at 14,000 g for 10 min at 4°C DNA was isolated by adding an equal volume of phenol:chloroform:isoamylalcohol (25:24:1), vortexing, and spinning at 12,000 g for 5 min at room temperature. The aqueous phase was isolated and a second round of phenol:chloroform:isoamylalcohol purification was performed. Following collection of aqueous phase, chloroform:isoamylalcohol (24:1) was added, vortexed and spun at 12,000 g for 5 min at room temperature. The aqueous phase was then

collected and DNA was ethanol precipitated and recovered in 40 µl TE buffer. This recovered material was then treated with RNase for 30 minutes (0.2mg/ml) and re-cleaned with Phenol followed by Chloroform:isoamylalcohol (24:1) and subsequently ethanol precipitated. Next, DNA was recovered in 40 µl TE buffer. For deriving input DNA control, RNase and proteinase K digestion, reverse crosslinking and purification of 10 % input material and real time PCR analysis was performed as described for ChIP samples.

Quantitative RT-PCR

mRNA levels were quantified as described previously (*49*). In brief, total RNA was prepared using Trizol (Invitrogen) or SurePrep TrueTotal RNA Purification Kit (Fisher Scientific), reverse transcribed with First Strand cDNA Synthesis Kit (Fermentas), and transcripts were quantified by PCR using SYBR green PCR MasterMix (ABI) on ViiA7 PCR machine (life technologies). Human or mouse Ctcf and Rpl19 primers were used for normalization. Sequences of all primers used in this study are provided as Table S4.

Reagents and antibodies

Reagents used in the present study were TGF-β (rhTGF-β1 240-B, R&D systems). DMEM (21969-035, Invitrogen), RPMI-1640 (R0883, Sigma), PBS (D8537, Sigma), Trypsin (25300-054,Invitrogen), Opti-MEM (31985-047, Invitrogen), FBS (10270, Invitrogen), Glutamine (25030-024, Invitrogen), MEM NEAA (100x) (11140-035, Invitrogen), Lipofectamine 2000 (11668, Invitrogen), Lipofectamine RNAiMax (13778-150, Invitrogen), Trizol (15596026, Invitrogen), reverse transcriptase kit (K1612, Thermo Fischer), SYBR-green PCR MasterMix (4334973,Invitrogen) and Bradford reagent (5000205, BioRad), Protease inhibitor cocktail (04693132001, Roche), Phosphatase inhibitor cocktail (04906837001, Roche).

Antibodies used for Western blot were phospho-JNK (4688, Cell Signaling), JNK (sc474, Santa Cruz) pSmad2 (3108, Cell signaling), Smad2 (5339, Cell Signaling) and Lamin B (sc2616, Santa Cruz), c-Jun antibody (c-Jun Antibody (H-79) Santa Cruz), Phospho-c-Jun (Ser63) II (Cell Signaling),

Antibodies used for immunofluorescence were E-Cadherin (13-1900, Invitrogen and 610182, BDTransduction Laboratories), N-Cadherin (610921, BD Transduction Laboratories), ZO-1 (617300, Invitrogen), Fibronectin (F-3648, Sigma-Aldrich), ,Paxillin (610052, BD, Transduction Laboratories) and p-c-Jun (9261, Cell Signaling)., Atf2 (CREB-2 (C-200) Santa Cruz), Green Fluorescent Protein (2BScientific, GFP-1020), TBR1 (abcam, ab31940), SATB2 (abcam, ab34735), Alexa Fluor-488 goat anti mouse IgG (H+L) (A11029, Invitrogen); Alexa Fluor-568 goat anti rabbit IgG (H+L) A11011 Invitrogen; and Alexa Fluor-633 goat anti rat IgG (H+L) A21094, Invitrogen, Alexa Fluor 633 Phalloidin (A22284, Invitrogen) was used to stain F-actin,

Antibodies used for immunohistochemistry were p-Smad2 (3108, Cell Signaling) and p-JNK (4668, Cell Signaling).

For chemical inhibition experiments we used JNK inhibitors (SP600125, S5567, Sigma; JNK-IN-8, 420150, Merck Millipore), Actinomycin D (A1410, Sigma) and Cycloheximide (C7698, Sigma), Akt inhibitor (B2311 Sigma), PI3 kinase inhibitor Wortmannin (W3144 Sigma), TGF-β receptor blocker (SB-431542).

Inhibitor studies

Cells were seeded at same densities and pretreated 30 min prior to TGF-β treatment with signaling pathway inhibitors or their solvent DMSO at the final concentrations of 10μM SP600125 or 2μM IN-8 and analyzed at indicated time points. Same treatment scheme was employed for inhibitors of transcription (Actinomycin-D; 0.2 μg/ml) and translation (Cycloheximide; 5 μg/ml). To access the effect of later inhibition of JNK during TGF-β induced EMT NMuMG cells were treated starting at day 1 or day 3 after TGF-β stimulation; medium, inhibitor and TGF-β was freshly added every second day. To test the effect of blocking AKT and Pi3k signaling on JNK signaling, NMuMG cells were pretreated with the Akt inhibitor (Akt Inhibitor IV) and Pi3k kinase inhibitor (Wortmannin) for 30 minutes and then induced with TGF-β. This was followed by harvesting the cells 24 hours later for western blot analysis of p-JNK levels.

Immunofluorescence assay

Cells were grown on coverslips, fixed with 4% paraformaldehyde in PBS and permeabilized with 0.2% Triton X-100 for 15 minutes at room temperature. Subsequently, cells were blocked with 10% goat serum, 5% FCS and 0.5% BSA in PBS for 20 minutes and incubated with primary antibodies at 4°C overnight then incubated with fluorochrome-labeled secondary antibody or Phalloidin-633 for 1 hour at room temperature. The coverslips were counterstained with Hoechst mounted with immomount and imaged with a confocal laser-scanning microscope. Data were processed with ImageJ software.

Immunoblotting

Cells were lysed in RIPA buffer and protein concentrations were quantified by using Bradford reagent. Equal amounts of proteins (30 µg) were boiled in 6 × SDS–PAGE loading buffer run on a polyacrylamide gel and transferred to a PVDF membrane and probed with respective antibodies.

Wounding migration assay

Cells were seeded at equal densities and following either treatment with DMSO or JNK inhibitor, a scratch wound was generated using a 10 µl pipette tip on confluent cell monolayers growing in six-well culture plates. Cells were then washed with fresh medium to remove floating cells. Bright field images were taken at 20x magnification after 24 hours of wounding.

Migration and invasion assays

Migration assays were performed as previously described [197]. In brief, 10^4 cells were seeded in 2% FBS/DMEM (Sigma) in the upper chamber of a 24-well transwell migration insert (pore size: 8µm; Falcon BD). The lower chamber was filled with 20% FBS/DMEM. After 16 hours of incubation at 37°C, cells in the upper chamber were removed with a cotton swap, and the cells that had traversed the membrane were fixed in 4% paraformaldehyde/PBS and quantified by DAPI staining using a fluorescent microscope and

ImageJ software. For invasion assays, transwells were coated with 0.5mg/ml matrigel in serum free media overnight. 1.5×10^5 cells were used for the assays.

Immunohistochemistry

Immunohistochemical analysis of 10 invasive and 5 non-invasive breast cancer samples were performed on 4-µm-thick sections according to standard procedures. Antigen retrieval reactions were performed in a steamer using citrate buffer, pH 6.0, for 40 minutes. All slides were incubated with the peroxidase block reagent supplied by Dako (Carpinteria, California, USA). Sections were then stained with respective antibodies in a 1:25 dilution at 4°C overnight. Slides were then incubated with a polymeric biotin-free visualization system (EnVision™; Dako, Carpinteria, California, USA) for 30 minutes at room temperature, followed by 5 minutes incubation with the chromogen diaminobenzidine (DAB) and finally lightly counterstained with hematoxylin. All series included appropriate positive and negative controls that gave adequate results. All slides were mounted and then were evaluated under a Leica light microscope (Leica Microsystem, Wetzler, Germany) by one of the authors trained in histological and immunohistochemical diagnostics (M.S.). Both phospho-SAPK/JNK and phospho-Smad2 showed nuclear staining.

RNA-Seq

RNA samples from the TGFβ-induced EMT time course in NMuMG cells were depleted from rRNA prior to library preparation using the Ribo-Zero kit (Epicentre) following the manufacturer's instructions. RNA-Seq libraries for all samples were generated using TruSeq kit from Illumina. In all cases, 50bp reads and single-end sequencing was derived using Illumina HiSeq2000 platform. Reads were aligned to mouse genome (mm9) using TopHat [177] (version 2.0.9) with default parameters using RefSeq transcript file for coding genes from UCSC. In case of lincRNA expression analysis, reads were aligned to lincRNA coordinates collated from different databases. The aligned reads were then provided as an input to HTSeq_count utility from HTSeq package. The raw read count files obtained from HTSeq-count were then processed for

differential expression using DESeq package and normalized read count with a FDR rate of 0.1 for coding and non-coding RNA were generated after library size normalization [178].

Cell cycle analysis

MDA-MB-231 cells were pulse-treated with 10 µM BrdU for 30 minutes. Then cells were trypsinized and fixed for 15 minutes at room temperature using CytofixCytoperm buffer (BD Biosciences). After two washes with Perm wash buffer (BD Biosciences) the cells were permeabilized with Cytoperm permeabilization buffer (BD Biosciences) for 10 minutes at 4°C, washed with Perm wash buffer (BD Biosciences) and subjected to DNase (300µg/ml in PBS) digest for 1 hour at 37°C. The staining with APC-BrdU antibody was performed using BrdU Flow Kit (BD Biosciences) for 20 minutes at room temperature. DNA was counterstained with 7-AAD along with 30 min RNase (0.1 mg/ml final concentration) digest. Samples were subsequently measured using the BD LSRFortessa Cell Analyzer with BD FACSDiva software and plots were generated using FlowJo.

Time-lapse video microscopy

Time-lapse video microscopy and single cell tracking of NMuMG cells depleted with control siRNA or with siRNA against transcription factors was performed with a cell observer (Zeiss) at a constant temperature of 37^{o}C and 7% CO_2. Phase contrast images were acquired every 5 minutes for 4 days using a 20x phase contrast objective (Zeiss), an AxioCamHRm camera and a Zeiss AxioVision 4.7 software. Single-cell tracking was performed using a self-written computer program (TTT)[253],.

In utero electroporation

Experiments were carried out essentially as described (Saito, 2006). All experimental procedures were conducted in accordance with European, national, and institutional guidelines for animal care. Timed-pregnant (Theiler stages 20 = E12.5) C57BL/6 mice were anaesthetized with Isoflurane (2.5 % via mask, Forene®, Abbod) and analgesized subcutaneously with carprofen (4 mg/kg bodyweight, Rimadyl, Pfizer).

After opening the abdominal cavity the embryos were carefully exposed and 1 µl colored solution containing 1 µg of the p.SUPER-GFP shRNA plasmid expressing shRNAs against Tsc22d1 or control shRNA injected into one of the lateral ventricles. With specialized platinum electrodes (Nepagene CUY 650P) the appropriate voltage was applied (50 ms, interval 950 ms, 5 pulses). Then, the uterine horns were replaced after electroporation in the abdominal cavity.

Immunofluorescence assay to assess neuronal migration in mouse brain

Isolated E16.5 embryonic brains were fixed immediately for 24 h in 4% PFA in PBS at 4°C. Brains were then cryoprotected in 10% sucrose for 2 h then in 30% sucrose (in PBS) overnight, embedded in Tissue-Tek, stored at −20°C and cryosectioned 12 µm. Sections on coverslips were preblocked with 2%BSA, 0.5% Triton (in PBS) for 1h. Primary antibodies (Satb2, 1:500, Abcam (ab34735), Tbr1, 1:500, Abcam (ab31940) and anti-GFP (chicken, Aves Labs, 1:1000) were applied in blocking solution overnight at 4°C. Fluorescent 11 secondary antibodies were applied according to the manufacturer's protocol (life technologies). The coverslips were counterstained with Hoechst mounted with immomount and imaged with a confocal laser- canning microscope Leica (SP5). Data were processed with ImageJ software.

Annexin V FACS

MDA-MB-231 cells were trypsinized and washed twice with cold PBS. One million cells were resuspended in 100 µl of annexin binding buffer (0.01 M HEPES, pH 7.4; 0.14 M NaCl; 2.5 mM $CaCl_2$) and stained adding 5µl Annexin V-APC (550474 BD Biosciences) for 15 minutes at room temperature in the dark. After two washes with annexin binding buffer samples were measured using the BD LSRFortessa Cell Analyzer with BD FACSDiva software.

ACKNOWLEDGEMENTS

We would like to thank members of the Tiwari lab for cooperation and critical feedback during the progress of the project. Support by the Core Facilities of the Institute of Molecular Biology (IMB), Mainz, is gratefully acknowledged, especially the microscopy, cytometry, genomics and bioinformatics core facilities. We thank Gerhard Christofori (University of Basel, Basel, Switzerland) for providing NMuMG cells. We also thank Christian Beisel, Ina Nissen and Manuel Kohler (Laboratory of Quantitative Genomics at D-BSSE, ETH, Zürich, Switzerland) for Illumina sequencing. Research in the laboratory of V.K.T. is supported by the Wilhelm Sander Stiftung 2012.009.1 and 2012.009.2, EpiGeneSys RISE1 program, Marie Curie CIG 322210 and Deutsche Forschungsgemeinschaft (DFG) Grant TI 799/1-1.

AUTHOR CONTRIBUTIONS

S.K.S. designed the study, performed experiments, analyzed data and wrote the manuscript. A.G. designed the study, performed experiments and analyzed data. N.T., S.G. and N.S. performed experiments. S.T., J.T. and F.O. analyzed data. B.B. and R.N. provided tools. M.S. provided clinical samples and analyzed data. V.T. designed the study, analyzed data and wrote the manuscript. All authors read and approved the final manuscript.

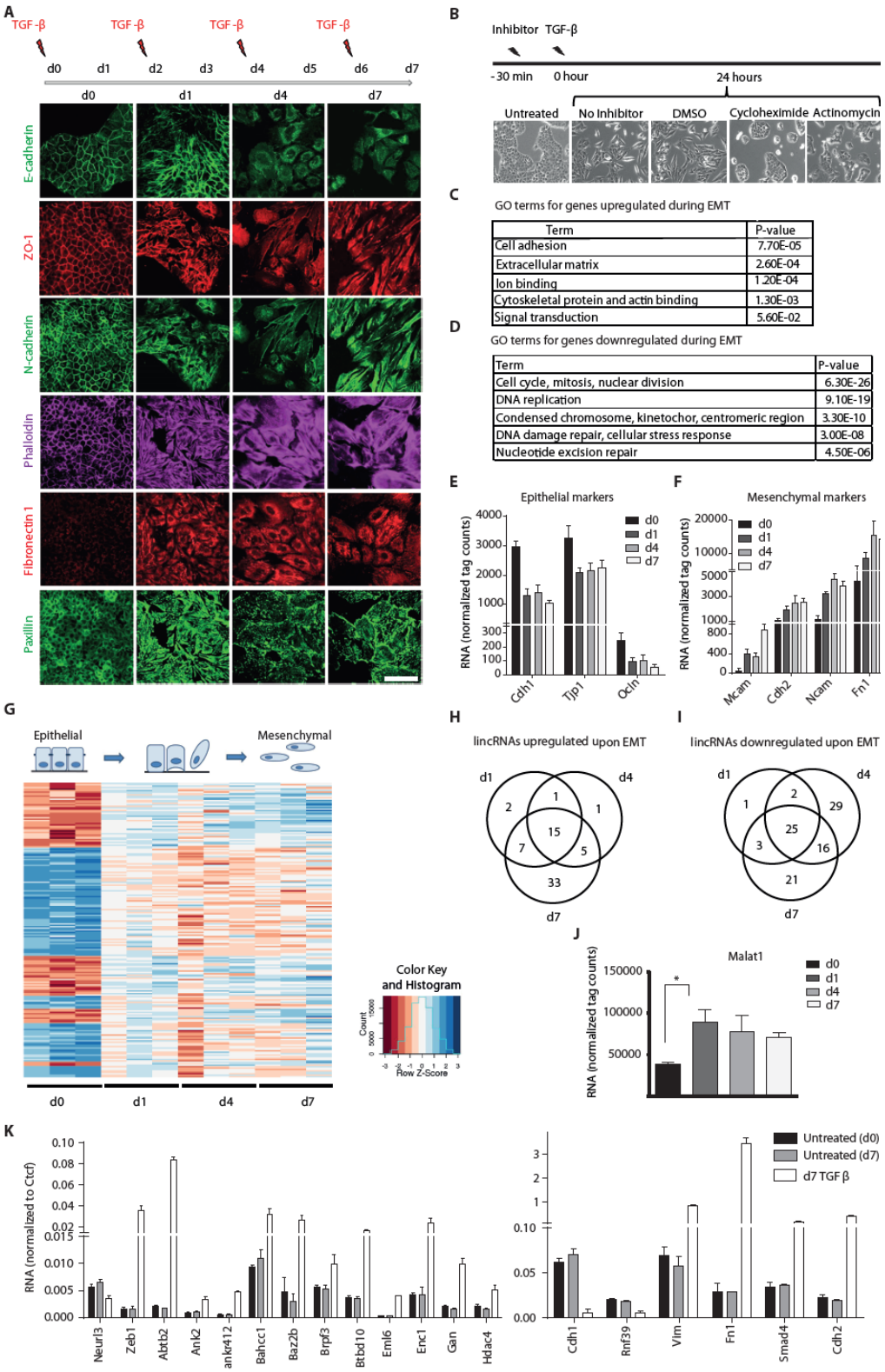
COMPETING INTERESTS

The authors declare no competing financial interests.

ACCESSION NUMBERS

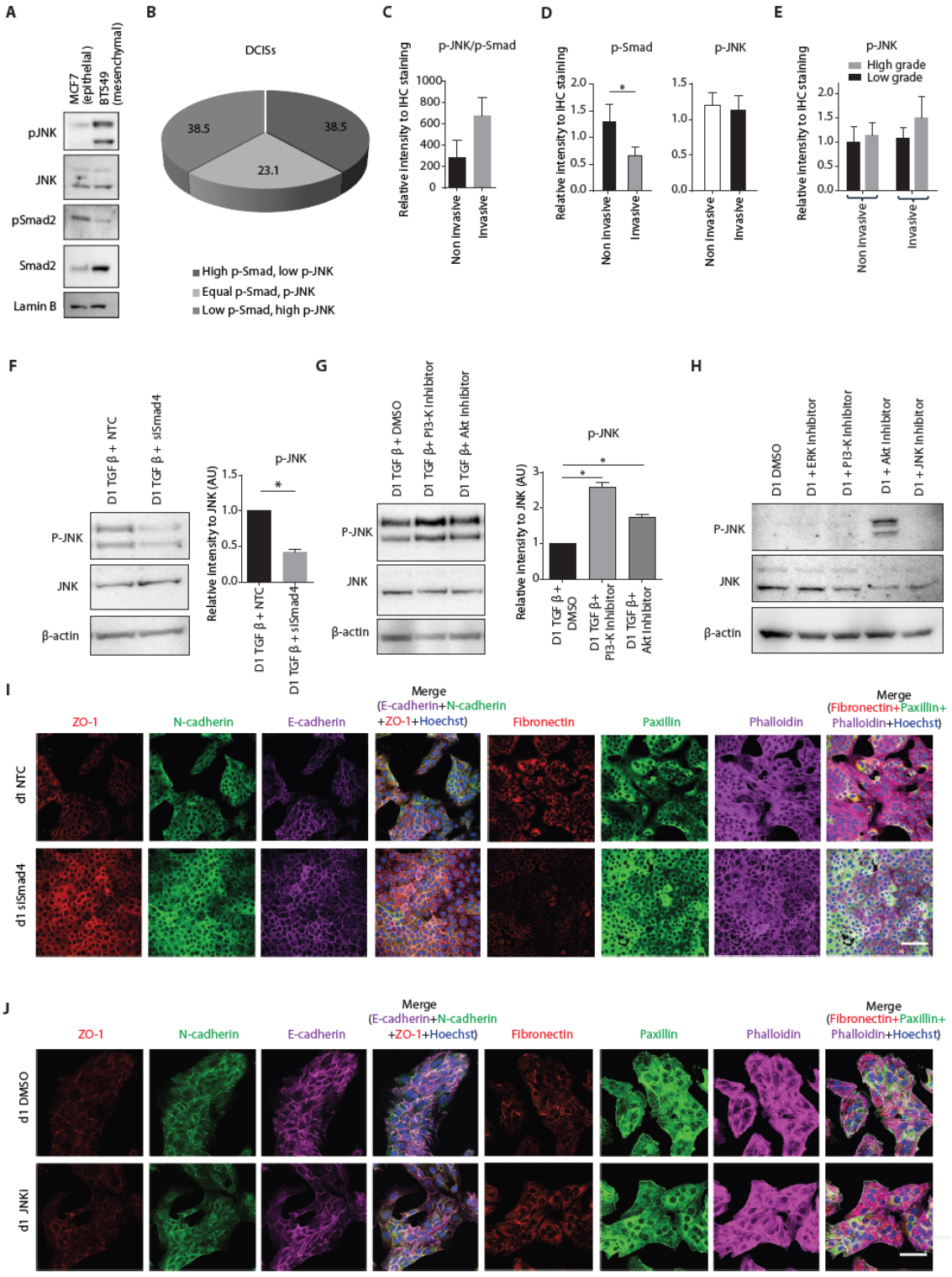
All the next generation sequencing datasets used in this study have already been submitted to GEO and will be publically available under accession number GSE54133.

Supplementary figures



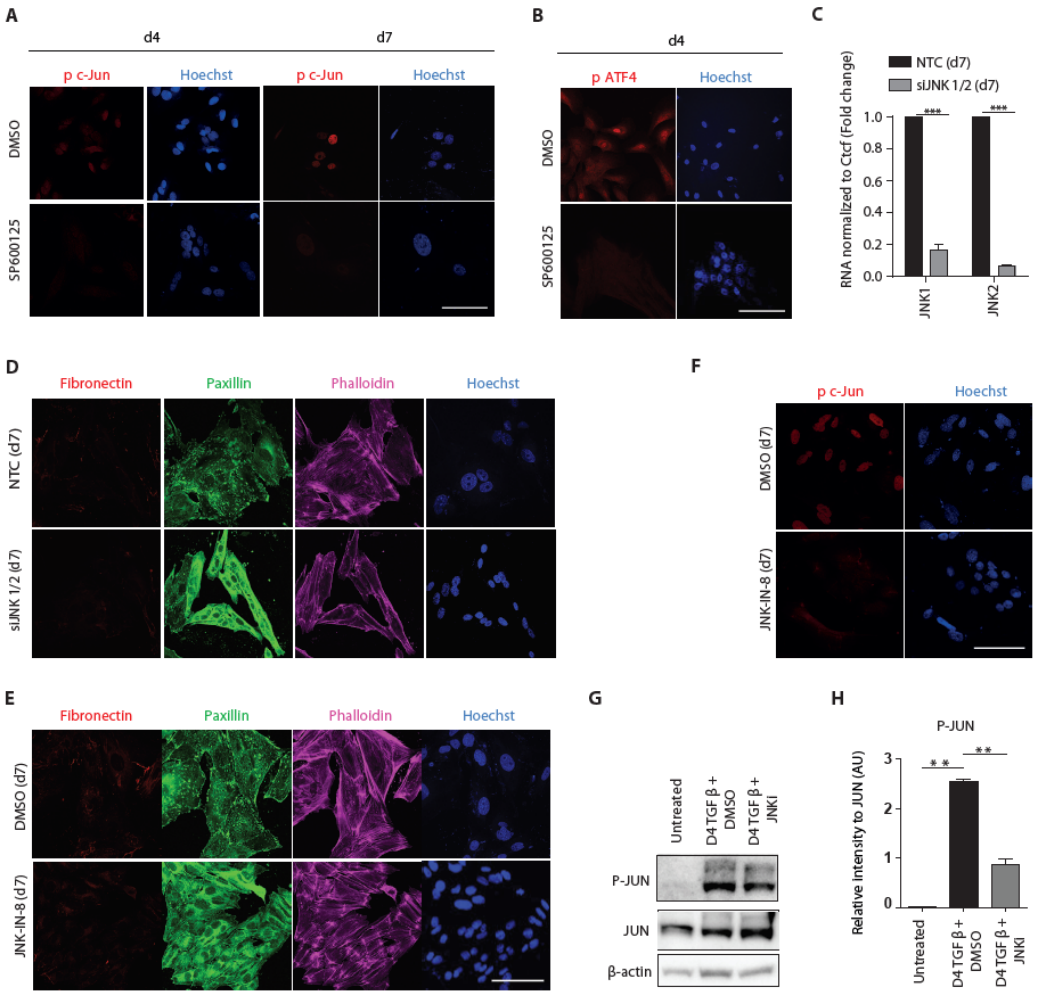
Supplementary figure S1. TGF-β-induced EMT involves stepwise transcriptome reprogramming. (A) Immunofluorescence microscopy for changes in the localization and expression levels of marker proteins during different time points (untreated (d0), day1 (d1), day 4 (d4) and day 7 (d7)) of TFG-β induced EMT).

Staining was performed with antibodies against the epithelial markers E-cadherin and ZO-1, the mesenchymal markers N-cadherin and Fibronectin 1, with Phalloidin to visualize the actin cytoskeleton and against Paxillin to detect focal adhesion plaques. Scale bar, 100 µm, 63X magnification. Treatment scheme is depicted above. (B) Phase contrast images of NMuMG cells pretreated with DMSO, Actinomycin D or Cycloheximide 30 min before TGF-β stimulation for 24 hours. Scale bar, 100 µm, 20X magnification. (C) GO enrichment analysis was performed for genes upregulated during EMT shown in Fig. 1B with DAVID functional classification tool; the plot shows representative categories of the top 5 clusters. (D) Same as C for Fig. 1C downregulated genes. (E-F) RNA represented as normalized tag counts for hallmark epithelial (E) and mesenchymal (F) markers. (G) Similar analysis as shown in (Figure 1A), but limited to lincRNA differentially expressed compared to day 0. (H-I) Same analysis as shown in Figure 1B and C, but only for lincRNAs. (J) Normalized tag counts for a known EMT driving lincRNA (Malat1) during EMT time course. (K) mRNA levels for several genes identified as deregulated in the RNA-Seq analysis of TGFβ-induced EMT in NMuMG cells and performed RT-qPCRs in NMuMG cultured under similar conditions but without exposure to TGFβ were measured by qRT-PCRs relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates. * $p<0.05$ Student's t-test. (Results in this figure contributed by me)

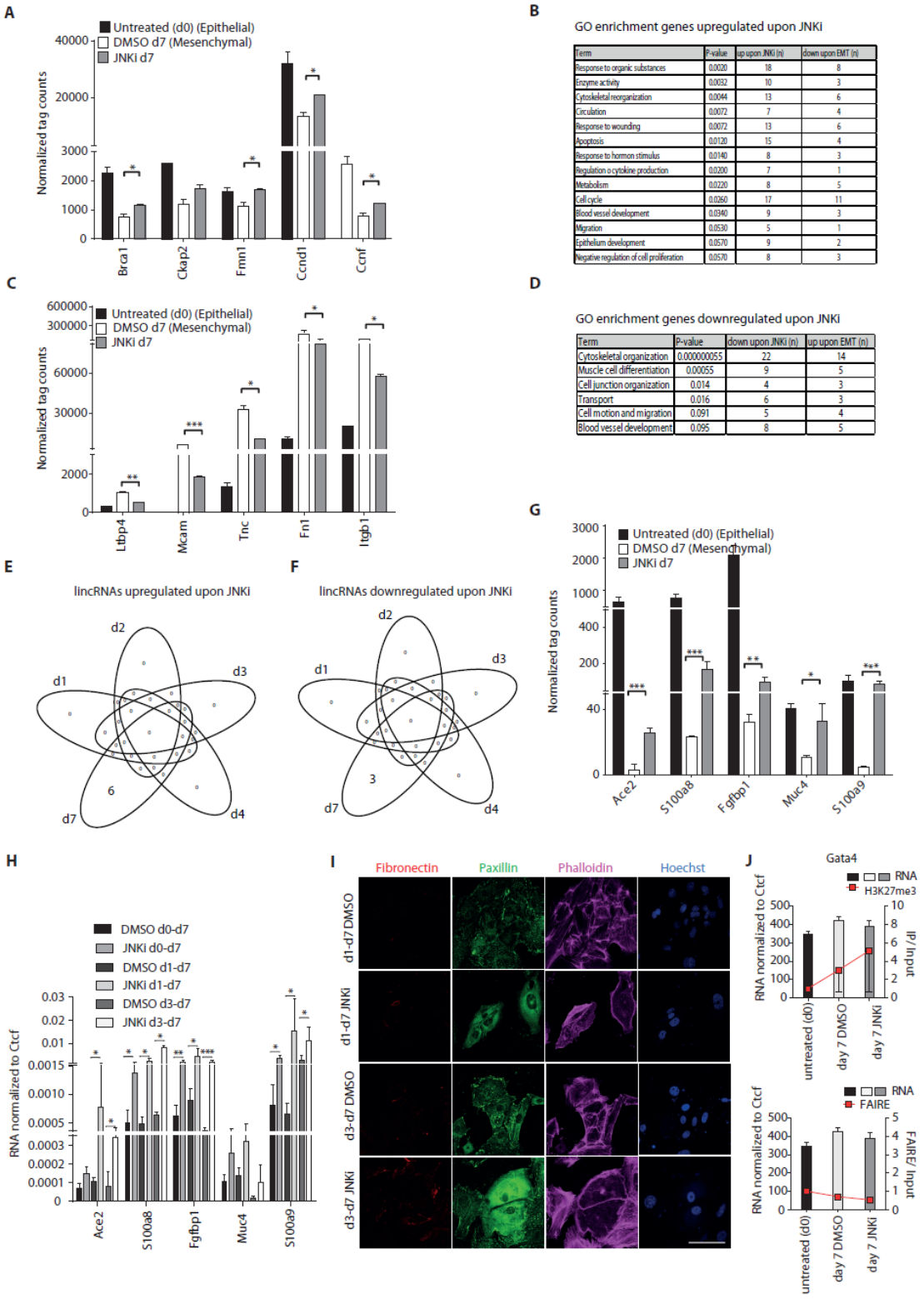


Supplementary figure S2. Smad signaling but not JNK pathway is required for the onset of EMT. **(A)** Immunoblots for total JNK and Smad as well as their phosphorylated forms in human epithelial (MCF7) and mesenchymal (BT549) breast cancer cells. Lamin B serves as loading control. **(B)** Similar analysis as in Fig 1G was performed in non-invasive tumor samples (n=14). **(C)** A relative score is assigned for ratio between pJNK and pSMAD2 and staining in each clinical sample based on the IHC staining intensity (0 to 3 where 0 is lowest and 3 is highest, but to calculate ratio, instead of 0 we have assigned a very small value

0.001 for calculation purpose). Bar plot on the right side shows the ratio between pJNK and pSMAD2 in non-invasive (n=15) and invasive (n=34) clinical samples. y axis represents relative staining intensity. p=0.06 Student's t-test **(D)** Similar analysis as in (C) but plotted values ware ratio between pJNK and pSMAD2 and staining in each clinical sample based on the IHC staining intensity (0 to 3 where 0 is lowest and 3 is highest). Bar plot on the right side shows the relative intensity of pSmad2 or pJNK in non-invasive (n=15) and invasive (n=34) clinical samples. y axis represents relative staining intensity. **(E)** Similar analysis as in (C) for pJNK staining in each clinical sample based on the IHC staining intensity and sub divided based on grades. For invasive tumors grade 1 and 2 considered as low grade while grade 3 considered as high grade tumor samples. Bar plot on the right side shows the relative intensity of pJNK in non-invasive and invasive clinical samples. y axis represents relative staining intensity. **(F)** Representative immunoblots (n=2) showing JNK and phosphorylation state of JNK in NMuMG cells depleted with Smad4 for four days and induced with TGF-β for 24 hours. Actin serves as loading control. The right panel depicts the quantification of immunoblots showing a decrease in JNK signaling upon Smad4 knockdown during TGF-β-induced EMT in NMuMG cells. These analysis show decrease pJNK signaling upon Smad depletion during TGF-β induced EMT **(G)** Similar analysis as in (F) for phospho-JNK was performed but instead of Smad4 depletion, cells were pretreated with PI3-Kinase inhibitor and Akt inhibitor. These results show increased JNK activation upon Akt/PI3 Kinase inhibition. **(H)** Representative immunoblots (n=2) showing JNK and phosphorylation state of JNK in NMuMG cells treated with ERK inhibitor, PI3-Kinase inhibitor and Akt inhibitor for 24 hours. Actin serves as loading control.This analysis show only Akt inhibitor can induce JNK signaling in TGF-β untreated epithelial cells **(I)** Immunofluorescence microscopy in Smad4 depleted cells for changes in the localization and expression levels of depicted marker proteins after 1 day of TGF-β-induced EMT. Scale bar, 100 μm, 40X maginfication. **(J)** NMuMG cells pretreated for 30 minutes with DMSO or JNK inhibitor were stimulated with TGF-β for 1 day and immunofluorescence assay was performed for marker proteins. Scale bar, 100 μm, 40X magnification. * $p<0.05$ Student's *t*-test. (Results in this figure contributed by me)



Supplementary figure S3. Alternate ways of interfering with JNK signaling validate the significance of JNK pathway for mesenchymal fate. **(A)** NMuMG cells treated for four or seven days with TGF-β and DMSO or JNKi were stained for the phosphorylated form of the JNK target c-Jun (p c-Jun). Scale bar 100µm, 63X magnification. **(B)** HMEC cells treated for four days with TGF-β in the presence of DMSO or JNKi were stained for the phosphorylated form of the JNK target ATF4 (pATF4). Scale bar 100µm, 63X magnification. **(C)** mRNA levels of JNK1 and JNK2 were measured by qRT-PCR relative to Ctcf after seven days of JNK1 and JNK2 double knock down in NMuMG cells and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates. **(D)** Immunofluorescence analysis of cells treated in (C) for EMT marker proteins. Scale bar 100µm, 63X magnification. **(E)** Immunoflurescence analysis for EMT marker proteins of NMuMG cells treated with JNK-IN-8 or DMSO for seven days during TGF-β induced EMT. Scale bar 100µm, 63X magnification. **(F)** Immunoflourescence analysis in cells treated as in (E) for phospho c-Jun (p c-Jun). Scale bar 100µm, 63X magnification. **(G)** Representative immunoblots (n=2) showing phosphorylation state of JNK substrate c-JUN during TGF-β-induced EMT upon chemical inhibition of JNK activity in NMuMG cells. B- Actin serves as loading control. **(H)** Quantification of immunoblots in (G) and its graphical representation. Y axis represents intensity normalized to c-Jun level. Mean and SEM is plotted from two independent biological replicates. ** p<0.01, *** *p*< 0.001, Student's *t*-test**.** (Results in this figure contributed by me)

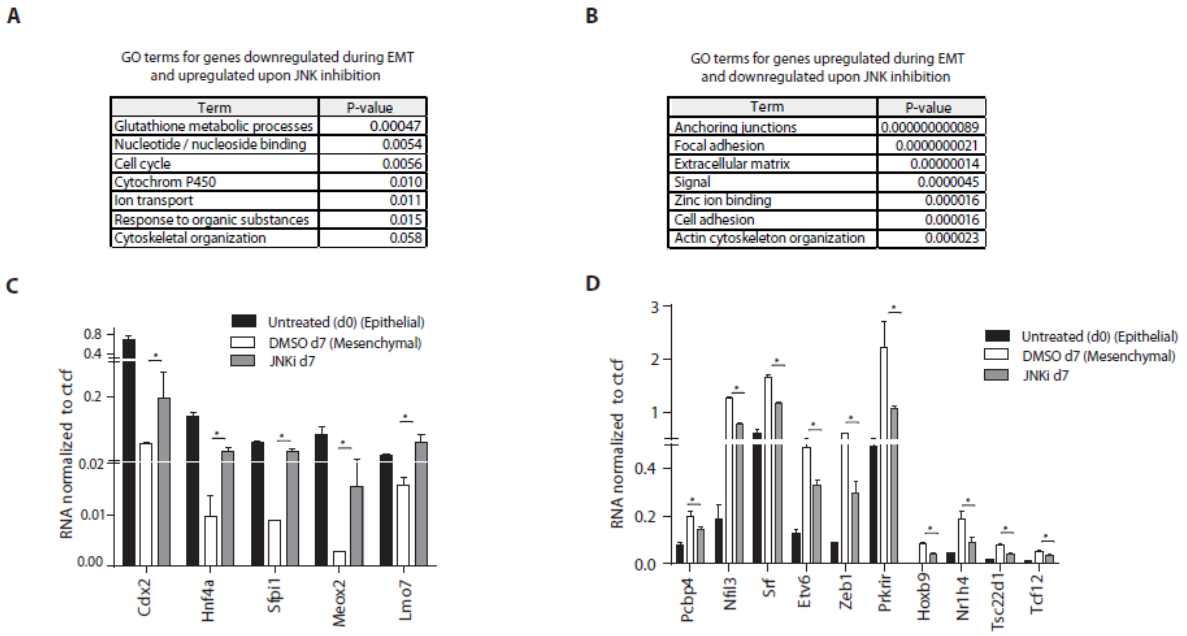


Term	P-value	up upon JNKi (n)	down upon EMT (n)
Response to organic substances	0.0020	18	8
Enzyme activity	0.0032	10	3
Cytoskeletal reorganization	0.0044	13	6
Circulation	0.0072	7	4
Response to wounding	0.0072	13	6
Apoptosis	0.0120	15	4
Response to hormon stimulus	0.0140	8	3
Regulation o cytokine production	0.0200	7	1
Metabolism	0.0220	8	5
Cell cycle	0.0260	17	11
Blood vessel development	0.0340	9	3
Migration	0.0530	5	1
Epithelium development	0.0570	9	2
Negative regulation of cell proliferation	0.0570	8	3

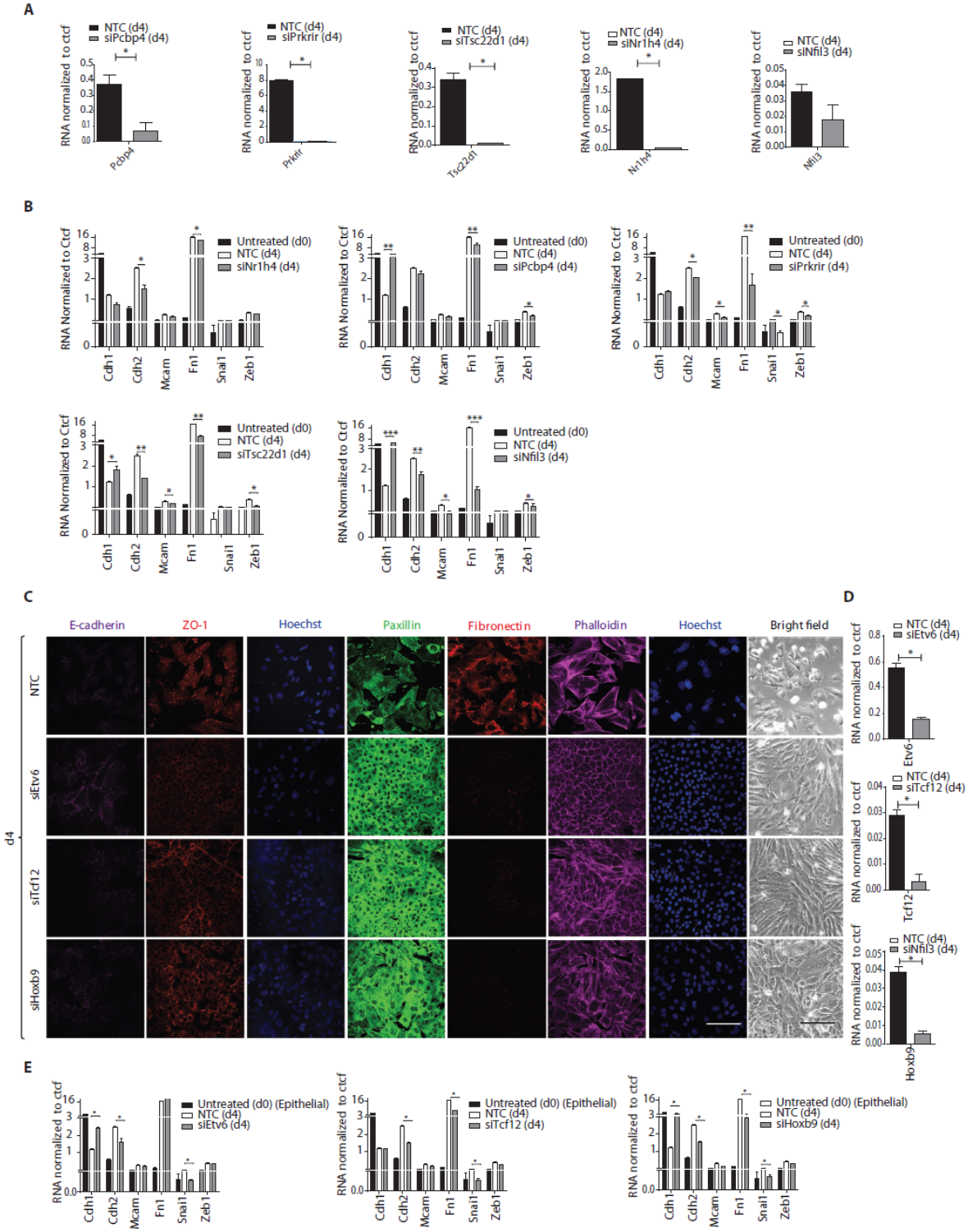
Term	P-value	down upon JNKi (n)	up upon EMT (n)
Cytoskeletal organization	0.000000055	22	14
Muscle cell differentiation	0.00055	9	5
Cell junction organization	0.014	4	3
Transport	0.016	6	3
Cell motion and migration	0.091	5	4
Blood vessel development	0.095	8	5

Supplementary figure S4. Inhibition of JNK signaling significantly reverts gene expression changes acquired during TGF-β-induced EMT. **(A)** RNA levels represented by normalized tag counts derived from RNA-seq data for key epithelial genes that are upregulated upon JNKi at day 7 (d7). **(B)** GO enrichment analysis using DAVID functional classification tool for genes upregulated following JNK inhibition at day 7. Top enriched categories are represented with corresponding P-values. The numbers of genes that are

downregulated during EMT and are reversed by JNK inhibitor are depicted in the second last and last columns respectively. **(C)** Same as in (A) for key mesenchymal genes that are upregulated during EMT and downregulated upon JNKi at day 7 (d7). **(D)** Same as in (B) for genes that are upregulated during EMT and downregulated upon JNKi at day 7 (d7). **(E-F)** Venn diagrams same as Figure 2B and C, but for lincRNAs. **(G)** RNA levels (normalized tag counts) for genes depicted in Fig. 2E and F. **(H)** mRNA levels for genes depicted in Fig. 2E and F were measured by qRT-PCR relative to Ctcf after seven days of TGF-β induced EMT with JNKi or DMSO treatment starting at later time points, i.e. 1 day or 3 days after initial TGF-β induction. Mean and SEM is plotted from three independent biological replicates. **(I)** Immunofluorescence for EMT marker proteins of cells treated as in (H). Scale bar 100µm, 63X magnifications. **(J)** Same as in Fig. 2E and F for a control locus (Gata4 promoter) that is not changing in RNA levels during EMT. (Results in this figure contributed by me)

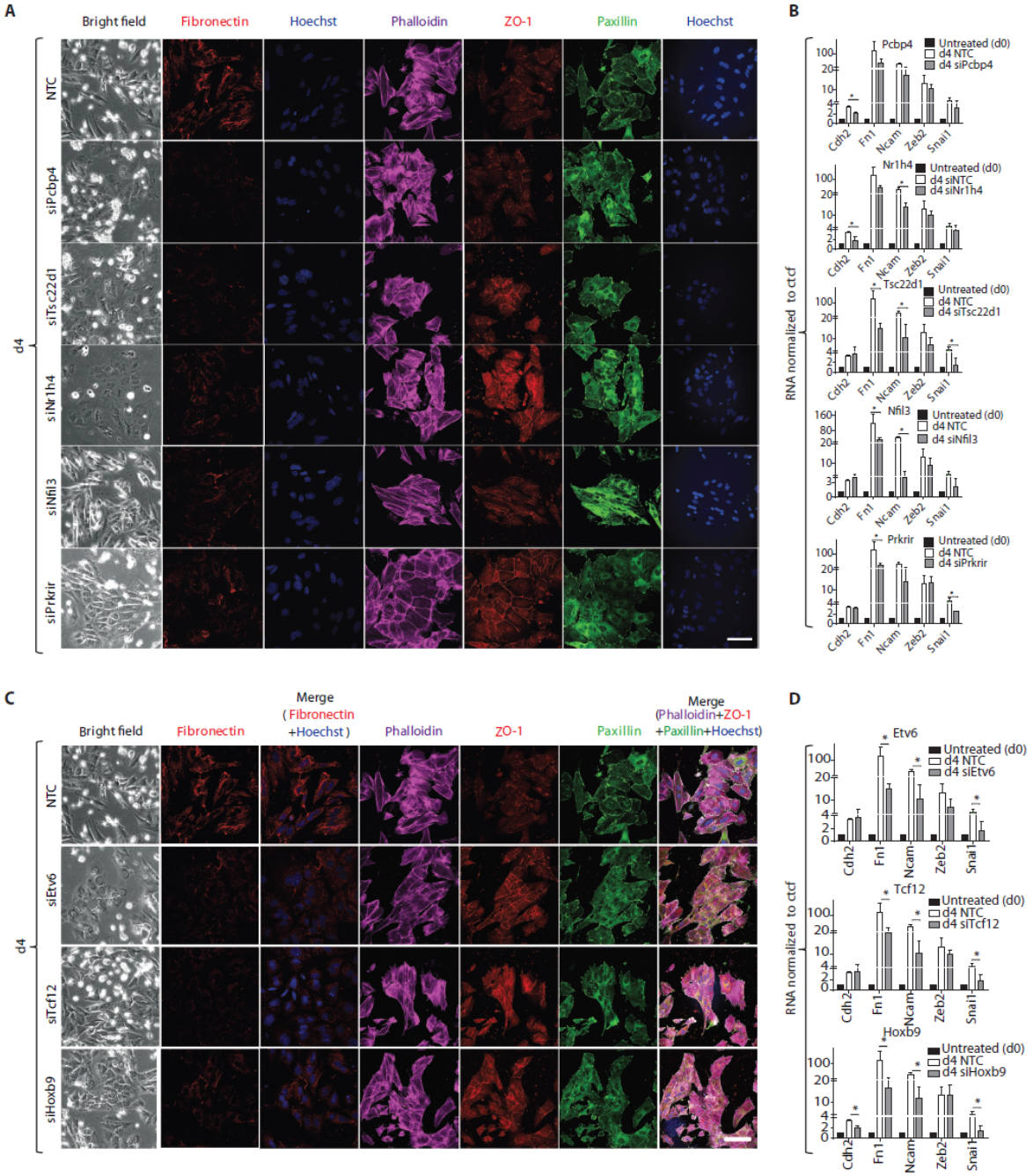


Supplementary figure S5. Regulatory factors employed by JNK signaling regulate mesenchymal fate. **(A-B)** GO term analysis of genes upregulated upon JNKi and downregulated during normal EMT (A) as well as genes downregulated upon JNKi and upregulated during normal EMT (B). GO-term enrichment analysis was performed as in Supplementary Figure 4B. **(C)** mRNA levels for factors depicted in Fig. 3A were measured by qRT-PCR relative to Ctcf in NMuMG cells treated for seven days with TGF-β and DMSO or JNKi. Mean and SEM is plotted from two independent biological replicates. **(D)** Same as in (C) for factors in Figure 3B. Mean and SEM is plotted from two independent biological replicates. * $p<0.05$, Student's *t*-test. (Results in this figure contributed by me)



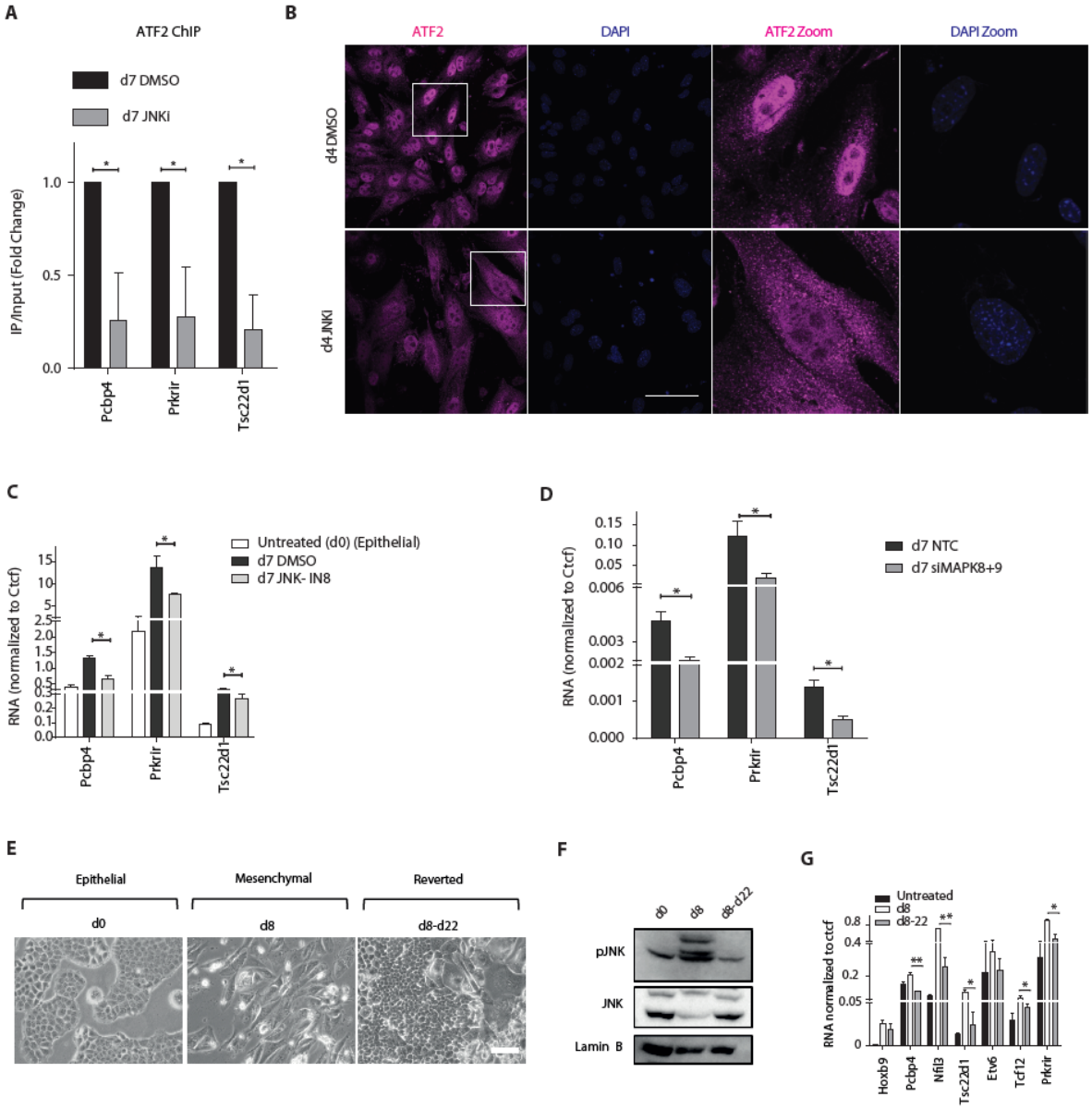
Supplementary figure S6. JNK-induced factors partly implicated in EMT also critically regulate EMT. **(A)** mRNA levels for transcription factors were measured by qRT-PCR relative to Ctcf to access knock down efficiency in NMuMG cells used for experiments depicted in Fig. 3C. **(B)** mRNA levels for classical EMT markers in cells treated as in Fig 3C were measured by qRT-PCR relative to Ctcf and SEM is plotted from 3 independent biological replicates. **(C)** Representative bright field images and immunofluorescence microscopy analysis of changes in the localization and expression levels of marker proteins after four days

(d4) of siRNA mediated depletion of three transcription factors (Etv6, Tcf12 and Hoxb9) using single siRNAs during TGF-β-induced EMT. Scale bar, 100 µm, 40X magnification. **(D)** RT-qPCR for knock down efficiency of analyzed transcription factors in same cells as in (C). **(E)** mRNA levels for classical EMT markers in cells treated as in (C) were measured by RT-qPCR relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from two independent biological replicates. * p<0.05, ** p<0.01, *** *p*< 0.001, Student's *t*-test. (Results in this figure contributed by me)



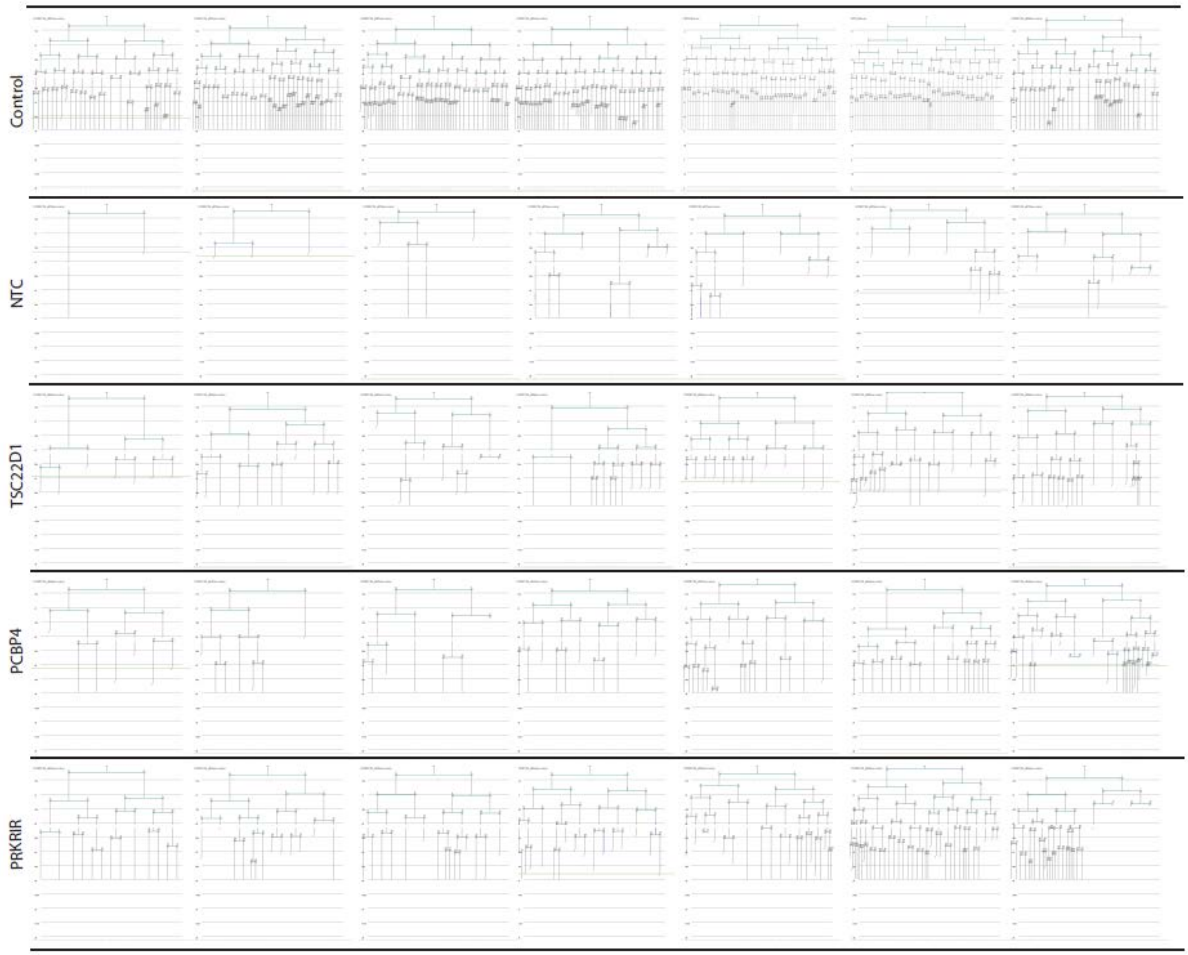
Supplementary figure S7. Additional validation of JNK induced factors by alternate siRNA pool. **(A)** Immunofluorescence analysis as shown in Fig. 3C but knock down of candidate factors was performed using SMART pool siRNAs instead of single siRNAs. **(B)** mRNA levels of EMT markers were measured by qRT-PCR relative to Ctcf for cells treated as in Fig S7A. Mean and SEM is plotted from two independent

biological replicates.**(C-D)** Similar validation as in A and B was performed for JNK-induced factors partly implicated in EMT. * p<0.05, Student's *t*-test.

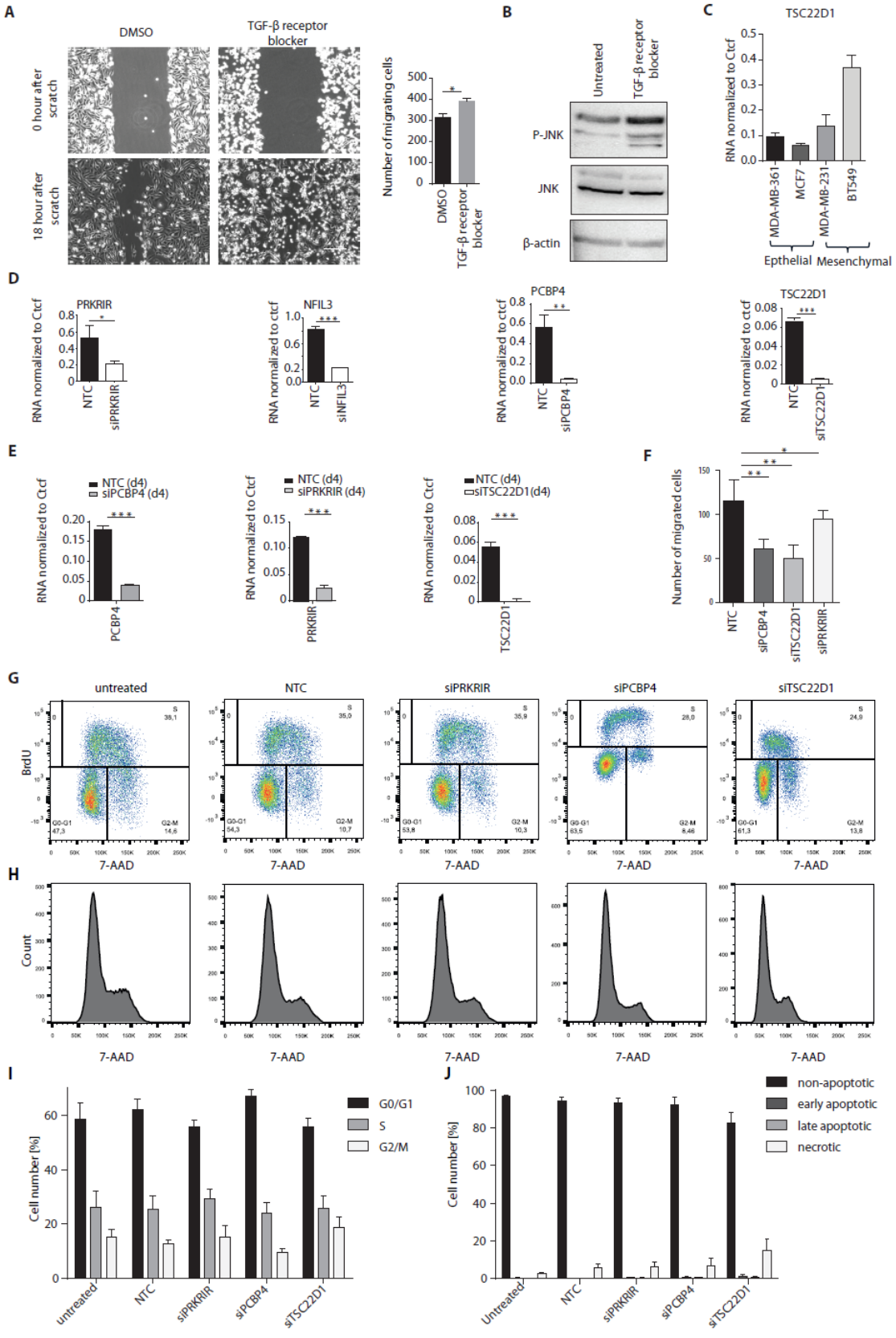


Supplementary figure S8. Alternate interference with JNK signaling validates the transcriptional dependence of transcription factors on JNK. **(A)** ChIP assay using anti-ATF2 antibody following JNK inhibition in NMuMG cells undergoing EMT for four days was performed. Quantitative PCRs were performed for promoters of JNK-dependent transcription factors and enrichments are plotted on the y-axis as ratio of precipitated DNA (bound) to total input DNA and then further divided by the same obtained in the empty vector transfected cells. Similar analysis for control gene promoters such as Hoxb8 and Hprt showed no detectable enrichment signal. SEM is derived from independent biological replicates. **(B)** NMuMG cells treated for four days with TGF-β in the presence of DMSO or JNKi as in (A) were stained for the ATF2 localization. Zoomed nuclei in the right part were shown for visualization purpose. Scale bar 100µm, 63X magnification. **(C)** NMuMG cells were treated with DMSO or JNK inhibitor JNK-IN8 for seven days during TGF-β induced EMT. mRNA levels for transcription factors were measured by RT-qPCR relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from two independent biological replicates.

(D) NMuMG cells were transfected with siRNA SMART pools for JNK1 and JNK2 or NTC and treated with TGF-β for 7 days. mRNA levels for transcription factors were measured by RT-qPCR relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates. **(E)** Bright field images of epithelial NMuMG cells (d0), cells that have undergone TGF-β-induced EMT for 8 days (d8) and cells that have reverted to an epithelial state from d8 mesenchymal state after TGF-β withdrawal for 22 days (d8-22) (MET). Scale bar, 100 µm, 20X magnification. **(F)** Immunoblot for total JNK and phospho-JNK (pJNK) levels in cells shown in (E). Lamin B serves as loading control. **(G)** mRNA levels for transcription factors in cells treated as in (Figure 3B) were measured by qRT-PCRs relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from three independent biological replicates. * $p<0.05$, ** $p<0.01$, Student's *t*-test. (Results in this figure contributed by me)

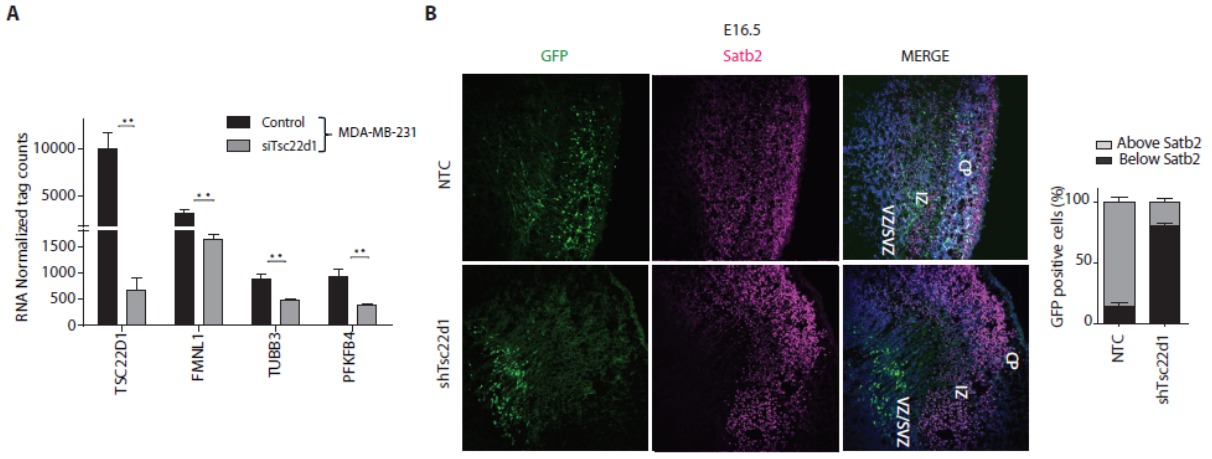


Supplementary figure S9. Newly identified transcription factors PRKRIR, PCBP4 and TSC22D1 rescue TGF-β induced cell cycle arrest. Lineage trees tracked by live cell imaging on NMuMG cells under control, Non treated control (NTC) and siRNA-mediated depletion of factors (X cell death, ? Cell could not be tracked till the end). (Results in this figure partially contributed by me)



Supplementary figure S10. siRNA-mediated efficient knockdown of transcription factors results in

transcriptional and phenotypic remodeling without affecting cell-cycle or cell viability. **(A)** Scratch assay in MDA-MB-231 cells treated for four days with DMSO or TGF-β receptor blocker. Scale bar, 200 µm; 20 X magnification. Mean and SEM is plotted from three independent biological replicates. **(B)** Representative immunoblots (n=3) showing JNK and phosphorylation state of JNK in MDA MB 231 treated similarly as (A). β-actin serves as loading control. On right quantification of immunoblots and its graphical representation showing decreased JNK activation upon Smad4 knock down during TGF-β-induced EMT in NMuMG. **(C)** mRNA levels of TSC22D1 was measured by RT-qPCRs relative to CTCF and plotted on the y-axis from two epithelial cell lines MCF7 and MDA MB 361 and two mesenchymal cell lines MDA MB 231 and BT549. Mean and SEM is plotted from two independent biological replicates. **(D)** Following transfection with control (NTC) or siRNAs against the four candidates (PRKRIR, NFIL3, PCBP4 and TSC22D1) for four days in MDA-MB-231 cells, mRNA levels of the target genes were measured by RT-qPCRs relative to Ctcf and plotted on the y-axis. Mean and SEM is plotted from two independent biological replicates. **(E)** Same analysis as in (D) for cells transfected with single siRNAs. Mean and SEM is plotted from two independent biological replicates. **(F)** Same analysis as in Fig. 5E for cells transfected with single siRNAs. Mean and SEM is plotted from two independent biological replicates. **(G)** Cell cycle was monitored in MDA-MB-231 cells that were non-treated or transfected with siRNAs as in (D) by measuring BrdU incorporation (x-axis) and total DNA content determined by 7-AAD (y-axis). Percentage of cells in G0/G1-phase, S-phase and G2-M-phase are presented in the plots. **(H)** Representative histograms displaying number of cells (y-axis) with respective DNA content determined by 7-AAD (x-axis) for cell cycle analysis described in (D). **(I)** Bar plots showing percentage for cell populations depicted in (G). Mean and SEM is plotted from three independent biological replicates. **(J)** Induction of apoptosis was determined in MDA-MB-231 cells treated like described in (G) by annexin V staining. Bar plots show percentage for cell populations positive for annexin V only (early apoptotic), for annexin V and PI DNA counterstain (late apoptotic), only for PI (necrotic) or negative for both stainings (non-apoptotic). Mean and SEM is plotted from three independent biological replicates. * p<0.05, ** p<0.01, *** *p*< 0.001, Student's *t*-test. (Results in this figure partially contributed by me)



Supplementary figure S11. Expression of TSC22D1-target genes is downregulated upon TSC22D1 knockdown in MDA MB 231 cells. **(A)** RNA levels of TSC22D1 and its target genes is represented by normalized tag counts (y-axis) derived using RNA-seq data from TSC22D1 depleted MDA MB 231 cells. **(B)** Similar analysis as in (Figure 7I) but with respect to Satb2 staining. Error bars represent SEM from independent biological replicates. ** p<0.01, Student's *t*-test.

Supplementary Table S1. Detailed description of clinical samples (age, sex, tumor size, grades, lymph node migration status, ER PR HER status etc.) used in this study for immunohistochemistry.

Histo-Nr.	p-JNK	p-SMAD2	age	histology	Grade (G)	size (mm)	T	N	ER	PR	HER2
980155	0	0	60	ductulo-lobular	2	16	1c	0	+	+	+
980805	3	0	48	ductal	2	16	1c	0	+	+	+
971606	1	0	64	ductal	2		2	0	+	+	+
980809	2	0	50	ductulo-lobular	2	15	2	0	+	+	-
990994	2	1	73	ductal	2	16	2	0	+	+	-
980316	1	0	50	ductal	3	35	4	0	-	-	-
970901	3	0	63	ductal	2	25	2	0	+	+	-
970068	3	1	68	tubular	1	7	1b	0	+	+	-
970068	2	1	68	tubular	1	7	1b	0	+	+	-
980873	3	0	72	tubular	1	15	1c	0	+	+	-
960847	1	0	80	tubular	1	20	1c	0	+	+	-
960847	1	0	80	tubular	1	20	1c	0	+	+	-
20522	3	1		ductal	2	35	2	0	-	+	-
851004	0	1	43	ductal	2	10	1b	0	+	+	-
851940	1	0	46	ductal	2	10	1b	0			
860369	0	0	39	ductal	2	15	1c	0	-	+	-
861947	0	2	66	ductal	2	19	1c	0	+	-	-
871140	2	0	39	ductal	2	25	2	0	+	+	-
871240	0	3	61	ductal	2	26	2	0	-		
871840	0	0	39	ductal	2	25	2	0	+	+	-
881424	0	3	83	tubulär	1	17	1c	0	+	+	
901882	0	2	47	ductal	2	15	1c	0	-	-	+
910969	0	1	70	ductal	1	10	1b	0	+	+	-
930593	3	3	69	ductal	3	15	1c	0	-	-	-
931486	1	1	61	ductal	3	18	1c	0	+	+	-
940414	1	0	70	ductal	2	12	1c	0	+	+	-
940453	0	0	79	ductal	1	24	2	0	+	+	-
940880	2	1	56	ductal	3	28	2	0	-	-	-
941306	0	0	47	lobular	2	23	2	0	+	+	-
950460	0	0	78	ductal	2	40	2	0	+	+	-
970813	0	0	67	ductal	2	10	1b	0	+	+	-
971606	2	0	64	ductal	2	20	1c	0	+	+	-
980316	2	0	49	ductal	3	35	2	0	+	+	-
981842	0	1	56	ductal	3						
20639	0	0		DCIS			is				
20772	2	3		DCIS	low-grade	22	is		+	+	+
20785	2	1		DCIS	high-grade	15	is		+	+	-
20945	0			DCIS	low-grade	11	is				
21074	2	1		DCIS	high-grade	44	is		+	+	-
21305	2	0		DCIS							
11241	1	3	59	DCIS	high-grade	25	is		-	-	+
10948	1	0		DCIS	high-grade	23	is				
11241	1	2		DCIS	high-grade	25	is		-	-	+
20421	1	3		DCIS	low-grade	12	is		+	-	-
10254	1	3	71	DCIS	high-grade	16	is		+	+	-
11234	1	1	45	DCIS	Low-grade	40	is		+	+	-
31388	1	0	46	DCIS	low-grade	5	is		+	+	-
30826			67	DCIS	high-grade	9	is		-	-	-

Supplementary Table S2. A number of genes differentially expressed upon depletion of novel transcription factors in mesenchymal breast cancer cells line (MDA-MB-231) are previously implicated in cancer biology in similar contexts.

Genes differentially regulated upon PRKRIR knock down in MDA-MB-231 cell line

id	log2 FoldChange	Reference
IL8	-1.951702107	1: Fernando RI, Castillo MD, Litzinger M, Hamilton DH, Palena C. IL-8 signaling plays a critical role in the epithelial-mesenchymal transition of human carcinoma cells. Cancer Res. 2011 Aug 1;71(15):5296-306. doi: 10.1158/0008-5472.CAN-11-0156. Epub 2011 Jun 8. PubMed PMID: 21653678; PubMed Central PMCID: PMC3148346.
CSF3	-1.908163695	Kowanetz M, Wu X, Lee J, Tan M, Hagenbeek T, Qu X, Yu L, Ross J, Korsisaari N, Cao T, Bou-Reslan H, Kallop D, Weimer R, Ludlam MJ, Kaminker JS, Modrusan Z, van Bruggen N, Peale FV, Carano R, Meng YG, Ferrara N. Granulocyte-colony stimulating factor promotes lung metastasis through mobilization of Ly6G+Ly6C+ granulocytes. Proc Natl Acad Sci U S A. 2010 Dec 14;107(50):21248-55. doi: 10.1073/pnas.1015855107. Epub 2010 Nov 16. PubMed PMID: 21081700; PubMed Central PMCID: PMC3003076.
FRS3	-1.784673809	Valencia T, Joseph A, Kachroo N, Darby S, Meakin S, Gnanapragasam VJ. Role and expression of FRS2 and FRS3 in prostate cancer. BMC Cancer. 2011 Nov 11;11:484. doi: 10.1186/1471-2407-11-484. PubMed PMID: 22078327; PubMed Central PMCID: PMC3231952.
RANBP3	-1.028845602	Dai F, Shen T, Li Z, Lin X, Feng XH. PPM1A dephosphorylates RanBP3 to enable efficient nuclear export of Smad2 and Smad3. EMBO Rep. 2011 Oct 28;12(11):1175-81. doi: 10.1038/embor.2011.174. PubMed PMID: 21960005; PubMed Central PMCID: PMC3207100.
FLOT2	-1.022561755	Berger T, Ueda T, Arpaia E, Chio II, Shirdel EA, Jurisica I, Hamada K, You-Ten A, Haight J, Wakeham A, Cheung CC, Mak TW. Flotillin-2 deficiency leads to reduced lung metastases in a mouse breast cancer model. Oncogene. 2013 Oct 10;32(41):4989-94. doi: 10.1038/onc.2012.499. Epub 2012 Nov 12. PubMed PMID: 23146906.
TWF1	-1.012274777	Bockhorn J, Dalton R, Nwachukwu C, Huang S, Prat A, Yee K, Chang YF, Huo D, Wen Y, Swanson KE, Qiu T, Lu J, Park SY, Dolan ME, Perou CM, Olopade OI, Clarke MF, Greene GL, Liu H. MicroRNA-30c inhibits human breast tumour chemotherapy resistance by regulating TWF1 and IL-11. Nat Commun. 2013;4:1393. doi: 10.1038/ncomms2393. PubMed PMID: 23340433; PubMed Central PMCID: PMC3723106.
ARNTL2	-0.990786259	Mazzoccoli G, Pazienza V, Panza A, Valvano MR, Benegiamo G, Vinciguerra M, Andriulli A, Piepoli A. ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer. J Cancer Res Clin Oncol. 2012 Mar;138(3):501-11. doi: 10.1007/s00432-011-1126-6. Epub 2011 Dec 24. PubMed PMID: 22198637.

CXCL2	-0.963136522	Acharyya S, Oskarsson T, Vanharanta S, Malladi S, Kim J, Morris PG, Manova-Todorova K, Leversha M, Hogg N, Seshan VE, Norton L, Brogi E, Massagué J. A CXCL1 paracrine network links cancer chemoresistance and metastasis. Cell. 2012 Jul 6;150(1):165-78. doi: 10.1016/j.cell.2012.04.042. PubMed PMID: 22770218; PubMed Central PMCID: PMC3528019.
HTRA1	-0.935380065	Esposito V, Campioni M, De Luca A, Spugnini EP, Baldi F, Cassandro R, Mancini A, Vincenzi B, Groeger A, Caputi M, Baldi A. Analysis of HtrA1 serine protease expression in human lung cancer. Anticancer Res. 2006 Sep-Oct;26(5A):3455-9. PubMed PMID: 17094466. ; Mullany SA, Moslemi-Kebria M, Rattan R, Khurana A, Clayton A, Ota T, Mariani A, Podratz KC, Chien J, Shridhar V. Expression and functional significance of HtrA1 loss in endometrial cancer. Clin Cancer Res. 2011 Feb 1;17(3):427-36. doi: 10.1158/1078-0432.CCR-09-3069. Epub 2010 Nov 23. PubMed PMID: 21098697; PubMed Central PMCID: PMC3057564.
GPR116	-0.922929776	Tang X, Jin R, Qu G, Wang X, Li Z, Yuan Z, Zhao C, Siwko S, Shi T, Wang P, Xiao J, Liu M, Luo J. GPR116, an adhesion G-protein-coupled receptor, promotes breast cancer metastasis via the Gαq-p63RhoGEF-Rho GTPase pathway. Cancer Res. 2013 Oct 15;73(20):6206-18. doi: 10.1158/0008-5472.CAN-13-1049. Epub 2013 Sep 5. PubMed PMID: 24008316.
FASN	-0.920360159	Agostini M, Almeida LY, Bastos DC, Ortega RM, Moreira FS, Seguin F, Zecchin KG, Raposo HF, Oliveira HC, Amoêdo ND, Salo T, Coletta RD, Graner E. The Fatty Acid synthase inhibitor orlistat reduces the growth and metastasis of orthotopic tongue oral squamous cell carcinomas. Mol Cancer Ther. 2014 Mar;13(3):585-95. doi: 10.1158/1535-7163.MCT-12-1136. Epub 2013 Dec 20. PubMed PMID: 24362464.
SRGN	-0.918073523	Li XJ, Ong CK, Cao Y, Xiang YQ, Shao JY, Ooi A, Peng LX, Lu WH, Zhang Z, Petillo D, Qin L, Bao YN, Zheng FJ, Chia CS, Iyer NG, Kang TB, Zeng YX, Soo KC, Trent JM, Teh BT, Qian CN. Serglycin is a theranostic target in nasopharyngeal carcinoma that promotes metastasis. Cancer Res. 2011 Apr 15;71(8):3162-72. doi: 10.1158/0008-5472.CAN-10-3557. Epub 2011 Feb 2. PubMed PMID: 21289131.
NET1	-0.915848912	Gilcrease MZ, Kilpatrick SK, Woodward WA, Zhou X, Nicolas MM, Corley LJ, Fuller GN, Tucker SL, Diaz LK, Buchholz TA, Frost JA. Coexpression of alpha6beta4 integrin and guanine nucleotide exchange factor Net1 identifies node-positive breast cancer patients at high risk for distant metastasis. Cancer Epidemiol Biomarkers Prev. 2009 Jan;18(1):80-6. doi: 10.1158/1055-9965.EPI-08-0842. PubMed PMID: 19124484; PubMed Central PMCID: PMC2669542.
SDC1	-0.896291757	Péterfia B, Füle T, Baghy K, Szabadkai K, Fullár A, Dobos K, Zong F, Dobra K, Hollósi P, Jeney A, Paku S, Kovalszky I. Syndecan-1 enhances proliferation, migration and metastasis of HT-1080 cells in cooperation with syndecan-2. PLoS One. 2012;7(6):e39474. doi: 10.1371/journal.pone.0039474. Epub 2012 Jun 26. PubMed PMID: 22745764; PubMed Central PMCID: PMC3383727.
ADAMTS15	0.807423717	Viloria CG, Obaya AJ, Moncada-Pazos A, Llamazares M, Astudillo A, Capellá G, Cal S, López-Otín C. Genetic inactivation of ADAMTS15 metalloprotease in human colorectal cancer. Cancer Res. 2009 Jun 1;69(11):4926-34. doi:

		10.1158/0008-5472.CAN-08-4155. Epub 2009 May 19. PubMed PMID: 19458070.
IFIT2	0.915392682	Lai KC, Liu CJ, Chang KW, Lee TC. Depleting IFIT2 mediates atypical PKC signaling to enhance the migration and metastatic activity of oral squamous cell carcinoma cells. Oncogene. 2013 Aug 8;32(32):3686-97. doi: 10.1038/onc.2012.384. Epub 2012 Sep 17. PubMed PMID: 22986528.
ELF3	1.071053635	Shatnawi A, Norris JD, Chaveroux C, Jasper JS, Sherk AB, McDonnell DP, Giguère V. ELF3 is a repressor of androgen receptor action in prostate cancer cells. Oncogene. 2014 Feb 13;33(7):862-71. doi: 10.1038/onc.2013.15. Epub 2013 Feb 25. PubMed PMID: 23435425.
KIAA0101	-1.668986284	Jain M, Zhang L, Patterson EE, Kebebew E. KIAA0101 is overexpressed, and promotes growth and invasion in adrenal cancer. PLoS One. 2011;6(11):e26866. doi: 10.1371/journal.pone.0026866. Epub 2011 Nov 11. PubMed PMID: 22096502; PubMed Central PMCID: PMC3214018.
IL6	-1.440239691	Yadav A, Kumar B, Datta J, Teknos TN, Kumar P. IL-6 promotes head and neck tumor metastasis by inducing epithelial-mesenchymal transition via the JAK-STAT3-SNAIL signaling pathway. Mol Cancer Res. 2011 Dec;9(12):1658-67. doi: 10.1158/1541-7786.MCR-11-0271. Epub 2011 Oct 5. PubMed PMID: 21976712; PubMed Central PMCID: PMC3243808. ; Tawara K, Oxford JT, Jorcyk CL. Clinical significance of interleukin (IL)-6 in cancer metastasis to bone: potential of anti-IL-6 therapies. Cancer Manag Res. 2011;3:177-89. doi: 10.2147/CMR.S18101. Epub 2011 May 18. PubMed PMID: 21625400; PubMed Central PMCID: PMC3101113.

Genes differentially regulated upon PCBP4 knock down in MDA-MB-231 cell line

id	log2 Fold Change	Reference
EGR1	-1.505824757	Tarcic G, Avraham R, Pines G, Amit I, Shay T, Lu Y, Zwang Y, Katz M, Ben-Chetrit N, Jacob-Hirsch J, Virgilio L, Rechavi G, Mavrothalassitis G, Mills GB, Domany E, Yarden Y. EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF. FASEB J. 2012 Apr;26(4):1582-92. doi: 10.1096/fj.11-194654. Epub 2011 Dec 23. PubMed PMID: 22198386; PubMed Central PMCID: PMC3316897.
RHOU	-1.358245804	Sequeira L, Dubyk CW, Riesenberger TA, Cooper CR, van Golen KL. Rho GTPases in PC-3 prostate cancer cell morphology, invasion and tumor cell diapedesis. Clin Exp Metastasis. 2008;25(5):569-79. doi: 10.1007/s10585-008-9173-3. Epub 2008 May 7. PubMed PMID: 18461284. ; Schiavone D, Dewilde S, Vallania F, Turkson J, Di Cunto F, Poli V. The RhoU/Wrch1 Rho GTPase gene is a common transcriptional target of both the gp130/STAT3 and Wnt-1 pathways. Biochem J. 2009 Jun 26;421(2):283-92. doi: 10.1042/BJ20090061. PubMed PMID: 19397496; PubMed Central PMCID: PMC2908995.
MAP3K3	-1.337204043	Hasan R, Sharma R, Saraya A, Chattopadhyay TK, DattaGupta S, Walfish PG, Chauhan SS, Ralhan R. Mitogen activated protein kinase kinase kinase 3 (MAP3K3/MEKK3) overexpression is an early event in esophageal tumorigenesis and is a predictor of poor disease prognosis. BMC Cancer.

		2014 Jan 2;14:2. doi: 10.1186/1471-2407-14-2. PubMed PMID: 24383423; PubMed Central PMCID: PMC3890584.
CDK2AP1	-1.274874696	Tsuji T, Ibaragi S, Shima K, Hu MG, Katsurano M, Sasaki A, Hu GF. Epithelial-mesenchymal transition induced by growth suppressor p12CDK2-AP1 promotes tumor cell local invasion but suppresses distant colony growth. Cancer Res. 2008 Dec 15;68(24):10377-86. doi: 10.1158/0008-5472.CAN-08-1444. PubMed PMID: 19074907; PubMed Central PMCID: PMC2605670.
CCND1	-1.262724653	Li Z, Wang C, Jiao X, Lu Y, Fu M, Quong AA, Dye C, Yang J, Dai M, Ju X, Zhang X, Li A, Burbelo P, Stanley ER, Pestell RG. Cyclin D1 regulates cellular migration through the inhibition of thrombospondin 1 and ROCK signaling. Mol Cell Biol. 2006 Jun;26(11):4240-56. PubMed PMID: 16705174; PubMed Central PMCID: PMC1489104.
CXCL3	-1.252325649	See AL, Chong PK, Lu SY, Lim YP. CXCL3 is a Potential Target for Breast Cancer Metastasis. Curr Cancer Drug Targets. 2014 Mar 5. [Epub ahead of print]c PubMed PMID: 24605943.
CHCHD2	-1.179780173	Seo M, Lee WH, Suk K. Identification of novel cell migration-promoting genes by a functional genetic screen. FASEB J. 2010 Feb;24(2):464-78. doi: 10.1096/fj.09-137562. Epub 2009 Oct 7. PubMed PMID: 19812375.
S100A2	-1.106498653	Bulk E, Sargin B, Krug U, Hascher A, Jun Y, Knop M, Kerkhoff C, Gerke V, Liersch R, Mesters RM, Hotfilder M, Marra A, Koschmieder S, Dugas M, Berdel WE, Serve H, Müller-Tidow C. S100A2 induces metastasis in non-small cell lung cancer. Clin Cancer Res. 2009 Jan 1;15(1):22-9. doi: 10.1158/1078-0432.CCR-08-0953. PubMed PMID: 19118029.
TUBB3	-1.056159314	Hwang JE, Hong JY, Kim K, Kim SH, Choi WY, Kim MJ, Jung SH, Shim HJ, Bae WK, Hwang EC, Lee KH, Lee JH, Cho SH, Chung IJ. Class III β-tubulin is a predictive marker for taxane-based chemotherapy in recurrent and metastatic gastric cancer. BMC Cancer. 2013 Sep 23;13:431. doi: 10.1186/1471-2407-13-431. PubMed PMID: 24053422.
SRGN	-1.047275507	Li XJ, Ong CK, Cao Y, Xiang YQ, Shao JY, Ooi A, Peng LX, Lu WH, Zhang Z, Petillo D, Qin L, Bao YN, Zheng FJ, Chia CS, Iyer NG, Kang TB, Zeng YX, Soo KC, Trent JM, Teh BT, Qian CN. Serglycin is a theranostic target in nasopharyngeal carcinoma that promotes metastasis. Cancer Res. 2011 Apr 15;71(8):3162-72. doi: 10.1158/0008-5472.CAN-10-3557. Epub 2011 Feb 2. PubMed PMID: 21289131.
ENC1	-1.039270142	Hernandez MC, Andres-Barquin PJ, Martinez S, Bulfone A, Rubenstein JL, Israel MA. ENC-1: a novel mammalian kelch-related gene specifically expressed in the nervous system encodes an actin-binding protein. J Neurosci. 1997 May 1;17(9):3038-51. PubMed PMID: 9096139.; Fujita M, Furukawa Y, Tsunoda T, Tanaka T, Ogawa M, Nakamura Y. Upregulation of the ectodermal-neural cortex 1 (ENC1) gene, a downstream target of the beta-catenin/T-cell factor complex, in colorectal carcinomas. Cancer Res. 2001 Nov 1;61(21):7722-6. PubMed PMID: 11691783 ;
AREG	-1.035163082	Bolin C, Tawara K, Sutherland C, Redshaw J, Aranda P, Moselhy J, Anderson R, Jorcyk CL. Oncostatin m promotes mammary tumor metastasis to bone and osteolytic bone degradation. Genes Cancer. 2012 Feb;3(2):117-30. doi:10.1177/1947601912458284. PubMed PMID: 23050044; PubMed Central PMCID: PMC3463924.

PDRG1	-1.023042141	Jiang L, Luo X, Shi J, Sun H, Sun Q, Sheikh MS, Huang Y. PDRG1, a novel tumor marker for multiple malignancies that is selectively regulated by genotoxic stress. Cancer Biol Ther. 2011 Mar 15;11(6):567-73. Epub 2011 Mar 15. PubMed PMID: 21193842; PubMed Central PMCID: PMC3087961.
CXCL2	-0.996713841	Acharyya S, Oskarsson T, Vanharanta S, Malladi S, Kim J, Morris PG, Manova-Todorova K, Leversha M, Hogg N, Seshan VE, Norton L, Brogi E, Massagué J. A CXCL1 paracrine network links cancer chemoresistance and metastasis. Cell. 2012 Jul 6;150(1):165-78. doi: 10.1016/j.cell.2012.04.042. PubMed PMID: 22770218; PubMed Central PMCID: PMC3528019.
TOMM34	-0.984857815	Aleskandarany MA, Negm OH, Rakha EA, Ahmed MA, Nolan CC, Ball GR, Caldas C, Green AR, Tighe PJ, Ellis IO. TOMM34 expression in early invasive breast cancer: a biomarker associated with poor outcome. Breast Cancer Res Treat. 2012 Nov;136(2):419-27. doi: 10.1007/s10549-012-2249-4. Epub 2012 Oct 4. PubMed PMID: 23053644.
PHLDA1	-0.97241514	Sakthianandeswaren A, Christie M, D'Andreti C, Tsui C, Jorissen RN, Li S, Fleming NI, Gibbs P, Lipton L, Malaterre J, Ramsay RG, Phesse TJ, Ernst M, Jeffery RE, Poulsom R, Leedham SJ, Segditsas S, Tomlinson IP, Bernhard OK, Simpson RJ, Walker F, Faux MC, Church N, Catimel B, Flanagan DJ, Vincan E, Sieber OM. PHLDA1 expression marks the putative epithelial stem cells and contributes to intestinal tumorigenesis. Cancer Res. 2011 May 15;71(10):3709-19. doi: 10.1158/0008-5472.CAN-10-2342. Epub 2011 May 10. PubMed PMID: 21558389.
AURKA	-0.960256717	1: Zhang H, Chen X, Liu B, Zhou L. Effects of stable knockdown of Aurora kinase A on proliferation, migration, chromosomal instability, and expression of focal adhesion kinase and matrix metalloproteinase-2 in HEp-2 cells. Mol Cell Biochem. 2011 Nov;357(1-2):95-106. doi: 10.1007/s11010-011-0879-1. Epub 2011 May 18. PubMed PMID: 21590355.
MARCKSL1	-0.955459	1: Björkblom B, Padzik A, Mohammad H, Westerlund N, Komulainen E, Hollos P, Parviainen L, Papageorgiou AC, Iljin K, Kallioniemi O, Kallajoki M, Courtney MJ, Mågård M, James P, Coffey ET. c-Jun N-terminal kinase phosphorylation of MARCKSL1 determines actin stability and migration in neurons and in cancer cells. Mol Cell Biol. 2012 Sep;32(17):3513-26. doi: 10.1128/MCB.00713-12. Epub 2012 Jul 2. PubMed PMID: 22751924; PubMed Central PMCID: PMC3421996.
C9orf140	-0.933623277	1: Weng YR, Yu YN, Ren LL, Cui Y, Lu YY, Chen HY, Ma X, Qin WX, Cao W, Hong J, Fang JY. Role of C9orf140 in the promotion of colorectal cancer progression and mechanisms of its upregulation via activation of STAT5, β-catenin and EZH2. Carcinogenesis. 2014 Apr 8. [Epub ahead of print] PubMed PMID: 24608043.
CDC20	-0.903530861	1: Chang DZ, Ma Y, Ji B, Liu Y, Hwu P, Abbruzzese JL, Logsdon C, Wang H. Increased CDC20 expression is associated with pancreatic ductal adenocarcinoma differentiation and progression. J Hematol Oncol. 2012 Apr 4;5:15. doi: 10.1186/1756-8722-5-15. PubMed PMID: 22475564; PubMed Central PMCID: PMC3350393.

FASN	-0.881406134	1: Agostini M, Almeida LY, Bastos DC, Ortega RM, Moreira FS, Seguin F, Zecchin KG, Raposo HF, Oliveira HC, Amoêdo ND, Salo T, Coletta RD, Graner E. The Fatty Acid synthase inhibitor orlistat reduces the growth and metastasis of orthotopic tongue oral squamous cell carcinomas. Mol Cancer Ther. 2014 Mar;13(3):585-95.doi: 10.1158/1535-7163.MCT-12-1136. Epub 2013 Dec 20. PubMed PMID: 24362464.
TPX2	-0.835015046	1: Wei P, Zhang N, Xu Y, Li X, Shi D, Wang Y, Li D, Cai S. TPX2 is a novel prognostic marker for the growth and metastasis of colon cancer. J Transl Med. 2013 Dec 17;11:313. doi: 10.1186/1479-5876-11-313. PubMed PMID: 24341487; PubMed Central PMCID: PMC3878622.
PLAGL2	-0.821711901	1: Sekiya R, Maeda M, Yuan H, Asano E, Hyodo T, Hasegawa H, Ito S, Shibata K, Hamaguchi M, Kikkawa F, Kajiyama H, Senga T. PLAGL2 regulates actin cytoskeletal architecture and cell migration. Carcinogenesis. 2014 Mar 27. [Epub ahead of print] PubMed PMID: 24675530.
RFX7	-0.816916051	1: Yau C, Esserman L, Moore DH, Waldman F, Sninsky J, Benz CC. A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. Breast Cancer Res. 2010;12(5):R85. doi: 10.1186/bcr2753. Epub 2010 Oct 14. PubMed PMID: 20946665; PubMed Central PMCID: PMC3096978.
BCL9	-0.799978961	1: Mani M, Carrasco DE, Zhang Y, Takada K, Gatt ME, Dutta-Simmons J, Ikeda H, Diaz-Griffero F, Pena-Cruz V, Bertagnolli M, Myeroff LL, Markowitz SD, Anderson KC, Carrasco DR. BCL9 promotes tumor progression by conferring enhanced proliferative, metastatic, and angiogenic properties to cancer cells. Cancer Res. 2009 Oct 1;69(19):7577-86. doi: 10.1158/0008-5472.CAN-09-0773. Epub 2009 Sep 8.PubMed PMID: 19738061.
IL17RC	-0.794724505	1: Zhang Q, Liu S, Ge D, Zhang Q, Xue Y, Xiong Z, Abdel-Mageed AB, Myers L, Hill SM, Rowan BG, Sartor O, Melamed J, Chen Z, You Z. Interleukin-17 promotes formation and growth of prostate adenocarcinoma in mouse models. Cancer Res. 2012 May 15;72(10):2589-99. doi: 10.1158/0008-5472.CAN-11-3795. Epub 2012 Mar 28.PubMed PMID: 22461511; PubMed Central PMCID: PMC3665158.
WDR5	-0.786896096	1: Wu MZ, Tsai YP, Yang MH, Huang CH, Chang SY, Chang CC, Teng SC, Wu KJ.Interplay between HDAC3 and WDR5 is essential for hypoxia-induced epithelial-mesenchymal transition. Mol Cell. 2011 Sep 2;43(5):811-22. doi: 10.1016/j.molcel.2011.07.012. PubMed PMID: 21884981.
TXNRD1	-0.728249476	paper from MarcusCadenas C, Franckenstein D, Schmidt M, Gehrmann M, Hermes M, Geppert B, Schormann W, Maccoux LJ, Schug M, Schumann A, Wilhelm C, Freis E, Ickstadt K, Rahnenführer J, Baumbach JI, Sickmann A, Hengstler JG. Role of thioredoxin reductase 1 and thioredoxin interacting protein in prognosis of breast cancer.Breast Cancer Res. 2010;12(3):R44. doi: 10.1186/bcr2599. Epub 2010 Jun 28. PubMed PMID: 20584310; PubMed Central PMCID: PMC2917039.
SUZ12	-0.716149719	Fan Y, Shen B, Tan M, Mu X, Qin Y, Zhang F, Liu Y. TGF-β-Induced Upregulation of malat1 Promotes Bladder Cancer Metastasis by Associating with suz12. Clin Cancer Res. 2014 Mar 15;20(6):1531-41. doi: 10.1158/1078-0432.CCR-13-1455. Epub 2014 Jan 21. PubMed PMID: 24449823.

ERP29	-0.703540375	1: Yuan LW, Liu DC, Yang ZL. Correlation of S1P1 and ERp29 expression to progression, metastasis, and poor prognosis of gallbladder adenocarcinoma. Hepatobiliary Pancreat Dis Int. 2013 Apr;12(2):189-95. PubMed PMID: 23558074.
CRKL	-0.686060455	1: Shen Q, Rahn JJ, Zhang J, Gunasekera N, Sun X, Shaw AR, Hendzel MJ, Hoffman P, Bernier A, Hugh JC. MUC1 initiates Src-CrkL-Rac1/Cdc42-mediated actin cytoskeletal protrusive motility after ligating intercellular adhesion molecule-1. Mol Cancer Res. 2008 Apr;6(4):555-67. doi: 10.1158/1541-7786.MCR-07-2033. PubMed PMID: 18403635.
RUNX2	-0.686041494	Chimge NO, Baniwal SK, Little GH, Chen YB, Kahn M, Tripathy D, Borok Z, Frenkel B. Regulation of breast cancer metastasis by Runx2 and estrogen signaling: the role of SNAI2. Breast Cancer Res. 2011;13(6):R127. doi: 10.1186/bcr3073. Epub 2011 Dec 9. PubMed PMID: 22151997; PubMed Central PMCID: PMC3326569.
G6PD	-0.641711365	Du W, Jiang P, Mancuso A, Stonestrom A, Brewer MD, Minn AJ, Mak TW, Wu M, Yang X. TAp73 enhances the pentose phosphate pathway and supports cell proliferation. Nat Cell Biol. 2013 Aug;15(8):991-1000. doi: 10.1038/ncb2789. Epub 2013 Jun 30.PubMed PMID: 23811687; PubMed Central PMCID: PMC3733810.
EXTL3	0.705143899	Karibe T, Fukui H, Sekikawa A, Shiratori K, Fujimori T. EXTL3 promoter methylation down-regulates EXTL3 and heparan sulphate expression in mucinous colorectal cancers. J Pathol. 2008 Sep;216(1):32-42. doi: 10.1002/path.2377.PubMed PMID: 18543267
PPARA	0.715963012	Goetze S, Eilers F, Bungenstock A, Kintscher U, Stawowy P, Blaschke F, Graf K, Law RE, Fleck E, Gräfe M. PPAR activators inhibit endothelial cell migration by targeting Akt. Biochem Biophys Res Commun. 2002 May 24;293(5):1431-7. PubMed PMID: 12054675.
BAK1	0.716362153	Shi XB, Xue L, Yang J, Ma AH, Zhao J, Xu M, Tepper CG, Evans CP, Kung HJ, deVere White RW. An androgen-regulated miRNA suppresses Bak1 expression and induces androgen-independent growth of prostate cancer cells. Proc Natl Acad Sci U S A. 2007 Dec 11;104(50):19983-8. Epub 2007 Dec 3. PubMed PMID: 18056640; PubMed Central PMCID: PMC2148409
CDK8	0.727009255	Gu W, Wang C, Li W, Hsu FN, Tian L, Zhou J, Yuan C, Xie XJ, Jiang T, Addya S, Tai Y, Kong B, Ji JY. Tumor-suppressive effects of CDK8 in endometrial cancer cells. Cell Cycle. 2013 Mar 15;12(6):987-99. doi: 10.4161/cc.24003. Epub 2013 Mar 1. PubMed PMID: 23454913; PubMed Central PMCID: PMC3637357.
VGLL4	0.773129051	Zhang W, Gao Y, Li P, Shi Z, Guo T, Li F, Han X, Feng Y, Zheng C, Wang Z, Li F, Chen H, Zhou Z, Zhang L, Ji H. VGLL4 functions as a new tumor suppressor in lung cancer by negatively regulating the YAP-TEAD transcriptional complex. Cell Res. 2014 Mar;24(3):331-43. doi: 10.1038/cr.2014.10. Epub 2014 Jan 24. PubMed PMID: 24458094; PubMed Central PMCID: PMC3945886.
EBPL	0.824555421	
CMTM4	0.853446723	Wang Y, Li J, Cui Y, Li T, Ng KM, Geng H, Li H, Shu XS, Li H, Liu W, Luo B, Zhang Q, Mok TS, Zheng W, Qiu X, Srivastava G, Yu J, Sung JJ, Chan AT, Ma D, Tao Q, Han W. CMTM3, located at the critical tumor suppressor locus 16q22.1, is silenced by CpG methylation in carcinomas and inhibits tumor cell growth through inducing apoptosis. Cancer Res. 2009 Jun

		15;69(12):5194-201. doi: 10.1158/0008-5472.CAN-08-3694. Epub 2009 Jun 9. PubMed PMID: 19509237.
BTN3A1	0.858748958	Le Page C, Marineau A, Bonza PK, Rahimi K, Cyr L, Labouba I, Madore J, Delvoye N, Mes-Masson AM, Provencher DM, Cailhier JF. BTN3A2 expression in epithelial ovarian cancer is associated with higher tumor infiltrating T cells and a better prognosis. PLoS One. 2012;7(6):e38541. doi: 10.1371/journal.pone.0038541. Epub 2012 Jun 7. PubMed PMID: 22685580; PubMed Central PMCID: PMC3369854.
TLR3	0.867317442	Hsu WM, Huang CC, Wu PY, Lee H, Huang MC, Tai MH, Chuang JH. Toll-like receptor 3 expression inhibits cell invasion and migration and predicts a favorable prognosis in neuroblastoma. Cancer Lett. 2013 Aug 19;336(2):338-46.doi: 10.1016/j.canlet.2013.03.024. Epub 2013 Mar 27. PubMed PMID: 23541683.
BTN3A2	0.886326508	Le Page C, Marineau A, Bonza PK, Rahimi K, Cyr L, Labouba I, Madore J, Delvoye N, Mes-Masson AM, Provencher DM, Cailhier JF. BTN3A2 expression in epithelial ovarian cancer is associated with higher tumor infiltrating T cells and a better prognosis. PLoS One. 2012;7(6):e38541. doi: 10.1371/journal.pone.0038541. Epub 2012 Jun 7. PubMed PMID: 22685580; PubMed Central PMCID: PMC3369854.
IFIT2	0.925200999	Lai KC, Liu CJ, Chang KW, Lee TC. Depleting IFIT2 mediates atypical PKC signaling to enhance the migration and metastatic activity of oral squamous cell carcinoma cells. Oncogene. 2013 Aug 8;32(32):3686-97. doi: 10.1038/onc.2012.384. Epub 2012 Sep 17. PubMed PMID: 22986528.
DPP4	1.069833421	Wesley UV, Tiwari S, Houghton AN. Role for dipeptidyl peptidase IV in tumor suppression of human non-small cell lung carcinoma cells. Int J Cancer. 2004 May 10;109(6):855-66. PubMed PMID: 15027119.
RECK	1.378304473	Oh J, Seo DW, Diaz T, Wei B, Ward Y, Ray JM, Morioka Y, Shi S, Kitayama H,Takahashi C, Noda M, Stetler-Stevenson WG. Tissue inhibitors of metalloproteinase 2 inhibits endothelial cell migration through increased expression of RECK.Cancer Res. 2004 Dec 15;64(24):9062-9. PubMed PMID: 15604273

Genes differentially regulated upon TSC22D1 knock down in MDA-MB-231 cell line

id	log2FoldChange	Reference
DUSP2	2.142071997	Lin, S. C., Hsiao, K. Y., & Tsai, S. J. (2013). Regulation of epithelial-mesenchymal transition and apoptosis by dual specificity phosphatase-2. The FASEB Journal, 27, 1043-4.
LAMA1	1.840433451	Virtanen, I., Gullberg, D., Rissanen, J., Kivilaakso, E., Kiviluoto, T., Laitinen, L. A., ... & Ekblom, P. (2000). Laminin α1-chain shows a restricted distribution in epithelial basement membranes of fetal and adult human tissues. *Experimental cell research*, *257*(2), 298-309.
ITGB2	1.772590541	Ahn, J., Yoon, Y., Yeu, Y., Lee, H., & Park, S. (2013). Impact of TGF-b on breast cancer from a quantitative proteomic analysis. Computers in biology and medicine, 43(12), 2096-2102.

LHX6	1.66906375	Liu, W., Jiang, X., Han, F., Li, Y., Chen, H., Liu, Y., ... & Liu, J. (2013). LHX6 acts as a novel potential tumour suppressor with epigenetic inactivation in lung cancer. *Cell death & disease*, *4*(10), e882.
PTPRR	1.543831062	Su, P. H., Lin, Y. W., Huang, R. L., Liao, Y. P., Lee, H. Y., Wang, H. C., ... & Lai, H. C. (2013). Epigenetic silencing of PTPRR activates MAPK signaling, promotes metastasis and serves as a biomarker of invasive cervical cancer. Oncogene, 32(1), 15-26.
FAM123B	1.366626252	Moisan, A., Rivera, M. N., Lotinun, S., Akhavanfard, S., Coffman, E. J., Cook, E. B., ... & Bardeesy, N. (2011). The< i> WTX</i> Tumor Suppressor Regulates Mesenchymal Progenitor Cell Fate Specification. Developmental cell, 20(5), 583-596.
BCL7A	1.358495287	Carbone, A., Bernardini, L., Valenzano, F., Bottillo, I., De Simone, C., Capizzi, R., ... & Amerio, P. (2008). Array-based comparative genomic hybridization in early-stage mycosis fungoides: Recurrent deletion of tumor suppressor genes BCL7A, SMAC/DIABLO, and RHOF. *Genes, Chromosomes and Cancer*, *47*(12), 1067-1075.
OBSCN	1.225076876	Chattopadhyay, I., Singh, A., Phukan, R., Purkayastha, J., Kataki, A., Mahanta, J., ... & Kapur, S. (2010). Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma exposed to tobacco and betel quid from high-risk area in India. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, *696*(2), 130-138.
FLRT2	1.168923749	Egea, J., Erlacher, C., Montanez, E., Burtscher, I., Yamagishi, S., Heß, M., ... & Klein, R. (2008). Genetic ablation of FLRT3 reveals a novel morphogenetic function for the anterior visceral endoderm in suppressing mesoderm differentiation. Genes & development, 22(23), 3349-3362.
CPEB2	1.160623208	Nairismägi, M. L., Vislovukh, A., Meng, Q., Kratassiouk, G., Beldiman, C., Petretich, M., ... & Groisman, I. (2012). Translational control of TWIST1 expression in MCF-10A cell lines recapitulating breast cancer progression. Oncogene, 31(47), 4960-4966.
PHLPP2	-1.230621153	Ghalali, A., Ye, Z. W., Högberg, J., & Stenius, U. (2014). PTEN and PHLPP Crosstalk in Cancer Cells and in TGFβ-Activated Stem Cells. Journal of Biological Chemistry, jbc-M113.
MIR100HG	-1.265606705	Chen, D., Sun, Y., Yuan, Y., Han, Z., Zhang, P., Zhang, J., ... & Ma, L. (2014). miR-100 Induces Epithelial-Mesenchymal Transition but Suppresses Tumorigenesis, Migration and Invasion. *PLoS genetics*, *10*(2), e1004177.
S1PR1	-1.286333644	ROSENFELDT, H. M., HOBSON, J. P., MACEYKA, M., OLIVERA, A., NAVA, V. E., MILSTIEN, S., & SPIEGEL, S. (2001). EDG-1 links the PDGF receptor to Src and focal adhesion kinase activation leading to lamellipodia formation and cell migration. *The FASEB Journal*, *15*(14), 2649-2659.
DBF4	-1.299525234	Nambiar, S., Mirmohammadsadegh, A., Hassan, M., Mota, R., Marini, A., Alaoui, A., ... & Hengge, U. R. (2007). Identification and functional characterization of ASK/Dbf4, a novel cell survival gene in cutaneous melanoma with prognostic relevance. *Carcinogenesis*, *28*(12), 2501-2510.
ADAM10	-1.309098448	Gavert, N., Sheffer, M., Raveh, S., Spaderna, S., Shtutman, M., Brabletz, T., ... & Ben-Ze'ev, A. (2007).

		Expression of L1-CAM and ADAM10 in human colon cancer cells induces metastasis. *Cancer research*, *67*(16), 7703-7712.
WNT9A	-1.309618635	Person, A. D., Garriock, R. J., Krieg, P. A., Runyan, R. B., & Klewer, S. E. (2005). Frzb modulates Wnt-9a-mediated β-catenin signaling during avian atrioventricular cardiac cushion development. *Developmental biology*, *278*(1), 35-48.
TNFRSF11A	-1.322120035	Palafox, M., Ferrer, I., Pellegrini, P., Vila, S., Hernandez-Ortega, S., Urruticoechea, A., ... & González-Suárez, E. (2012). RANK induces epithelial–mesenchymal transition and stemness in human mammary epithelial cells and promotes tumorigenesis and metastasis. *Cancer research*, *72*(11), 2879-2888.
PKN3	-1.350450711	Unsal-Kacmaz, K., Ragunathan, S., Rosfjord, E., Dann, S., Upeslacis, E., Grillo, M., ... & Klippel, A. (2012). The interaction of PKN3 with RhoC promotes malignant growth. *Molecular oncology*, *6*(3), 284-298.
PBK	-1.536166529	Ayllon, V., & O'connor, R. (2007). PBK/TOPK promotes tumour cell proliferation through p38 MAPK activity and regulation of the DNA damage response. *Oncogene*, *26*(24), 3451-3461.
CMKLR1	-1.575483083	Yoshimura, T., & Oppenheim, J. J. (2011). Chemokine-like receptor 1 (CMKLR1) and chemokine (C–C motif) receptor-like 2 (CCRL2); Two multifunctional receptors with unusual properties. *Experimental cell research*, *317*(5), 674-684.
NGFR	-1.685707784	Isohata, N., Aoyagi, K., Mabuchi, T., Daiko, H., Fukaya, M., Ohta, H., ... & Sasaki, H. (2009). Hedgehog and epithelial-mesenchymal transition signaling in normal and malignant epithelial cells of the esophagus. *International Journal of Cancer*, *125*(5), 1212-1221.
EGR1	-1.929342421	Tarcic, G., Avraham, R., Pines, G., Amit, I., Shay, T., Lu, Y., ... & Yarden, Y. (2012). EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF. The FASEB Journal, 26(4), 1582-1592.
FOXM1	-2.002905673	Bao, B., Wang, Z., Ali, S., Kong, D., Banerjee, S., Ahmad, A., ... & Sarkar, F. H. (2011). Over-expression of FoxM1 leads to epithelial–mesenchymal transition and cancer stem cell phenotype in pancreatic cancer cells. *Journal of cellular biochemistry*, *112*(9), 2296-2306.
S100A2	-2.214097398	Naz, S., Bashir, M., Ranganathan, P., Bodapati, P., Santosh, V., & Kondaiah, P. (2014). Protumorigenic actions of S100A2 involve regulation of PI3/Akt signaling and functional interaction with Smad3. *Carcinogenesis*, *35*(1), 14-23.
CTBP1	-0.816191774	Zhang, X. L., Huang, C. X., Zhang, J., Inoue, A., Zeng, S. E., & Xiao, S. J. (2013). CtBP1 is involved in epithelial-mesenchymal transition and is a potential therapeutic target for hepatocellular carcinoma. Oncology reports, 30(2), 809-814.

CTNNB1	-0.80314676	Stewart, C. A., Wang, Y., Bonilla-Claudio, M., Martin, J. F., Gonzalez, G., Taketo, M. M., & Behringer, R. R. (2013). CTNNB1 in Mesenchyme Regulates Epithelial Cell Differentiation during Müllerian Duct and Postnatal Uterine Development. Molecular Endocrinology, 27(9), 1442-1454.

Supplementary Table S3. The table contains details of all siRNAs employed in this study.

Gene Name	SMART pools siRNA sequences
Nr1h4	UGCCAGGAGUGCCGGCUAA GAUUUGUGCCGGACGGGAU GUGUAAAUCUAAACGGCUA GAAACUUCCUGCCGGACAU
Pcbp4	GCACAUAUUAAGAUCGGGA GGGAGAGACUGUAAAGCGA GGAUGCUGAUGCACGGGAA AAAUCAAGGAGAUCCGAGA
Prkrir	AAGUUGAGAAUGAGCGCUA ACAAAUGUUAGGUGGAAUA GGGAUUGAAUAUGGAGUAU AUUUAAUGGUGGACACAUA
Etv6	GCAUUAAGCAGGAACGAAU UGAUGAGAGUCACGUGCUA GAUCGACAUUAAAGCGACA CCAAAGAGGAUUUCCGCUA
Nfil3	CCGCACAAGCUUCGGAUUA GGAAGUUGCAUCUCAGUCA CAAAGACGGUGGCUAUAAG AGAAAGACGCCAUGUAUUG
Tsc22d1	GCUGAAGUCUUGCGAAGUA GUAUAGUCGUUCUGCGCAU GCUCAGGAUCAACCGCAUA GUGUGGUAGCUAUCGACAA
Smad4	UCAGGUGGCUGGUCGGAAA GAGUGCAGUUGGAAUGUAA GCAAUUGAGAGUUUGGUAA UCAGUAUGCGUUUGACUUA
Tcf12	CUAUCCAUGUGCUACGAAA CUUGAAUGACAGUCGAUUA GCAGCAACUUCACGAGCAU ACACAAGUCUUCCGCCAAU
Hoxb9	CCAAGUUUCCUUCGGGCCA UCAUAAGUCACGAGAGCGA GUUAGAAGCUAGCUAUACU GGACAAUAAAAUUUGCGAA

Supplementary Table S4. This table provides details of all primers used in this study.

Ctcf	Mouse	F	gcgctatcatgatcccaactttgt
	Mouse	R	gcatgtcttgccattgtgttcc
Atf4	Mouse	F	atggcaggtgtgacagtctgg
	Mouse	R	cgctctaggatcaggggtctgc
Pgcp	Mouse	F	tccctggaatcaagaataacagtca
	Mouse	R	tgctccatctctaaaaagcatagagtg
Cnnm	Mouse	F	aaggtcaacgaaaatcccagaagt
	Mouse	R	tgcactctgtcgctccgttct
Plekhh	Mouse	F	cagcttccagtattgccttttgc
	Mouse	R	ccttccttgaccttcgggattta
Rheb	Mouse	F	gggaaagggagcagaaagaaca
	Mouse	R	ggccatgtcatgcctagtaaaca
Trip6	Mouse	F	ctaggctgggctcaaactcacac
	Mouse	R	acaggatgaggctggtgaaagc
Cadm4	Mouse	F	cagtctttccctgcccacattc
	Mouse	R	cgtgtgctgtgagtgtgagcttt
Lpp	Mouse	F	ccactcctccccttttcagtgc
	Mouse	R	tctcttccatagaagggcagtatcc
Ltbp1	Mouse	F	ggccaacccccctctgcattac
	Mouse	R	cagggcaaggcagattaaagagc
Fn1	Mouse	f	agcagaggcttatatgggaca
	Mouse	r	tgacgtcacccggactct
Cdh2	Mouse	F	ggtggaggagaagaagaccagga
	Mouse	R	tggcatcaggctccacagtatct
Ncam	Mouse	F	aagcacacagagcccaacgag
	Mouse	R	aaggcagcatgtcctcgacagt
Zeb2	Mouse	F	ccagaggaaacaaggatttcagg
	Mouse	R	cggagtctgtcatgtcatctaggc
Snai1	Mouse	F	tctctaggccctggctgcttc
	Mouse	R	cagcaaaagcacggttgcagt
Nr1h4	Mouse	F	agatggggatgttggctgaatgta
	Mouse	R	cgtgctgcttcacatttttcctta
Pcbp4	Mouse	F	ggggacaatgtgggactgagc
	Mouse	R	actgtggctcagctagctctcg
Nfil3	Mouse	F	accagggagcagaaccacgat
	Mouse	R	ttgtccggcacagggtaaatct
Srf	Mouse	F	tgacagcagtggggaaaccaa
	Mouse	R	gtgctgggtgctgtctggatt
Etv6	Mouse	F	tgatccctttcgctgtgagacat
	Mouse	R	gcgtgtatgaaattcgttcctgct
Tcf12	Mouse	F	catttgcaagatgcaatgtccttc
	Mouse	R	gtctaagcggtcttccattcgaga
Prkrir	Mouse	F	gacctggccttcttcaggttcc
	Mouse	R	tgtttggcacaaagtcgataatgtt
Tsc22d1	Mouse	F	gggagtagcagcagcctagcag

	Mouse	R	gctaccacacttgcaccagagc
Zeb1	Mouse	F	gtgatccagccaaacggaaacc
	Mouse	R	tctttttgggtggcgtggagt
Cdh1	Mouse	F	cccgggacaatgtgtattactatga
	Mouse	R	gcagctggctcaaatcaaagtcc
Ocln	Mouse	F	catgtccgtgaggccttttga
	Mouse	R	tggtgcataatgattgggtttga
Smad4	Mouse	F	gagaacattggatggacgact
	Mouse	R	cacagacgggcatagatcac
CDH1	Human	F	tcttggaggaattcttgctttgc
	Human	R	accgctctcctccgaagaaaca
CLDN3	Human	F	aacctgcatggactgtgaaacct
	Human	R	ggtggtcaagtattggcggtcac
CDH2	Human	F	ggccgtcatcacagtgacagat
	Human	R	ccctgttctcaggaacttcacca
ZEB1	Human	F	ccaacagaccagacagtgttaccag
	Human	R	tcttgcccttcctttcctgtgt
MCAM	Human	F	ccccattcctcaagtcatctggta
	Human	R	gtctgggacgactgaatgtggac
HOHB9	Human	F	gacaaagagaggccggatcaaac
	Human	R	tctggtatttggtgtagggacagc
PCBP4	Human	F	ccaagatcaaggagatccgagaga
	Human	R	tgtggagttggggagcaggtc
NFIL3	Human	F	atgcagctgagaaaaatgcagacc
	Human	R	ccacattgctactggcatcaagag
SRF	Human	F	ctgcctaccagcttcaccctcat
	Human	R	gtgcacttgaatggcctgcac
ETV6	Human	F	ccctgcgccactactacaaactaaa
	Human	R	tcatgatttcatctggggttttca
TCF12	Human	F	gaactttaatcggggtggttggat
	Human	R	gttgttgctggggattcatcttc
PRKRIR	Human	F	gaaacaaatgacaaagtttgaatggaa
	Human	R	cacaaaggaatgtaaacagcaacga
TSC22D1	Human	F	ctgacgacacccctggtggat
	Human	R	cgattttgttgtcaatagctaccacac
CTCF	Human	F	gggcttgagagctgggttctatt
	Human	R	cttcgactgcatcaccttccatt

FBXO32 promotes microenvironment underlying Epithelial-Mesenchymal Transition via CtBP1 during tumour metastasis and brain development

Sanjeeb Kumar Sahu[1], Neha Tiwari[2], Abhijeet Pataskar[1], Yuan Zhuang[1], Marina Borisova[1], Mustafa Diken[3], Susanne Strand[4], Petra Beli[1], Vijay K. Tiwari[1,*]

1. Institute of Molecular Biology (IMB), Ackermannweg 4, 55128 Mainz, Germany
2. Institute of Physiological Chemistry, University Medical Center, Johannes Gutenberg University, 55131 Mainz, Germany
3. TRON - Translational Oncology at the University Medical Center of the Johannes Gutenberg University gGmbH, Freiligrathstrasse 12, 55131 Mainz, Germany
4. Department of Internal Medicine I, University Medical Center, Johannes Gutenberg University, Obere Zahlbacher Straße 63, 55131 Mainz, Germany

*Correspondence should be addressed to: v.tiwari@imb-mainz.de

Nat Commun. 2017 Nov 15;8(1):1523. doi: 10.1038/s41467-017-01366-x.

ABSTRACT

The set of events that convert adherent epithelial cells into migratory cells are collectively known as epithelial-mesenchymal transition (EMT). EMT is involved during development, for example, in triggering neural crest migration, and in pathogenesis such as metastasis. Here we discover FBXO32, an E3 ubiquitin ligase, to be critical for hallmark gene expression and phenotypic changes underlying EMT. Interestingly, FBXO32 directly ubiquitinates CTBP1, which is required for its stability and nuclear retention. This is essential for epigenetic remodeling and transcriptional induction of CTBP1 target genes, which create a suitable microenvironment for EMT progression. FBXO32 is also amplified in metastatic cancers and its depletion in a NSG mouse xenograft model inhibits tumor growth and metastasis. In addition, FBXO32 is essential for neuronal EMT during brain development. Together, these findings establish that FBXO32 acts as an upstream regulator of EMT by governing the gene expression program underlying this process during development and disease.

Epithelial cells remain in close contact with their neighbors and maintain an apical-basal axis of polarity by the sequential arrangement of adherent junctions, desmosomes, and tight junctions [188]. Following the induction of epithelial to mesenchymal transition (EMT) program, cells undergo molecular and phenotypic remodeling that involves changes like cytoskeletal reorganization and loss of cell-cell junctions. This allows epithelial cells to escape from their original location by acquiring a migratory, mesenchymal identity [189]. Such a dramatic change in cell fate is essential for key processes during embryonic development, such as embryo implantation, embryonic layer formation, gastrulation and neural tube formation. In adults, this change in cell fate is important for processes like tissue regeneration and wound healing [188,189,191,196]. However, aberrant activation of the EMT program is associated with disease phenotypes such as organ fibrosis [193] and tumor metastasis [196,254].

A number of signaling pathways, such as TGF-β, FGF, EGF, HGF, Wnt/β-catenin and Notch are known to induce EMT [188,255]. Among these, TGF-β is known to be the most potent and a prototypic inducer of EMT in various contexts, including development and cancer metastasis, whereas the others have more context-specific functions [189,198]. A large number of evidences have established that the microenvironment plays a critical role during initiation and progression of EMT[256]. However, the effectors, through which TGF-β mediates remodeling of the microenvironment to promote EMT, remain poorly explored. It is also well established that EMT relies on defined, genome-wide transcriptional reprogramming [242,257]. A number of transcription factors are implicated in this process, including SNAIL, ZEB, and several basic helix-loop-helix proteins [197,198,257]. Importantly, these transcription factors modulate gene expression in partnership with co-regulator proteins such as CTBPs [258-260]. Several studies have reported that CtBPs form complexes with a variety of epigenetic regulators or corepressor complexes that recruit epigenetic regulators [258,260,261]. CtBPs are also known to undergo dynamic posttranslational modifications which influence their stability or subcellular localization [259]. Recent studies have suggested dispensability of the established key EMT

transcription factors in driving metastasis and further vouched for a need of exploring more potent factors driving this process [214,262,263].

Advances in proteomics have begun to highlight the role of post-translational modifications during EMT and their enormous complexity and regulatory potential [198,264]. Ubiquitination via the ubiquitin-proteasome system governs diverse cellular processes, such as cell proliferation, cell cycle progression, transcription and apoptosis [265]. F-box proteins, the substrate-recognition subunits of Skp1–Cullin1–F-box protein E3 ligase complex, play pivotal roles in ubiquitination and subsequent activation or degradation of target proteins, depending on the lysine residue of ubiquitin (Ub) that participates in the formation of poly-ubiquitin chains [266,267]. Recent studies have shed light on the biological functions attributed by these F-box proteins [268]. FBXO32 (also known as Mafbx/Atrogin1) was first identified as a muscle-specific F-box protein, and further studies indicated its importance during heart development and muscle homeostasis [269,270]. This protein is also induced upon stress, for instance, during serum starvation and hypoxia, and it functions as an apoptosis regulator [271,272]. In a previous study, *FBXO32* promoter was shown to be repressed by DNA methylation in ovarian cancer and its expression was correlated with shorter progression-free survival [273]. FBXO32 has been implicated in the regulation of transcription factor stability (e.g. Myc) and localization (e.g. Foxo1 and Foxo3a) [270,274]. A very recent study showed an association between EMT and *Fbxo32* in tumors with acquired platinum resistance [275]. However, despite these advances and indications of its role in cancer, the function of FBXO32 in EMT progression, metastasis and its contributions to the gene regulatory circuitry underlying these processes remain completely unknown.

Here we discover FBXO32 to be required in various contexts of EMT including in disease and development. We report that FBXO32 is substantially induced during EMT and plays a critical role in the transition towards mesenchymal identity by governing the required gene expression program. We show that FBXO32-dependent K63-linked ubiquitination of CtBP1 is required for its nuclear retention, which is essential for mediating transcriptional changes via epigenetic reprogramming of EMT target genes. These

include various chemokines, chemokine receptors and matrix metalloproteinases, thus promoting a suitable environment promoting EMT progression. FBXO32 is also highly amplified in a large panel of cancers, and its depletion severely impairs the metastatic properties of cancer cells both *in vitro* and *in vivo*. Furthermore, depletion of FBXO32 during brain development impairs neuronal EMT. Together, these observations establish FBXO32 among the most potent regulator of EMT by governing the gene expression program that underlies this process during both development and disease.

RESULTS

FBXO32 is essential for Epithelial-Mesenchymal Transition

To identify so far unknown regulators of EMT, we employed an established model system in which, normal murine mammary gland epithelial cells (NMuMG) undergo TGF-β-induced EMT with high synchrony and purity [197,257] (Supplementary Fig. 1a-b). To examine global alterations in the transcription of coding and noncoding genes during EMT, we analyzed genome-wide transcription profiles obtained via RNA-seq from NMuMG cells treated with TGF-β at several consecutive time points representing untreated, early (day 1), intermediate (day 4), and late (day 7) stages of EMT [257]. To identify new regulators of EMT that function in the nuclear compartment, we performed a stepwise selection approach to obtain candidate genes that are differentially expressed during EMT, contain a nuclear localization signal, and are aberrantly expressed in various cancers (using the NextBio, cBioPortal, and Oncomine databases). We further analyzed the expression patterns of these candidates in other established model systems of EMT, including TGF-β-induced EMT in a mouse Ras-transformed breast epithelial cell line (EpRAS) and human primary mammary epithelial cells (HMECs) (Fig 1a, Supplementary Fig. 1c-f). We retained only those candidates that showed consistent expression patterns. These candidates were finally tested for their functional role using loss-of-function experiments using a pool of four siRNAs targeting the same gene, followed by extensive phenotypic and molecular characterization. Among these candidates, *FBXO32* emerged as the most potent regulator of EMT.

Fbxo32 showed strong transcriptional induction during TGF-β-induced EMT in NMuMG cells (Fig 1b and 1d, Supplementary Fig. 2a). FBXO25, the closest evolutionarily conserved member to FBXO32 (based on sequence and protein domain similarity), remained transcriptionally unchanged during EMT (Supplementary Fig. 2b-c). Furthermore, *FBXO32* induction was dependent on TGF-β SMAD axis during onset of EMT (Supplementary Fig. 2d), in line with previous observations [273,276]. Moreover, among all of the F-box-containing proteins (n=69), only *Fbxo32* showed significantly strong upregulation during EMT

(Supplementary Fig. 2e). In line with these observations, a similar transcriptional induction of *Fbxo32* was observed in other EMT models in mouse (EpRas cells) as well as human (primary human mammary epithelial cells (HMEC) and MCF7 cells) (Fig 1c and 1e, Supplementary Fig. 2f-g).

Next, we investigated the function of *Fbxo32* during TGF-β-induced EMT in mouse (NMuMG) cells using loss-of-function approaches both by independent siRNAs and a pool of siRNAs to ensure no off-target effects. Such depletion of Fbxo32 during EMT led to a strong blockage in acquiring a mesenchymal state in all tested cell types. This phenotype was confirmed by an immunofluorescence assay that showed noticeable retention in the epithelial markers (e.g. E-cadherin and ZO-1) at the membrane, inability to acquire proper levels of crucial mesenchymal markers (e.g. N-cadherin, Fibronectin), and failure to undergo cytoskeletal remodeling involving cortical actin, stress fibers and focal adhesion formation (e.g. Phalloidin and Paxillin) (Fig 1f). Such blockage in undergoing EMT upon *Fbxo32* depletion was further validated at RNA and protein levels for several key EMT markers (Fig 1g, Supplementary Fig. 2h). In support of *Fbxo32*-dependancy of these changes, we could rescue the phenotypic and molecular alterations upon Fbxo32 depletion by overexpressing a siRNA-resistant *FBXO32* (Supplementary Fig. 2i-j). Importantly, all key findings were reproduced in established human EMT model systems ((TGF-β-induced EMT in human mammary epithelial cells (HMEC) and MCF7 cells)) (Fig 1h, Supplementary Fig. 3a-d). Together, these observations reveal *FBXO32* is critical for phenotypic and transcriptional changes underlying EMT.

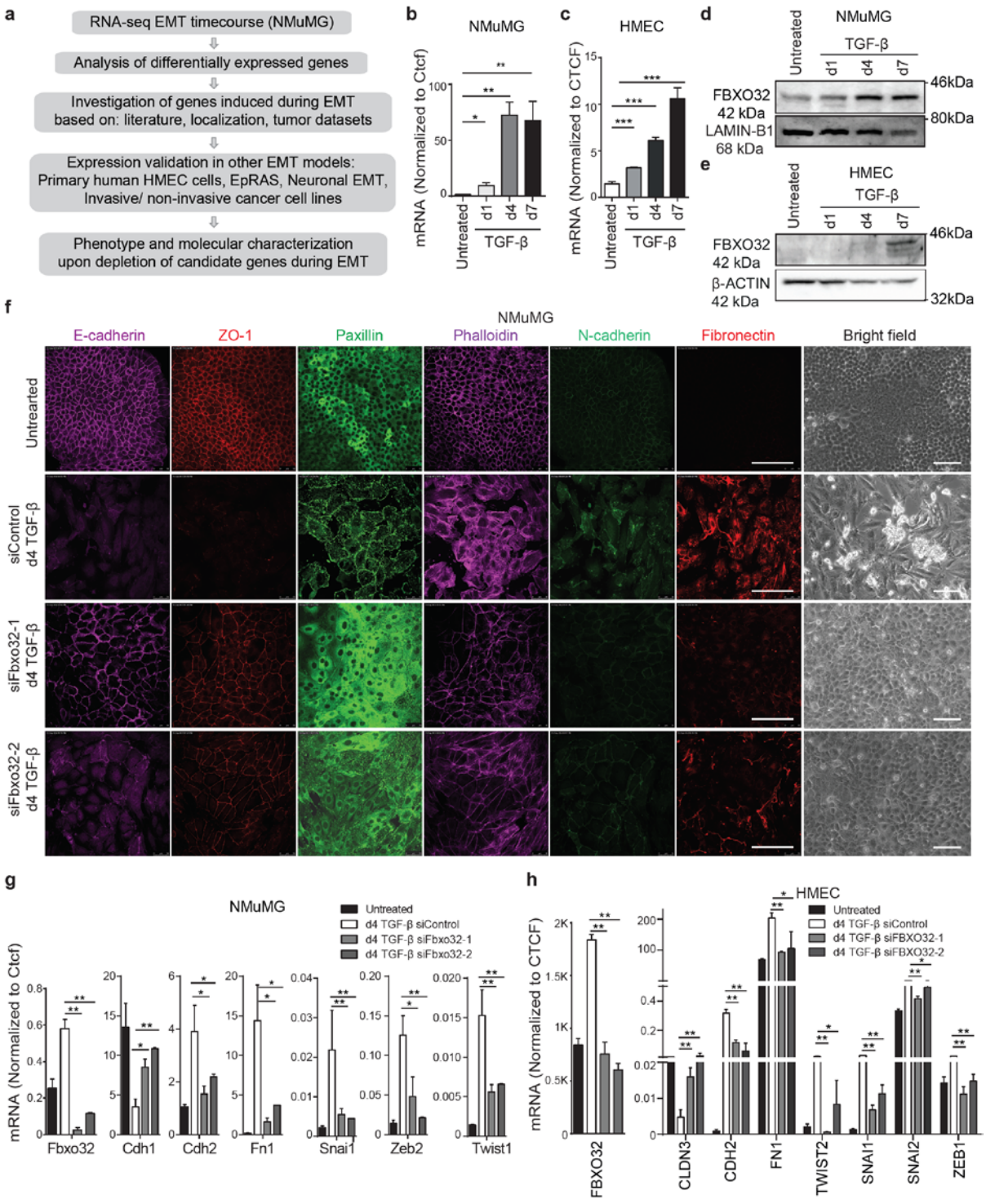


Fig 1. FBXO32 is an essential regulator of phenotypic and transcriptional changes that are hallmark of EMT.

(a) Schematic representation of *in vitro* EMT induction and the identification of new regulators. **(b)** Using RT-PCR, the levels of Fbxo32 mRNA in NMuMG cells during TGF-β-induced EMT were measured relative to Ctcf and plotted on the y-axis. **(c)** FBXO32 mRNA level was measured by RT-PCR in primary human breast epithelial cells (HMEC) undergoing TGF-β-induced EMT as described in Fig1b. **(d-e)** Western blot analysis of FBXO32 in mouse NMuMG cell (d) and human HMECs (e) undergoing TGF-β-induced EMT.

Lamin and β-Actin acted as a loading control. (**f**) Representative bright-field and immunofluorescence images showing the localization and expression levels of EMT marker proteins after 4 days (d4) of siRNA-mediated depletion of FBXO32 compared to non-targeting control (siControl) in NMuMG cells. Staining was performed to assess the expression with antibodies against the epithelial markers E-cadherin and ZO1, the mesenchymal marker N-cadherin, Fibronectin-1, Phalloidin (to visualize the actin cytoskeleton) and Paxillin (to detect focal adhesion plaques). Scale bar, 100 µm. **(g-h)** Using RT-PCR, the levels of FBXO32 and key EMT markers in NMuMG (g) and HMEC (h) cells transfected with either control siRNA or independent siRNAs against FBXO32 during TGF-β-induced EMT were measured relative to CTCF and plotted on the y-axis. All experiments were performed in biological triplicates unless otherwise specified. Error bars represent the SEM of three independent biological replicates (n=3). *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)
See also Supplementary figures 1-3.

FBXO32 induces genes that promote microenvironment underlying EMT

Intrigued by the observed critical role of FBXO32 in phenotypic and molecular changes during EMT, we decided to investigate genome-wide transcriptional alterations following FBXO32 depletion by a pool of four siRNAs during EMT in HMEC (human mammary epithelial cells). A computational analysis revealed that many genes were significantly differentially expressed, with more genes downregulated (n=231) than upregulated (n=102), suggesting a potential activating function of FBXO32 (Fig 2a). This included several established hallmark EMT genes such as CDH1, TJP3, OCLN, CLDN1/3/11 (all epithelial makers) and CDH2, FN1, VIM, ZEB1/2, SNAI1/2, TWIST1/2, (all mesenchymal makers) that were upregulated and downregulated respectively (Fig 2b). Interestingly, a deeper investigation of the downregulated genes uncovered enrichment of important cytokines (interleukin and CXCL family members), cytokine receptors (CXCR and TGFBR1), extracellular matrix components (laminins and collagens) and matrix metalloproteinases (MMPs, e.g. MMP1/3/9/10), which are known to play critical roles in EMT (Fig 2b). In line with these findings, an analysis of the secretome in these cells showed that the functional protein levels of these secretory components, including several MMPs and cytokines, were strongly downregulated upon FBXO32 depletion (Fig 2c-d). The downregulated genes from RNA-seq experiments showed enrichment for ontologies related to cell migration, cell motility and extracellular matrix organization, while the upregulated genes showed enrichment for biological functions such as proliferation, epithelium development and cell adhesion (Fig 2e-f). Interestingly, when we analyzed the

expression dynamics of these misregulated genes in epithelial cells, mesenchymal cells and cells depleted of FBXO32, we found that FBXO32 depleted cells retained a more epithelial-like gene expression profile and were unable to acquire a mesenchymal transcriptome (Fig 2g). A number of these genes were additionally validated for their expression changes independently by RT-PCRs (Fig 2h-i). These findings suggest that FBXO32 plays a critical role in EMT by regulating the underlying gene expression program especially of secretory cytokine and cellular matrix associated factors that are essential to promote a suitable microenvironment for EMT progression.

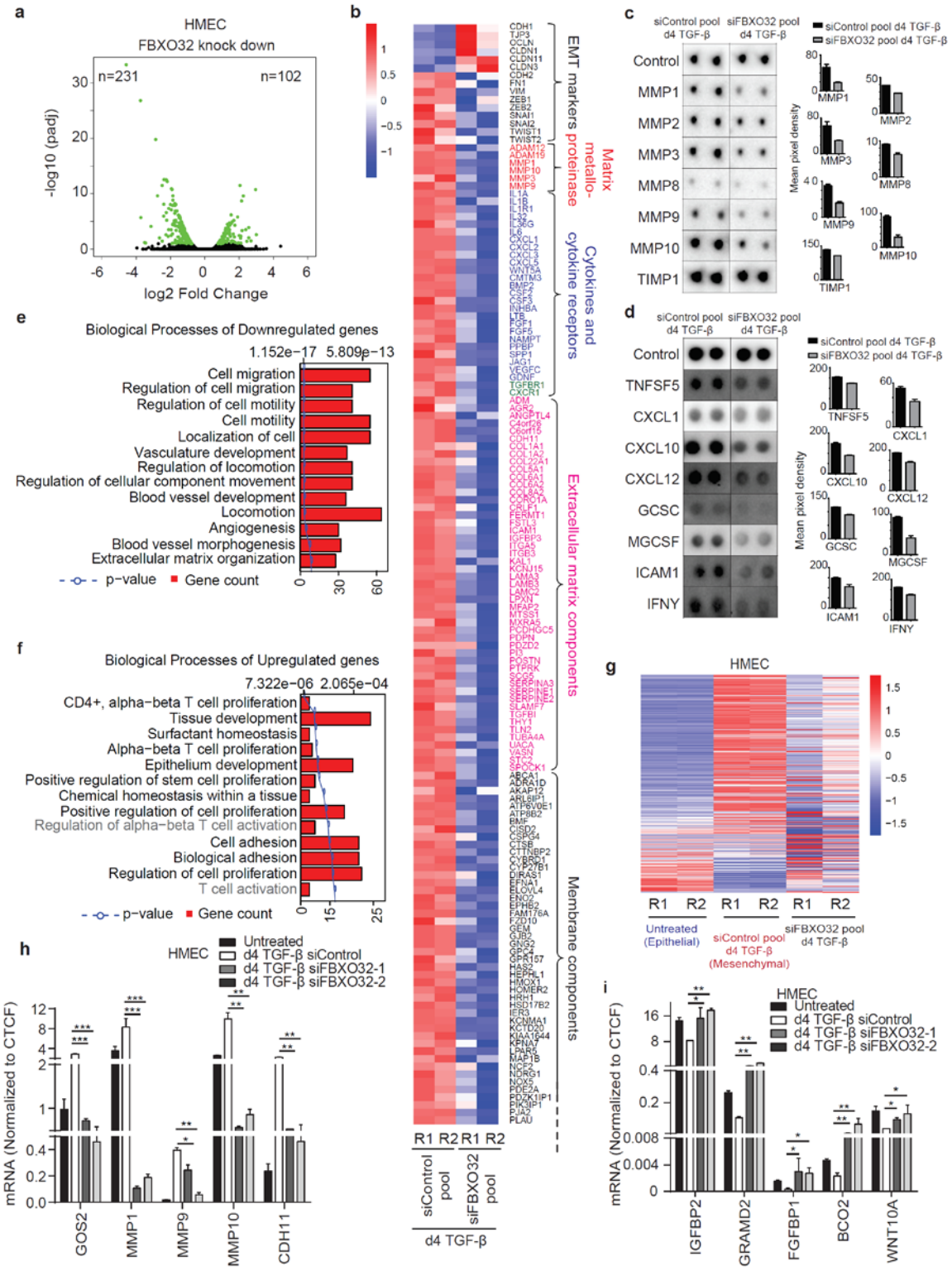


Fig 2. FBXO32 induces genes critical for promoting a suitable microenvironment for EMT progression. **(a)** Volcano plots showing significantly differentially expressed genes upon FBXO32 depletion in HMECs undergoing EMT. **(b)** Heat map showing standardized expression (Z-scores) of key downregulated genes along with EMT markers, upon FBXO32 depletion in HMEC cells undergoing EMT for 4 days. **(c-d)** Representative immunoblot (n=3) of an individual membrane array for human secretory MMPs (c) and

cytokines (d) from HMEC cells upon FBXO32 depletion, and bar-graph showing their quantifications on the right side. Control wells act as loading controls in our analysis to allow comparison of different blots. **(e-f)** GO analysis of downregulated (e) and upregulated genes (f) upon FBXO32 knockdown in HMECs that underwent EMT for 4 days. **(g)** Heat map showing standardized expression (Z-scores) of differentially expressed genes upon FBXO32 depletion in Epithelial HMEC cells, HMEC that underwent EMT for 4 days and treated with control siRNA-pool or siRNA-pool against FBXO32. **(h-i)** The mRNA levels of downregulated genes (h) and upregulated genes (i) during EMT in control and FBXO32 knockdown HMECs that underwent EMT for 4 days were measured relative to CTCF via RT-PCR, and the results were plotted on the y-axis. Error bars represent the SEM of three independent biological replicates (n=3). *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)
See also Supplementary figures 2-3.

FBXO32 ubiquitinates CtBP1 to promote its nuclear retention

Prompted by the critical role of FBXO32 in all the studied contexts of EMT, we sought to gain mechanistic insights into its function. While FBXO32 depletion resulted in transcriptome changes, given the known function of F-box-containing family members as a component of the ubiquitin ligase complex, and with strong nuclear localization (Supplementary Fig. 4a-b), we hypothesized that its gene regulatory function likely involves regulating the activity of other transcriptional regulators via post-translational modifications. We therefore performed immunoprecipitation of FBXO32, followed by liquid chromatography-tandem mass spectrometry, to search for its possible interactors and substrates (Fig 3a). Among the most enriched candidates detected by this analysis we identified C-terminal binding protein1 (CtBP1), which is known to play a key role in promoting EMT [260,277,278]. This interaction between FBXO32 and CtBP1 was further validated in independent co-immunoprecipitation assays both following their overexpression and at the endogenous level (Fig 3b-c, Supplementary Fig. 4c-d). Intriguingly, this interaction was abrogated by deletion of the large F-box domain (115 aa, as predicted by InterPro which encompasses the core F-box domain of 45 aa) (Supplementary Fig. 4d). Since CtBP1 is expressed at very similar levels during EMT and upon FBXO32 depletion (Supplementary Fig. 4e-g), we hypothesized that FBXO32 regulates its function via post-translational modifications.

FBXO32-mediated ubiquitination is known to cause both activation of signaling cascades and proteasome-mediated degradation [270,274]. The overexpression of FBXO32 in human cells resulted in an increased

ubiquitination of CtBP1, which can be blocked by ubiquitin E1 inhibitor UBEI-41/ PYR-41 [279] (Fig 3d, Supplementary Fig. 4h). Moreover, FBXO32 overexpression led to an increase in CtBP1 protein levels, whereas FBXO32 depletion reduced the CtBP1 protein levels in both human and mouse EMT model systems (Fig 3e and Supplementary Fig. 4i-k). Due to a poor transfection efficiency of primary HMEC cells, here we used HMLE cells which originate from HMEC cells and show a similar dynamics of EMT and FBXO32 induction (Supplementary Fig. 4l) [280,281]. We investigated whether the FBXO32-dependent ubiquitination of CtBP1 has consequences on the subcellular distribution of CtBP1. Strikingly, although CtBP1 is highly localized to the nucleus under normal cellular conditions, FBXO32 depletion resulted in CtBP1 localizing exclusively to the cytoplasm in both human and mouse cells (Fig 3f and Supplementary Fig. 5a). The degradation of proteins is known to be mediated via K48-linked poly-ubiquitin chains while K63-linked ubiquitination is associated with the activation of signaling pathways [282]. It is known that FBXO32 can perform both K48 and K63-linked ubiquitination, which results in proteosome-mediated degradation or nuclear localization of target proteins respectively [270,274]. Our further analyses showed that FBXO32 mediates K63-linked ubiquitination, but not K48-linked ubiquitination of CtBP1 (Fig 3g, Supplementary Fig. 5b). Moreover previous finding also indicates that additional non-K63 lysines are needed for full ubiquitination as it partially rescued the ubiquitination phenotype (Supplementary Fig. 5b). Together, these observations establish that FBXO32 interacts with and modifies CtBP1 to promote its stability and nuclear retention.

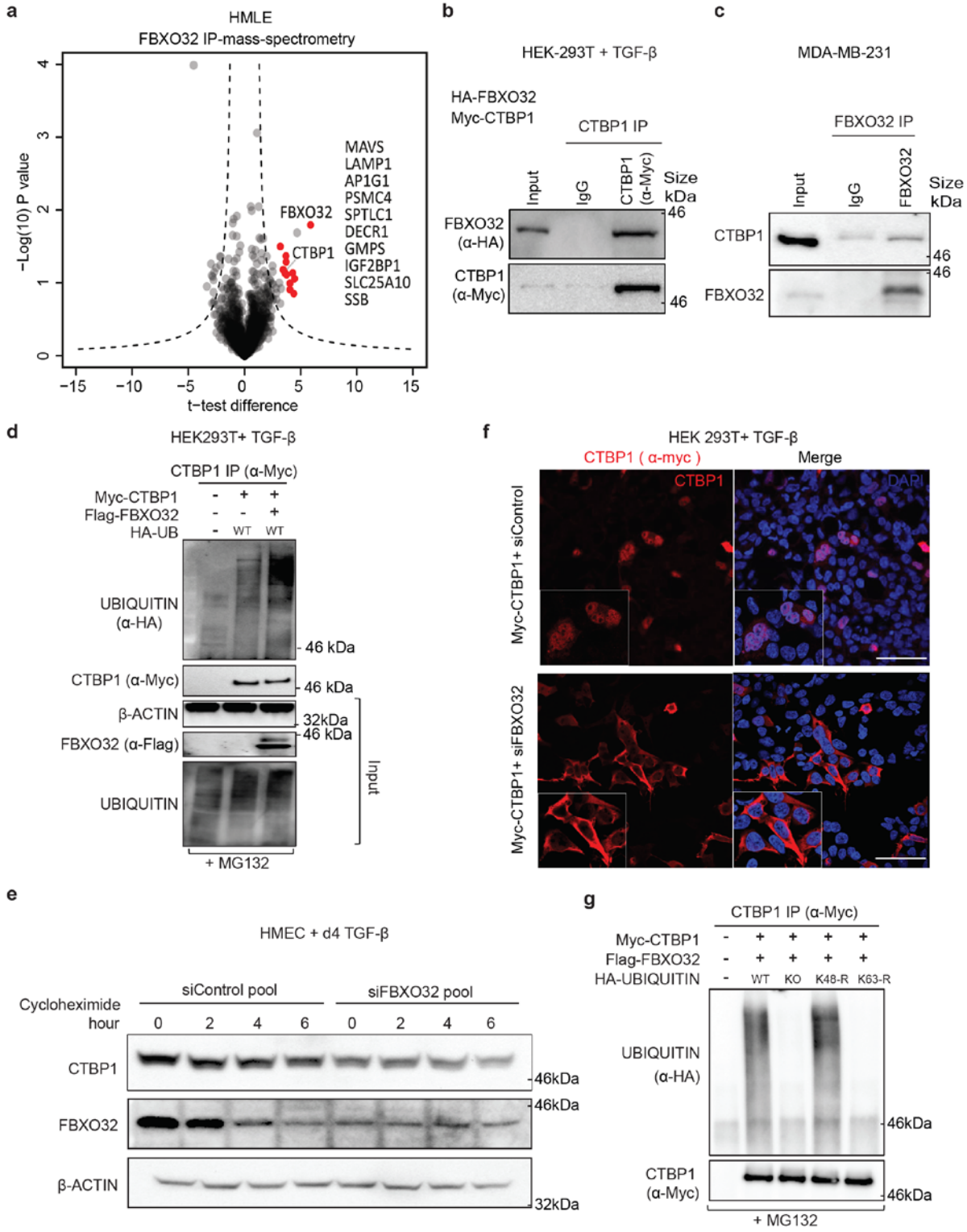


Fig 3. FBXO32 mediated K63-linked ubiquitination of CtBP1 is required for its stability and nuclear retention.

(a) Analysis of FBXO32 interaction partners by affinity purification-mass spectrometry. Volcano plot of three replicate showing selected significantly enriched proteins in FBXO32 immunoprecipitated samples.

HMLE cells were transfected with Flag-tagged FBXO32 or an empty vector and were induced with TGF-β for 4 days. FBXO32 and interaction partners were immunoprecipitated using Flag affinity resin, followed by their identification via LC-MS/MS. **(b)** Western blot of immunoprecipitated samples to validate FBXO32 and CtBP1 interactions. HEK293T cells were co-transfected with Myc-CtBP1 and Flag-HA-FBXO32 and anti-CtBP1 IP was performed. Immunoblot was performed to detect FBXO32 and CtBP1. **(c)** Western blot of immunoprecipitated samples to validate FBXO32 and CtBP1 endogenous interactions in MDA-MB-231 cell. FBXO32 IP was performed and Immunoblot was performed to detect CtBP1 and FBXO32. **(d)** HEK293 were transfected with various tagged construct as mentioned in the figure. Four hours before immunoprecipitation MG132 was added for all conditions. Western blot for CtBP1 (α-Myc) immunoprecipitated samples was performed to detect ubiquitin (α-HA) and CTBP1. β-Actin, FBXO32 and Ubiquitin were detected in Input samples as controls. **(e)** Western blot to access protein stability, showing CtBP1 and FBXO32 protein levels at various time points after cyclohexamide treatment in HMEC cells transfected with control siRNA or siRNA against FBXO32. β-Actin acted as a loading control. **(f)** Immunofluorescence image showing CtBP1 localization in control and FBXO32-depleted HEK293T cells. Magnified images of the nuclei are provided in the lower left corner of each image. Scale bar, 100 µm. **(g)** A similar analysis as in (d), but co-transfected with modified ubiquitin to demonstrate FBXO32-mediated K63-linked ubiquitination of CtBP1. Here, K63-R represents ubiquitin with lysine at position 63 modified to arginine. Similarly, in K48-R, lysine at position 63 modified to arginine. (Results in this figure contributed by me)
See also Supplementary figures 4-6.

Nuclear CtBP1 mediates epigenetic remodeling of EMT genes

CtBP1 can repress or activate transcription by recruiting distinct classes of epigenetic regulators [283]. To substantiate our findings regarding a FBXO32-dependent role of CtBP1 in promoting EMT, we depleted CtBP1 during EMT in human primary breast epithelial cells. Such loss of CtBP1 led to an impaired EMT similar to that observed following FBXO32 depletion (Supplementary Fig. 6a). Encouraged by these findings, we performed genome-wide gene expression profiling (RNA-seq) following CtBP1 ablation in HMECs undergoing TGF-β-induced EMT. Similar to FBXO32 knockdown, CtBP1 depletion led to a larger number of downregulated genes compared to upregulated genes (Fig 4a). Importantly, a comparative analysis of the genes downregulated upon CtBP1 depletion showed a highly significant overlap with the genes downregulated upon FBXO32 depletion (Fig 4b). The overlapping genes included MMPs and Cytokines, which have established functions in signaling, cytoskeletal rearrangement and cell migration [211,284-286] (Fig 4b, Supplementary Fig. 6b). Depletion of CtBP1 in FBXO32 overexpression background resulted in a reduced induction of its target genes (Supplementary Fig. 6c). These findings validated FBXO32 as an upstream regulator of CtBP1 during EMT.

Given our observations that FBXO32 mediated ubiquitination of CtBP1 is required for its stability and nuclear retention, we tested whether this is linked to CtBP1 chromatin binding and epigenetic regulatory function at its target genes. In line with our observations, we found that CtBP1 depletion impairs EMT (Supplementary Fig. 6a) and reduces expression of genes which were downregulated following FBXO32 knockdown (Fig 4b-c). Importantly, loss of FBXO32 led to a significant reduction in CtBP1 occupancy at promoters of key EMT relevant genes (Fig 4d). Since CtBP1 was previously reported to mediate gene regulation via epigenetic mechanisms [258,261], we explored whether CtBP1 depletion influences chromatin state of its target loci. In line with the changes in gene expression, CtBP1 depletion led to a significant decrease in chromatin accessibility at the majority of tested target promoters (Fig 4e). Further analysis of active histone modifications H3K4me3 and H3K27ac at these CtBP1 target promoters revealed a significant decrease in their enrichment upon FBXO32 depletion (Fig 4f and Supplementary Fig. 6d). Moreover, this also accompanied an increase in the repressive mark H3K9me3 at these sites (Fig 4g). Interestingly, CtBP1 knockdown also lead to similar epigenetic changes at the FBXO32 promoter resulting in its transcriptional downregulation (Fig 4c-g). This indicates a possible feed-forward loop between FBXO32 and CtBP1 that promotes EMT. All together these findings led us to conclude that FBXO32 mediates transcriptional activation of key EMT genes, via CtBP1 targeting to their promoters, and subsequent epigenetic remodeling that generates a transcriptionally permissive state.

We investigated whether the genes regulated by FBXO32 and CtBP1 are enriched for specific transcription factor binding motifs. For this, we performed motif enrichment analysis at the promoter regions of the genes that were deregulated following the depletion of FBXO32 or CtBP1. This investigation at the promoter of downregulated genes upon FBXO32 depletion identified NFKB, activating transcription factor 3 (ATF3) and BATF as the top motifs, whereas similar analysis for downregulated genes upon CtBP1 knockdown identified E2F6, BATF and ATF3 as the top motifs (Fig 4h-i). Since ATF3 was enriched in both FBXO32 and CtBP1 targets, we attempted to further validate these motif predictions using publically

available genome-wide binding (ChIP-seq) datasets for ATF3 in human cells. Remarkably, majority of the FBXO32 and CtBP1 regulated gene loci showed experimental evidence of ATF3 occupancy (Fig 4j and Supplementary Fig. 7a). Further analysis showed that while ATF3 pre-occupied these target sites in epithelial cells, the stability of its binding during EMT strongly relied on FBXO32 and CtBP1 (Fig 4k-l). Moreover, ATF3 depletion during EMT led to a downregulation of these genes (Supplementary Fig. 7b). These findings imply that ATF3 may cooperate with CtBP1 in gene regulation at distinct loci during EMT. Altogether, these range of findings conclude that the FBXO32-dependent nuclear localization of CtBP1 is essential for CtBP1 binding to its targets, and the epigenetic remodeling required for the gene expression changes, underlying EMT.

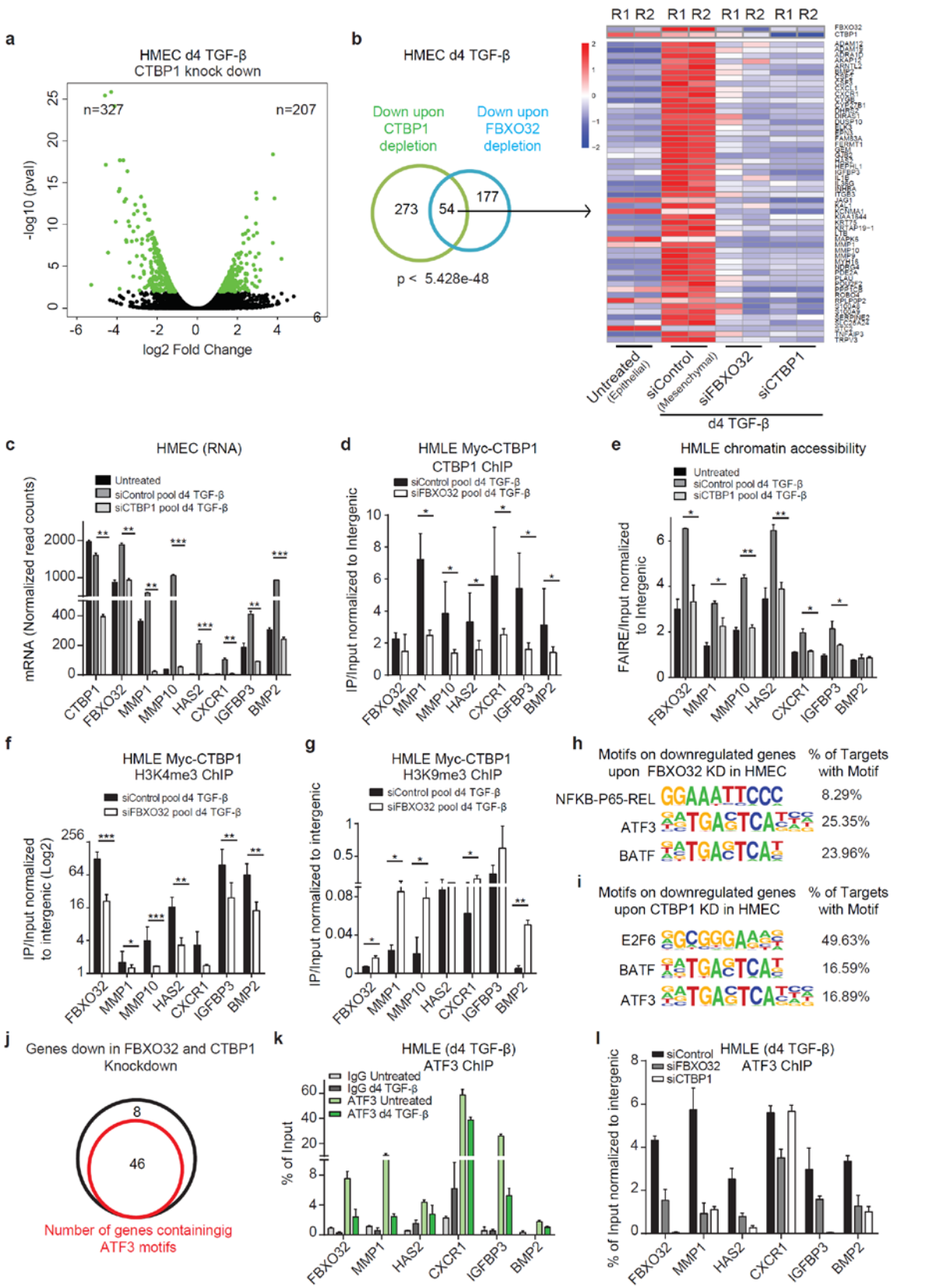


Fig 4. FBXO32-dependent nuclear localization of CtBP1 is essential for epigenetic and transcriptional remodeling of critical EMT genes.

(a) Differential expression analysis showing significantly differentially expressed genes upon CtBP1 depletion in HMECs undergoing EMT for 4 days. (b) A venn diagram showing the overlap of downregulated genes upon CtBP1 and FBXO32 depletion in HMECs undergoing EMT for 4 days and on the right hand, heat map showing their expression along with epithelial HMEC. (c) RNA-seq data showing the mRNA levels of key EMT-associated deregulated genes upon CtBP1 depletion. The y-axis represents the normalized tag count. (d) ChIP assay using an anti-Myc antibody to assess the binding of Myc-CtBP1 to its target promoter in HMLE cells induced with TGF-β for 4 days and transfected with either control siRNA or siRNA against FBXO32. q-PCR was performed for the indicated gene promoters, and gene enrichment was plotted on the y-axis as the ratio of precipitated DNA (bound) to total input DNA and normalized with intergenic region. (e) FAIRE assay was performed to assess the chromatin accessibility of EMT-associated deregulated genes in HMLE cells induced with TGF-β for 4 days and transfected with control siRNA or siRNA against CtBP1. Quantitative PCR was performed for the indicated gene promoters, and gene enrichment was plotted on the y-axis as the ratio of precipitated DNA (bound) to total input DNA and normalized with intergenic region. (f-g) A similar analysis as in (d) but ChIP performed for the H3K4me3 mark (f) and H3K9me3 mark (g). (h-i) Enriched motifs at the promoters of downregulated genes upon FBXO32 and CtBP1 knockdown (h) and (I) respectively. (j) A Venn diagram showing the presence of ATF3 motifs on or in close proximity to the promoters of genes that were downregulated upon FBXO32 and CtBP1 knockdown in HMEC. (k) A similar analysis as in (d) but with ChIP performed in HMLE cell for ATF3 and gene enrichment was plotted on the y-axis as % of total input. IgG were run as a control. (l) ATF3 ChIP were performed in HMLE cells undergoing TGF-β induced EMT and depleted for FBXO32 or CtBP1 and gene enrichment was plotted on the y-axis as % of total input and normalized to intergenic region. For all above experiments the error bars represent the SEM of three independent biological replicates (n=3). *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)
See also Supplementary figure 7.

Neuronal migration during brain development requires Fbxo32

We next asked whether FBXO32 functions similarly during developmental EMT. In the developing brain during cortical development, radial glial cells in the ventricular zone (VZ) undergo asymmetric division and daughter cells migrate toward the cortical plate (CP), passing through the sub-ventricular zone (SVZ) via an EMT-like mechanism [287] (Fig 5a). Interestingly, analysis of laser micro-dissected samples from the VZ, SVZ and CP of the E14.5 mouse cortex revealed expression changes in several EMT genes that closely mimicked classical EMT (Fig 5b). Further analysis of these datasets revealed a significant transcriptional induction of *Fbxo32* during neuronal EMT (Fig 5c). These results were independently validated using FAC-sorted CD133+ (VZ population) and NCAM+ cells (neuronal populations) (Fig 5d). In addition, the transcriptional dynamics of *Fbxo32* was confirmed by RNA *in situ* hybridization datasets in the developing mouse cortex (Fig 5e). Such *Fbxo32* upregulation during neuronal EMT *in vivo* also accompanied the induction of previously identified Fbxo32-dependent EMT-relevant genes (Supplementary Fig. 8).

To test the role of FBXO32 during cortical development, we performed *in utero* electroporation of mouse cortex [239] at E12.5 with GFP reporter plasmids containing either control shRNA or validated shRNA against Fbxo32 and the mice were sacrificed at E16.5 for further analysis (Fig 5f). While cells electroporated with control shRNA participated normally in neurogenesis and populated throughout the cortex four days post-electroporation, Fbxo32-depleted cells resided almost exclusively below the SATB2 positive layer, which marks neurons (Fig 5g-h). These observations were further confirmed by a binning analysis of the electroporated cortices (Fig 5g and i). Such retention of *Fbxo32* depleted cells in the VZ/SVZ suggested a potential defect in cellular migration that is known to involve EMT. Moreover, FBXO32 was also found to be induced during human *in vitro* neurogenesis (Fig 5J) [288]. Overall, these observations indicate FBXO32 as a critical player in neuronal EMT during brain development.

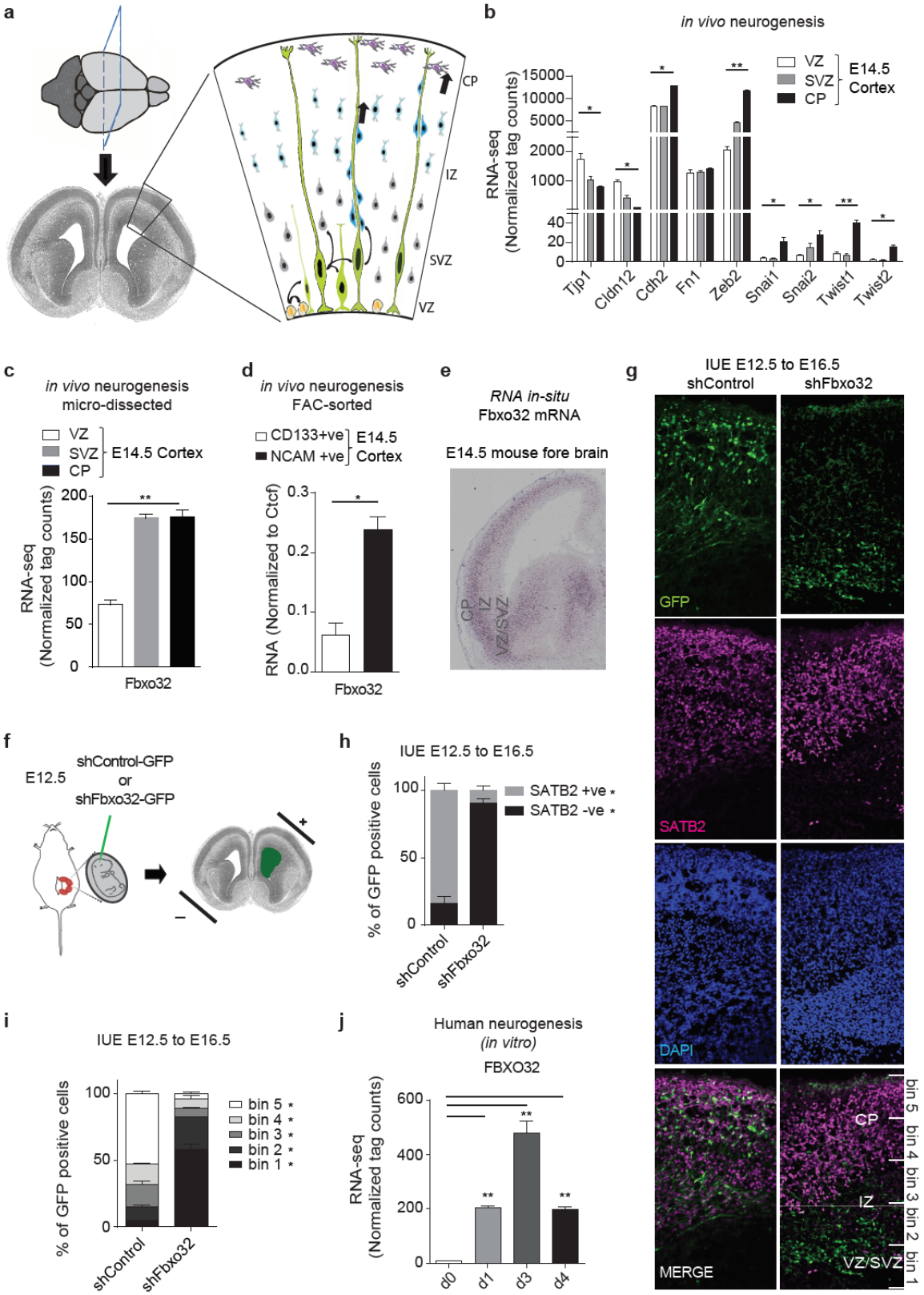


Figure 5. FBXO32 is required for neuronal EMT during brain development

(a) Schematic representation of cellular morphology changes in a developing mouse cortex (E14.5) across its various layers viz. ventricular zone (VZ), sub-ventricular zone (SVZ), intermediate zone (IZ) and cortical

plate (CP). **(b)** Normalized tag counts for key EMT markers in RNA-seq data derived from the ventricular zone (VZ), the sub-ventricular zone (SVZ), and the cortical plate (CP) of E14.5 mouse cortex. **(c)** The levels of Fbxo32 mRNA are shown as the average normalized tag counts derived from the RNA-seq data used in (b). **(d)** RT-PCR analysis of Fbxo32 in FAC-sorted cells from E14.5 mouse cortex using antibodies against CD133 and NCAM surface markers. The mRNA levels were measured relative to Ctcf via RT-PCR, and the results were plotted on the y-axis. Error bars represent the SEM of independent biological replicates (n=3). **(e)** The expression pattern of Fbxo32 mRNA as visualized by *in situ* hybridization in E14.5 cortex. The data were extracted from the publically available data set "gene paint." **(f)** Graphical representation of *in utero* electroporation (IUE). **(g)** *In utero* electroporation was performed at E12.5 using plasmids containing GFP and with control shRNA or shRNA against Fbxo32, and the mice were sacrificed at E16.5 for further analyses. A representative image (n=3) of the immunofluorescence analysis performed with anti-GFP and anti-SATB2 antibodies showing the retention of GFP-positive cells below the SATB2 layer in a brain electroporated with Fbxo32 shRNA compared to a control brain. **(h)** A bar plot showing quantification of the migrated GFP-positive cells in control and Fbxo32-depleted mouse brains with respect to SATB2 staining. The y-axis represents the percentage of cells on or below the Satb2-stained region. The error bars represent the SEM of three independent biological replicates. *p<0.05, Student's t-test. **(i)** The cortex was divided into five equal bins as shown in (g), and the number of electroporated cells was counted in each bin. The percentage of cells in each bin was plotted on the y-axis. Error bars reflect the S.E.M. of quantifications from independent brains, and significance was calculated using a t-test comparing shControl and shFbxo32 electroporated brains (* < 0.05) (n=3). **(j)** The expression of FBXO32 during human neurogenesis *in vitro* is depicted using normalized tag counts from RNA-seq data for various stages of this process in a previous study [288]. (Results in this figure contributed by me)
See also Supplementary figure 8.

Tumor aggressiveness correlates with FBXO32 expression levels

Prompted by our observations of a critical function of FBXO32 during EMT, we investigated its expression in a wide range of human tumors. A large-scale analysis of several well-characterized human clinical expression datasets revealed an exceptionally strong amplification of *FBXO32* in a majority of human tumors (Fig 6a). Computational analyses revealed significantly higher expression of *FBXO32* in various tumors compared to the matched normal tissues (Fig 6b-c and Supplementary Fig. 9a-d). In support of our observations, *FBXO32* levels correlated positively with the levels of established mesenchymal markers and negatively with the levels of epithelial markers in different types of tumor examined (Supplementary Fig. 9e-f). A survival analysis of the clinical breast tumor datasets revealed a significant correlation between higher *FBXO32* expression and poor relapse-free and metastasis-free survival (Fig 6d-e). Moreover, *FBXO32* expression showed the expected correlation with the levels of established EMT markers as well as the identified FBXO32-induced EMT genes in multiple cancers, corroborating its critical role during human tumor metastasis (Fig 6f-i). A deeper analysis of various tumor datasets (breast, colon, gastric, lung

and ovarian tumor) further validated these findings (Supplementary Fig. 9g-l, Supplementary Fig. 10a-c). In addition, transcriptome analysis of a large set of cancer cell lines confirmed an enhanced expression of *FBXO32* in mesenchymal cancer cell lines as compared to epithelial counterparts (Fig 6j). To ultimately validate these findings in clinical samples, we collected non-invasive and invasive breast tumor samples (n=20 each) and analyzed the expression of *FBXO32* in these samples. In line with our previous observations, *FBXO32* expression was significantly higher in the invasive tumors compared to the non-invasive tumors (Fig 6k). Overall, these comprehensive set of observations extend our findings to the clinic and suggest a role for *FBXO32* in promoting tumorigenicity and metastasis in humans.

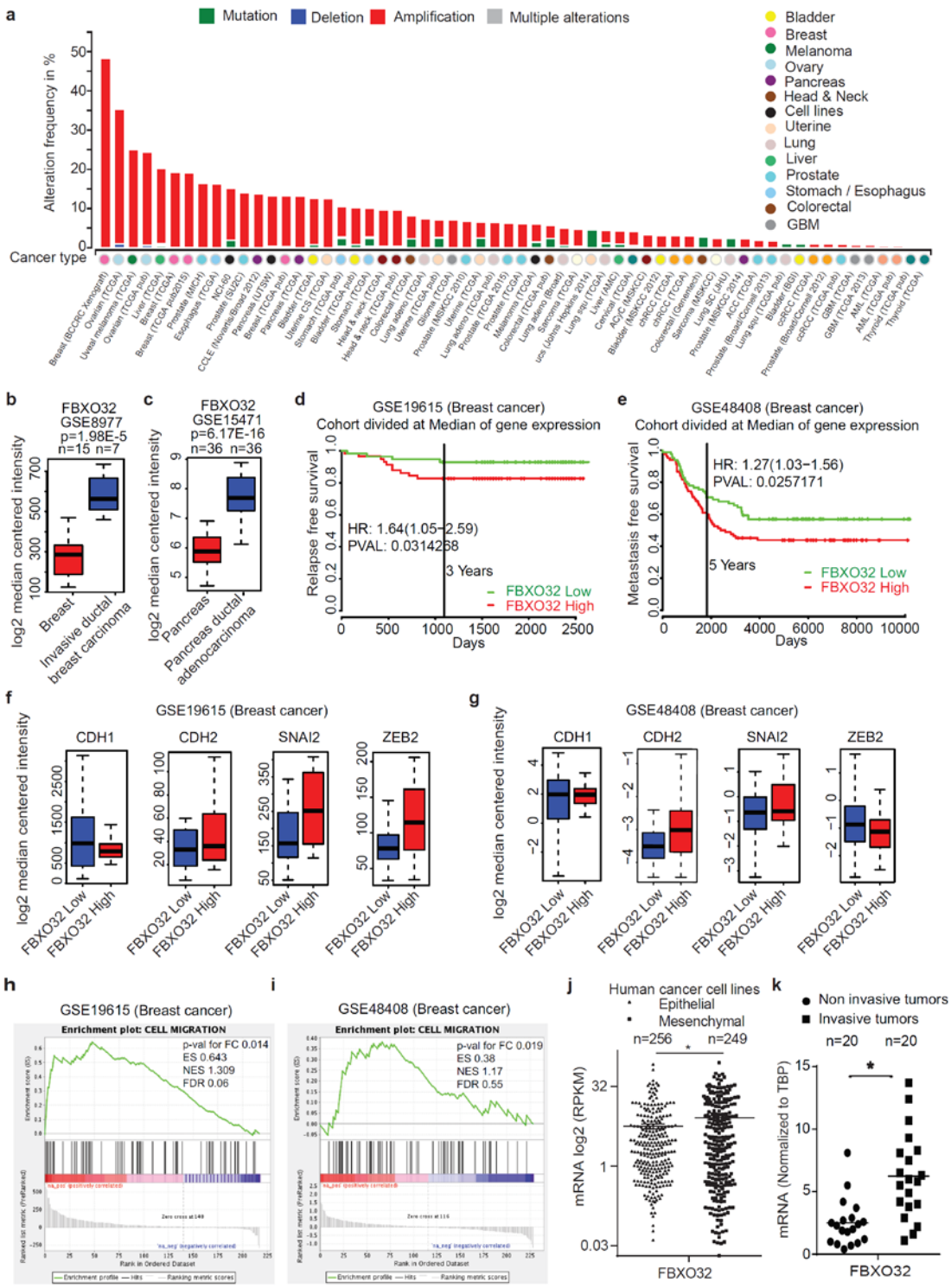


Fig 6. FBXO32 expression correlates with tumor aggressiveness.

(a) Bar plot showing the frequencies of mutations, deletions, amplifications and multiple alterations in well-characterized transcriptomes and genomic datasets of different tumors of various origins available from CBioPortal. The x-axis shows various types of tumors, and the y-axis represents the percentage of tumors with a specific defect. (b-c) Box plots showing expression level of FBXO32 in normal vs tumor samples in two well-characterized clinical tumor datasets for breast (b) and pancreas (c) origin. Box plots

represent the 25th to 75th quartiles with the bold horizontal line representing the median value. (d-e) A Kaplan-Meier analysis was performed for the Breast tumor datasets. Survival curves showing decreased relapse free survival (d) and decreased metastasis free survival (e), which was significantly correlated with higher FBXO32 expression. (f-g) Box plots showing levels of key EMT markers expression in the clinical breast tumor databases used in Fig 5d-e respectively, and split based on expression of FBXO32. For all such analyses here and later, we considered the first and forth quartile of patient samples ordered based on the median expression of FBXO32. The best probe was considered for analysis, if provided by the manufacturers. Box plots represent the 25th to 75th quartiles with the bold horizontal line representing the median value. (h-i) Gene set enrichment analysis showing correlation between genes downregulated upon FBXO32 depletion and the genes associated to the specified GO terms, as a function of fold change between expression levels of these genes in tumors with high FBXO32 and low FBXO32 expression in the characterized tumor datasets in Fig 5d-e respectively. P value for fold change, Enrichment score (ES), Normalized Enrichment Score (NES) and FDR q-value were provided in the figure (j) The levels of FBXO32 mRNA in epithelial (n=256) and mesenchymal (n=249) cancer cell lines are shown as the FPKM derived from the publically available RNA-seq data (C Klijn et.al, 2015) and the results were plotted on the y-axis. (k) The levels of FBXO32 mRNA in non-invasive (DCIS) (n=20) and invasive tumor samples (n=20) were measured relative to TBP via RT-PCR, and the results were plotted on the y-axis. *p<0.05, Student's t-test. (Results in this figure contributed by me)
See also Supplementary figures 9-10.

FBXO32 promotes tumorigenicity and metastasis in mouse model

To functionally assess the contribution of FBXO32 to the mesenchymal state in a disease context, we investigated FBXO32 expression in certain tumor cell lines that are routinely used for such studies. In line with our earlier observations (Fig 6j), FBXO32 expression was significantly higher in human mesenchymal cancer cell lines (MDA-MB-231 and BT547) compared to human epithelial cancer cell lines (MCF7 and MDA-MB-361) (Fig 7a-b). Next, we depleted FBXO32 in these cell lines and assessed their proliferation and colony-forming capacity. While the epithelial cell lines did not exhibit any obvious effects, the mesenchymal cells showed significantly reduced proliferation (Supplementary Fig. 11a). Moreover, depletion of FBXO32 in the mesenchymal breast cancer cell line (MDA-MD-231) led to a significant reduction in their migration and invasion potential (Fig 7c-d). Encouraged by these observations, we performed global gene expression profiling (RNA-seq) in FBXO32-depleted MDA-MB-231 cells. Similar to our previous findings during EMT in HMEC, FBXO32 depletion led to a significant downregulation of 704 genes and an upregulation of only 280 genes (Fig 7e). The downregulated genes included many migration-relevant genes, such as cytokines (e.g. IL6/8/18, CXCL2/3) (Supplementary Fig. 11b).

MDA-MB-231 cells have been reported to form tumors and metastasize to distant organs in mouse models 47. To further investigate the role of FBXO32 in primary tumor growth and metastasis formation, we generated MDA-MB-231 cells (containing GFP and a luciferase reporter) with stable integration of a control shRNA or validated shRNA against FBXO32. Our analysis showed no defects in proliferation or signs of apoptosis in these cells stably depleted of FBXO32 (Supplementary Fig. 11c-e). Similarly, in vitro sphere formation assay did not reveal any obvious alterations in cell proliferation and colony forming ability (Supplementary Fig. 11f-g). These cells (pool of cells with stable integration) were xenografted into the mammary fat pad of NOD SCID gamma (NSG) mice (n=6), and tumor growth and metastasis to distant organs were quantified at regular intervals. The primary tumor growth was severely reduced in the absence of FBXO32 over time (Fig 7f-g). Importantly, the FBXO32-depleted tumor cells were unable to migrate to distant organs, as measured by in vivo luciferase activity at day 50 (Fig 7h-k). These data were verified by analyzing the parental tumor cells using GFP staining in the lungs, which is known to be the preferred destination of these tumor cells following metastasis (Fig 7l). Gene expression analysis of primary tumors depleted of FBXO32 also validated downregulation of previously identified FBXO32-dependent genes (Fig 7m). Hematoxylin and eosin (H&E) staining on these tumors showed a more compact organization of cells in FBXO32 depleted primary tumors as compared to the control tumors, which may be linked to the reduced metastatic behavior of FBXO32-depleted tumor cells (Fig 7n). We also observed a clear decrease in the levels of secretory cytokines such as IL8, NGFR and GM-CSF in FBXO32-depleted tumors which may be associated with their compromised migratory potetial (Fig 7o). Furhermore, MDA-MB-231 cells in culture as well as primary tumors from mouse xenograft experiments showed strongly reduced CtBP1 levels in the nuclear compartment following FBXO32 depletion (Supplementary Fig. 11h-i). Alltogether these in vivo observations establish a critical role for FBXO32 in promoting tumorigenicity and metastasis.

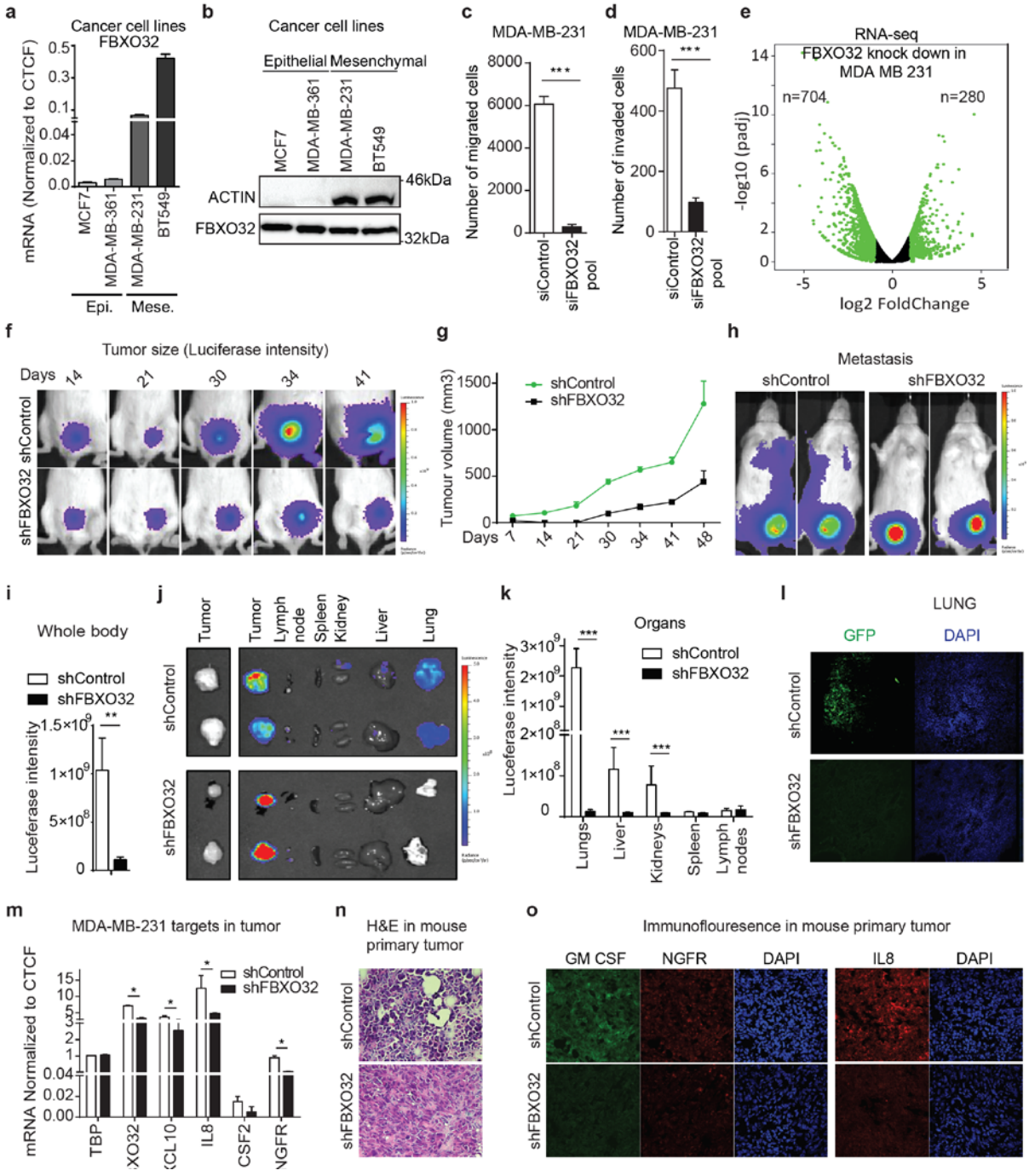


Fig 7. FBXO32 is crucial for the maintenance of mesenchymal identity and promotes tumorigenicity and metastasis in vivo.

(a) The levels of FBXO32 mRNA in two human epithelial cell lines (MCF7 and MDA-MB-361) and two mesenchymal cell lines (MDA-MB-231 and BT549) were measured relative to CTCF using RT-qPCR. (b) Western blot analysis of FBXO32 in cell lines described in (a). β-ACTIN acted as a loading control. (c-d) Migration (c) and invasion (d) assays in MDA-MB-231 cells transfected with either control siRNA or siRNA against FBXO32 for 4 days. (e) Volcano plots showing significantly differentially expressed genes upon FBXO32 depletion in the MDA-MB-231 cell line. (f-g) Representative luciferase images of tumor-bearing mice at various time points in the mice (f) and line graph (g) showing quantification of tumor size growth

in mice (n=6 per condition) analyzed at various time points after mammary fat pad injection of MDA-MB-231 cells stably expressing either control shRNA or shRNA against FBXO32 alongwith GFP and a luciferase reporter. The x-axis represents days post-injection, and the y-axis represents the size of tumors in mm3. (h-i) Representative whole-body luciferase scanning (h) along with quantification (i) showing primary and metastatic tumors (from f-g) (n=6 per condition). The primary tumor in the control mouse gets bigger and become necrotic leading to a decrease in the fluorescence intensity at the site of transplant. (j-k) Similar analysis as in (h-i) showing luciferase scanning image (j), along with quantifications (k) of various organs from those mice to detect metastatic tumors. (l) Lung sections were collected from the mice in the above experiment after sacrifice, and immunofluorescence images were obtained using an anti-GFP antibody to detect migrated cells. (m) mRNA levels of key genes deregulated upon FBXO32 depletion was validated in primary tumors obtained from the mouse used in Fig 6i were measured relative to CTCF. (n-o) H&E (n) and immunofluorescence images (o) for key cytokines were measured in primary tumor obtained from the mouse used in (h). Scale bar, 100 µm. Error bars represent the SEM of three independent biological replicates (n=3). ***p<0.001, Student's t-test. (Results in this figure contributed by me)
See also Supplementary figure 11.

DISCUSSION

The differentiation of epithelial cells into motile mesenchymal cells, a process known as EMT, is integral in development and wound healing and it contributes pathologically to fibrosis and cancer progression. EMT involves action of transcription factors such as SNAIL, ZEB and TWIST, the functions of which are finely regulated at the transcriptional, translational and post-translational levels. These key transcription factors utilize various cofactors, such as CtBPs, to recruit chromatin-modifying complexes to modulate the expression of genes [258,261,277,278]. In tissues, EMT is regulated via signaling by soluble growth factors (e.g. TGF-β and cytokines), composition and structure of extracellular matrix (ECM), and ECM remodeling enzymes (e.g. MMPs) [211,289]. Recent discoveries have shed light on the importance of post-translational modifications, such as ubiquitination, during EMT [290,291]. The K48 linked poly-ubiquitination of target proteins leads to proteasome-mediated degradation [292], whereas K63-linked ubiquitination has been implicated in the activation of several signaling pathways [293]. However, little is known about how the ubiquitin pathway contributes to the epigenetic reprogramming that drives the gene expression program underlying EMT.

A genome-wide transcriptome analysis, in a previously characterized and established cellular model system of EMT using primary mammary epithelial cells, identified strong transcriptional induction of F-box protein, *FBXO32* [197,257]. FBXO32 generally acts as a substrate-recognition subunit of the SKP1–Cullin1–F-box protein E3 ligase complex for ubiquitination and was shown to have a role in muscle homeostasis [269,294]. We found that *FBXO32* is directly modulated by TGF-β SMAD axis of EMT progression (Supplementary Fig. 2d) [276] and robustly induced during various contexts of EMT, irrespective of species and cell type. We found that FBXO32 depletion results in a phenotypic blockage of EMT in various established human and mouse cellular models, which was accompanied by a failure to gain expression of core EMT driver TFs, such as SNAIL, ZEB and TWIST, suggesting a critical requirement for FBXO32 in mediating the EMT process. Extending these findings to cancer metastasis as an example of disease EMT,

we found that FBXO32 was strongly amplified in metastatic cancers, and its depletion in a NSG mouse xenograft model significantly inhibited tumor growth and metastasis. During gastrulation, pluripotent epithelial epiblast cells ingress to form the primary mesoderm through EMT [198]. Similar phenomena have been observed during later embryogenesis and organ development, such as cortical neurogenesis, where the neural stem cells from the ventricular zone undergo changes in polarity and ultimately migrate towards the cortex [295]. Our findings of an impaired neuronal EMT during brain development following FBXO32 depletion *in vivo* imply that FBXO32 is a critical player and member of the core EMT machinery that mediates EMT in both developmental and disease contexts. These observations are also in line with the concept that the transition of epithelial cells into mesenchymal cells, both in development and under pathological conditions, follows a conserved program [190].

FBXO32 contains a class II PDZ domain that interacts with specific sequences at the C-terminal of its target proteins, and it has two nuclear localization signals [296]. PZD domain helps in localizing cellular elements, and regulating cellular pathways. These aspects of FBXO32 suggest that it may target transcription factors or other proteins in the nucleus for ubiquitination. Our global proteomics approach identified many FBXO32 interactors and substrates, among which CtBP1 was very strongly enriched. CtBP1 lacks nuclear localization signal and it is assumed that its hetero-dimerization with other proteins (e.g. CTBP2, KLF3) allows its translocation to the nucleus [259]. While a previous study indicated that the Pak1-mediated phosphorylation of CtBP1 causes its cytoplasmic retention [297], nothing is known about what results in its nuclear retention. Our comprehensive investigation revealed that FBXO32-dependent K63-mediated ubiquitination of CtBP1 is essential for its nuclear retention. We also find that a deletion within FBXO32 encompassing the core F-box domain results in a loss of its interaction with CtBP1. It is possible that the amino acids that were deleted in addition to the core F-box play an essential role in the interaction of FBXO32 with its substrate CtBP1. Such deletion may also affect binding of other factors directly or indirectly which in turn could cause a loss of FBXO32 interaction with CtBP1. Future work should involve

a more fine mapping of amino acid residues that are involved in FBXO32 function including its substrate recognition. It is known that CtBP1 functions in cooperation with other chromatin remodelers such as KDM1A and transcription factors including SNAIL, ZEB [258]. Our experimental evidences revealed that FBXO32-dependent nuclear localization of CtBP1 is essential for its binding at the promoters of key EMT genes and subsequent epigenetic remodeling including histone modifications and chromatin accessibility, thereby generating the gene expression program that drives EMT.

Among the most interesting observations following depletion of FBXO32 during EMT was the finding that the expression pattern of not only the established hallmark EMT genes but also a majority of EMT relevant genes was closely retained in an epithelial-like state. Interestingly, these data also revealed that MMPs (e.g. *MMP1*, *MMP9*, and *MMP10*), which degrade and modify the ECM and play a crucial role in EMT, require FBXO32 for their transcriptional induction during EMT. MMP1 cleaves and activates protease-activated receptor-1 (PAR-1), leading to increased migration and invasion of breast cancer cells [211]. Similarly, MMP9 mediated proteolytic activation of latent TGF-β and interleukin-8, results in a positive feed-forward loop for EMT progression [211]. An inflammatory tumor microenvironment is known to play a pivotal role in the development of cancer [298]. Notably, inflammatory cytokines such as *CXCL1/2/3/5*, *IL-1A/1B/32/6*, and many of their receptors, such as *IL1R1, CXCR1* and *TGFBR1*, were also strongly induced by FBXO32. Altogether, these evidences emphasize the essential function of FBXO32 in establishing the microenvironment required to initiate EMT.

Further computational and experimental analysis also discovered that the genes activated by FBXO32 via CtBP1 are also targeted by ATF3. *ATF3* is induced by a variety of signals, including TGF-β, and is known to play an integral role in upregulating TGF-β target genes, including *SNAIL*, *SLUG* and *TWIST*, to enhance cell motility [299]. These data suggest that other branches of signaling induced by TGF-β during EMT may converge and/or cooperate with FBXO32-CtBP1 axis to modulate expression of key EMT genes. It will be

interesting to further dissect this network and unravel its relevance in fine-tuning the gene expression program underlying EMT.

We further discovered that FBXO32 not only promotes EMT but also plays a critical role in the maintenance of mesenchymal identity by regulating a distinct set of genes in each case. Since EMT plays a critical role in tumorigenicity and metastasis, we interrogated a large panel of human cancers for *FBXO32* expression and revealed a very strong induction of *FBXO32* in various tumor types compared to the matched normal tissues. Moreover, an elevated *FBXO32* expression significantly correlated with poor relapse-free, overall, and metastasis-free survival in patients with tumors of various origins including breast, colon, lung, gastric and ovary. Moreover, *FBXO32* levels correlated positively with the levels of established mesenchymal markers and identified FBXO32-induced EMT genes, while it negatively correlated with the levels of epithelial markers in all studied tumor types. Of note, within tumors, *FBXO32* was more highly expressed in invasive tumors compared to non-invasive tumors. These findings are in full agreement with our observations of significantly reduced metastatic potential of breast cancer cells *in vitro* and *in vivo* following FBXO32 depletion. These range of findings and validations, in several model systems and primary cells from human and mouse, establish a critical role of FBXO32 in promoting tumorigenicity and metastasis in cancers.

TGF-β signaling pathways exert tumor suppressor effects in normal cells and early carcinomas. As tumors develop and progress, these protective and cytostatic effects of TGF-β are often lost due to cellular dynamics which can selectively shut down the tumor suppressive axis and TGF-β signaling then switches to promote cancer progression, invasion, and tumor metastasis [300,301]. This could also explain why few previous studies found FBXO32 to function as a tumor suppressor in gastric and urothelial carcinoma downstream of TGF-β signaling[275,276] while our own observations made in multiple model systems including primary EMT, metastatic cancer cell lines and clinical samples conclude that FBXO32 functions as a promoter of tumorigenicity and metastasis. Interestingly, various components of TGF-β signaling

pathways are known to be primarily mutated in distinct tumor types such as colon, ovarian, and uterian carcinomas, but strongly amplified in breast, lung, pancreas, esophagus and stomach carcinomas [301]. Our analysis showed a massive amplification of *FBXO32* in several tumor types which was very strong for most breast tumor datasets.

Altogether, our findings provide the first report on FBXO32 as a highly potent and new regulator of EMT in both development (e.g. neurogenesis) and disease (e.g. cancer metastasis) contexts of EMT. We show that FBXO32 plays a critical role in the progression and maintenance of mesenchymal identity by regulating the expression of key mesenchymal genes including MMPs, chemokines, chemokine receptors and the extracellular matrix components. The FBXO32-mediated ubiquitination of CtBP1 promotes its nuclear localization, which is essential for its chromatin remodeling function at the target key EMT genes to mediate the gene expression program that drives EMT (Fig 8). Together, these findings establish the function of FBXO32 as a critical regulator of EMT by governing the gene expression program underlying this process during development and disease.

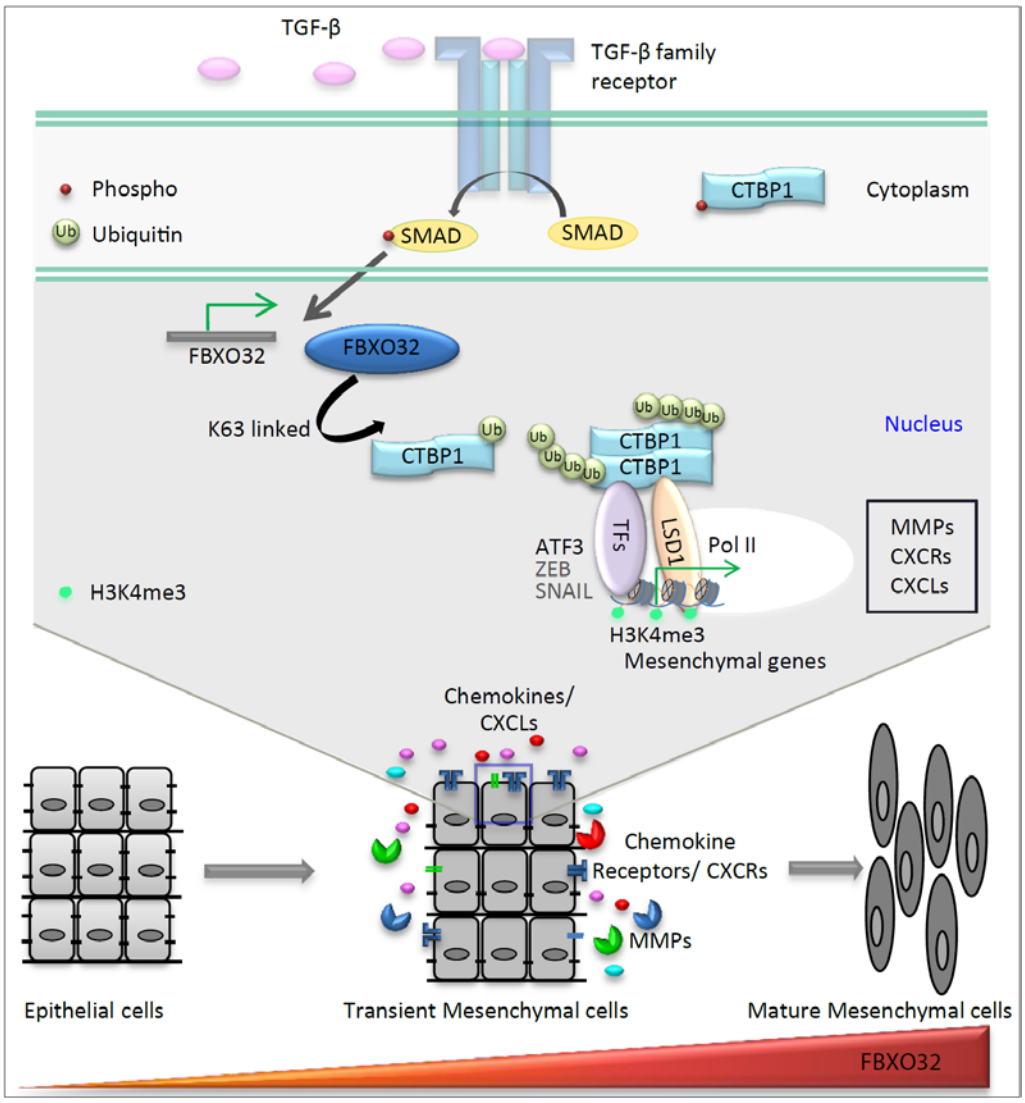


Fig 8. FBXO32 acts as a master regulator of EMT by governing the gene expression program underlying this process.

The model showing roles of FBXO32 during EMT. The onset of EMT crucially relies on SMAD-mediated transcriptional changes, and FBXO32 is one of the ubiquitin ligases upregulated during this process. FBXO32-mediated-K63 ubiquitination of CtBP1 leads to the nuclear retention of CtBP1 and so it can be recruited by various transcription factors like ATF3 to assist in chromatin remodeling. This leads to alterations in the transcription of key EMT and migration promoting genes, such as Matrix metalloproteinases (MMPs), chemokines and chemokine receptors. Such function of FBXO32 is essential for EMT in both development (e.g. neurogenesis) and disease (e.g. cancer metastasis) contexts, establishing this as a component of the EMT core machinery. Moreover, FBXO32 mediated induction of cytokine secretion acts as a feed forward loop for the EMT and further enhances the epithelial to mesenchymal transition process.

METHODS

Reagents

The reagents used in the present study were TGF-β (rhTGF-β1 240-B, R&D Systems), DMEM (21969-035, Invitrogen), RPMI-1640 (R0883, Sigma), PBS (D8537, Sigma), trypsin (25300-054, Invitrogen), Opti-MEM (31985-047, Invitrogen), FBS (10270, Invitrogen), glutamine (25030-024, Invitrogen), MEM NEAA (100x) (11140-035, Invitrogen), Lipofectamine 2000 (11668, Invitrogen), Lipofectamine RNAiMAX (13778-150, Invitrogen), Trizol (15596026, Invitrogen), reverse transcriptase kits (K1612, Thermo Fischer), SYBR-Green PCR MasterMix (4334973, Invitrogen), Bradford reagent (5000205, BioRad), a protease inhibitor cocktail (04693132001, Roche), and a phosphatase inhibitor cocktail (04906837001, Roche).

Inhibitors used in the study

For chemical inhibition experiments, we used JNK inhibitors, SP600125, (10 μM) (S5567, Sigma), ERK-signaling inhibitors UO126 (25μM) (Milipore) and p38i SCIO-469, (5 nM), (3528, r&d systems). Ubiquitination pathway inhibitor PYR-41 or 4[4-(5-Nitro-furan-2-ylmethylene)-3,5-dioxo-pyrazolidin-1-yl]-benzoic acid ethyl ester (N2915, Sigma) was added to cell culture medium to a final concentration of 10 μM, 4 hour prior to collection of cells.

Cells were seeded at required densities and pretreated 30 min prior to TGF-β treatment with signaling pathway inhibitors or their solvent control DMSO at the final concentrations, mentioned above and analyzed at indicated time points. For SMAD pathway blocking we have used on target plus smart pool siRNA (Dharmacon) targeting Smad4. This was followed by harvesting the cells 24 hours later for RNA analysis.

Antibodies used for Western blot, Immunoprecipitation and Immunofluorescence assays

The antibodies used for immunofluorescence were against E-cadherin (13-1900, Invitrogen and 610182, BD Transduction Laboratories), N-cadherin (610921, BD Transduction Laboratories), ZO-1 (617300,

Invitrogen), fibronectin (F-3648, Sigma-Aldrich), paxillin (610052, BD Transduction Laboratories), GFP (chicken, Aves Labs), SATB2 (Abcam, ab34735), beta-actin (C4, sc-47778, Santa Cruz), FBXO32 (sc-33782, Santa Cruz) (ab168372), ZEB1 (sc-25388 (H-102) Santa Cruz), ATF-3 (sc-188 (C-19) Santa Cruz), Cleaved caspase 3 (Asp175, 9661S, cell signaling), FLAG tag M2 (F1804, Sigma), HA tag monoclonal antibody (c15200190, Diagenode), Myc tag (ab9106, Abcam), ubiquitin mouse monoclonal (sc8017, Santa Cruz), Lamin B (sc-6216 (C-20), Santa Cruz), Histone H3 tri methyl K4, (ab12209, Abcam), Histone H3 tri methyl K9, (ab8898, Abcam), Histone H3 acetyl K27, (ab4729, Abcam), Alexa Fluor-488 goat anti-mouse IgG (H+L) (A11029, Invitrogen); Alexa Fluor-568 goat anti-rabbit IgG (H+L) (A11011, Invitrogen), Alexa Fluor-633 goat anti-rat IgG (H+L) (A21094, Invitrogen), and Alexa Fluor-633 phalloidin (A22284, Invitrogen), which was used to stain F-actin.

Cell culture

HMECs were obtained from Lonza and cultured according to the manufacturer's guidelines. Other cell lines were obtained from ATCC and cultured in the following media: MDA-MB-231: DMEM, 10% FBS; MDA-MB-361: DMEM, 20% FBS; BT549: RPMI-1640, 10% FBS, 0.001 mg/ml bovine insulin; MCF7: DMEM, 10% FBS, 0.01 mg/ml bovine insulin; HEK293T: the same medium as NMuMG. HMLE cells were a kind gift from Christina Scheel (Helmholtz Zentrum Munich) and cultured in LONZA primary cell culture medium. A subclone of NMuMG cells was grown in DMEM supplemented with 10% FBS, 2 mM L-glutamine and 1X non-essential amino acids[176]. All cells were cultured at 37°C with 7% CO_2 in a humid incubator. For TGF-β time-course experiments, NMuMG, HMEC and HMLE cells were treated with 2, 5 and 5 ng/ml TGF-β, respectively, for the indicated times. TGF-β was replenished and the medium was changed every 2 days.

siRNA mediated knockdown

For all siRNA-mediated knockdown experiments, cells were seeded at the same starting density and transfected every second day with ON-TARGET plus single siRNA or SMART pool siRNAs (i.e., a mixture of 4 siRNAs provided as a single reagent) (Dharmacon). For siRNA transfections, Lipofectamine RNAiMAX

(Invitrogen, 13778-150) was used according to the manufacturer's instructions. For experiments during TGF-β-induced EMT, TGF-β induction was performed at the same time that siRNA was added to avoid indirect effects due to loss of protein function. The sequences of the siRNAs are provided in Supplementary Data 1.

Mammosphere culture

For primary sphere formation, single cells were plated in ultralow attachment plates (Corning) at a density of 20,000 viable cells/mL in serum-free mammary epithelial growth medium composed of DMEM supplemented with 1% L-glutamine, 1% penicillin/streptomycin, 30% F12 , 2% B27 (Invitrogen), 20 ng/mL EGF (peprotech, AF-100-15) and 20 ng/mL bFGF (peprotech, 100-18B). The medium was made semi-solid by the addition of 0.5% Methylcellulose (Sigma) to prevent cell aggregation. Mammospheres were collected by gentle centrifugation (200 g) after 10 days and dissociated enzymatically (10 min in 0.05% trypsin, 0.53 mM EDTA, Invitrogen) and mechanically, using a fire-polished Pasteur pipette. The cells obtained from dissociation were sieved through a 40-µm sieve and analyzed microscopically for single-cellularity. For the secondary sphere formation, 1000 cells/ml were plated and colonies were monitored by a microscope at a regular interval.

Immunofluorescence assay

Cells were grown on coverslips, fixed with 4% paraformaldehyde in PBS and permeabilized with 0.2% Triton X-100 for 15 minutes at room temperature. Subsequently, the cells were blocked with 10% goat serum, 5% FCS and 0.5% BSA in PBS for 20 minutes and were incubated with primary antibodies at 4°C overnight. The cells were then incubated with fluorochrome-labeled secondary antibody or Phalloidin-633 for 1 hour at room temperature. The coverslips were counterstained with Hoechst, mounted with immu-mount and imaged using a confocal laser-scanning microscope. The data were processed by ImageJ software.

Immunohistochemistry for analysis of mouse cortex

Isolated E16.5 embryonic brains were immediately fixed for 6 hours in 4% PFA in PBS at 4°C. The brains were then cryoprotected in 10% sucrose for 2 hours and then in 30% sucrose (in PBS) overnight, embedded in Tissue-Tek, stored at -20°C and cryo-sectioned at 12 µm. Sections on coverslips were preblocked with 2% BSA and 0.5% Triton (in PBS) for 1 hour. Primary antibodies (Satb2, 1:500, and anti-GFP (chicken, Aves Labs, 1:1000)) were applied in blocking solution overnight at 4°C. Fluorescent secondary antibodies were applied according to the manufacturer's protocol (Life Technologies). The coverslips were counterstained with Hoechst, mounted with immomount and imaged using a confocal laser-scanning microscope (Leica SP5). The data were processed with ImageJ software.

Quantitative Real time PCR

mRNA levels were quantified as previously described [197]. In brief, total RNA was prepared using Trizol (Invitrogen) or a SurePrep TrueTotal RNA Purification Kit (Fisher Scientific) and was reverse transcribed with a First Strand cDNA Synthesis Kit (Fermentas). The transcripts were quantified via PCR using SYBR green PCR MasterMix (ABI) on a ViiA7 PCR machine (Life Technologies). Human or mouse CTCF/Ctcf and TBP/Tbp primers were used for normalization. The sequences of all of the primers used in this study are listed in Supplementary Data 1.

Immunoblotting

Cells were lysed in RIPA buffer, and protein concentrations were quantified using Bradford reagent. Equal amounts of protein (30 µg) were boiled in 6× SDS-PAGE loading buffer, subjected to polyacrylamide gel electrophoresis, transferred to a PVDF membrane and probed with the appropriate antibodies. All uncropped western blot images can be found in Supplementary Figure 12.

Chromatin Immunoprecipitation assay

Cells were cross-linked in medium containing 1% formaldehyde for 10 minutes at room temperature, neutralized with 0.125 M glycine, scraped, and rinsed twice with PBS. The pellets were suspended in buffer L1 (50 mM Hepes KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% glycerol, 5% NP-40, and 0.25% Triton-X 100) and incubated for 10 minutes at 4°C. The cells were then suspended in buffer L2 (200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0) for 10 minutes at room temperature. Finally, the pellet was suspended in buffer L3 (1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, 100 mM NaCl, 0.1% Na-deoxycholate, 0.17 mM N-lauroyl sarcosine) containing protease inhibitors and was incubated at 4°C for 3 hours following sonication using Bioruptor plus (Diagenode). Sixty micrograms of chromatin was incubated overnight at 4°C with 2 µg of the antibodies targeting H3K4me3, H3K9me3, H3K27ac (Abcam) or Myc (Santa Cruz) and then incubated with preblocked beads for 4 hours. Finally, the beads were washed twice with L3 and once with 1 ml of DOC buffer (10 mM Tris (pH 8.0), 0.25 M LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA), and the bound chromatin was eluted in 1% SDS/0.1 M NaHCO3. This followed treatment with RNase A (0.2 mg/ml) for 30 minutes at 37°C and then with proteinase K (50 µg/ml) for 2.5 hours at 55°C. The crosslinking was reversed at 65°C overnight with gentle shaking. The DNA was purified by phenol-chloroform extraction followed by ethanol precipitation and was recovered in TE buffer.

Mass spectrometry analysis

Peptide fractions were analyzed using a quadrupole Orbitrap mass spectrometer (Q Exactive Plus, Thermo Scientific) equipped with a UHPLC system (EASY-nLC 1000, Thermo Scientific), as described [302]. Peptide samples were loaded onto C18 reversed-phase columns and eluted for 2 hours with a linear gradient of acetonitrile from 8 to 40% containing 0.1% formic acid. The mass spectrometer was operated in data-dependent mode with automatic switching between MS and MS2 acquisition. Survey full scan MS spectra (m/z 300 – 1650) were acquired in the Orbitrap. The 10 most intense ions were sequentially isolated and fragmented by higher-energy C-trap dissociation (HCD) [303]. Peptides with unassigned charge states or

charge states less than +2 were excluded from fragmentation. The fragment spectra were acquired in the Orbitrap mass analyzer.

Peptide identification: raw data files were analyzed using MaxQuant (development version 1.5.2.8) [304]. Parent ion and MS2 spectra were searched against a database containing 88,473 human protein sequences obtained from the UniProtKB released in December 2013 using the Andromeda search engine [305]. The spectra were searched with a mass tolerance of 6 ppm in MS mode, 20 ppm in HCD MS2 mode, strict trypsin specificity and allowing up to 3 miscleavages. Cysteine carbamidomethylation was searched as a fixed modification, whereas protein N-terminal acetylation and methionine oxidation were searched as variable modifications. The dataset was filtered based on posterior error probability (PEP) to arrive at a false discovery rate below 1%, estimated using a target-decoy approach [306].

***In vivo* tumorigenicity assays and imaging**

NOD *scid* gamma (NSG) mice were bred and maintained under SPF conditions in the Translational Animal Research Center at the University Medical Center Mainz. We did preliminary experiments to determine the need for mice sample size and duration of experiment. The MDA-MB-231 cells with stable expression of GFP and a luciferase reporter were used for generating stable cell lines with either control shRNA or shRNA against FBXO32. Briefly, cells were counted and resuspended in a 1:1 (v/v) mixture of PBS and Matrigel (BD Biosciences). Ten-week-old female mice were injected unilaterally with 2×10^6 cells in 100 µl of 50:50 Matrigel/PBS into the fourth abdominal fat pad via subcutaneous injection at the base of the nipple. Tumor growth was monitored externally using Vernier calipers for up to 48 days. The maximum size of the tumor that was permitted in the mice was 2700 mm^3 according to ethical approval policy of Translational Animal Research Center at the University Medical Center Mainz. The tumor volume was calculated as follows: tumor volume (mm^3) = length × $(width)^2$ × 0.5. Necropsies were performed to identify macro-metastases. Primary tumors and organs were immediately frozen in liquid nitrogen and stored at -80°C until use.

In vivo bioluminescence imaging of tumor-bearing mice and their organs was performed at day 50 using an IVIS Lumina imaging system (Perkin Elmer). Briefly, mice were anesthetized with isoflurane and an aqueous solution of D-luciferin-K^+ salt (150 mg/kg body weight) (Perkin Elmer) was injected intraperitoneally. Five minutes after the injection, the mouse was placed onto the imaging chamber of IVIS, and photons acquired with an integration time of 10 s were presented as color-scaled images using IVIS Living Image Software (version 4.3.1) (Perkin Elmer). For organ imaging, mice were sacrificed after luciferin injection, and the dissected organs were imaged as described above.

Animal maintenance and experiments were performed under an approved protocol in accordance with the animal care guidelines of Johannes Gutenberg University.

***In utero* electroporation**

In utero electroporation experiments were performed essentially as previously described[239] . All experimental procedures were conducted in accordance with European, national, and institutional guidelines for animal care. Timed-pregnant (Theiler stage 20=E12.5) C57BL/6 mice were anesthetized with isoflurane (2.5% via mask, Forene®, Abbod), and carprofen (4 mg/kg body weight, Rimadyl, Pfizer) was administered subcutaneously as an analgesic. After the abdominal cavity was opened, the embryos were carefully exposed, and 1 µl of colored solution containing 1 µg of the p.SUPER-GFP shRNA plasmid expressing shRNA against Fbxo32 or control shRNA was injected into one of the lateral ventricles. Using specialized platinum electrodes (Nepagene CUY 650P), the appropriate voltage was applied (50 ms, interval 950 ms, 5 pulses). After electroporation, the uterine horns were returned to the abdominal cavity.

Cell survival assay

Cells were pre-depleted with siRNA (a pool of 4 siRNAs targeting the same gene, SMARTpool, Dharmacon) for two days, and 40,000 cells per 6 cm were plated and cultured for 6 days. Fresh medium along with siRNA was added every other day. The cells were counted, or bright-field images were obtained, or the

cells were fixed with 4% PFA. To visualize the fixed colonies, the cells were stained for 10 minutes with a solution containing crystal violet and 10% ethanol, and the plates were washed in water until the excess dye was washed away. Photographs of the plates were obtained for later analysis.

Wounding migration assay

Cells depleted of FBXO32 were seeded at equal densities, and a scratch wound was generated using a 10-µl pipette tip on confluent cell monolayers grown in six-well culture plates. The cells were then washed with fresh medium to remove floating cells. Bright-field images were obtained at 20x magnification after 18 hours of wounding.

Migration and invasion assays

Migration assays were performed as previously described [197]. Briefly, 10,000 cells were seeded in 2% FBS/DMEM (Sigma) into the upper chamber of a 24-well Transwell migration insert (pore size 8 µm; Falcon BD). The lower chamber was filled with 20% FBS/DMEM. After 16 hours of incubation at 37°C, the cells in the upper chamber were removed using a cotton swab, and the cells that had traversed the membrane were fixed in 4% paraformaldehyde/PBS and quantified by DAPI staining using a fluorescence microscope and ImageJ software. For the invasion assays, transwells were coated with 0.5 mg/ml Matrigel in serum-free medium overnight, and 1.5×105 cells were used for the assays.

IP and ubiquitylation assays

HEK-293T cells were co-transfected in various combinations with expression vectors for HA-tagged Ubiquitin, Myc-tagged CtBP1 and Flag-tagged FBXO32 using Lipofectamine® 2000 Transfection Reagent (Thermo scientific). For inhibition of ubiquitination pathway, UBEI/PYR-41 was added at a final concentration of 10 µM for 4 hours. All experimental cells for ubiquitination assay were treated with 10 µM of MG132 for 4 hours prior to lysis. Cells were lysed in lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM EGTA, 1.5 mM MgCl2, 1% Glycerol, 1% Triton X-100) and centrifuged for one hour to collect

the soluble fraction. Tagged proteins were immunoprecipitated for overnight at 4°C with the appropriate antibody (5 µg/IP). Pre-blocked beads were added to the lysate and incubated for 3 hours at 4°C. The bound beads were washed two times with lysis buffer and two times with wash buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.1% Triton X-100) and analyzed by immunoblot analysis. For all IP experiments, immunoprecipitated material was eluted with an equal volume of elution buffer and same amount of samples were loaded on gel for immunoblot analysis.To assess ubiquitylation, 293T cells were transfected with various tagged expression vectors in different combinations (as outlined in the manuscript figures) in the presence of TGF-β for 4 days. Proteins lysate were precipitated with anti-Myc antibody to pull down CtBP1 and analyzed by immunoblot using anti-HA antibody to detect Ubiquitin.

RNA sequencing data analysis

RNA-seq data were generated using Illumina sequencing. The reads were aligned to the mouse genome (mm9) using TopHat [179] (version 2.0.9) with the default options. After library size normalization using DESeq [178], expression was quantified and expressed in reads/kilobase of transcript per million mapped reads (RPKM) using cufflink (version 2.1.1) [179]. A differential expression analysis was performed using the DESeq package with an FDR cutoff of 0.1 [178]. Gene set enrichment analysis was performed using GSEA Preranked module [307].

DATA availability

The authors declare that all data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author upon reasonable request. RNA-seq data generated during this study have been deposited in the Genbank nucleaotide database (GEO) under accession code GSE77408. The proteomic mass spectroscopy data have been submitted as an excel file (Supplementary Data 2). In addition to these datasets, we had also used a number of publically available datasets (GEO accession code GSE54133, GSE62146, GSE8977, GSE15471, GSE19615, GSE48408,

GSE19615, GSE48408, GSE30611, GSE95983, GSE32863, GSE31210, GSE6764, GSE62254, GSE31210, GSE9891, GSE28814, GSE28722).

AUTHOR CONTRIBUTIONS

S.K.S. designed and performed experiments, analyzed data and wrote the manuscript. A. P. performed computational analysis. N.T. and Y.Z. performed experiments. M.B. and P.B. assisted with mass-spectrometry experiments and analysis. S.S. and M.D. provided help with *in vivo* tumorigenicity assays and imaging. V.T. designed and supervised the study, analyzed data and wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to thank the members of the Tiwari lab for their cooperation and critical feedback throughout the course of this study. The support from the Core Facilities of the Institute of Molecular Biology (IMB), Mainz, is gratefully acknowledged, especially the microscopy, cytometry, genomics and bioinformatics core facilities. We would especially like to thank Prof. Robert Nitsch and Nikolai Schmarowski from Institute for Microscopic Anatomy and Neurobiology University Medical Center, Johannes Gutenberg University in Mainz for help with the *in utero* electroporation experiments and Prof. Marcus Schmidt from Department of Obstetrics and Gynecology, Johannes Gutenberg University in Mainz for providing clinical samples. We are also grateful to Dr. Christina Scheel (Helmholtz Zentrum München) for providing HMLE cells. This study was exclusively supported by the Wilhelm Sander Stiftung 2012.009.1 and 2012.009.2 to V.K.T. The research conducted in the laboratory of P.B. is supported by the Emmy Noether grant BE 5342/1-1 and the Marie Curie grant CIG 630763.

COMPETING INTERESTS

The authors declare no competing financial interests

Supplementary Figures

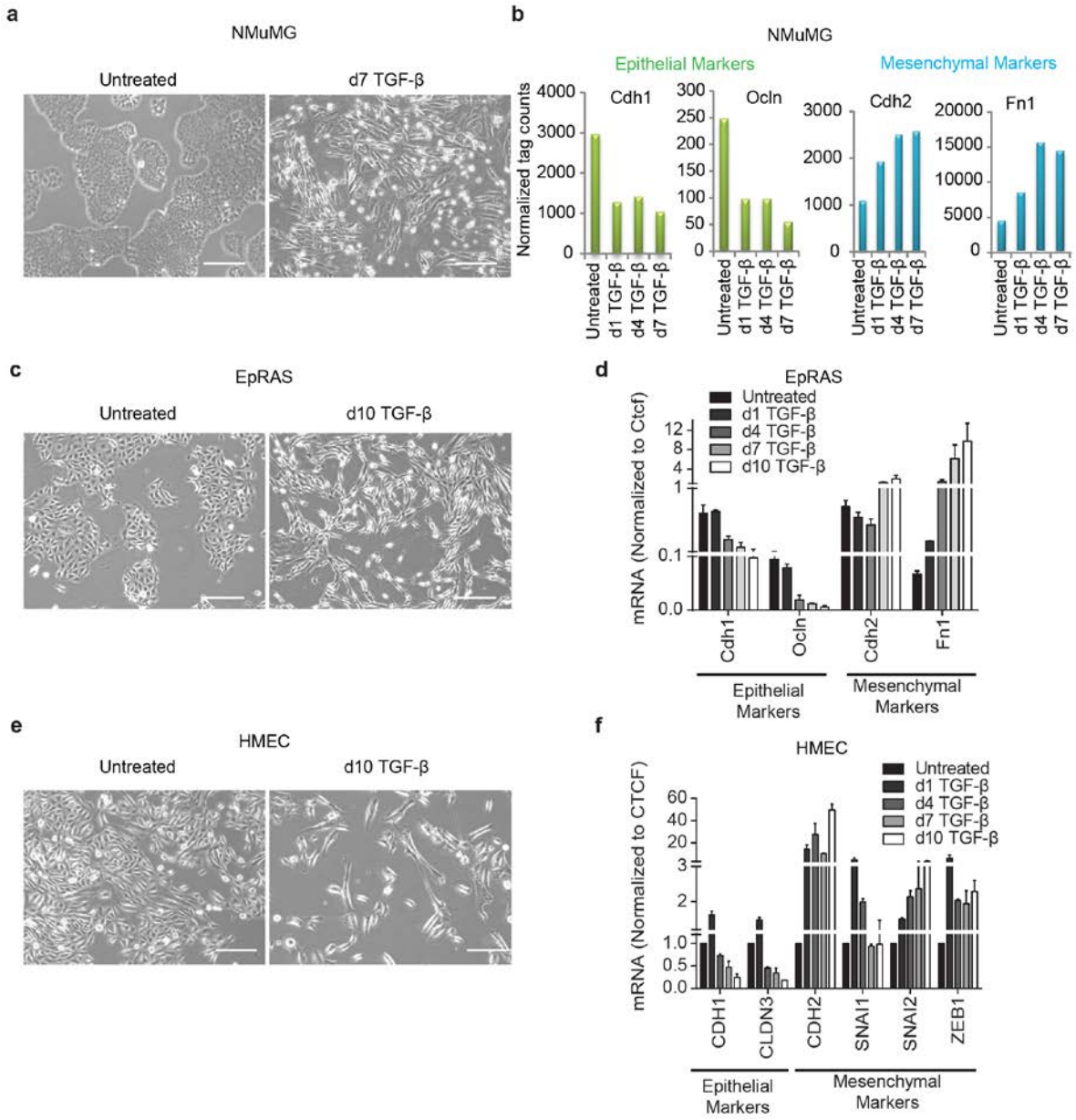


Fig S1. Model systems for EMT

(Related to Fig 1)

(a) Representative bright field images of NMuMG cells undergoing TGF-β induced EMT. (b) RNA-seq data showing mRNA level of key EMT relevant genes in NMuMG cells undergoing EMT, the y-axis represents the normalized tag count. (c) Representative bright field images of mouse EpRAS cell line undergoing TGF-β induced EMT. (d) mRNA levels of key EMT markers during EpRAS cell line undergoing TGF-β induced EMT measured by RT-PCR relative to Ctcf and plotted on y-axis. (e) Representative bright field images Human Primary Mammary Epithelial Cells (HMEC) undergoing TGF-β induced EMT. (f) mRNA levels of key EMT markers during HMEC cell undergoing TGF-β induced EMT measured by RT-PCR relative to CTCF and plotted on y-axis. (Results in this figure contributed by me)

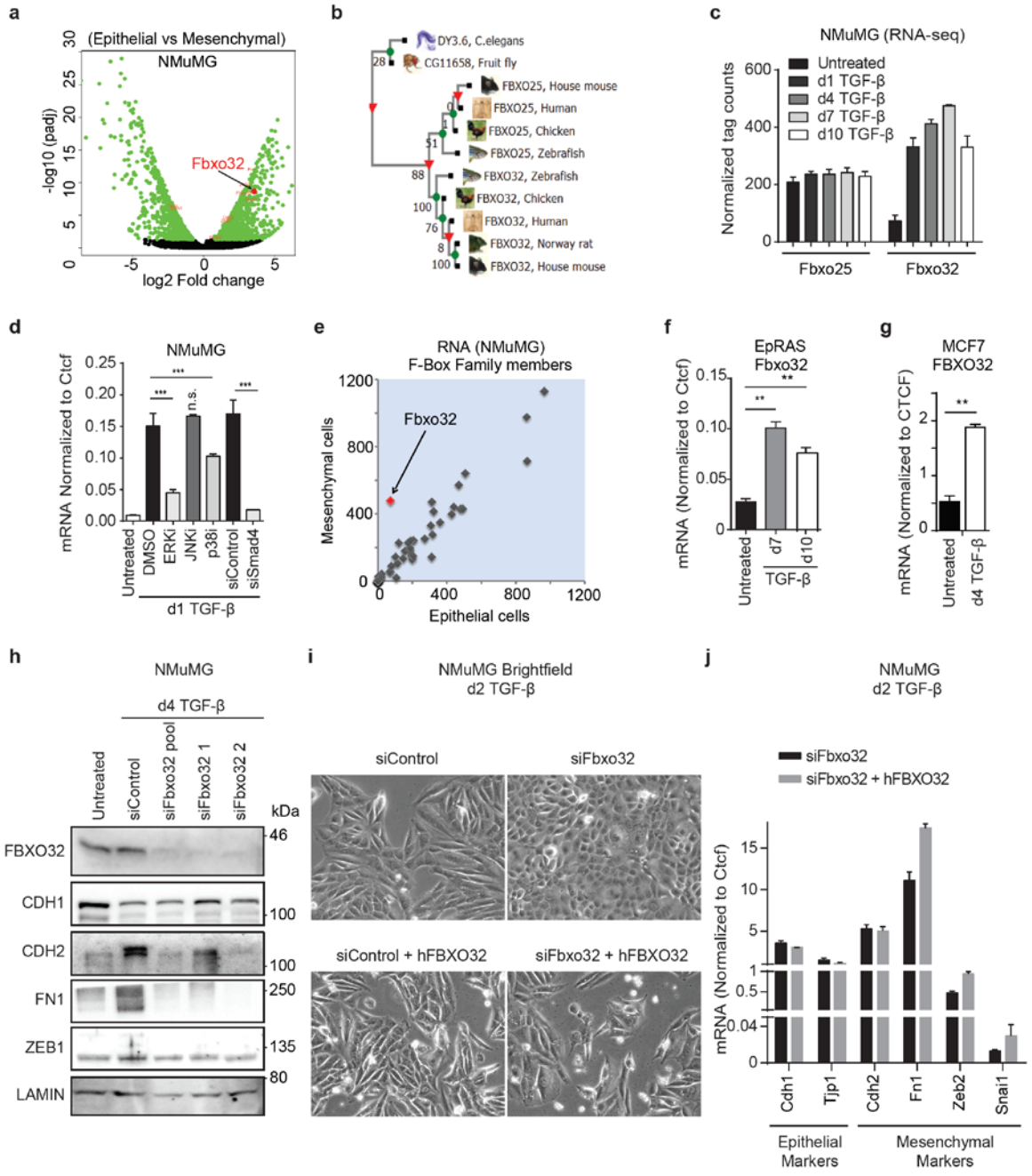


Fig S2. Fbxo32 is induced during EMT

(Related to Fig 1-2)

(a) Volcano plot showing significantly differentially expressed genes in both epithelial and mesenchymal NMuMG cells. (b) A phylogenetic tree showing FBXO32 and the closest evolutionarily conserved member of it FBXO25 (according to sequence similarity and protein domains). (c) RNA-seq data showing mRNA level of Fbxo32 and Fbxo25 during TGF-β induced EMT, y-axis represents normalized tag count. (d) Using RT PCR mRNA levels of Fbxo32 measured relative to CTCF in NMuMG cells undergoing TGF-β induced EMT and treated with inhibitors of various Map Kinases or depleted for Smad4. (e) Comparison of the differential expression, quantified as normalized tag counts, of F-box family members in epithelial and mesenchymal NMuMG cells. (f) mRNA level for Fbxo32 during TGF-β induced EMT in EpRAS were measured by RT-PCRs relative to Ctcf and plotted on y-axis. (g) mRNA level for FBXO32 during TGF-β

induced EMT in MCF7 were measured by RT-PCRs relative to CTCF and plotted on y-axis. (h) Western blot analysis of FBXO32 along with key EMT markers Cdh1, Cdh2, FN1, and Zeb1 in NMuMG cells undergoing TGF-β-induced EMT and depleted for Fbxo32, either with pool of siRNAs or single siRNA. Lamin acted as a loading control. (i) Representative bright-field images showing the morphology of NMuMG cell undergoing TGF-β-induced EMT and upon siRNA mediated depletion of Fbxo32 and rescue by overexpression of siRNA resistant human FBXO32 for two days. (j) Using RT-PCR, mRNA levels of key EMT markers in in above condition as mentioned in (i) were measured relative to Ctcf and plotted on the y-axis. Error bars represent the SEM of three independent biological replicates. *$p<0.05$, **$p<0.01$, ***$p<0.001$, Student's t-test. (Results in this figure contributed by me)

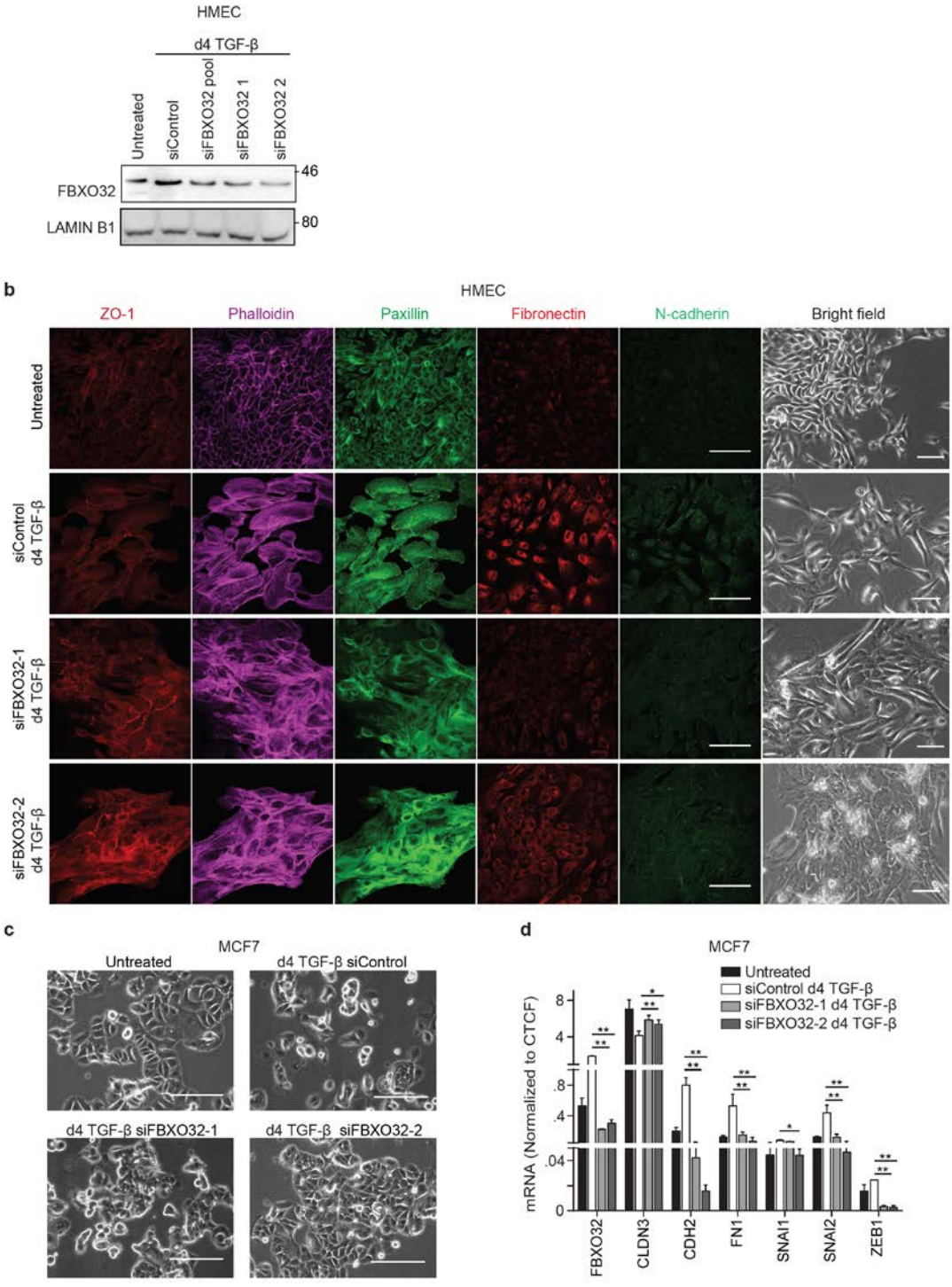


Fig S3. FBXO32 is critical for human EMT. Related to Figure 1-2

(a) Western blot analysis of FBXO32 in HMEC cells undergoing TGF-β-induced EMT and depleted for FBXO32, either with pool of siRNAs or single siRNA. Lamin acted as a loading control. (b) Representative bright-field and immunofluorescence images showing the localization and expression levels of EMT marker proteins after four days (d4) of independent siRNAs-mediated depletion of FBXO32 compared to non-targeting control (siControl) in HMEC. Staining was performed to assess the expression with antibodies against the epithelial marker ZO1, the mesenchymal marker N-cadherin, Fibronectin-1, Phalloidin (to visualize the actin cytoskeleton) and Paxillin (to detect focal adhesion plaques). Scale bar,

100 µm. (c) Representative bright-field images showing the morphology after four days (d4) of independent siRNAs-mediated depletion of FBXO32, compared to siControl in MCF7 undergoing TGF- β-induced EMT. (d) Using RT-PCR, mRNA levels of key EMT markers in MCF7 cells transfected with either control siRNA or independent siRNAs against FBXO32 during TGF-β-induced EMT were measured relative to CTCF and plotted on the y-axis. Error bars represent the SEM of three independent biological replicates. *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)

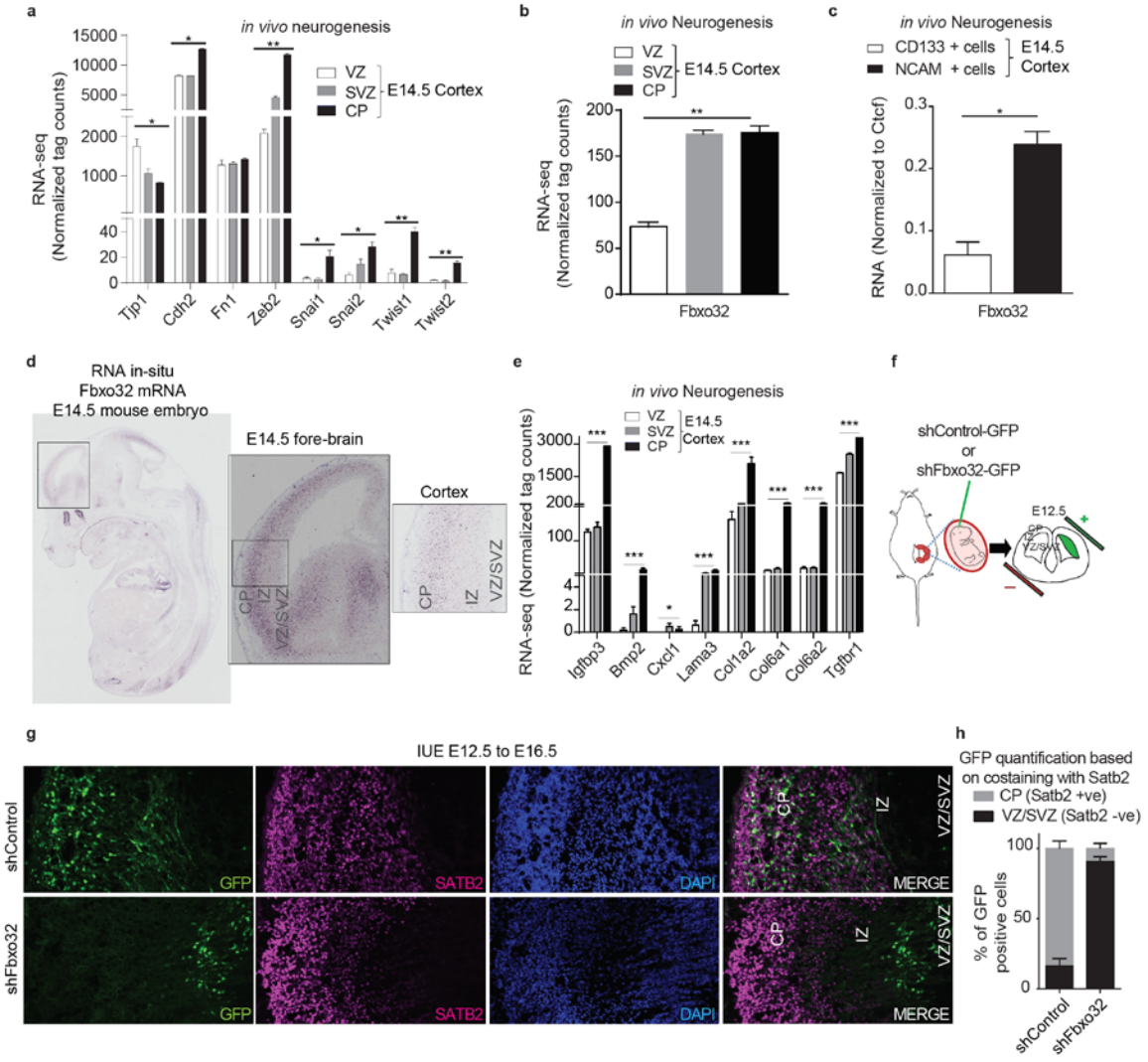


Fig S4. FBXO32 is required for neuronal EMT.

(a) Normalized tag counts for key EMT markers in RNA-seq data derived from the ventricular zone (VZ), the sub-ventricular zone (SVZ), and the cortical plate (CP) of E14.5 mouse cortex. (b) The levels of Fbxo32 mRNA are shown as the average normalized tag counts derived from the RNA-seq data used in (a). (c) RT-PCR analysis of Fbxo32 in FAC-sorted cells from E14.5 mouse cortex using CD133, a marker specific for cells from the ventricular zone and NCAM, a marker specific for cells from the cortical zone. The mRNA levels were measured relative to Ctcf via RT-PCR, and the results were plotted on the y-axis. Error bars represent the SEM of independent biological replicates. (d) The expression pattern of Fbxo32 mRNA, visualized via in situ hybridization, in E14.5 whole-mouse embryo showing the cortical localization of Fbxo32 in the developing neocortex. The data were extracted from the publically available data set "gene paint." (e) Normalized tag counts derived from the in vivo neurogenesis RNA-seq data used in (a) showing

the mRNA levels of genes that were downregulated in Fbxo32-ablated HMEC. (f) Graphical representation of in utero electroporation (IUE). (g) In utero electroporation was performed at E12.5 using plasmids containing GFP and with control shRNA or shRNA against Fbxo32, and the mice were sacrificed at E16.5 for further analyses. A representative image of the immunofluorescence analysis performed with anti-GFP and anti-Satb2 antibodies showing the retention of GFP-positive cells below the Satb2 layer in a brain electroporated with Fbxo32 shRNA compared to a control brain. (h) A bar plot showing quantification of the migrated GFP-positive cells in control and Fbxo32-depleted mouse brains with respect to Satb2 staining. The y-axis represents the percentage of cells on or below the Satb2-stained region. The error bars represent the SEM of three independent biological replicates. *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)

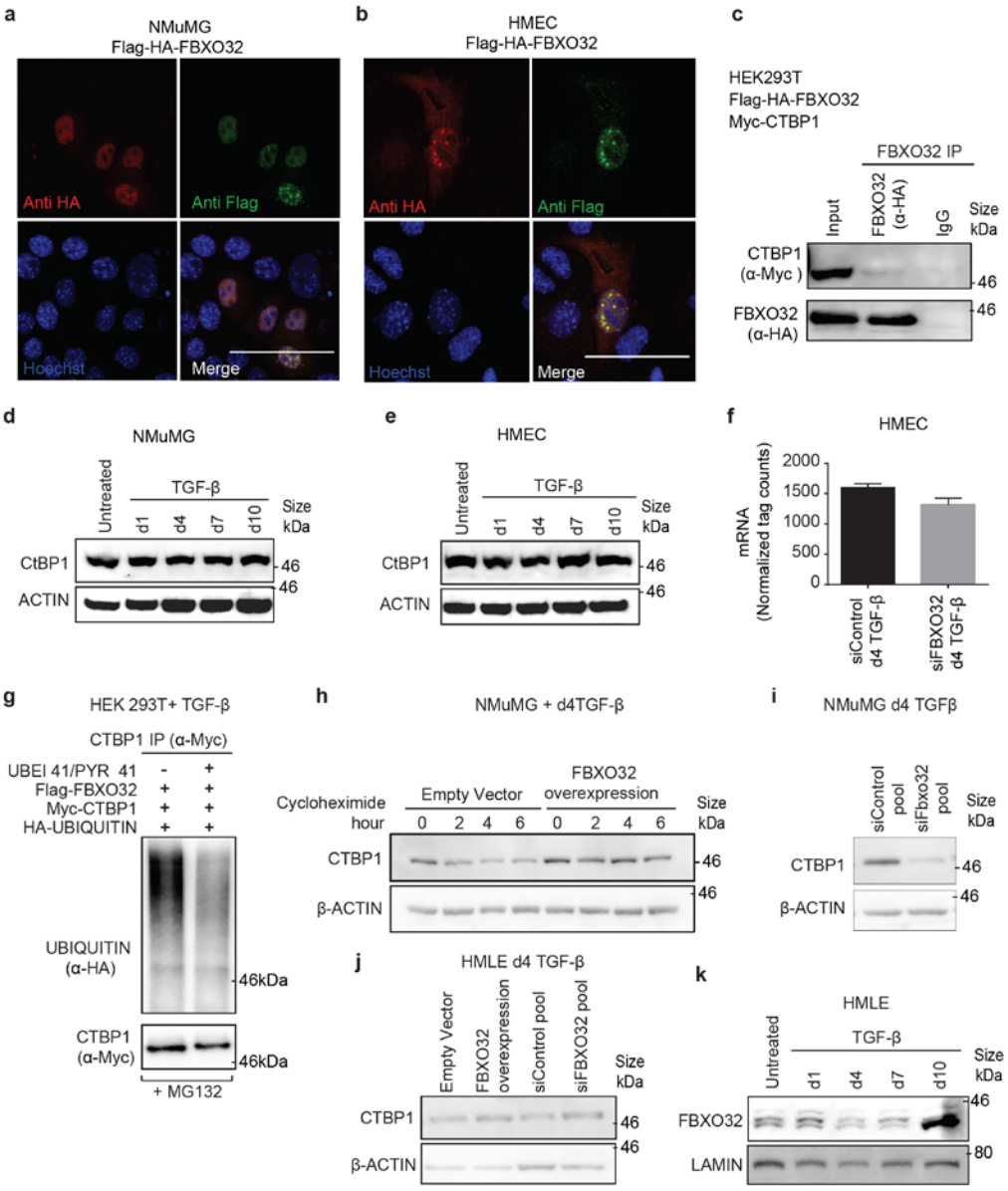


Fig S5. FBXO32 mediates stability of CTBP1 (Related to Fig 3)
(a-b) immunofluorescence images showing localization of FBXO32 in NMuMG (a) and HMEC (b) cells transfected with Flag-HA-FBXO32. (c) Western blot of immunoprecipitated samples to validate FBXO32

and CTBP1 interactions. HEK293T cells were co-transfected with Myc-CTBP1 and Flag-HA-FBXO32 and anti-FBXO32 IP was performed. Immunoblot was performed to detect CTBP1 and FBXO32. (d-e) Western blot analysis of CTBP1 in mouse NMuMG cell (d) and human HMECs (e) undergoing TGF-β-induced EMT. β-Actin acted as a loading control. (f) RNA-seq data showing mRNA level of CTBP1 in HMEC cell undergoing TGF-β induced EMT for 4 days, y-axis represents normalized tag count. (g) HEK293 cells were transfected with various tagged construct as mentioned in the figure. Four hour before immunoprecipitation MG132 was added to all conditions. Ubiquitination pathway blocker UBEI41/ PYR 41 was also added four hour before IP as mentioned in the figure. Western blot of CTBP1 (α-Myc) immunoprecipitated samples was performed to detect ubiquitin (α-HA) and CTBP1. (h) Western blot showing CTBP1 protein levels at various time points after cyclohexamide treatment in NMuMG cells stably expressing empty vector or FBXO32 overexpression construct. β-Actin acted as a loading control (i) Western blot showing CTBP1 protein stability in NMuMG cells transfected with control siRNA or siRNA against Fbxo32. β-ACTIN act as loading control. (j) Western blot showing CTBP1 protein levels in HMLE cells transfected with either empty vector or FBXO32 overexpression construct or control siRNA or siRNA against FBXO32. β-ACTIN act as loading control. (k) Western blot analysis of FBXO32 in human HMLE cells undergoing TGF-β-induced EMT. LAMIN-B1 acted as a loading control.

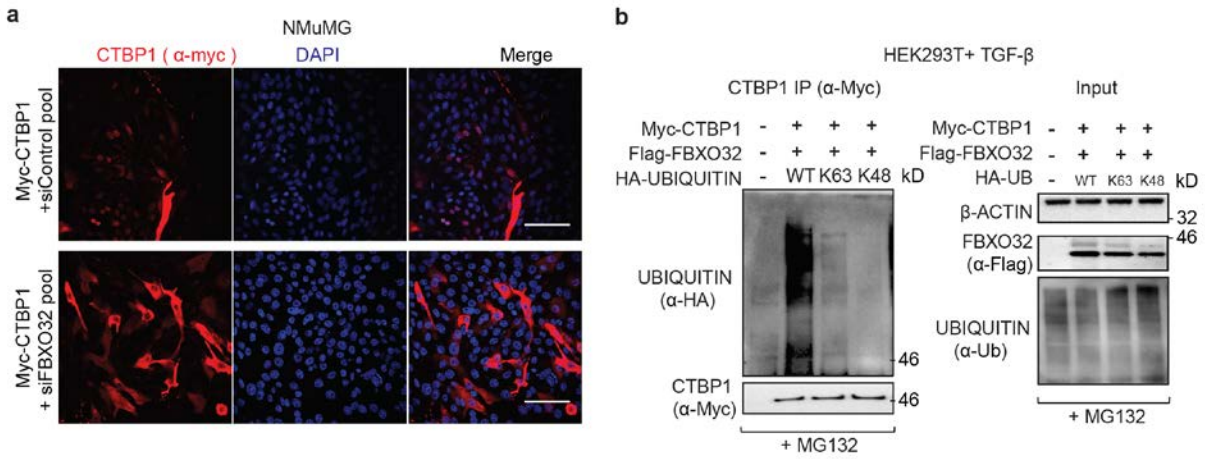


Fig S6. FBXO32 mediated K63 ubiquitination leads to nuclear retention of CTBP1. Related to Figure 3
(a) Immunofluorescence image showing CTBP1 localization in control and Fbxo32-depleted NMuMG cells undergoing EMT. Scale bar, 100 µm. (b) A similar analysis as in (Figure 3g), but co-transfected with modified ubiquitin to demonstrate FBXO32-mediated K63-linked ubiquitination of CTBP1. Here, K63 represents ubiquitin with all of the lysines modified to arginines, except the lysine at position 63. Similarly, in K48, all of the lysines were modified to arginines, except the lysine at position 48. (Results in this figure contributed by me)

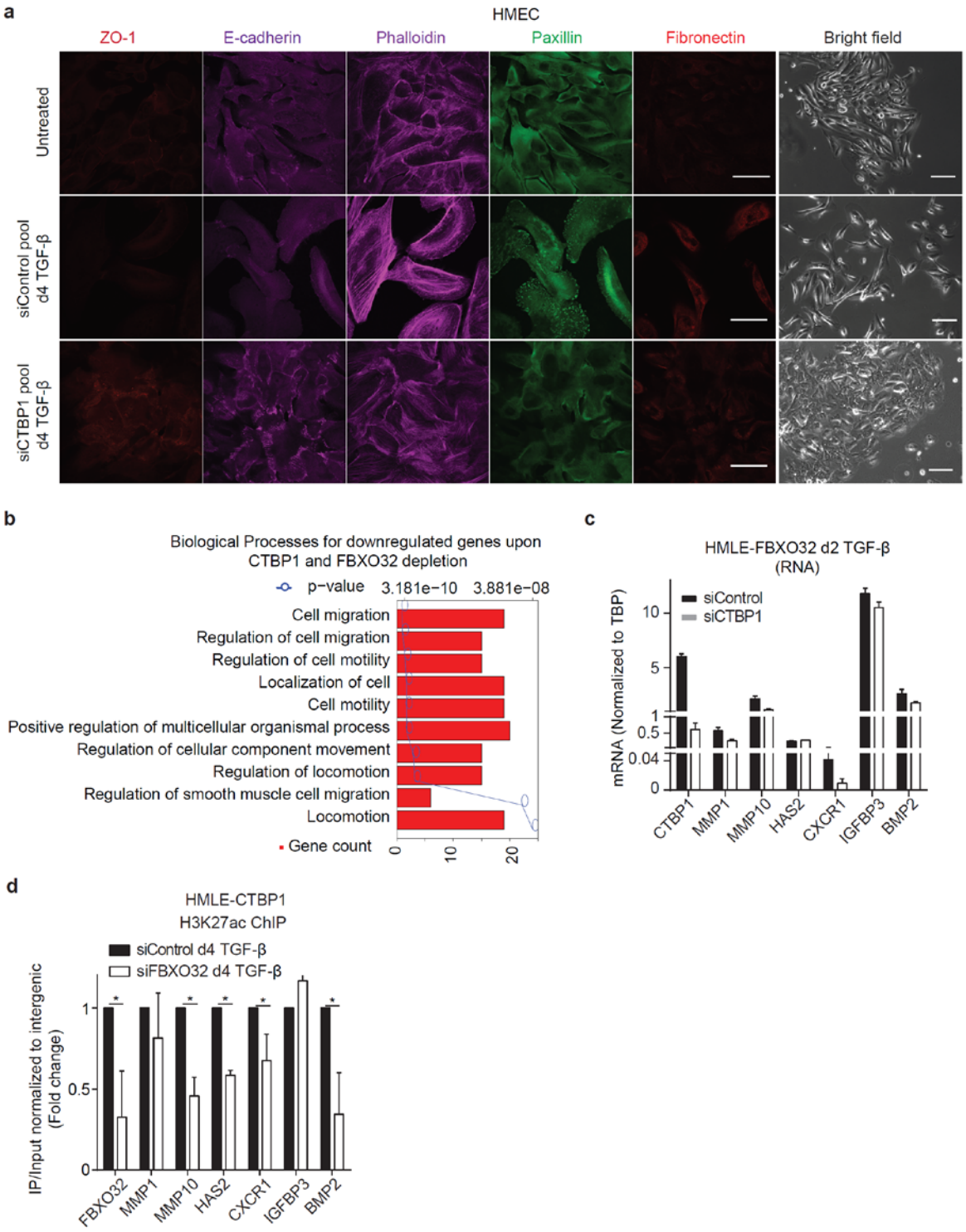


Fig S7. CTBP1 depletion results in a reversal of EMT in human primary epithelial cells. (Related to Fig 4) (a) Representative bright field and immunofluorescence images for localization and expression levels of EMT marker proteins after four days of siRNA-mediated depletion of CTBP1 compared to control siRNA in HMEC. Staining was performed with antibodies against epithelial markers ZO1 and E-cadherin, mesenchymal markers Fibronectin-1, with Phalloidin to visualize the actin cytoskeleton and against Paxillin to detect focal adhesion plaques. Scale bar, 100 µm. (b) GO analysis of genes downregulated upon CTBP1 and FBXO32 knockdown in HMEC undergoing EMT. (c) Using RT-PCR, the levels of FBXO32 and

CTBP1 target genes in HMLE cells undergoing TGF-β-induced EMT for two days (d2) and transfected with either control siRNA or siRNAs against CTBP1 in FBXO32 overexpression background were measured relative to CTCF and plotted on the y-axis. (d) ChIP assay using anti-H3K27ac antibody following expression of Myc-CTBP1 in HMLE cells induced by TGF-β for 4 days and transfected with control siRNA or siRNA against FBXO32. Quantitative PCRs were performed for indicating gene promoters and enrichments are plotted on the y-axis as the ratio of precipitated DNA (bound) to total input DNA and normalized to intergenic region (in fold change to control siRNA). SEM is derived from independent biological replicates. For all above experiments the error bars represent the SEM of three independent biological replicates. *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)

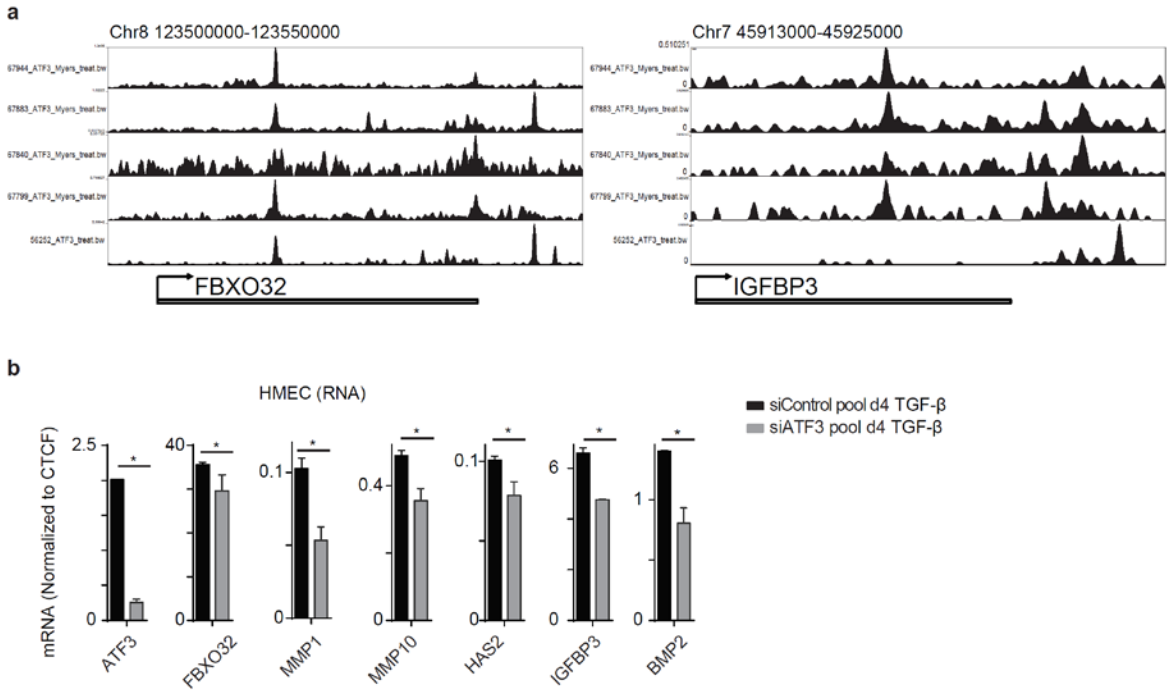


Fig S8. FBXO32 mediates transcriptional alterations via distinct transcription factors. (Related to Fig 4) (a) Genome browser tracks derived from a publicly available ATF3 ChIP-seq dataset (ENCODE/HAIB) showing binding of ATF3 at promoters and/or distal regions of genes downregulated upon FBXO32 and CTBP1 knockdown during EMT in human primary epithelial cells. (b) mRNA levels of key EMT-associated deregulated genes upon ATF3 depletion for 4 days in HMECs undergoing TGF-β-induced EMT were measured relative to CTCF via RT-PCR, and the results were plotted on the y-axis. Error bars represent the SEM of three independent biological replicates. *p<0.05, Student's t-test. (Results in this figure contributed by me)

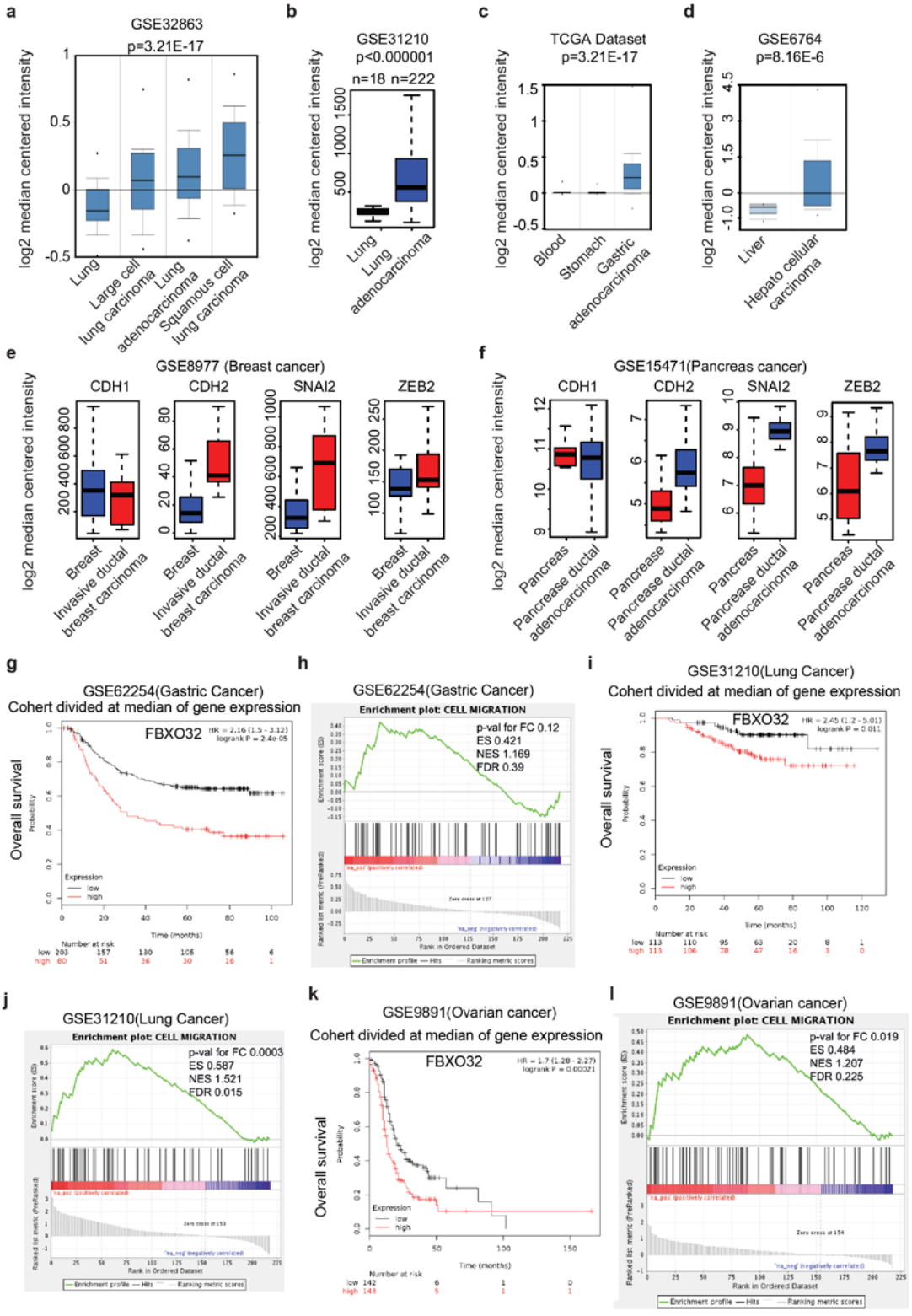


Fig S9. FBXO32 expression correlates with tumor aggressiveness. (Related to Fig 5)
(a-d) Box plots showing level of FBXO32 expression in normal vs tumor samples in various well characterized clinical tumor databases, in Lung (a-b), in Stomach (c), and in liver (d). (e-f) Box plots showing

level of key EMT markers in normal vs tumor condition in breast tumor datasets used in Fig 5b-c respectively. (g) A Kaplan-Meier analysis was performed for the Colon tumor dataset and survival curves showing decreased metastasis free survival, which was significantly correlated with higher FBXO32 expression. (h) A gene set enrichment analysis shows correlation between genes downregulated upon FBXO32 depletion and the genes associated to the specified GO terms as a function of fold change between expression levels of these genes in high FBXO32 and low FBXO32 expression in characterized breast tumor datasets used in Fig S7g. (i-j) Similar analysis as in Fig S7g-h, but in lung carcinoma dataset. (k-l) Similar analysis as in Fig S7g and h, but in ovarian carcinoma dataset. P value for fold change, Enrichment score (ES), Normalized Enrichment Score (NES) and FDR q-value were provided in the figure for all gene set enrichment analysis. (Results in this figure contributed by me)

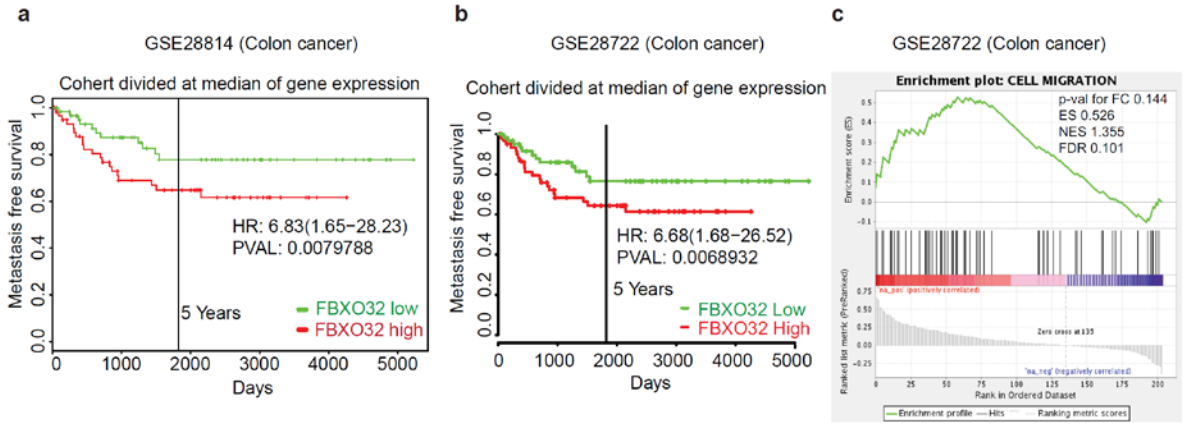


Fig S10. FBXO32 expression correlates with tumor aggressiveness. (Related to Fig 5)
(a) A Kaplan-Meier analysis was performed for the Colon tumor dataset GSE28814 and survival curves showing decreased metastasis free survival which was significantly correlated with higher FBXO32 expression. (b) Similar analysis as in Fig S8a but in another Colon tumor dataset GSE28722. (c) Gene set enrichment analysis showing correlation between genes downregulated upon FBXO32 depletion and the genes associated to the specified GO terms as a function of fold change between expression levels of these genes in tumor with high FBXO32 and low FBXO32 expression in the characterized tumor datasets in Fig S8b. p-value for fold change, Enrichment score (ES), Normalized Enrichment Score (NES) and FDR q-value were provided in the figure. (Results in this figure contributed by me)

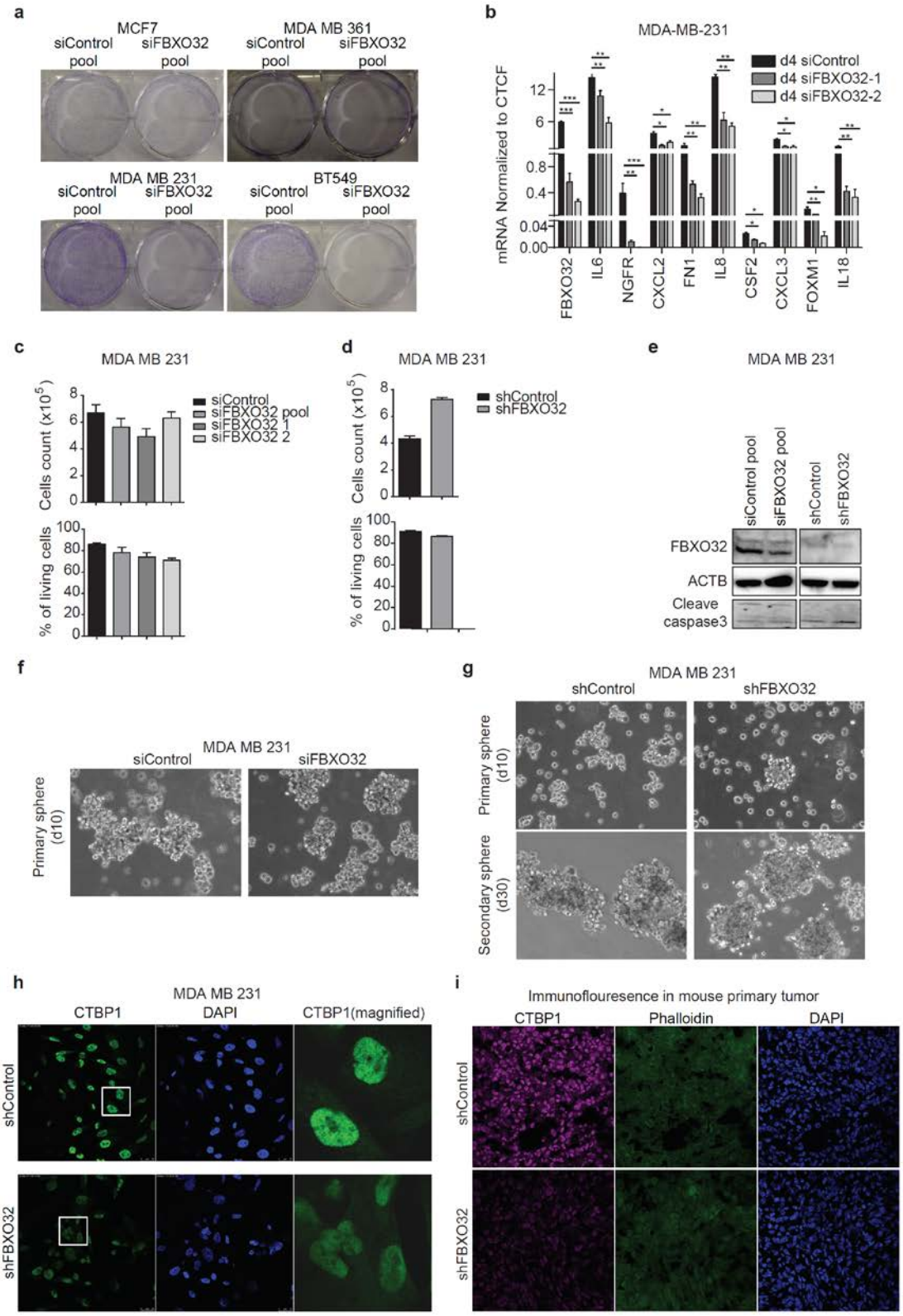


Fig S11. FBXO32 is crucial for the maintenance of mesenchymal identity in breast cancer cells. (Related to Fig 6) (a) Colony forming assay in two epithelial cell lines (MDA-MB-361 and MCF7) and two mesenchymal cell lines (MDA-MB-231 and BT549) showing colony formation ability in control and FBXO32

knock down conditions. (b) Using RT PCR mRNA levels of FBXO32 and key downregulated genes in control and FBXO32 knock down MDA-MB-231 cells for 4 days with independent siRNA were measured relative to CTCF and plotted on the y-axis. (c-d) Total number of cells and percentage of living cell were calculated for MDA-MB-231 upon depletion of FBXO32 by siRNA (c) and shRNA (d). (e) Western blot analysis of FBXO32 along with apoptotic marker Cleaved Caspase-3 in MDA-MB-231 cells depleted for FBXO32, either with pool of siRNAs or shRNA. B-Actin acted as a loading control. (f-g) Bright field images of MDA-MB-231 cells depleted for FBXO32 by siRNA (f) or shRNA (g), showing primary and secondary mammosphere, formed by sphere forming assay in presence of methylcellulose and absence of serum. (h) Immunofluorescence image in MDA-MB-231 cells showing CTBP1 localization in stably expressing shControl or shFBXO32 construct. Magnified images of the nuclei are provided in the right. (i) Immunofluorescence images for CTBP1 were measured in primary tumor obtained from the mouse injected with MDA-MB-231 cell containing shControl or shFBXO32 construct. Error bars represent the SEM of three independent biological replicates. *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)

Supplemental Tables

Table S1. Mouse RT primers

Ctcf	CACACACACAGGTACTCGTCCTCA CCACTGGTCACAAAGGCCATATC
Tbp	GTTTCTGCGGTCGCGTCATTT TGGGTTATCTTCACACACCATGAA
Fbxo32	AGTGAGGACCGGCTACTGTG GATCAAACGCTTGCGAATCT
Cdh1	GAGACAGGCTGGCTGAAAGTGAC TGACACGGCATGAGAATAGAGGA
Cdh2	GGTGGAGGAGAAGAAGACCAGGA TGGCATCAGGCTCCACAGTATCT
Fn1	CGGAGAGAGTGCCCCTACTA CGATATTGGTGAATCGCAGA
Twist1	TCAGCTACGCCTTCTCCGTCTG TGTCCATTTTCTCCTTCTCTGGAA
Snai1	TCTCTAGGCCCTGGCTGCTTC CAGCAAAAGCACGGTTGCAGT
Zeb2	CCAGAGGAAACAAGGATTTCAGG CGGAGTCTGTCATGTCATCTAGGC

Human RT primers

CTCF	GGGCTTGAGAGCTGGGTTCTATT CTTCGACTGCATCACCTTCCATT
TBP	GTGAACATCATGGATCAGAACAACA AAGATAGGGATTCCGGGAGTCAT

FBXO32	GGTGTATCGGATGGAGACGATTCT TCAGTGAAGGTGAGGCCTTTGAA
CLDN3	AACCTGCATGGACTGTGAAACCT GGTGGTCAAGTATTGGCGGTCAC
CDH2	GGCCGTCATCACAGTGACAGAT CCCTGTTCTCAGGAACTTCACCA
FN1	GCGAGAGTGCCCCTACTACACTG AATGTTGGTGAATCGCAGGTCA
TWIST2	GTCCATGTCCGCCTCCCACTA CAGCATCATTCAGAATCTCCTCCT
SNAI2	CCTGGTTGCTTCAAGGACACATT TGTTGCAGTGAGGGCAAGAAAA
SNAI1	CAGCGAGCTGCAGGACTCTAATC AGGATCTCCGGAGGTGGGATG
ZEB1	CCAACAGACCAGACAGTGTTACCAG TCTTGCCCTTCCTTTCCTGTGT
G0S2	GCAGCACGCCTCCTAGGAACT GGTCTGTCTCTGTCTACTGCGTCTC
CXCL3	CCATGGTTCAGAAAATCATCGAAA TCTGGTAAGGGCAGGGACCAC
MMP1	CCCCAAAAGCGTGTGACAGTAAG AAGGGATTTGTGCGCATGTAGAA
MMP10	TTTCTGCATTTTGGCCCTCTCTT AACGGTGTCCCTGCTGTTAACTT
CDH11	CCCTCAGCAAGACCACCGTACA GTGGTAGGCACAGGAGAATGCAG
IGFBP2	CTGGAGGAGCCCAAGAAGCTG AGGGAGTAGAGGTGCTCCAGAGG
GRAMD2	AATGCAGAGGGGAGGGCTCAT GGCCCAGAGCATGGAAATTCA
FGFBP1	GGGAAAAGGAGAACTCAGCACTTT ACAACACTGTGGCACGTTACTCAC
BCO2	CAGAATTCACTATTTCCTTGCCACTG AAACCGAAGCTTGTATGCACTCG
WNT10A	CTGTGGGCTCTAGGACTGACTGG CAACTGAACTGTCTGGGCTTCG
NGFR	CCTCATCCCTGTCTATTGCTCCA GTTCTGCTTGCAGCTGTTCCAC
FAM123B	CAGGAATAGGATCCTTACCACCTGA CTGCTTCTGCAGCCACATATCCT

IFIT2	GTGGCAGAAGAGGAAGATTTCTGA TTTTAGTTGCCGTAGGCTGCTCTC
CCND1	GTGCATCTACACCGACAACTCCAT ACTTGAGCTTGTTCACCAGGAGCA
CTBP1	CACGAGTCGGAACCCTTCAGC CACGAGTCGGAACCCTTCAGC

Human ChIP Primers

FBXO32	GGCAGTAGCTGCCGCAGTATTTA GAAAACAAGCCGAGCCCATAAAC
MMP1	CCCCAAAAGCGTGTGACAGTAAG AAGGGATTTGTGCGCATGTAGAA
MMP10	TTTCTGCATTTTGGCCCTCTCTT AACGGTGTCCCTGCTGTTAACTT
HAS2	CCTCCTCCAACTTAAGGGGGTCT AGTCCACACCTCCCTCTCCACTT
CXCR1	CAGATGACACCTCCCTTCTGAGC TGGACCCTGGCAGTCTCTAATCA
IGFBP3	TCGTCTACAAGAACCAAGGTGTGC AGGGAGACCTCACCCCGAGAG
BMP2	TGGATCCCACGTCTATGCTATGC AGGGTCAGGGTCTGGCCTCTTAT
CXCL1	GAGACACAACGCTCTTCCTCCAA ACTCTGGGATATTCGCCTTCTGC
INTERGENIC	CTACGTTCTCTATGGGGGTGTGC TGGGTTAAGAATTTGAGGGTAAATGAA

Table S2. **Mouse siRNA**

Fbxo32	ACAAGGAGGUAUACAGUAA UGUUGGAGCUGAUAGCAAA GCAGAGAGUCGGCAAGUCU CCUCAAGACUUUAUCAAUU
Ctbp1	GCAGCGGGUUUGACAAUAU AGAAUCAUCGUCCGAAUUG AUACCUAUCUGAUGGAAUC CCAUACGAGUGACCAGUUG

Human siRNA

FBXO32	GCAGAUCCGCAAACGAUUA GUACACUGGUCCAAAGAGU GUGCUGGUCGGGAACAUUA CAACUGAACAUCAUGCAGA
CTBP1	GGAUAGAGACCACGCCAGU GAGCAGGCAUCCAUCGAGA UGAAGAACUGUGUCAACAA

	AUGAGAAGGUCCUGAACGA

Supplemental Reagents and Experimental Procedures

Reagents

The reagents used in the present study were TGF-β (rhTGF-β1 240-B, R&D Systems), DMEM (21969-035, Invitrogen), RPMI-1640 (R0883, Sigma), PBS (D8537, Sigma), trypsin (25300-054, Invitrogen), Opti-MEM (31985-047, Invitrogen), FBS (10270, Invitrogen), glutamine (25030-024, Invitrogen), MEM NEAA (100x) (11140-035, Invitrogen), Lipofectamine 2000 (11668, Invitrogen), Lipofectamine RNAiMAX (13778-150, Invitrogen), Trizol (15596026, Invitrogen), reverse transcriptase kits (K1612, Thermo Fischer), SYBR-Green PCR MasterMix (4334973, Invitrogen), Bradford reagent (5000205, BioRad), a protease inhibitor cocktail (04693132001, Roche), and a phosphatase inhibitor cocktail (04906837001, Roche).

Inhibitor used in the study:

For chemical inhibition experiments, we used JNK inhibitors, SP600125, (10 μM) (S5567, Sigma), ERK-signaling inhibitors UO126 (25μM) (Milipore) and p38i SCIO-469, (5 nM), (3528, r&d systems).

Cells were seeded at required densities and pretreated 30 min prior to TGF-β treatment with signaling pathway inhibitors or their solvent control DMSO at the final concentrations, mentioned above and analyzed at indicated time points. For SMAD pathway blocking we have used on target plus smart pool siRNA (Dharmacon) targeting Smad4. This was followed by harvesting the cells 24 hours later for RNA analysis.

Ubiquitination pathway inhibitor PYR-41 or 4[4-(5-Nitro-furan-2-ylmethylene)-3,5-dioxo-pyrazolidin-1-yl]-benzoic acid ethyl ester (N2915, Sigma) was added to cell culture medium to a final concentration of 10 μM, 4 hour prior to collection of cells.

Antibodies used for Western blot, IP and IF

The antibodies used for immunofluorescence were against E-cadherin (13-1900, Invitrogen and 610182, BD Transduction Laboratories), N-cadherin (610921, BD Transduction Laboratories), ZO-1 (617300, Invitrogen), fibronectin (F-3648, Sigma-Aldrich), paxillin (610052, BD Transduction Laboratories), GFP (chicken, Aves Labs), SATB2 (Abcam, ab34735), beta-actin (C4, sc-47778, Santa Cruz), FBXO32 (sc-33782, Santa Cruz) (ab168372), ZEB1 (sc-25388 (H-102) Santa Cruz), ATF-3 (sc-188 (C-19) Santa Cruz), Cleaved caspase 3 (Asp175, 9661S, cell signaling), FLAG tag M2 (F1804, Sigma), HA tag monoclonal antibody (c15200190, Diagenode), Myc tag (ab9106, Abcam), ubiquitin mouse monoclonal (sc8017, Santa Cruz), Lamin B (sc-6216 (C-20), Santa Cruz), Histone H3 tri methyl K4, (ab12209, Abcam), Histone H3 tri methyl K9, (ab8898, Abcam), Histone H3 acetyl K27, (ab4729, Abcam), Alexa Fluor-488 goat anti-mouse IgG (H+L) (A11029, Invitrogen); Alexa Fluor-568 goat anti-rabbit IgG (H+L) (A11011, Invitrogen), Alexa Fluor-633 goat anti-rat IgG (H+L) (A21094, Invitrogen), and Alexa Fluor-633 phalloidin (A22284, Invitrogen), which was used to stain F-actin.

Mammosphere culture

For primary sphere formation, single cells were plated in ultralow attachment plates (Corning) at a density of 20,000 viable cells/mL in serum-free mammary epithelial growth medium composed of DMEM supplemented with 1% L-glutamine, 1% penicillin/streptomycin, 30% F12 , 2% B27 (Invitrogen), 20 ng/mL EGF (peprotech, AF-100-15) and 20 ng/mL bFGF (peprotech, 100-18B). The medium was made semi-solid by the addition of 0.5% Methylcellulose (Sigma) to prevent cell aggregation. Mammospheres were collected by gentle centrifugation (200 g) after 10 days and dissociated enzymatically (10 min in 0.05% trypsin, 0.53 mM EDTA, Invitrogen) and mechanically, using a fire-polished Pasteur pipette. The cells obtained from dissociation were sieved through a 40-µm sieve and analyzed microscopically for single-cellularity. For the secondary sphere formation, 1000 cells/ml were plated and colonies were monitored by a microscope at a regular interval.

Immunofluorescence assay

Cells were grown on coverslips, fixed with 4% paraformaldehyde in PBS and permeabilized with 0.2% Triton X-100 for 15 minutes at room temperature. Subsequently, the cells were blocked with 10% goat serum, 5% FCS and 0.5% BSA in PBS for 20 minutes and were incubated with primary antibodies at 4°C overnight. The cells were then incubated with fluorochrome-labeled secondary antibody or Phalloidin-633 for 1 hour at room temperature. The coverslips were counterstained with Hoechst, mounted with immu-mount and imaged using a confocal laser-scanning microscope. The data were processed by ImageJ software.

Immunofluorescence assay to assess neuronal migration in mouse brain

Isolated E16.5 embryonic brains were immediately fixed for 6 hours in 4% PFA in PBS at 4°C. The brains were then cryoprotected in 10% sucrose for 2 hours and then in 30% sucrose (in PBS) overnight, embedded in Tissue-Tek, stored at -20°C and cryo-sectioned at 12 µm. Sections on coverslips were preblocked with 2% BSA and 0.5% Triton (in PBS) for 1 hour. Primary antibodies (Satb2, 1:500, and anti-GFP (chicken, Aves Labs, 1:1000)) were applied in blocking solution overnight at 4°C. Fluorescent 11 secondary antibodies were applied according to the manufacturer's protocol (Life Technologies). The coverslips were counterstained with Hoechst, mounted with immomount and imaged using a confocal laser-scanning microscope (Leica SP5). The data were processed with ImageJ software.

Quantitative RT-PCR

mRNA levels were quantified as previously described (49). In brief, total RNA was prepared using Trizol (Invitrogen) or a SurePrep TrueTotal RNA Purification Kit (Fisher Scientific) and was reverse transcribed with a First Strand cDNA Synthesis Kit (Fermentas). The transcripts were quantified via PCR using SYBR green PCR MasterMix (ABI) on a ViiA7 PCR machine (Life Technologies). Human or mouse Ctcf and Tbp primers were used for normalization. The sequences of all of the primers used in this study are listed in Table 2.

Immunoblotting

Cells were lysed in RIPA buffer, and protein concentrations were quantified using Bradford reagent. Equal amounts of protein (30 µg) were boiled in 6× SDS-PAGE loading buffer, subjected to polyacrylamide gel electrophoresis, transferred to a PVDF membrane and probed with the appropriate antibodies.

ChIP assay

Cells were cross-linked in medium containing 1% formaldehyde for 10 minutes at room temperature, neutralized with 0.125 M glycine, scraped, and rinsed twice with PBS. The pellets were suspended in buffer L1 (50 mM Hepes KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% glycerol, 5% NP-40, and 0.25% Triton-X 100) and incubated for 10 minutes at 4°C. The cells were then suspended in buffer L2 (200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0) for 10 minutes at room temperature. Finally, the pellet was suspended in buffer L3 (1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, 100 mM NaCl, 0.1% Na-deoxycholate, 0.17 mM N-lauroyl sarcosine) containing protease inhibitors and was incubated at 4°C for 3 hours following sonication using Bioruptor plus (Diagenode). Sixty micrograms of chromatin was incubated overnight at 4°C with 2 µg of the antibodies targeting H3K4me3, H3K9me3, H3K27ac (Abcam) or Myc (Santa Cruz) and then incubated with preblocked beads for 4 hours. Finally, the beads were washed twice with L3 and once with 1 ml of DOC buffer (10 mM Tris (pH 8.0), 0.25 M LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA), and the bound chromatin was eluted in 1% SDS/0.1 M NaHCO3. This followed treatment with RNase A (0.2 mg/ml) for 30 minutes at 37°C and then with proteinase K (50 µg/ml) for 2.5 hours at 55°C. The crosslinking was reversed at 65°C overnight with gentle shaking. The DNA was purified by phenol-chloroform extraction followed by ethanol precipitation and was recovered in TE buffer.

Mass spectrometry (MS) analysis

Peptide fractions were analyzed using a quadrupole Orbitrap mass spectrometer (Q Exactive Plus, Thermo Scientific) equipped with a UHPLC system (EASY-nLC 1000, Thermo Scientific), as described [302]. Peptide samples were loaded onto C18 reversed-phase columns and eluted for 2 hours with a linear gradient of acetonitrile from 8 to 40% containing 0.1% formic acid. The mass spectrometer was operated in data-dependent mode with automatic switching between MS and MS2 acquisition. Survey full scan MS spectra (m/z 300 – 1650) were acquired in the Orbitrap. The 10 most intense ions were sequentially isolated and fragmented by higher-energy C-trap dissociation (HCD) [303]. Peptides with unassigned charge states or charge states less than +2 were excluded from fragmentation. The fragment spectra were acquired in the Orbitrap mass analyzer.

Peptide identification: raw data files were analyzed using MaxQuant (development version 1.5.2.8) [304]. Parent ion and MS2 spectra were searched against a database containing 88,473 human protein sequences obtained from the UniProtKB released in December 2013 using the Andromeda search engine [305]. The spectra were searched with a mass tolerance of 6 ppm in MS mode, 20 ppm in HCD MS2 mode, strict trypsin specificity and allowing up to 3 miscleavages. Cysteine carbamidomethylation was searched as a fixed modification, whereas protein N-terminal acetylation and methionine oxidation were searched as variable modifications. The dataset was filtered based on posterior error probability (PEP) to arrive at a false discovery rate below 1%, estimated using a target-decoy approach [306].

***In vivo* tumorigenicity assays and imaging**

NOD *scid* gamma (NSG) mice were bred and maintained under SPF conditions in the Translational Animal Research Center at the University Medical Center Mainz. Briefly, cells were counted and resuspended in a 1:1 (v/v) mixture of PBS and Matrigel (BD Biosciences). Ten-week-old female mice were injected unilaterally with 2×10^6 cells in 100 µl of 50:50 Matrigel/PBS into the fourth abdominal fat pad via subcutaneous injection at the base of the nipple. Tumor growth was monitored externally using Vernier

calipers for up to 48 days. The tumor volume was calculated as follows: tumor volume (mm^3) = length × $(width)^2 \times 0.5$. Necropsies were performed to identify macro-metastases. Primary tumors and organs were immediately frozen in liquid nitrogen and stored at -80°C until use.

In vivo bioluminescence imaging of tumor-bearing mice and their organs was performed at day 50 using an IVIS Lumina imaging system (Perkin Elmer). Briefly, mice were anesthetized with isoflurane and an aqueous solution of D-luciferin-K^+ salt (150 mg/kg body weight) (Perkin Elmer) was injected intraperitoneally. Five minutes after the injection, the mouse was placed onto the imaging chamber of IVIS, and photons acquired with an integration time of 10 s were presented as color-scaled images using IVIS Living Image Software (version 4.3.1) (Perkin Elmer). For organ imaging, mice were sacrificed after luciferin injection, and the dissected organs were imaged as described above.

Animal maintenance and experiments were performed under an approved protocol in accordance with the animal care guidelines of Johannes Gutenberg University.

***In utero* electroporation**

In utero electroporation experiments were performed essentially as previously described (Saito, 2006). All experimental procedures were conducted in accordance with European, national, and institutional guidelines for animal care. Timed-pregnant (Theiler stage 20=E12.5) C57BL/6 mice were anesthetized with isoflurane (2.5% via mask, Forene®, Abbod), and carprofen (4 mg/kg body weight, Rimadyl, Pfizer) was administered subcutaneously as an analgesic. After the abdominal cavity was opened, the embryos were carefully exposed, and 1 µl of colored solution containing 1 µg of the p.SUPER-GFP shRNA plasmid expressing shRNA against Fbxo32 or control shRNA was injected into one of the lateral ventricles. Using specialized platinum electrodes (Nepagene CUY 650P), the appropriate voltage was applied (50 ms, interval 950 ms, 5 pulses). After electroporation, the uterine horns were returned to the abdominal cavity.

Cell survival assay

Cells were pre-depleted with siRNA (a pool of 4 siRNAs targeting the same gene, SMARTpool, Dharmacon) for two days, and 40,000 cells per 6 cm were plated and cultured for 6 days. Fresh medium along with siRNA was added every other day. The cells were counted, or bright-field images were obtained, or the cells were fixed with 4% PFA. To visualize the fixed colonies, the cells were stained for 10 minutes with a solution containing crystal violet and 10% ethanol, and the plates were washed in water until the excess dye was washed away. Photographs of the plates were obtained for later analysis.

Wounding migration assay

Cells depleted of FBXO32 were seeded at equal densities, and a scratch wound was generated using a 10-µl pipette tip on confluent cell monolayers grown in six-well culture plates. The cells were then washed with fresh medium to remove floating cells. Bright-field images were obtained at 20x magnification after 18 hours of wounding.

Migration and invasion assays

Migration assays were performed as previously described (Tiwari et al, 2013). Briefly, 104 cells were seeded in 2% FBS/DMEM (Sigma) into the upper chamber of a 24-well Transwell migration insert (pore size 8 µm; Falcon BD). The lower chamber was filled with 20% FBS/DMEM. After 16 hours of incubation at 37°C, the cells in the upper chamber were removed using a cotton swab, and the cells that had traversed the membrane were fixed in 4% paraformaldehyde/PBS and quantified by DAPI staining using a fluorescence microscope and ImageJ software. For the invasion assays, transwells were coated with 0.5 mg/ml Matrigel in serum-free medium overnight, and 1.5×105 cells were used for the assays.

ZNF827-dependent splicing dynamics governs epithelial to mesenchymal transition

Sanjeeb Sahu[1], Eneritz Agirre[2], Mustafa Diken[3], Susanne Strand[4],

Reini F. Luco[2], Vijay K. Tiwari[1,*]

1. Institute of Molecular Biology (IMB), Mainz, Germany
2. Institute of Human Genetics - UMR9002 CNRS-University of Montpellier, Montpellier, France
3. TRON - Translational Oncology at the University Medical Center of the Johannes Gutenberg University gGmbH, Mainz, Germany
4. Department of Internal Medicine I, University Medical Center, Johannes Gutenberg University, Mainz, Germany

*Correspondence should be addressed to: v.tiwari@imb-mainz.de

Running title: ZNF827 regulates splicing program underlying EMT

Keywords: C2H2 zinc finger/EMT/Epigenetics/ Gene regulation/ Splicing

ABSTRACT

Epithelial to mesenchymal transition (EMT) is a multi-step process during which static epithelial cells become migratory mesenchymal cells. EMT is involved during gastrulation, neural crest migration, tissue regeneration and wound healing while its aberrant activation is associated with organ fibrosis and cancer metastasis. Despite advances, molecular machinery governing the alternate splicing repertoire underlying EMT remain largely unexplored. Here we find that ZNF827, a C2H2-zinc finger containing protein, is critical for molecular and phenotypic changes underlying EMT and is indispensable for tumor metastasis and neuronal migration *in vivo*. Further analysis revealed that these effects are linked to ZNF827-dependent hundreds of alternative splicing events within numerous genes during EMT. We show that ZNF827 mediates these responses by directly targeting these gene loci and modulating their epigenetic landscape and RNA Pol II kinetics. Furthermore, ZNF827 also facilitates the recruitment of core splicing components to nascent RNA. Taken together, these findings establish that ZNF827 is a critical regulator of EMT by defining the alternative transcriptome that underlies this process during development and disease.

INTRODUCTION

Epithelial-to-mesenchymal transition (EMT) is a multi-step morphogenetic process in which epithelial cells downregulate their epithelial properties, such as apical-basal polarity, tight and adherent junctions and acquire mesenchymal characteristics, such as reorganized actin cytoskeleton, as well as migratory and invasive behavior [198,214,308,309]. EMT is essential for various developmental processes, such as gastrulation and neural crest migration. In adults, EMT participates in wound healing and regeneration. Importantly, EMT mechanisms are involved in various pathological conditions, such as fibrosis and cancer metastasis [214,309,310]. The dynamic reprogramming of the cellular state during EMT is governed by numerous signaling pathways that respond to extracellular cues [197,257,311]. TGF-β is considered a prototypic EMT inducer, whereas other molecules, such as FGF, EGF, HGF, Wnt/beta-catenin and Notch, display more context-dependent action [219]. The Activation of these signaling cascades during EMT results in transcriptional, post-transcriptional, translational and post-translational modulation that is essential for driving EMT [198].

Recent advances in genomics have led to an increased appreciation of a role of alternative splicing (AS) during various biological process [312-314]. Alternative splicing is the process by which various combinations of splice sites in precursor mRNA (pre-mRNA) are selected to generate structurally- and functionally-distinct mature messenger mRNA species that give rise to additional layers of regulation in mRNA stability and translational efficiency. Alternative splicing also generates multiple protein variants with various structural and functional identities [198]. In this way, alternative splicing extensively increases the coding capacity of metazoan genomes [315-317]. Post-transcriptional RNA processing can modulate essential protein functions and help the cell adapt to various physiological conditions [318].

Few studies, mostly involving single gene analysis, have implied that EMT depends on a defined, alternatively spliced transcriptome [319-323]. However, a comprehensive, genomewide analysis of alternative transcriptome during EMT was lacking and the molecular machinery governing this process remained

largely uncharacterized. C2H2 Zinc Finger proteins constitute the largest class of annotated DNA-binding proteins (n = 740 for human and n = 780 for mouse) [324,325]. However, due to their complex structure and duplicated family members, the functions of a vast majority of these proteins remain unknown [326]. Interestingly, a set of these proteins were recently implicated in alternative splicing [327]. Here we find that one such protein, ZNF827, is indispensable for molecular and phenotypic changes that define EMT including during tumor metastasis and neuronal migration *in vivo*. The only other previous study on ZNF827 had found this to be functioning in telomere maintenance in a very small fraction of tumor types via recruitment of NuRD complex [328]. We find that ZNF827 is required for hundreds of alternative splicing events within numerous genes during EMT. Mechanistically, ZNF827 influences the recruitment of core splicing machinery to its target genomic sites by creating a specific epigenetic state that influences RNA Polymerase II occupancy, and thus kinetics, supporting a new role for transcriptional regulators in alternative splicing via chromatin remodeling [329,330]. In summary, ZNF827 is a critical regulator of the alternative transcriptome that underlies transition from an epithelial to a mesenchymal cell fate during development and disease.

RESULTS

ZNF827 is essential for phenotypic and molecular changes underlying EMT

To identify novel regulators of EMT, we employed untransformed normal murine mammary gland epithelial cells (NMuMG), as well as human HMLE cells exposed to TGF-β. These are established model for EMT *in vitro* [197,257,281,311] (Fig. S1a-b, S2a-b).. To identify genes that alter their expression pattern during TGF-β-induced EMT in both mouse and human cells, we performed high-coverage deep sequencing of the transcriptome (RNA-Seq) at several time points of TGF-β exposure (day 0, day 1, day 4, day 7, and day 10) spanning early, intermediate, and late stages of EMT [257] (Fig. S3a-b). We next performed a stepwise selection approach to obtain candidate genes that are differentially expressed during EMT and are required for this process (Fig 1a). We particularly focused on C2H2 Zinc finger proteins, given that their function in EMT remains largely unexplored. Our data showed that many members of this family are differentially expressed during EMT in both human and mouse cells (Fig. S3c-d). We further analyzed the expression patterns of selected candidates in other established model systems of EMT, including TGF-β-induced EMT in a mouse Ras-transformed breast epithelial cell line (EpRAS) and in human primary mammary epithelial cells (HMECs). We also examined expression patterns in neuronal EMT and various invasive/non-invasive cancer cell lines (data not shown). The candidates that showed consistent expression patterns in various EMT models were further tested for their functional role using loss-of-function experiments. This involved siRNA-mediated depletion of the target mRNA during EMT, followed by extensive phenotypic and molecular characterization (data not shown). This comprehensive screening identified ZNF827 (called Zfp827 in mouse) as a highly potent regulator of EMT. ZNF827 showed strong transcriptional induction during TGF-β-induced EMT in mouse (NMuMG), human primary breast epithelial cells (HMEC) and HMLE cells (Fig 1b-c, S3e). We also observed transcriptional induction of ZNF827 in an independent mouse EMT model (EpRas cells) (Fig S3f). In line with these observations, we detected an

increased expression of ZNF827 in mesenchymal cancer cell lines (MDA-MB-231 and BT549) as compared with epithelial cancer cell lines (MCF7 and MDA-MB-361) (Fig S3g).

Next, we investigated the function of ZNF827 during TGF-β-induced EMT in mouse (NMuMG) cells using loss-of-function approaches. We did this with two independent siRNAs (Fig 1d) and with a pool of siRNAs (Fig. S4a), to ensure specificity of phenotypic effects. Depletion of Zfp827 during EMT strongly inhibited acquisition of the mesenchymal state. Evidence for inhibition of EMT was confirmed by an immunofluorescence assay showing substantial retention of the epithelial markers (e.g. E-cadherin and ZO-1) at the membrane, inability to acquire proper levels of crucial mesenchymal markers (e.g. N-cadherin, Fibronectin), and failure to undergo cytoskeletal remodeling involving cortical actin, stress fibers and focal adhesion formation (e.g. Phalloidin and Paxillin) (Fig 1d). We further validated EMT inhibition at the RNA level for several key EMT markers for both mouse and human cells (Fig 1 e-f). Our gene expression analysis of EMT markers during EMT in human HMEC cells indicated that the majority of critical EMT regulators strongly change in expression by day 4, justifying our selection of this time point for the functional analysis of ZNF827 at an early stage of EMT (Fig S2a-b). Together, these data suggest that ZNF827 is an essential regulator of phenotypic and molecular changes underlying EMT.

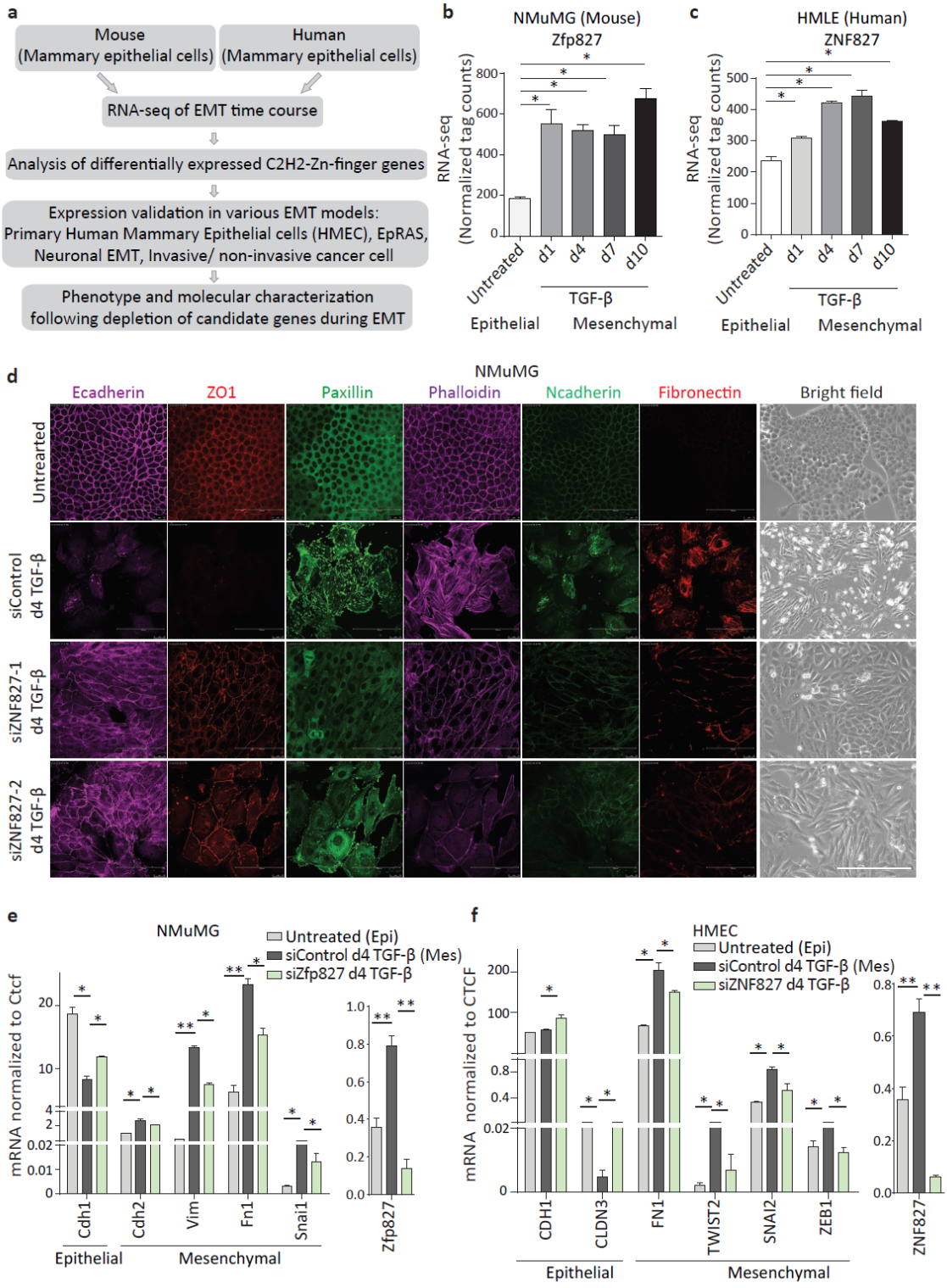


Fig 1. ZNF827 is critical for phenotypic and transcriptional changes during EMT.
(a) Schematic representation of *in vitro* EMT model system and the identification of novel Zinc finger proteins that are critical for EMT. **(b-c)** RNA-seq data showing mRNA level of ZNF827 in NMuMG cells (b) and HMLE cells (c), undergoing TGF-β-induced EMT, the y-axis represents the normalized tag count. **(d)** Representative bright-field and immuno-fluorescence images of the localization and expression levels of

EMT marker proteins after 4 days (d4) of TGF-β-induced EMT and siRNA-mediated depletion of ZNF827 compared to non-targeting control (siControl) in NMuMG cells. Staining was performed to assess the expression with antibodies against the epithelial markers E-cadherin and ZO1, the mesenchymal marker N-cadherin, Fibronectin-1, Phalloidin (to visualize the actin cytoskeleton) and Paxillin (to detect focal adhesion plaques). Scale bar: 100 µm. **(e-f)** RT-PCR measurement of key EMT marker levels in NMuMG (e) and HMEC (f) cells transfected with either control siRNA or siRNAs against ZNF827 during TGF-β-induced EMT. RNA levels are shown relative to CTCF mRNA levels. All experiments were performed in biological triplicates unless otherwise specified. Error bars represent the SEM of three independent biological replicates. *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)

ZNF827 is required to confer gene expression program hallmark of EMT

Based on our observation of the critical role of ZNF827 in phenotypic and molecular changes during EMT, we investigated the biological processes that this molecule regulates during EMT. ZNF827 possesses nine C2H2 zinc finger domains (Fig 2a). In contrast to the only other study involving this protein, which showed its telomere-specific localization and function [328], we found a homogenous distribution in the nuclear compartment, suggesting a wider nuclear role for this protein (Fig 2b). Pursuing this possibility, we performed genome-wide transcriptome analysis (RNA-seq) following ZNF827 depletion during EMT in human primary mammary epithelial cells (HMEC) on day 4. Computational analysis revealed many significantly differentially-expressed genes, with more genes upregulated (n = 135) than downregulated (n = 46) in ZNF827 knockdown condition as compared to control knockdown. This suggests a potential repressive function of ZNF827 (Fig 2c, S4b, and Table 2). The downregulated genes were enriched for functions related to regulation of cell migration, while the upregulated genes showed enrichment for epithelial development and subcellular protein localization (Fig 2d). Many of these differentially expressed genes revert to an epithelial state upon ZNF827 depletion. This phenomenon was validated by independent RT-qPCRs (Fig 2e-f). Here mesenchymal markers such as FN1, TWIST2, SNAI2 and ZEB1, PCDH19, GOS2, PLAT, IL1B and DCBLD1 were downregulated following ZNF827 depletion (Fig 1f, Fig 2e). Furthermore ZNF827 knockdown led to an upregulation of epithelial genes such as CLDN3, BCO2, RBKS, CXCL14 and WN10A (Fig 1f, Fig 2f). These findings suggest that ZNF827 plays a crucial role in EMT by regulating transcriptome changes critical for EMT progression.

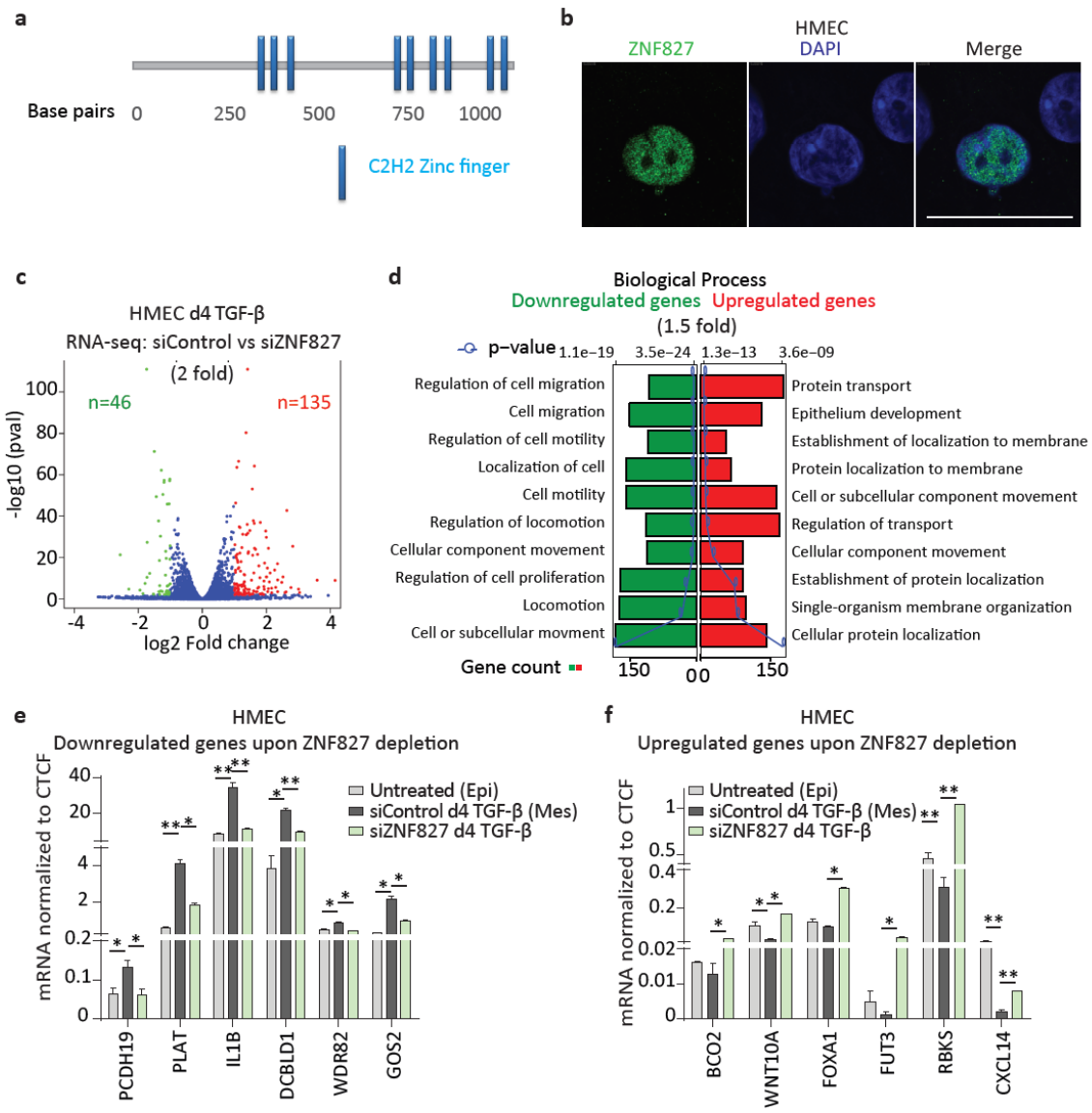


Fig 2. ZNF827 is essential for EMT gene expression program.

(a) Graphical representation of Zn finger domain organization of ZNF827. **(b)** Immunofluorescence image showing localization of ZNF827 in human primary breast epithelial cell HMEC. Scale bar: 100 µm. **(c)** Volcano plot- representation of RNA-seq results shows the number of genes with significantly altered expression levels upon ZNF827 depletion in HMECs undergoing TGF-β-induced EMT for 4 days. **(d)** GO analysis of downregulated and upregulated genes upon ZNF827 knockdown in HMECs that underwent EMT. **(e-f)** The mRNA levels of downregulated genes (e) and upregulated genes (f) upon ZNF827 depletion during EMT, in control and ZNF827 knockdown HMECs, were measured relative to CTCF via RT-PCR, and the results were plotted on the y-axis. Error bars represent the SEM of three independent biological replicates. **p*<0.05, ***p*<0.01, ****p*<0.001, Student's t-test. (Results in this figure contributed by me)

ZNF827 is induced during neurogenesis and is required for neuronal EMT *in vivo*

To further examine the role of ZNF827 (Zfp827 in mouse) underlying EMT *in vivo*, we next analyzed its requirement during developmental EMT events. During cortical development, radial glial cells in the

ventricular zone (VZ) undergo asymmetric division and daughter cells migrate toward the cortical plate (CP), passing through the sub-ventricular zone (SVZ) via an EMT-like mechanism [257,287]. Gene expression analysis of different stages of neurogenesis *in vitro* (neuronal differentiation of ES cells) and *in vivo* (VZ, SVZ and CP of the E14.5 mouse cortex) revealed transcriptional induction of Zfp827 during neuronal EMT (Fig S5a-b). These stages have been shown to exhibit transcriptional changes characteristic of EMT [257]. These results were further validated by analyzing existing single cell RNA-seq datasets for different subpopulations of the developing mouse cortex (Fig S5c).

To further examine the role of Zfp827 in neuronal migration during cortical development, we performed *in utero electroporation* of developing mouse cortex [239] with control and ZNF827-overexpressing plasmids containing an RFP reporter. Embryos were electroporated at E13.5 and the mice were sacrificed at E17.5 for further analysis (Fig S5d). In contrast to the control, the overexpression of ZNF827 led to a depletion of progenitor cells in the VZ and SVZ regions. Most ZNF827-overexpressing cells become multipolar, loose directionality and get confined to the intermediate zone of the mouse cortex (Fig S5e-g). These observations raise the possibility that ectopic overexpression of ZNF827 in progenitor cells gives rise to premature neuronal differentiation. We further validated these observations by shRNA-mediated depletion of Zfp827 during cortical development. This led to decreased neuronal migration (Fig S5h-j). Together, these observations suggest that ZNF827 plays a critical role in neurogenesis during brain development, potentially via regulating neuronal EMT.

ZNF827 promotes tumorigenicity and metastasis in the *NSG* mouse model

Prompted by extensive characterization of the molecular function of ZNF827 during EMT *in vitro* as well as during development, we next addressed the function of ZNF827 in disease conditions that involve EMT. Aberrant activation of the EMT program is associated with tumor metastasis [331]. ZNF827 expression is higher in human mesenchymal cancer cell lines (MDA-MB-231 and BT547) compared with epithelial cancer cell lines (MCF7 and MDA-MB-361) (Fig S3g), suggesting a potential function in mesenchymal

cancer cells. We depleted ZNF827 in mesenchymal breast cancer cells and assessed their migration ability. Depletion of ZNF827 in MDA-MD-231 cells led to a significant reduction in migration and invasion potential (Fig 4a-d). Monitoring of expression of key epithelial and mesenchymal genes in MDA-MB-231 cells upon ZNF827 depletion revealed changes towards the epithelial state-specific expression pattern (Fig 4e-f). MDA-MB-231 cells have been reported to form tumors and metastasize to distant organs in mouse models [332]. To investigate the role of ZNF827 in primary tumor growth and metastasis formation, we generated MDA-MB-231 cell lines (containing GFP and a luciferase reporter) that stably express control or ZNF827-targeting shRNA. These cells were injected into the mammary fat pad of Nude SCID Gamma (NSG) mice (n = 6), and tumor growth and metastasis formation of the xenograft was quantified at regular intervals for 48 days. Depletion of ZNF827 led to reduced primary tumor growth compared with control (Fig 4g). Migration of ZNF827-depleted tumor cells to distant organs was also impaired, as measured by *in vivo* luciferase activity (Fig 4h-i). These *in vivo* observations suggest a critical role for ZNF827 in promoting tumorigenicity and metastasis.

Prompted by our observations of a critical function of ZNF827 during EMT, to further validate these findings in human, we analyzed the expression of ZNF827 in non-invasive and invasive breast tumor samples (n = 20 each). In line with our previous observations, ZNF827 expression was significantly higher in the invasive tumors compared to the non-invasive tumors (Fig S6a). Large-scale analysis of several well-characterized human clinical expression datasets revealed a higher mutation rate of ZNF827 in a majority of human tumors (Fig S6b), while a few tumor types (including breast) also showed amplification. Taken together, these *in vivo* observations suggest a critical role for ZNF827 in promoting tumorigenicity and metastasis.

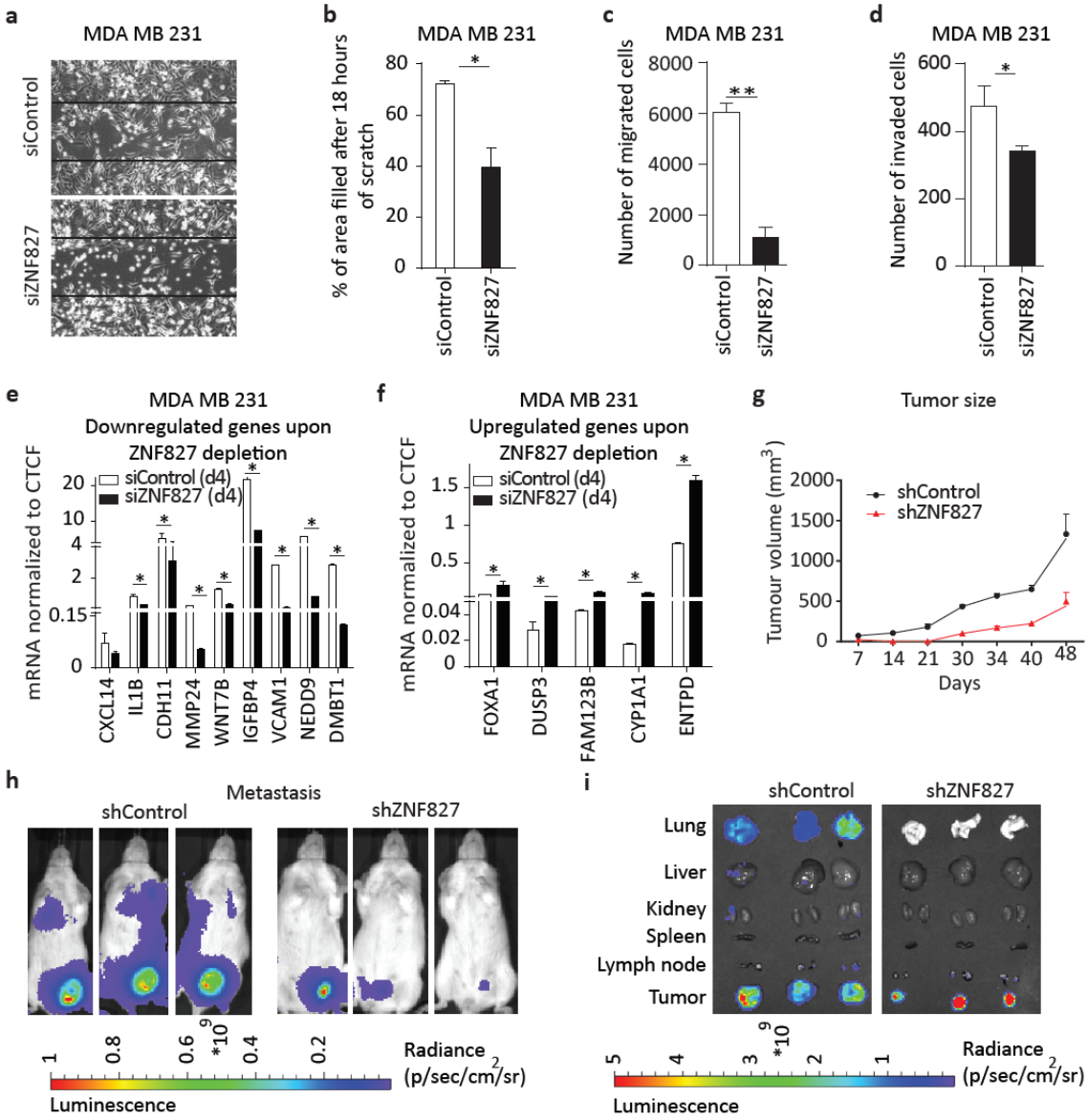


Fig 3. ZNF827 is crucial for tumorigenicity and metastasis *in vivo*.
(a-b) Scratch assay showing area filled after 18 hours of scratch by MDA MB-231 cells transfected with control or ZNF827-specific siRNA. **(c-d)** Bar graph showing results of migration (c) and invasion (d) assays in MDA-MB-231 cells transfected with either control siRNA or siRNA against ZNF827 for 4 days. **(e-f)** RT-PCR measurements of mRNA levels of key EMT relevant downregulated genes (e) and upregulated genes (f) in control and ZNF827 knockdown (4 days) MDA MB 231 cells. RNA levels are shown relative to CTCF mRNA levels. **(g)** Tumor size growth in mice (n=6 per condition) analyzed at various time points after mammary fat pad injection of MDA-MB-231 cells stably expressing either control shRNA or shRNA against ZNF827. The x-axis represents the days post-injection, and the y-axis represents the size of the tumors in mm^3. **(h)** Representative whole-body luciferase scanning showing primary and metastatic tumors in mice injected with MDA-MB-231 cells stably expressing either control shRNA or shRNA against ZNF827 (n=6 per condition), as in (g). **(i)** Various organs of the mice used in (h) scanned for luciferase to detect metastatic tumors. Error bars represent the SEM of three independent biological replicates. *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)

ZNF827 directly binds distinct genomic loci

To further dissect the molecular mechanisms underlying ZNF827 function during EMT, we investigated whether it binds distinct genomic regions. Towards this, we adopted HMLE cells (immortalized human mammary epithelial cells derived from HMEC) which have been routinely used as an EMT model system [281,333,334]. ZNF827 knockdown during EMT in these cells led to transcriptional changes similar to those observed in HMEC cells (Fig. S7, Fig. 2c, Table 4). We performed a chromatin immunoprecipitation (ChIP) assay for ZNF827 in HMLE cells undergoing TGF-β-induced EMT, followed by next-generation sequencing (ChIP-seq). A genome browser visualization of these data demonstrated ZNF827 enrichment at distinct loci in the genome (Fig 4a). Computational analysis revealed a preference for intronic regions (66%), followed by intergenic (11%), exonic (9%) and promoter (7%) regions (Fig 4b, table 3). Many of these targets were validated in independent ChIP-qPCR assays (Fig 4c). We next investigated whether the ZNF827-bound regions were misregulated in these cells following ZNF827 depletion. Comparison of transcriptionally deregulated genes upon ZNF827 depletion with ZNF827 bound loci revealed almost no overlap (Fig. 4d). These findings prompted us to look for mechanisms independent of direct transcriptional regulation that could explain ZNF827 function during EMT.

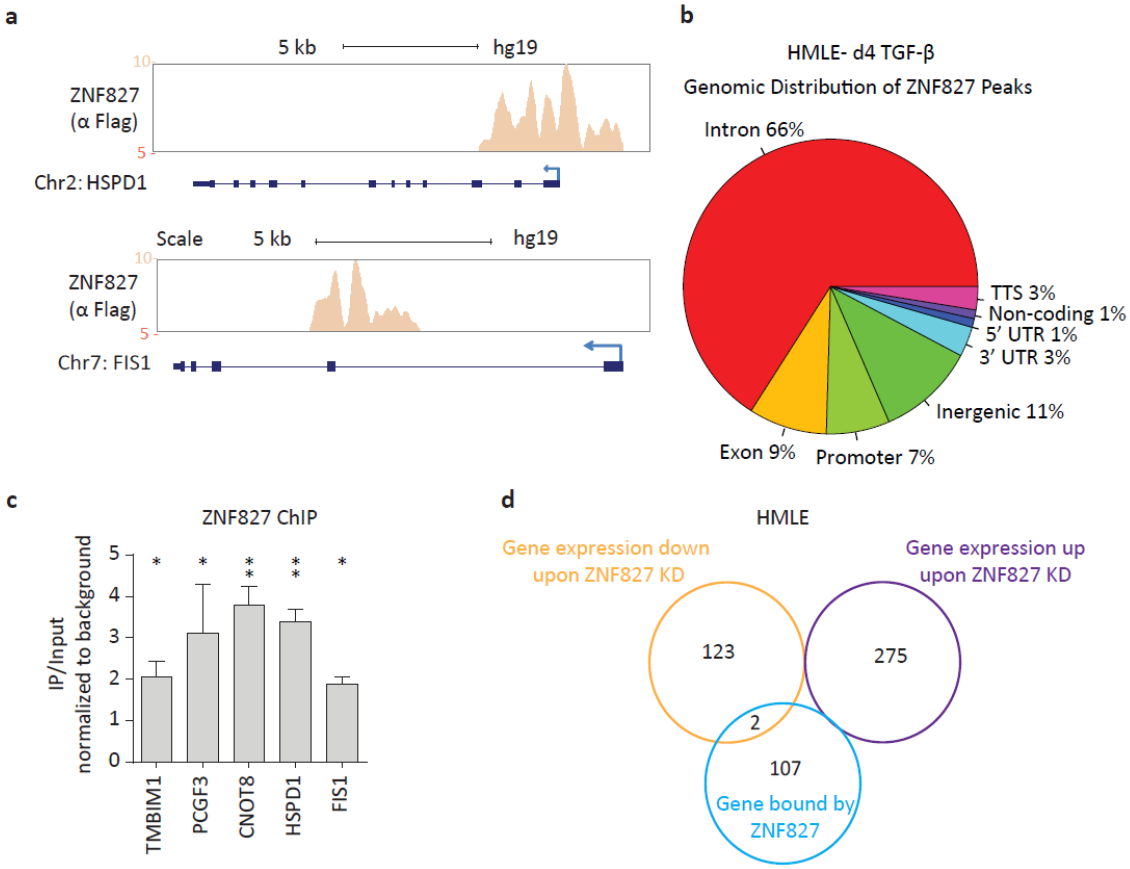


Fig 4. ZNF827 directly occupies distinct genomic loci.

(a) Genome browser tracks derived from ZNF827 ChIP-sequencing performed on Flag tagged ZNF827 showing binding of at promoters and intronic regions of genes in human primary epithelial cells undergoing TGF-β-induced EMT. **(b)** Pie chart representation of ChIP-seq results based on peaks obtained from Flag ChIP-seq, showing genomic distribution of ZNF827 bound regions. **(c)** Bar graph showing ChIP enrichment using an anti-Flag antibody to assess the binding of Flag-ZNF827 to its target region in HMLE cells induced with TGF-β for 4 days. Quantitative PCR of ChIP samples were performed for the indicated gene promoters or intronic region. The ratio of precipitated DNA (bound) to total input DNA was calculated, and plotted after normalization to an intergenic region as background. **(d)** A Venn diagram showing the overlap of ZNF827 bound region and the genes down- or up- regulated upon ZNF827 depletion in HMLE cell undergoing TGF-β-induced EMT for 4 days. Error bars represent the SEM of three independent biological replicates. (Results in this figure contributed by me)

EMT associated alternative transcriptome strongly relies on ZNF827

A recent study employed a high-throughput approach to reveal several novel splicing regulators [327]. Interestingly, C2H2 Zinc Finger proteins constituted a major fraction among these proteins [327]. Prompted by these findings, we next attempted to investigate whether ZNF827 plays a role in the alternate splicing program underlying EMT. A comprehensive analysis of the changes in alternative splicing that happen

during early EMT (day 4) by using our high-coverage, long reads and paired-end RNA-seq data identified a large set of high-confidence splicing events (n = 3415) comprising 1134 genes (Fig. 5a-b). This analysis specifically excluded the genes that were differentially expressed during EMT, and focused on genes that exclusively undergo transcript isoform changes using splicing with stringent criteria (delta PSI >0.2). The identified alternative splicing events included all known splicing types ((e.g. SE (Skipping Exon), MX (Mutually Exclusive Exon), RI (Retained intron), A5 (Alternate 5' Splice Site), A3 (Alternate 3' Splice Site), AF (Alternate First Exon) and AL (Alternate Last Exon)). We found that a large number of splicing events (n = 1974, 58% of all the EMT-dependent splicing events) were altered following knockdown of ZNF827 during TGF-β induced EMT at day 4 (Fig. 5b, Table 4). A substantial portion of EMT-driven alternative splicing events reverted to a normal epithelial state upon ZNF827 depletion (n = 1586 events in 924 genes; 80% of the ZNF827-dependent events) suggesting an important role for ZNF827 in the establishment of the splicing program linked to EMT (Fig 5b-c, S8a-f, Table 4). In support of a novel role of ZNF827 in EMT-specific splicing, several genes sensitive to ZNF827 are known to be important for EMT, including ABI2, CLSTN1, DKK3, ENAH, FLNB, ILK, KIF13A, PLOD2, PTPRF, and WASF1 [320]. Finally, we found that most of the splicing changes occurring during EMT and reverted upon ZNF827 knock down were exon skipping events, with no preference for inclusion or exclusion and alternative first exon, which goes very much in line with a co-transcriptional effect (Fig 5c-d, Table 4).

A Gene Ontology (GO) term analysis of alternatively spliced genes upon ZNF827 depletion showed enrichment for protein metabolism pathways (Fig 5e). This observation was novel and relevant given that significant changes in metabolism were noticed when cells change from an epithelial to a mesenchymal phenotype [335,336]. We further analyzed whether expression of any known splicing regulators was affected following ZNF827 knockdown. While the number of proteins known to be involved in splicing is large (n = 540), we found only five of these genes among those differentially expressed upon ZNF827 depletion and these are not known to be important for EMT (Fig 5f, Table 4). Nevertheless, we found a substantial

number of splicing factors (n = 52) and RNA binding proteins (n = 137) that relied on ZNF827 to undergo alternative splicing changes during EMT (Fig 5g). These results suggest that the global effects of ZNF827 on alternative splicing during EMT might result from both via direct regulation of splicing, as well as indirectly via modulation of splicing components.

We next investigated how ZNF827 binding at distinct genomic loci may be linked to the regulation of alternative splicing events. We compared genes bound by ZNF827 and those alternatively spliced upon ZNF827 depletion during EMT. Strikingly, this analysis revealed that almost all the genes bound by ZNF827 (n = 104 out of 109) undergo alternative splicing in a ZNF827-dependent manner, suggesting its direct role in splicing regulation (Fig 5h-i). We also independently validated ZNF827-dependent alternative splicing of selected bound genes (Fig S9a-b). Taken together, these findings imply that ZNF827 is a novel major regulator of the alternative splicing program during EMT that may involve its direct DNA binding function.

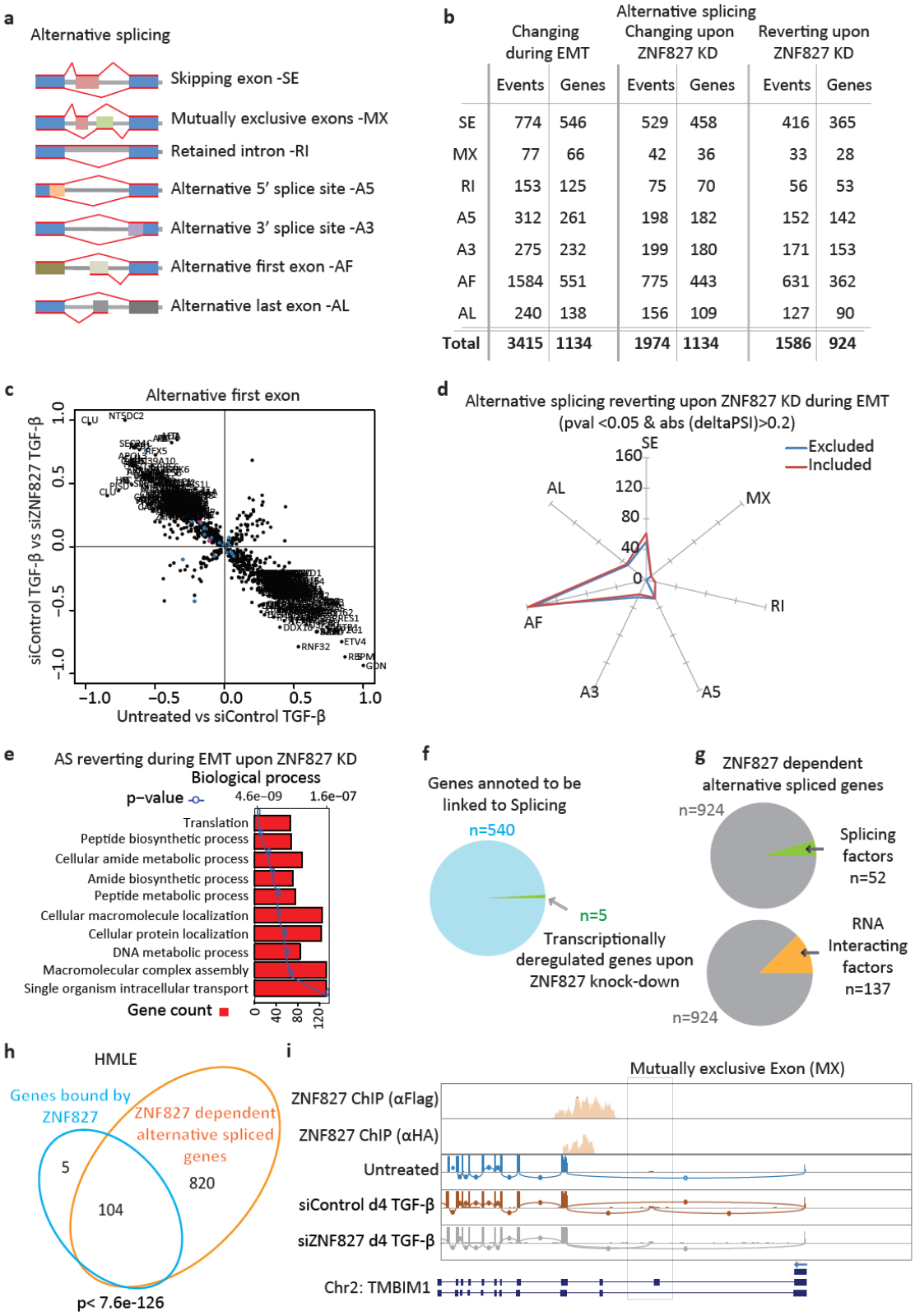


Fig 5. ZNF827 is required for hundreds of alternative splicing events within numerous genes during EMT. **(a)** Graphical representation of different alternative splicing events. **(b)** Table showing the number of alternative splicing events and the associated number of genes obtained from deep RNA-seq (long reads,

deep coverage, paired end): (1) during EMT, (2) dependent on ZNF827, and (3) dependent on ZNF827 during EMT and reverting to the original epithelial state. **(c)** Scatter plot showing alternative first exon events (AF) during EMT and reverting to epithelial state upon ZNF827 depletion. **(d)** Radar plot showing a comparison of number of inclusion or exclusion event of the EMT driven alternative splicing dependent on ZNF827. **(e)** GO analysis of genes undergoing alternative splicing and reverting to epithelial state upon depletion of ZNF827 during EMT. **(f)** Pie-chart showing overlap of genes annotated as splicing regulators and transcriptionally changing genes among the genes altered upon ZNF827 depletion during EMT. **(g)** Pie-chart showing number of splicing factors and RNA interacting factors, undergoing alternative splicing in ZNF827-dependent manner during EMT. **(h)** A Venn diagram showing the overlap of ZNF827 bound region and the genes undergoing alternative splicing in ZNF827 dependent manner in HMLE cell undergoing TGF-β induced EMT for 4 days. **(i)** Representative example showing browser track for ZNF827 binding and transcription dynamics of a mutually exclusive exon during EMT and upon ZNF827 depletion. (Results in this figure partially contributed by me)

ZNF827 employs epigenetic mechanisms to regulate alternative splicing

We were next interested to gain insights into the mechanisms through which ZNF827 regulates splicing during EMT. Towards this, we performed immunoprecipitation of ZNF827 in mesenchymal MDA-MB-231 cells after stable isotope labeling of amino acids in cell culture (SILAC), and subjected the samples to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Interestingly, the results of the quantitative proteomics studies revealed specific enrichment of splicing factors, RNA binding proteins and chromatin regulators in the ZNF827 immunoprecipitations compared to the control (Fig 6a-b). The identified proteins included core splicing components such as U2AF1/2 and chromatin remodelers e.g. HDAC1, suggesting potential cooperation among ZNF827, splicing machinery and epigenetic mechanisms.

Chromatin modifications such as histone acetylation have been shown to influence the splicing outcome by modulating the RNA Polymerase II elongation rate (the “kinetic model of co-transcriptional splicing”) [330,337-340]. This model suggests that slow elongation expands, and fast elongation compresses the ‘window of opportunity’ for recognition of upstream splice sites, thereby increasing or decreasing inclusion of alternative exons, respectively [338]. This prompted us to examine enrichment of HDAC1, active histone modification H3K27ac (acetylation of histone H3 at lysine 27), and RNA Polymerase II at the ZNF827 target regions (regions bound by ZNF827) following ZNF827 knockdown during EMT. Interestingly, ZNF827 depletion led to a significant decrease in the occupancy of HDAC1, accompanied by a significant increase

in the levels of H3K27ac, as well as Polymerase II occupancy at ZNF827 targets sites (Fig. 6c-d, S10a). Furthermore, a deeper analysis showed a ZNF827-dependent differential RNA Polymerase II processivity at its targets genes [341] (Fig. 6e-g, S10b-c). Given these evidences, we further analyzed the functional cooperativity between ZNF827 and HDAC machinery during EMT. Interestingly, chemical inhibition of HDAC activity during EMT phenocopied EMT reversal phenotype, that was seen upon depletion of ZNF827 (Figure S5c). Importantly further, this also mimicked aberrations in alternative splicing that were observed following ZNF827 depletion during EMT (Fig 6h, S10d-e). We next assessed whether ZNF827-dependent changes in epigenetic landscape influences the kinetics of recruitment of core splicing components. Interestingly, knockdown of ZNF827 resulted in a significant decrease in the binding of U2AF1 and 2 to target RNA loci (Fig. 6i-j). Taken together, these observations suggest that ZNF827 regulates alternative splicing events during EMT by modulating the epigenetic landscape and facilitating the recruitment of splicing factors (Fig 7).

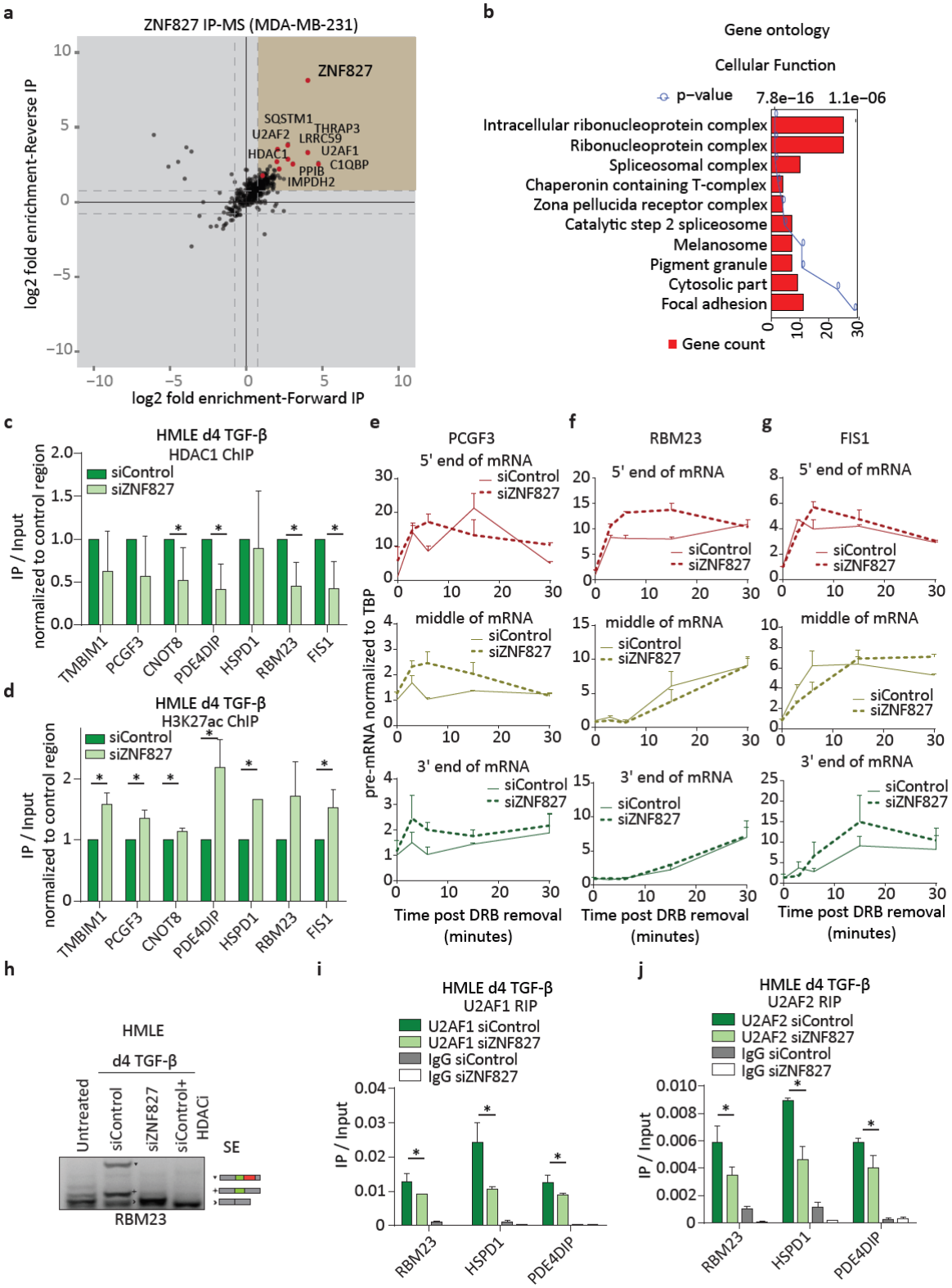


Fig 6. ZNF827 modulates epigenetic landscape and facilitates recruitment of core splicing machinery. **(a)** Scatterplot showing selected significantly enriched proteins in ZNF827 immuno-precipitated samples both in forward and reverse SILAC mass spectrometry performed in MDA-MB-231 cells. **(b)** Gene ontology analysis of the identified two fold enriched ZNF827 interacting partners. **(c-d)** Bar graph showing ChIP

enrichment for HDAC1 (c) and H3K27ac (d), on selected ZNF827 bound regions in HMLE cells undergoing TGF-β induced EMT and upon siRNA mediated depletion of ZNF827 for 4 days. q-PCR was performed for the indicated gene locus and enrichment was plotted on the y-axis as the ratio of precipitated DNA (bound) to total input DNA and normalized with ACTB upstream promoter region. **(e-g)** Kinetics of RNAPII-dependent transcription elongation were measured in cells treated with DRB, and post removal. Quantitative real-time RT-PCR was performed using primer sets specific for different parts of the indicated genes, PCGF3 (e), RBM23 (f), FIS1 (g) to measure the levels of pre-mRNA expression. Error bars indicates s.e.m. and primers were designed at Exon-Intron junction or Intron-Exon Junction of pre-mRNA. **(h)** Representative DNA gel showing PCR amplified SE changes in RBM23 during EMT, upon ZNF827 knockdown, and cells undergoing EMT when treated with the HDAC inhibitor Sodium-butyrate. **(i-j)** Bar graph showing RIP (RNA-immunoprecipitation) enrichment for U2AF1 (i), and U2AF2 (j) on selected regions in HMLE cells undergoing TGF-β induced EMT and upon siRNA mediated depletion of ZNF827 for 4 days. q-PCR was performed for the indicated gene at the RNA locus and enrichment was plotted on the y-axis as the ratio of precipitated RNA (bound) to total input RNA. Expression analysis indicate constant level of expression for tested genes in both conditions. Error bars represent the SEM of three independent biological replicates. *p<0.05, Student's t-test. (Results in this figure contributed by me)

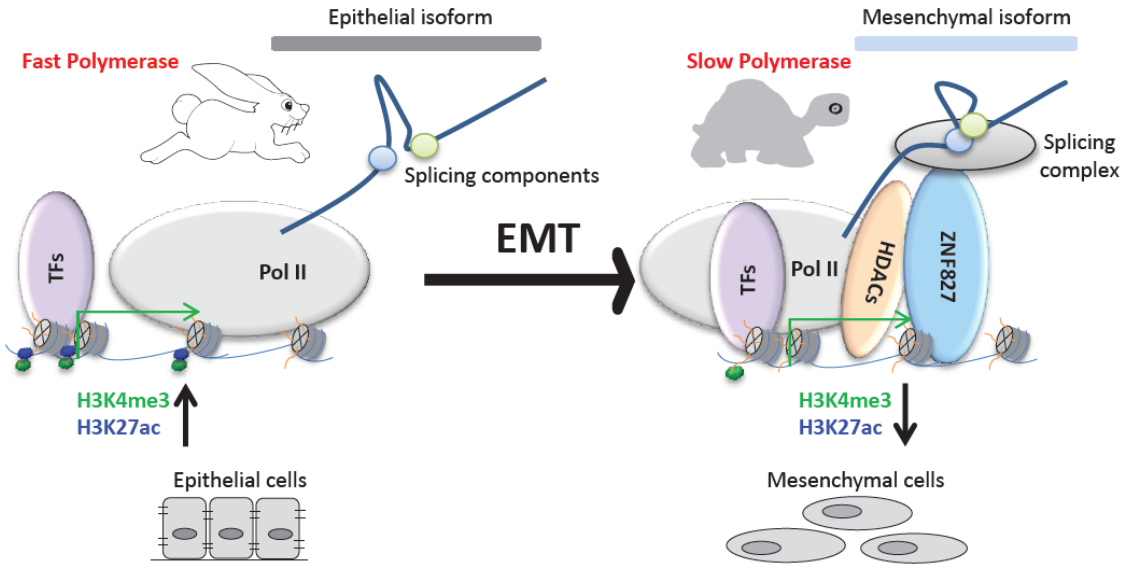


Fig 7. ZNF827 is critical regulator of alternative transcription program during EMT.
Schematic model showing mechanism of action of ZNF827 in regulation of alternative splicing program underlying EMT. Induction of EMT accompanies a significant upregulation of ZNF827. It functions in regulation of many EMT-related splicing events by directly targeting these gene loci and modulating their epigenetic landscape and RNA Pol II kinetics. Furthermore, ZNF827 also facilitates the recruitment of core splicing components to nascent RNA.

DISCUSSION

Our study identified ZNF827, a C2H2 Zinc finger protein, as a novel regulator of EMT in the context of both development and disease. We found ZNF827 to be robustly induced during various instances of EMT, including in primary breast epithelial cells *in vitro*, cancer metastasis, and neuronal migration during brain

development. ZNF827 loss-of-function strongly inhibits EMT progression *in vitro* and *in vivo.* Further molecular analysis revealed these effects to be linked with ZNF827-dependent regulation of a major set of alternative splicing events that accompany EMT. We show that ZNF827 mediates these responses by directly targeting these gene loci and modulating their epigenetic landscape and RNA Pol II kinetics as well as facilitating the recruitment of core splicing components. These findings establish that ZNF827 is a novel and major determinant of alternative splicing repertoire that underlies EMT during development and disease.

The only other study that has investigated ZNF827 function showed it to be a telomere maintenance factor, operating via recruitment of nucleosome remodeling NuRD complex (comprised of histone deacetylase) [328]. This group found ZNF827 activity to be restricted to a very small fraction of tumor types that lack enhanced telomerase activity [328]. In contrast, our study reveals a much larger, genome-wide function of ZNF827, in which it directly modulates hundreds of alternative splicing events within numerous genes during EMT. Mechanistically, ZNF827 interacts with epigenetic (e.g. HDAC1) and splicing (e.g. U2AF1/2) machinery and assists in their recruitment to the target chromatin and RNA sites respectively. ZNF827-dependent targeting of HDAC1 modulates the epigenetic landscape of the target sites, which in turn influences RNA Polymerase II kinetics. Altogether, these ZNF827 actions dictate the alternative splicing events underlying EMT. We suggest further investigation into the function of this C2H2 Zinc finger protein in distinct cell types during development and disease that has been largely unexplored so far.

Recent advances have led to an increased appreciation of alternative splicing events in cell-fate determination and function [314,342]. Almost all human genes (90%) are known to undergo alternative splicing to permit fine tuning of various biological processes. It is estimated that alternative splicing of selected exons and introns may yield at least 100,000 different proteins [343]. Our comprehensive analysis during EMT identified revealed a large number of such events (n = 3415) that involved 1134 genes. This

list included a majority of known EMT-associated alternative splicing events, while identifying a large number of novel set of events that should be further investigated.

Given the importance of alternative splicing repertoire in defining cell identify and function, investigators have put enormous efforts in recent years to identify molecular regulators of this process. Specific classes of zinc finger proteins possessing CCCH-type motifs, such as U2AF1 and MBNL proteins, are known to be critical for alternative splicing regulation [344,345]. However, while the family of C2H2 zinc finger proteins (n= 718 in humans) constitutes the largest class of DNA-binding proteins, only a small number of these proteins (e.g. TFIIIA, ZNF385a, CTCF, YY1, WT1) have been implicated in regulation of alternative splicing. A large-scale study identified C2H2 zinc finger proteins among the potent regulators of alternative splicing [327]. Our study, by identifying and comprehensively characterizing ZNF827 as a major regulator of alternative splicing during EMT, extends these observations and provides further impetus for intensive investigation into the function of C2H2 zinc finger proteins.

Epigenetic regulation and gene transcription play critical roles in regulating alternative splicing of nascent transcripts by various mechanisms, including recruitment of splicing components and modulation of polymerase elongation rate [338]. For example, a differential rate of RNA Polymerase II movement alters the kinetics of exposure of splicing components to splice sites in nascent transcripts, thereby causing differential splicing [315]. Here we show that ZNF827 directly targets many genomic loci, and recruits chromatin modifiers such as HDAC1. In so doing, ZNF827 facilitates a mild repressive chromatin environment, that in turn reduces RNA Polymerase II processivity, thereby regulating splicing. In further support of these findings, blockade of HDAC activity during EMT could mimic a EMT reversal phenotype as well as aberrations in alternative splicing that was observed following ZNF827 depletion. HDAC1 is known to regulate alternative splicing of distinct genes during EMT [337]. Our study extends this HDAC1 function to a large number of genes and further implicates ZNF827 as a DNA binding protein that recruits HDAC1 during EMT to wire alternative transcriptome.

Among the genes regulated at the alternative splicing level by ZNF827 during EMT, we found strong enrichment for protein and nucleic acid biosynthetic process-relevant genes. This observation is highly relevant given that several recent studies have implied a critical role of metabolites and the underlying regulatory networks during EMT [335,336,346-348]. Despite these recent studies, only a few of these metabolic regulatory components have been explored in the context of EMT [348]. Our findings show a massive reorganization in the alternative transcript repertoire of a large number of these genes during EMT. These data suggest a potentially critical function for many of these gene products in cell-fate transition including EMT.

In addition to these important biological findings, our study also provides several data resources for the scientific community. For example, for the first-time, we provide a comprehensive, deep transcriptome time-course in biological triplicates during TGF-β-induced EMT in human mammary epithelial cells, as well as the kinetics of C2H2 zinc finger proteins at various stages of this process. In addition, our long-read, paired-end, and high-coverage data allow us to perform a comprehensive analysis of all alternative splicing events during TGF-β-induced EMT, and to specifically reveal those events that rely on ZNF827.

Taken together, our findings provide the first report on ZNF827 as a highly potent regulator of EMT *in vitro* and during neuronal migration and tumor metastasis *in vivo*. We show that ZNF827 plays a critical role in the progression and maintenance of mesenchymal identity by regulating the alternative splicing of a large number of critical protein metabolic pathway genes. Furthermore, ZNF827 mediates these responses by directly binding and modulating the epigenetic landscape of these gene loci and facilitating the recruitment of core splicing machinery. These findings establish ZNF827 as a novel and important regulator of the alternative transcriptome that underlies epithelial to mesenchymal transition during development and disease. These exciting findings further warrant an investigation into the function of this largely non-explored C2H2 Zinc finger protein in various contexts of development and disease.

METHODS

Cell culture

HMECs were obtained from Lonza and cultured according to the manufacturer's guidelines. HMLE cells, a kind gift from Christina Scheel (Helmholtz Zentrum, Munich), were cultured in Lonza primary cell culture medium. A previously described sub-clone of NMuMG cells was grown in DMEM supplemented with 10% FBS, 2 mM L-glutamine and 1x non-essential amino acids (Maeda et al. 2005). Other cell lines were obtained from ATCC, and were cultured in the following media: MDA-MB-231: DMEM, 10% FBS; MDA-MB-361: DMEM, 20% FBS; BT549: RPMI-1640, 10% FBS, 0.001 mg/ml bovine insulin; MCF7: DMEM, 10% FBS, 0.01 mg/ml bovine insulin; HEK293T: same medium as NMuMG. All cells were cultured at 37°C with 7% CO_2 in a humid incubator. For TGF-β time-course experiments, NMuMG, HMEC, and HMLE cells were treated with 2, 5, and 5 ng/ml TGF-β, respectively, for the indicated times. TGF-β was replenished and the medium was changed every 2 days.

siRNA-mediated knockdown

For all siRNA-mediated knockdown experiments, cells were seeded at the same starting density and transfected every second day with ON-TARGET plus single siRNAs or SMART pool siRNAs (i.e., a mixture of 4 siRNAs provided as a single reagent) (Dharmacon). For siRNA transfections, Lipofectamine RNAiMAX (Invitrogen, 13778-150) was used according to the manufacturer's instructions. For experiments involving TGF-β-induced EMT, TGF-β induction was performed at the same time that siRNA was added to avoid indirect effects due to loss of protein function. The sequences of the siRNAs are displayed in Supplemental Table 3.

Immunofluorescence assay

Cells were grown on coverslips, fixed with 4% paraformaldehyde in PBS and permeabilized with 0.2% Triton X-100 for 15 minutes at room temperature. Subsequently, the cells were blocked with 10% goat

serum, 5% FCS and 0.5% BSA in PBS for 20 minutes and were incubated with primary antibodies at 4°C overnight. The cells were then incubated with fluorochrome-labeled secondary antibody or Phalloidin-633 for 1 hour at room temperature. The coverslips were counterstained with Hoechst, mounted with immu-mount and imaged using a confocal laser-scanning microscope. The data were processed by ImageJ software.

Immunofluorescence assay to assess neuronal migration in mouse brain

Isolated E16.5 embryonic brains were immediately fixed for 6 hours in 4% PFA in PBS at 4°C. The brains were then cryoprotected in 10% sucrose for 2 hours and then in 30% sucrose (in PBS) overnight, embedded in Tissue-Tek, stored at -20°C and cryo-sectioned at 12 µm. Sections on coverslips were preblocked with 2% BSA and 0.5% Triton (in PBS) for 1 hour. Primary antibodies (anti RFP or anti-GFP (chicken, Aves Labs, 1:1000)) were applied in blocking solution overnight at 4°C. Fluorescent secondary antibodies were applied according to the manufacturer's protocol (Life Technologies). The coverslips were counterstained with Hoechst, mounted with immomount and imaged using a confocal laser-scanning microscope (Leica SP5). The data were processed with ImageJ software.

***In vivo* tumorigenicity assays and imaging**

NOD *Scid* Gamma (NSG) mice were bred and maintained under SPF conditions in the Translational Animal Research Center at the University Medical Center Mainz. Briefly, cells were counted and suspended in a 1:1 (v/v) mixture of PBS and Matrigel (BD Biosciences). Ten-week-old female mice were injected unilaterally with 2×10^6 cells in 100 µl of 50:50 Matrigel/PBS into the fourth abdominal fat pad via subcutaneous injection at the base of the nipple. Tumor growth was monitored externally using Vernier calipers for up to 48 days. The tumor volume was calculated as follows: tumor volume (mm^3) = length × $(\text{width})^2 \times 0.5$. Necropsies were performed to identify macro-metastases. Primary tumors and organs were immediately frozen in liquid nitrogen and stored at -80°C until use.

In vivo bioluminescence imaging of the tumor-bearing mice and their organs was performed using an IVIS Lumina imaging system (Perkin Elmer). Briefly, mice were anesthetized with isoflurane and an aqueous solution of D-luciferin-K^+ salt (Perkin Elmer) (150 mg/kg body weight) was injected intraperitoneally. Five minutes after the injection, the mouse was placed onto the imaging chamber of IVIS, and photons acquired with an integration time of 10 second were presented as color-scaled images using IVIS Living Image Software (version 4.3.1, Perkin Elmer). For organ imaging, mice were sacrificed after luciferin injection, and the dissected organs were imaged as described above.

Animal maintenance and experiments were performed under an approved protocol in accordance with the animal care guidelines of Johannes Gutenberg University.

RNA-seq and data analysis

From HMLE, total RNA was extracted with Invitrogen PureLink RNA Mini Kit. Extracted RNA was quantified at OD260 nm with a ND-1000 spectrophotometer (Nanodrop Technology) and quality of the RNA was assessed with the RNA 6000 LabChip kit using a Bioanalyzer 2100 (Agilent Technologies). Total RNA Library Preparation was done with Illumina TruSeq Strand-Specific Ribo-Zero 'Gold' Depletion kit as per manufacture's instruction. Sequencing was performed with Illumina, NextSeq500 for Paired-End, 150-cycle. RNA-seq data were generated using Illumina sequencing. The reads were aligned to the mouse genome (mm9) or human genome (hg38) using TopHat (version 2.0.9) with the default options. After library size normalization using DESeq (Anders and Huber 2010), expression was quantified and expressed in reads/kilobase of transcript per million mapped reads (RPKM) using cufflink (version 2.1.1) (Trapnell et al. 2010). Differential expression analysis was performed using the DESeq package with an FDR cutoff of 0.1 (Anders and Huber 2010). Gene set enrichment analysis was performed using the GSEA Preranked module [307]. Detailed explanation for splicing analysis is provided in the Supplementary material.

ChIP assay

Cells were cross-linked in medium containing 1% formaldehyde for 10 minutes at room temperature, neutralized with 0.125 M glycine, scraped, and rinsed twice with PBS. The pellets were suspended in buffer L1 (50 mM Hepes KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% glycerol, 5% NP-40, and 0.25% Triton-X 100) and incubated for 10 minutes at 4°C. The cells were then suspended in buffer L2 (200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0) for 10 minutes at room temperature. Finally, the pellet was suspended in buffer L3 (1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, 100 mM NaCl, 0.1% Na-deoxycholate, 0.17 mM N-lauroyl sarcosine) containing protease inhibitors and was incubated at 4°C for 3 hours following sonication using Bioruptor plus (Diagenode). Sixty micrograms of chromatin was incubated overnight at 4°C with 2 µg of the antibodies targeting HDAC1, H3K27ac (Abcam) or Polymerase II (Santa Cruz) and then incubated with preblocked beads for 4 hours. Finally, the beads were washed twice with L3 and once with 1 ml of DOC buffer (10 mM Tris (pH 8.0), 0.25 M LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA), and the bound chromatin was eluted in 1% SDS/0.1 M NaHCO3. This followed treatment with RNase A (0.2 mg/ml) for 30 minutes at 37°C and then with proteinase K (50 µg/ml) for 2.5 hours at 55°C. The crosslinking was reversed at 65°C overnight with gentle shaking. The DNA was purified by phenol-chloroform extraction followed by ethanol precipitation and was recovered in TE buffer.

ChIP-seq data analysis

ChIP-seq reads were mapped to the reference genome sequence (hg19) using bowtie (Langmead, 2010) with parameters (-v 2 -best-strata m 1), keeping the best unique matches with at most two mismatches. Using Pyicos [349], all reads were extended to 200nt in the 5' to 3' direction and reads overlapping centromeres, gaps, satellites, peri-centromeric regions and low mappability regions were excluded. Finally, clusters were built from reads that overlap each other in genomic coordinates using Pyicos. Singletons (clusters with one read) were discarded. Significant ChIP-seq clusters were filtered based on a Poisson distribution using the clusters height from Pyicos and with a significant p-value < 0.01. The signal

was corrected to Hg19 genome size. Final selected clusters were recovered from the intersection of the different samples using bedtools2 [350] to remove possible false positives. Selected clusters were intersected with previously defined ZNF827 regulated alternative splicing events, using Bedtools2, recovering the coordinates of the alternative splicing event. Finally, the candidate clusters were ranked based on the summit. For visualization purposes, all cluster files were converted to bedgraph and normalized to Hg19 genome size using Deeptools 2 [351] bamCoverage. Genomic distribution of ZNF827 clusters was calculated using HOMER [185] *annotatePeaks.pl with Human (Hg19) genome.*

Mass spectrometry (MS) analysis

Peptide fractions were analyzed using a quadrupole Orbitrap mass spectrometer (Q Exactive Plus, Thermo Scientific) equipped with a UHPLC system (EASY-nLC 1000, Thermo Scientific), as described [302]. Peptide samples were loaded onto C18 reversed-phase columns and eluted for 2 hours with a linear gradient of acetonitrile from 8 to 40% containing 0.1% formic acid. The mass spectrometer was operated in data-dependent mode with automatic switching between MS and MS2 acquisition. Survey full scan MS spectra (m/z 300 – 1650) were acquired in the Orbitrap. The 10 most intense ions were sequentially isolated and fragmented by higher-energy C-trap dissociation (HCD) [303]. Peptides with unassigned charge states or charge states less than +2 were excluded from fragmentation. The fragment spectra were acquired in the Orbitrap mass analyzer.

Peptide identification: raw data files were analyzed using MaxQuant (development version 1.5.2.8) [304]. Parent ion and MS2 spectra were searched against a database containing 88,473 human protein sequences obtained from the UniProtKB released in December 2013 using the Andromeda search engine [305]. The spectra were searched with a mass tolerance of 6 ppm in MS mode, 20 ppm in HCD MS2 mode, strict trypsin specificity and allowing up to 3 miscleavages. Cysteine carbamidomethylation was searched as a fixed modification, whereas protein N-terminal acetylation and methionine oxidation were searched

as variable modifications. The dataset was filtered based on posterior error probability (PEP) to arrive at a false discovery rate below 1%, estimated using a target-decoy approach [306].

Alternative splicing RNA-seq analysis

A high coverage RNA-seq data set with Stranded paired-end, long reads (100 base pairs), average read depth of 150 million reads were generated from three biological replicates for alternative splicing analysis. RNA-Seq of HMLE cells (untreated), and day4 TGF-β-induced and transfected with either control siRNA or siZNF827 samples were mapped to human reference (Hg19) genome using STAR [352] version 2.3 with parameters (--outFilterMismatchNmax 2 --clip3pAdapterSeq AAAAAA –outSJfilterReads Unique --alignSJDBoverhangMin 3 --alignSJoverhangMin 5 --outSJfilterOverhangMin 30 12 12 12 --outSJfilterCountUniqueMin 5 1 1 1 --outSJfilterIntronMaxVsReadN 50000 100000 200000 --sjdbScore 2 --outFilterType BySJout --outSAMstrandField intronMotif --seedSearchStartLmax 50). The junction file used for STAR index genome contained the intron coordinates generated from the UCSC table browser. The aligned reads files were converted and processed to bam files using Samtools version 1.1 [353]. The signal was normalized to the human genome using bamcoverage from Deeptools version 2.0 [351] . Bam files were split into forward and reverse strand for visualization purposes. Transcripts were quantified using Salmon 0.7.2 [354], using k-mer length 31 and H sapiens GRCh37 ENSG75 annotation. Normalized TPM values were used as input for SUPPA 2.0. Differential splicing was calculated with SUPPA2, for all the comparisons between the three conditions, using three replicates for calculation and statistics. Changes in any type of alternative splicing from expressed genes were selected when a p-value < 0.05 between two conditions was found. DESeq2 [355] package was used for normalization and posterior calculation of differential expression. Differentially spliced events from both differentially expressed and non-differentially expressed genes were recovered. The events that significantly alter splicing (delta PSI > 0.2 and delta PSI Fisher exact test < 0.05) from untreated to TGF-β-induced siControl were assigned as EMT-dependent events. The events that significantly change splicing (delta PSI > 0.2 and delta PSI Fisher exact test < 0.05)

from TGF-β-induced siControl to TGF-β-induced siZNF827 were assigned as ZNF827 regulated events. Finally, to identify all EMT-splicing events dependent on ZNF827 (called reverting events), we selected the splicing events that changed in TGF-β-induced siControl compared to TGF-β-induced siZNF827 (delta PSI > 0.2 and delta PSI Fisher exact test < 0.05), but in the opposite direction from the changes observed in untreated vs. TGF-β-induced.

The splicing regulators list was retrieved from Biomart [356]and included all the ENSEMBL gene IDs classified as "RNA_splicing, RNA_cis_splicing, RNA-splicing or RNA_binding", with a function in the nucleus [357]. The list was curated to a more restrictive list based on literature.

RNA Pol II kinetics assay

We grew cells on 60-mm plates for TGF-β induction and siRNA mediated depletion and on day 3 (70–80% confluency) treated them with 100 uM 5,6-dichlorobenzimidazole 1-β-d-ribofuranoside, DRB (Sigma) in culture medium for 3 hours. We washed the cells twice with PBS to remove the DRB and then incubated them in fresh medium for various time periods. Following the incubation period, we lysed the cells directly and isolated the total RNA using a High Pure RNA isolation kit (Roche), as per the manufacturer's instructions.

DATA ACCESS

The RNA-seq data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE99722.

SUPPLEMENTAL INFORMATION

This manuscript includes seven supplemental figures and one supplementary table.

AUTHOR CONTRIBUTIONS

S.K.S. designed and performed experiments, analyzed data and wrote the manuscript. E. A. performed computational analysis. S.S. and M.D. provided help with *in vivo* tumorigenicity assays and imaging. R. L. provided critical input regarding alternative splicing analysis. V.T. designed the study, analyzed data and wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to thank the members of the Tiwari lab for their cooperation and critical feedback over the course of this study. The support from the Core Facilities of the Institute of Molecular Biology (IMB), Mainz, is gratefully acknowledged, especially the microscopy, genomics and bioinformatics core facilities. This study was supported by the Wilhelm Sander Stiftung 2012.009.1 and 2012.009.2 to V.K.T. and the Marie Curie (PCIG12-GA-2012-334000) and ATIP/AVENIR program to R.F.L.

CONFLICTS OF INTERESTS

The authors declare no competing financial interests.

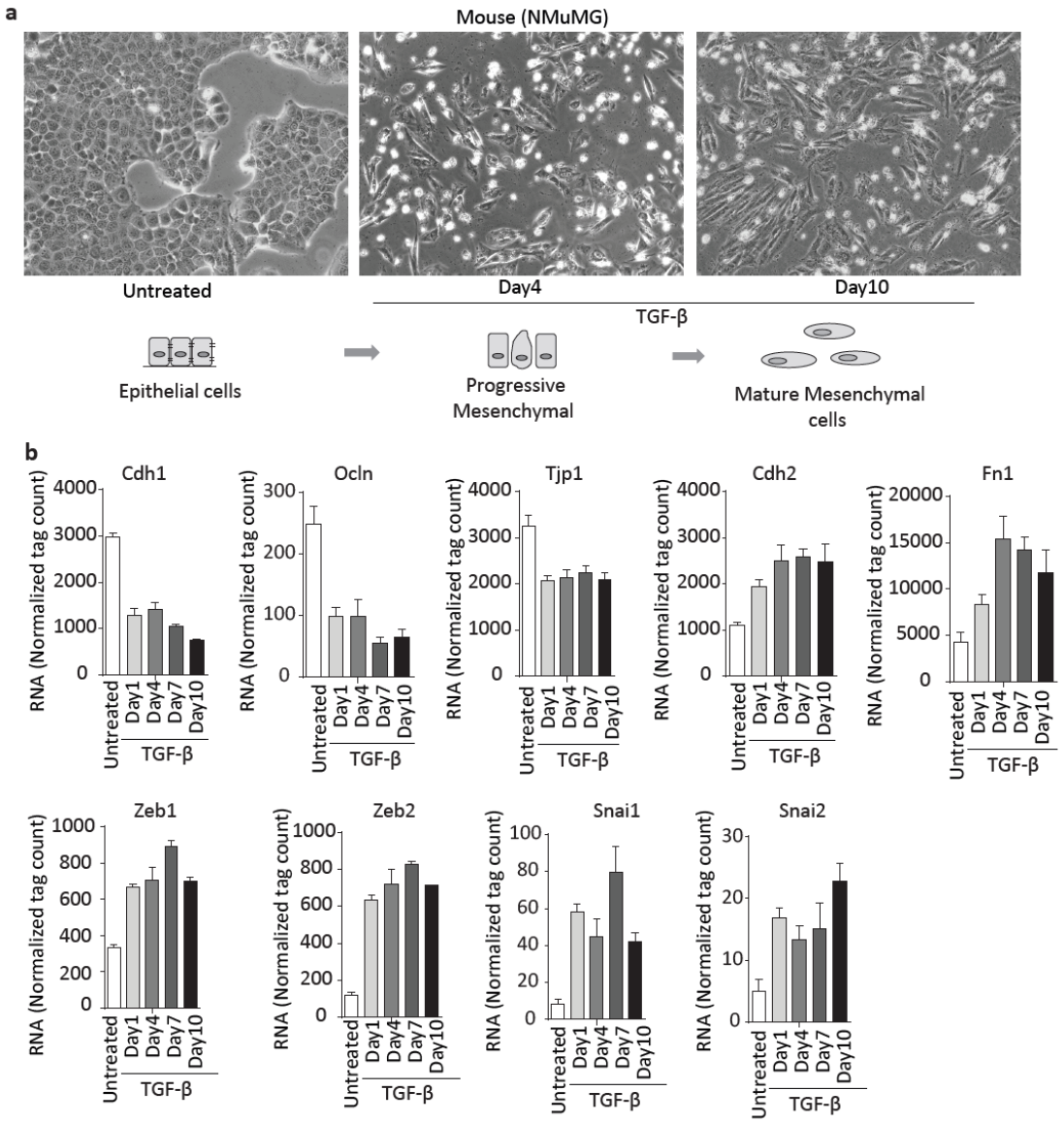


Fig S1. Model systems for EMT (Mouse).

(a) Representative bright field images of NMuMG cells undergoing TGF-β induced EMT. **(b)** RNA-seq data showing mRNA level of key EMT relevant genes in NMuMG cells undergoing EMT, the y-axis represents the normalized tag count. (Results in this figure contributed by me)

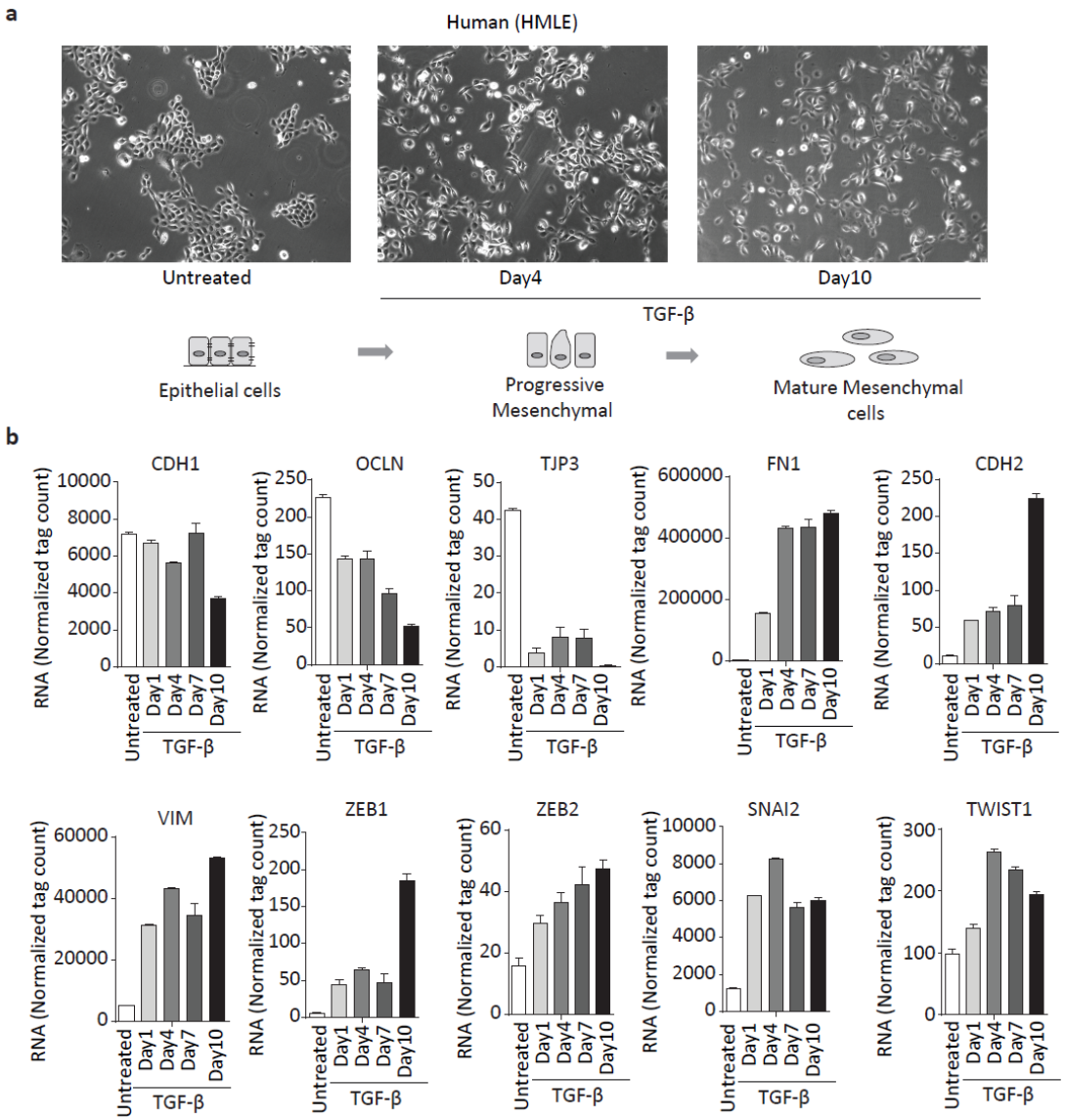


Fig S2. Model systems for EMT (Human).

(a) Representative bright field images of HMLE cells undergoing TGF-β induced EMT. **(b)** RNA-seq data showing mRNA level of key EMT relevant genes in HMLE cells undergoing EMT, the y-axis represents the normalized tag count. (Results in this figure contributed by me)

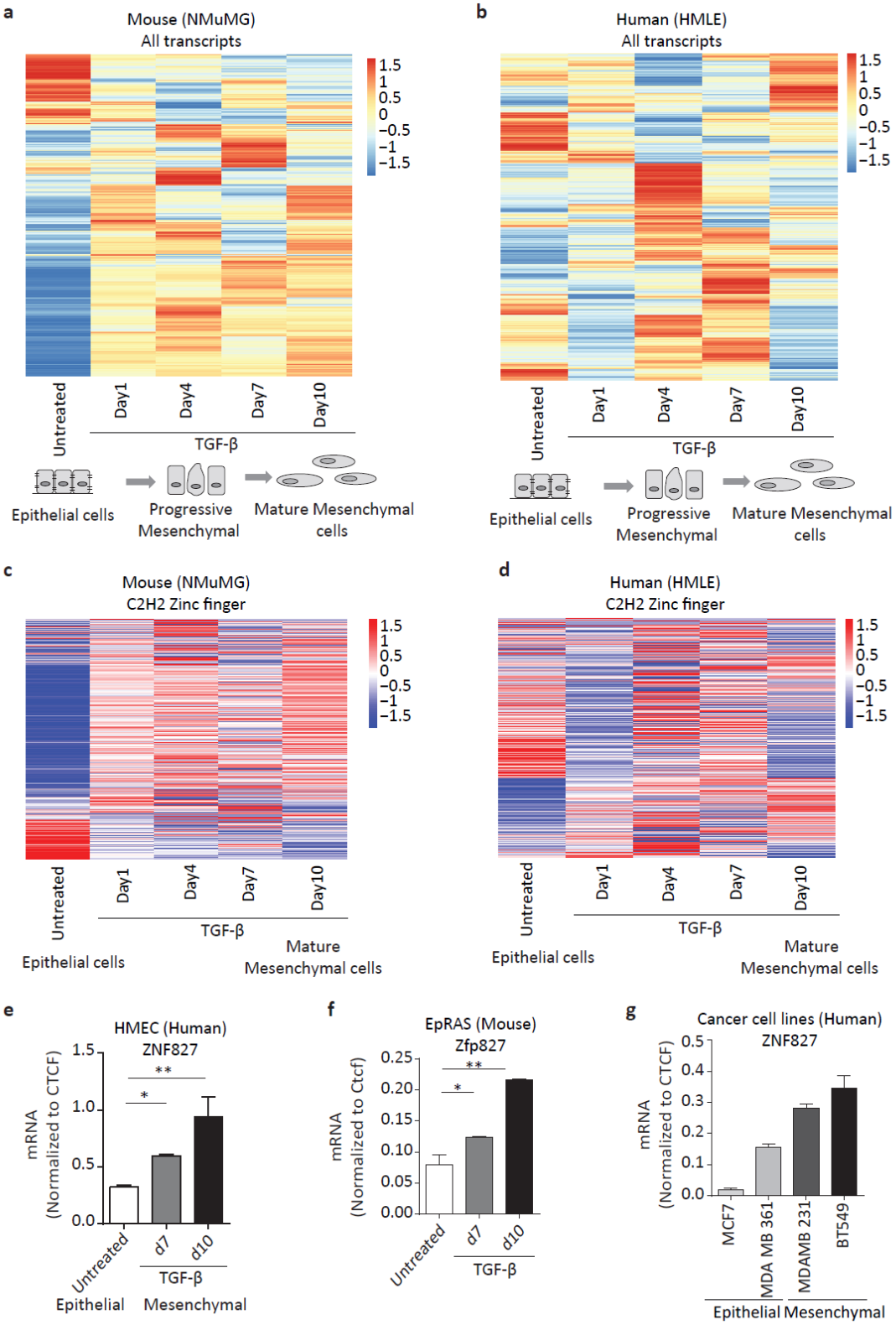


Fig S3. Dynamics of expression of zinc finger proteins during EMT.

(a-b) Heat map showing standardized expression (Z-scores) of all transcripts at various stages during TGF-β-induced EMT in mouse mammary epithelial cells (NMuMG) (a) and human mammary epithelial cells

(HMLE) (b). **(c-d)** Similar analysis as in (a-b), but focused on all C2H2 zinc finger-containing genes at various stages during TGF-β-induced EMT in NMuMG (c) and HMLE (d). **(e)** ZNF827 mRNA levels were measured by RT-PCR in primary human breast epithelial cells (HMEC) undergoing TGF-β-induced EMT. **(f)** Using RT-PCR, the levels of Zfp827 mRNA in EpRAS cells during TGF-β-induced EMT were measured relative to Ctcf and plotted on the y-axis. **(g)** ZNF827 mRNA level was measured by RT-PCR in two human breast cancer epithelial cell lines (MCF7 and MDA-MB-361) and two mesenchymal cell lines (MDA-MB-231 and BT547). Error bars represent the SEM of three independent biological replicates. *p<0.05, **p<0.01, ***p<0.001, Student's t-test. (Results in this figure contributed by me)

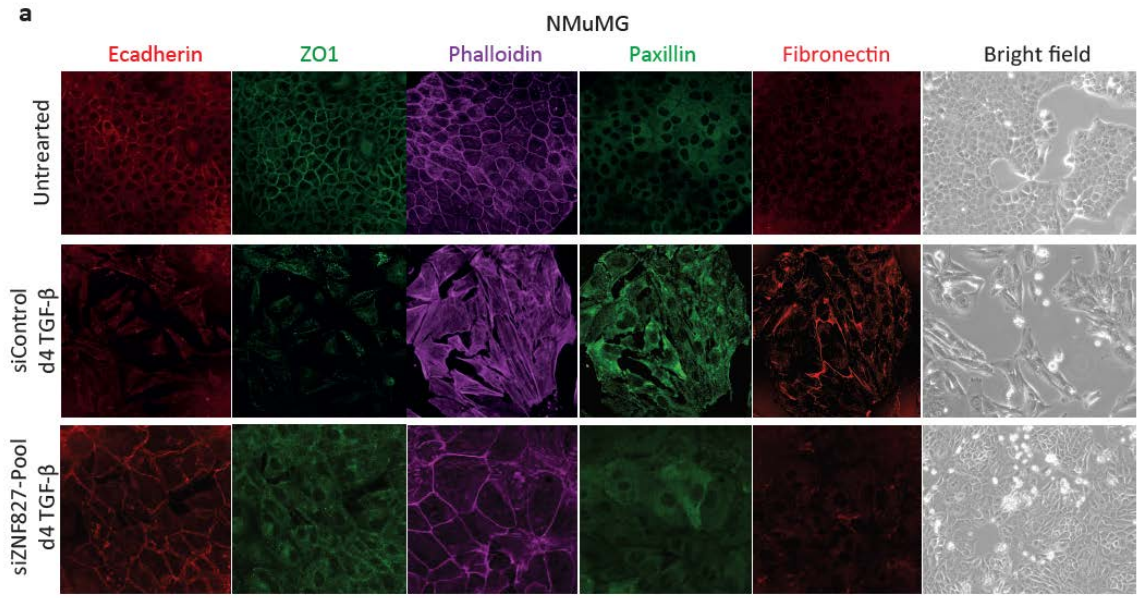


Fig S4. ZNF827 is required for phenotypic and molecular alteration during EMT.
Representative bright-field and immuno-fluorescence images showing the localization and expression levels of EMT marker proteins after 4 days (d4) of TGF-β-induced EMT and siRNA-mediated depletion of ZNF827 compared to non-targeting control (siControl) in NMuMG cells. Staining was performed to assess the expression with antibodies against the epithelial markers E-cadherin and ZO1, the mesenchymal marker Fibronectin-1, Phalloidin (to visualize the actin cytoskeleton) and Paxillin (to detect focal adhesion plaques). Scale bar: 100 µm. **(b)** Volcano plot- representation of RNA-seq results shows the number of genes with significantly altered expression levels (1.5 fold), upon ZNF827 depletion in HMECs undergoing TGF-β-induced EMT for 4 days. (Results in this figure contributed by me)

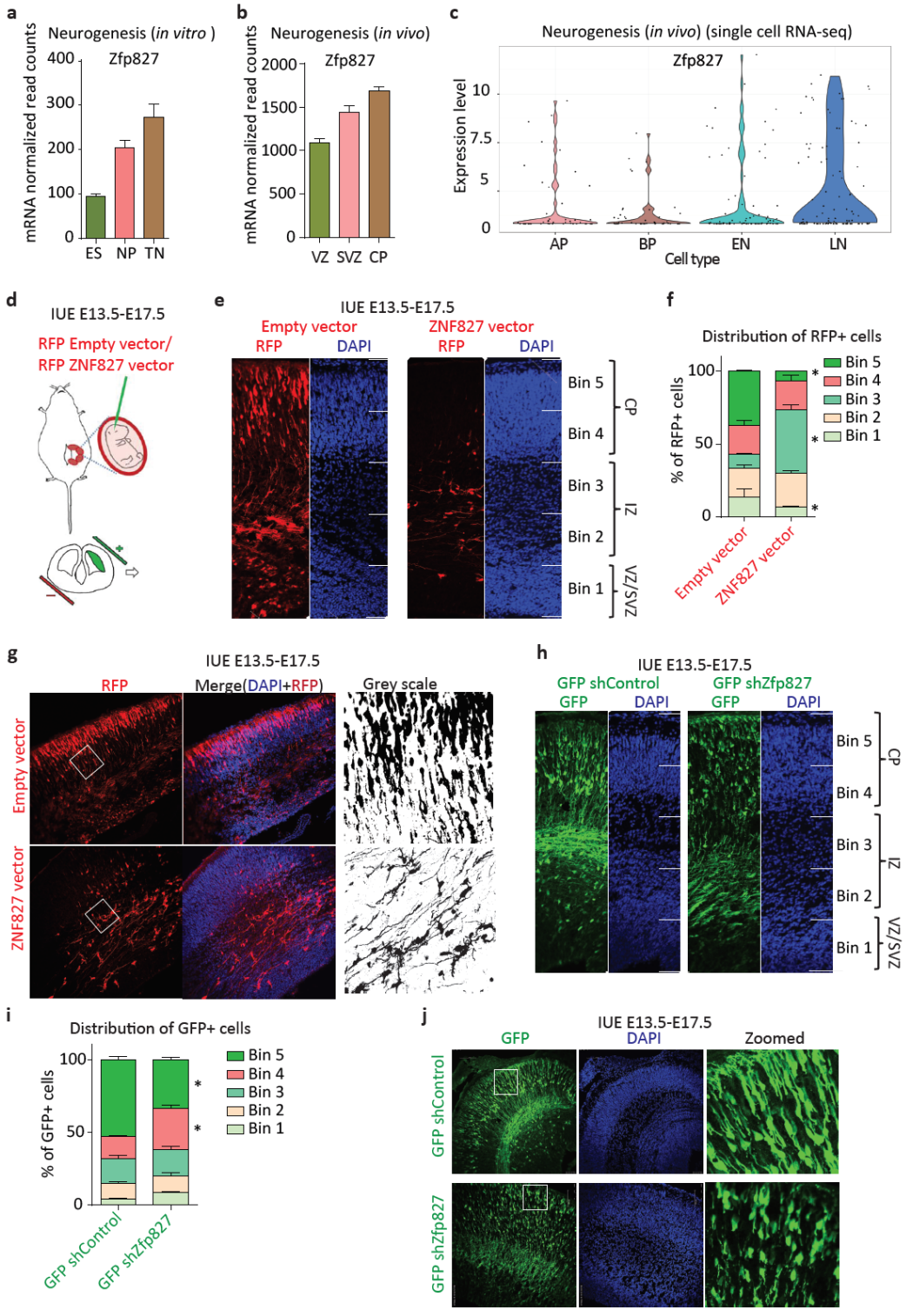


Fig S5. ZNF827 regulates neuronal EMT.

(a) Using RT-PCR, the levels of Zfp827 mRNA in embryonic stem cells (ES), neural progenitors (NP) and terminally differentiated neurons (TN) were measured relative to Ctcf and plotted on the y-axis. **(b)** The levels of Zfp827 mRNA are shown as the average normalized tag counts derived from the RNA-seq data obtained from the ventricular zone (VZ), the sub-ventricular zone (SVZ), and the cortical plate (CP) of E14.5

mouse cortex, in normalized tag counts **(c)** Expression pattern of Zfp827 in publically available single cell RNA-sequencing data derived from mouse cortex. **(d)** Graphical representation of *in utero electroporation* (IUE). **(e)** *IUE* was performed at E13.5 using plasmids containing only RFP or RFP with Zfp827 overexpression, and the mice were sacrificed at E17.5 for further analyses. A representative image of the immunofluorescence analysis performed with anti-RFP showing the retention of RFP-positive cells at various bins in the cortex electroporated with control RFP vector as compared to Zfp827 overexpression. **(f)** A bar plot showing quantification of the migrated RFP-positive cells in control and Zfp827 overexpressing mouse brains at various bins. The y-axis represents the percentage of cells at various bins. **(g)** An enlarged immunofluorescence image showing morphological difference of the RFP positive control cell and cell overexpression ZNF827. **(h-j)** Similar analysis as in e-g, but with plasmids containing GFP and with control shRNA or shRNA against Zfp827. The error bars represent the SEM of three independent biological replicates. *p<0.05, Student's t-test. (Results in this figure contributed by me)

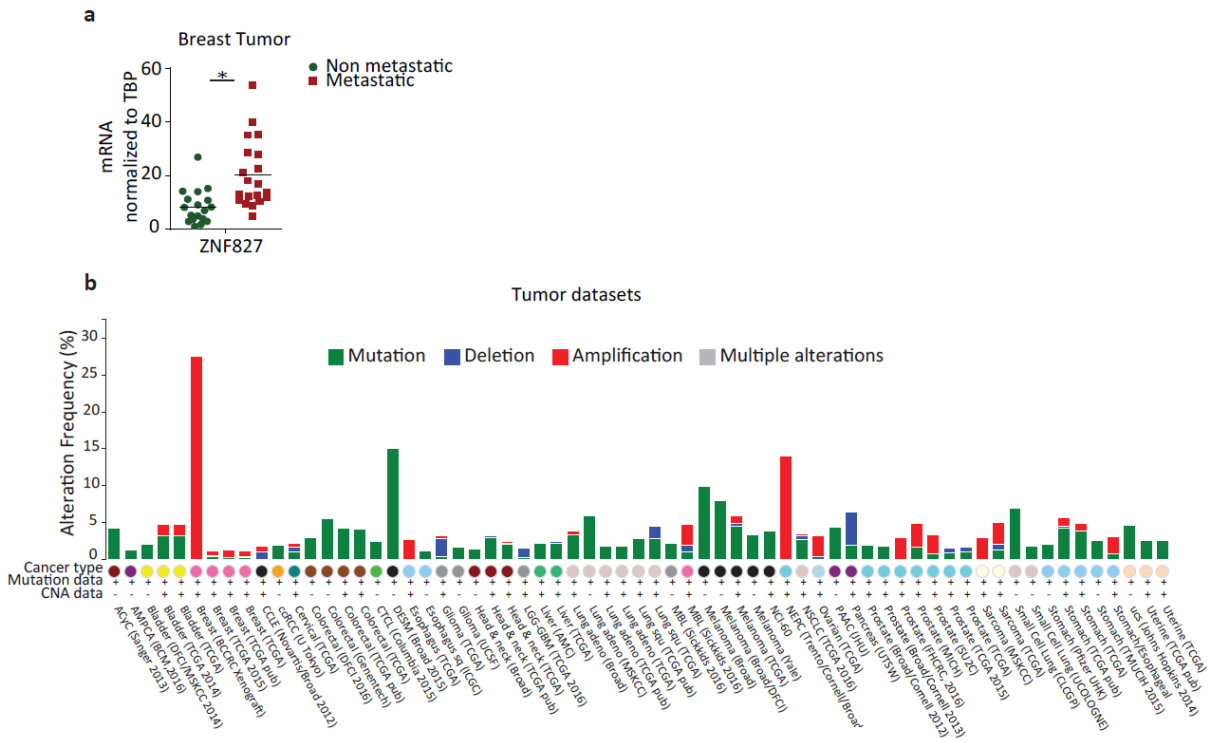


Fig S6. ZNF827 is amplified in breast tumors, but is frequently mutated in the majority of other tumor types.

(a) The levels of ZNF827 mRNA in non-invasive (DCIS) (n=20) and invasive breast tumor samples (n=20) were measured relative to TBP via RT-PCR, and the results were plotted on the y-axis. **(b)** Bar plot showing the frequencies of mutations, deletions, amplifications, and multiple alterations in well-characterized transcriptomes and genomic datasets of different tumors of various origins available from CBioPortal. A grey color box highlighted to show the frequent amplification of ZNF827 in available breast tumor datasets. The x-axis shows various types of tumors, and the y-axis represents the percentage of tumors with a specific defect. *p<0.05, Student's t-test. (Results in this figure contributed by me)

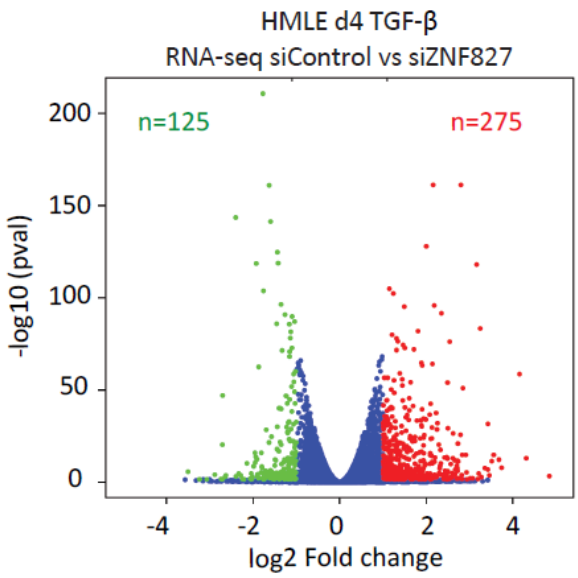


Fig S7. ZNF827-dependent transcriptome in HMLE cells undergoing EMT.
Volcano plot representation of RNA-seq results shows the number of genes with significantly altered expression levels upon ZNF827 depletion in HMLE cells undergoing TGF-β-induced EMT for 4 days. (Results in this figure contributed by me)

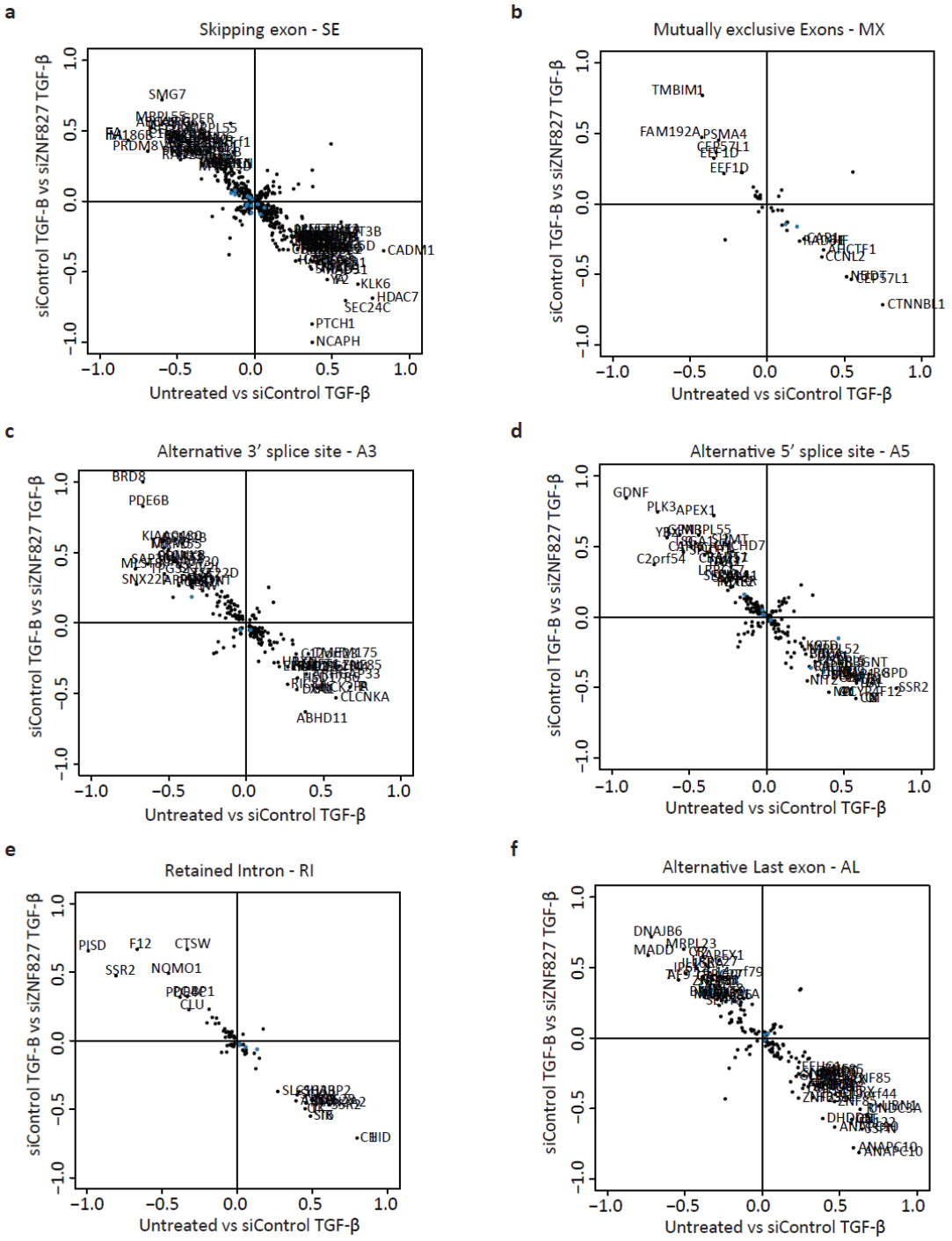


Fig S8. ZNF827 affects all types of alternative splicing events during EMT.

(a-f) Scatterplot showing skipping exon-SE (a), mutually exclusive exon-MX (b), alternative 3' splice site-A3 (c), alternative 5' splice site-A5 (d), retained intron-RI (e), and alternative last exon-AL (f), in genes during EMT and reverting back to its epithelial state upon ZNF827 depletion.

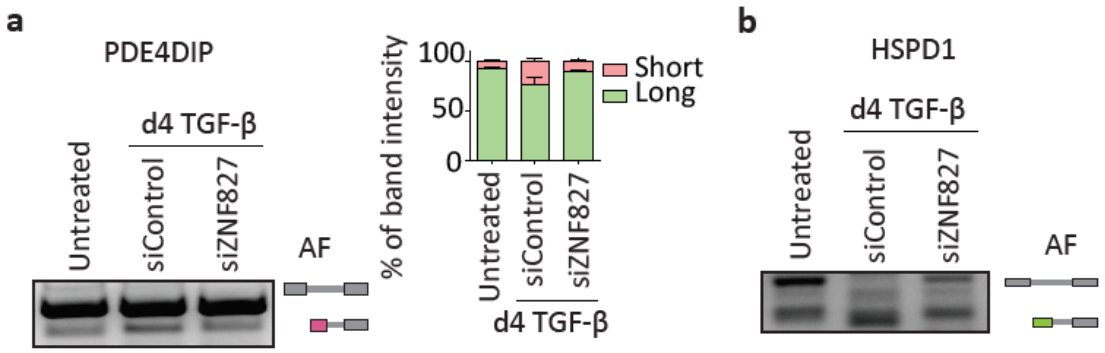


Fig S9. ZNF827 regulates EMT-associated splicing program
(a-b) Representative PCR assays detecting AF changes in PDE4DIP (a) and HSPD1 (b).

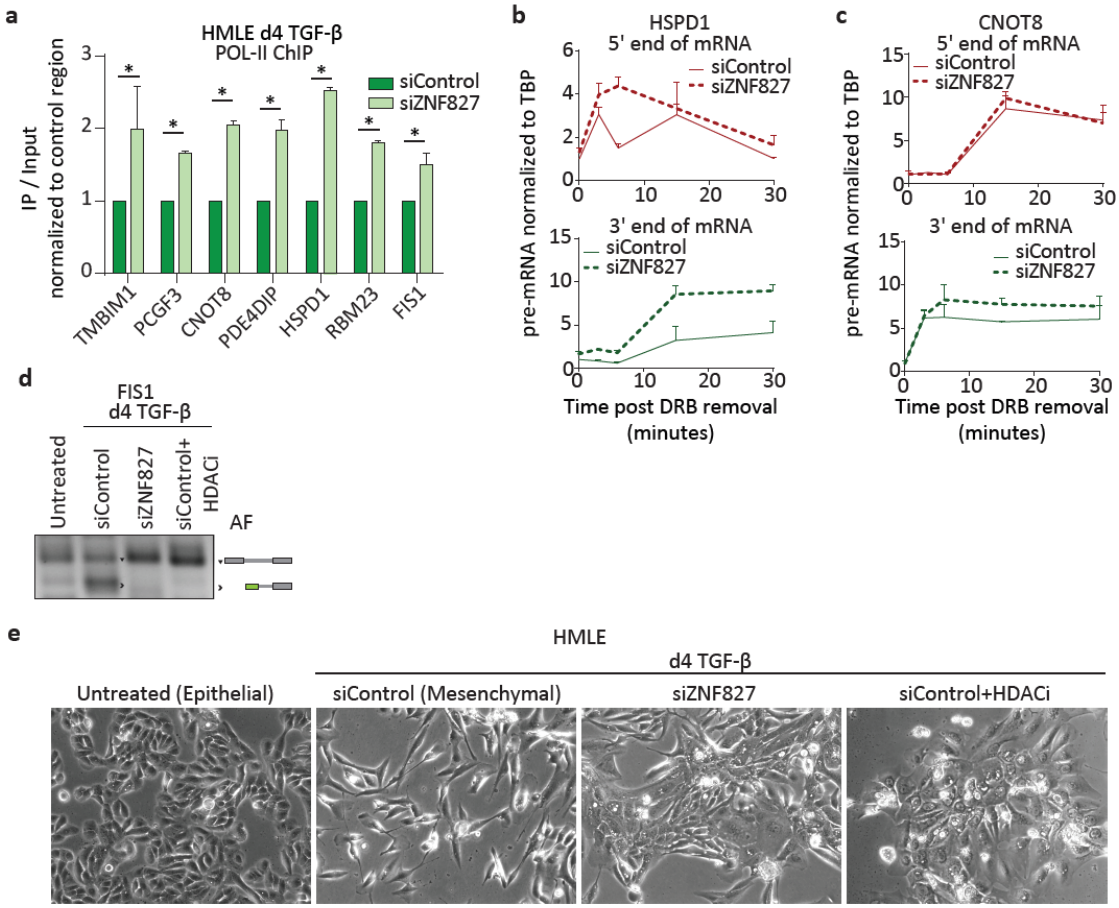


Fig S10. ZNF827 regulates EMT-associated splicing program via epigenetic modulation and splicing machinery recruitment at target sites.
(a) Bar graph showing ChIP enrichment for Pol-II, on selected ZNF827 bound regions in HMLE cells undergoing TGF-β induced EMT and upon siRNA mediated depletion of ZNF827 for 4 days. q-PCR was performed for the indicated gene locus and enrichment was plotted on the y-axis as the ratio of precipitated DNA (bound) to total input DNA and normalized with ACTB upstream promoter region. **(b-c)** Kinetics of RNAPII-dependent transcription elongation were measured in cells treated with DRB, and post removal. Quantitative real-time RT-PCR was performed using primer sets specific for different parts of the

indicated genes, HSPD1 (b), CNOT8 (c) to measure the levels of pre-mRNA expression. Error bars indicates s.e.m. and primers were designed at Exon-Intron junction or Intron-Exon Junction of pre-mRNA. **(d)** Representative DNA gel showing PCR amplified AF changes in FIS1 during EMT, upon ZNF827 knockdown, and cells undergoing EMT when treated with the HDAC inhibitor Sodium-butyrate. **(e)** Representative bright-field images showing phenotypic changes in HMLE cells after 4 days (d4) of TGF-β-induced EMT and siRNA-mediated depletion of ZNF827 compared to non-targeting control (siControl) as well as in cell treated with HDAC inhibitor. (Results in this figure contributed by me)

Supplemental Tables

Table S1. Mouse RT primers

Cdh1	GAGACAGGCTGGCTGAAAGTGAC TGACACGGCATGAGAATAGAGGA
Cdh2	GGTGGAGGAGAAGAAGACCAGGA TGGCATCAGGCTCCACAGTATCT
Ctcf	CACACACACAGGTACTCGTCCTCA CCACTGGTCACAAAGGCCATATC
Fn1	CGGAGAGAGTGCCCCTACTA CGATATTGGTGAATCGCAGA
Snai1	TCTCTAGGCCCTGGCTGCTTC CAGCAAAAGCACGGTTGCAGT
Tbp	GTTTCTGCGGTCGCGTCATTT TGGGTTATCTTCACACACCATGAA
Vim	TGCCAACCTTTTCTTCCCTGAAC TTTGAGTGGGTGTCAACCAGAGG
Zfp827	TCCAGTTTCCAAGGGAAGAG AATGGCTTCCCGCTTTCT

Human RT primers

CTCF	GGGCTTGAGAGCTGGGTTCTATT CTTCGACTGCATCACCTTCCATT
TBP	GTGAACATCATGGATCAGAACAACA AAGATAGGGATTCCGGGAGTCAT
ZNF827	GAGAATGTGAGAAAAGGGGGAAAAT TCCGCCTCTTCCTGCCTACTAAC
CLDN3	AACCTGCATGGACTGTGAAACCT GGTGGTCAAGTATTGGCGGTCAC
CDH2	GGCCGTCATCACAGTGACAGAT CCCTGTTCTCAGGAACTTCACCA
FN1	GCGAGAGTGCCCCTACTACACTG AATGTTGGTGAATCGCAGGTCA
TWIST2	GTCCATGTCCGCCTCCCACTA CAGCATCATTCAGAATCTCCTCCT

SNAI2	CCTGGTTGCTTCAAGGACACATT TGTTGCAGTGAGGGCAAGAAAA
CDH1	TCTTGGAGGAATTCTTGCTTTGC ACCGCTCTCCTCCGAAGAAACA
ZEB1	CCAACAGACCAGACAGTGTTACCAG TCTTGCCCTTCCTTTCCTGTGT
PCDH19	GCATCTCGGTTTCTCCCATTAGC CAATTCTCTTTCCCCTTAGGCTCA
PLAT	AACTTAAAGGAGGCCGGAGCTGT TTCACAGCGTCCCTTAAATTCAC
IL1B	CAAGCTGAGGAAGATGCTGGTTC TCCCATGTGTCGAAGAAGATAGGT
DCBLD1	GTTCCCTGTCAGACAAGCGATTT GCCCGTCAGGTTCAAAACTCAA
WDR82	TACAGGTGCTCAAGCCCTATTG CCCAATTCTTAAGGTGGCAAACG
GOS2	GCAGCACGCCTCCTAGGAACT GGTCTGTCTCTGTCTACTGCGTCTC
BCO2	CAGAATTCACTATTTCCTTGCCACTG AAACCGAAGCTTGTATGCACTCG
WNT10A	CTGTGGGCTCTAGGACTGACTGG CAACTGAACTGTCTGGGCTTCG
FOXA1	TAATCATTGCCATCGTGTGCTTG ACACCATGTCCAACTGTGGAAAG
FUT3	GTGAGCTGTCCTCATCCACTGCT CAAATGGCCTGAAGCAACTGAAG
CXCL14	GAGTCACCGAGTGGTTCTGCAT CGCCCTGGCCTCTTTTAAATC
RBKS	CAGAGGCAGTGGCAAGAGGAG TCCAGTTTTTGGCAAACGAGAAG

Human ChIP Primers

TMBIM1	TCAGCTCTTTGTCCTCCCTGCTA TTCTGAATCCCAGGCTCTCTGTG
PCGF3	ACAAAGGACAGCAGCACCCGTA GCTGGAGCAGCAGAGTCCTCA
CNOT8	CATCCAGCCTAGGAATTGGAGGA AGTGACGAGGAGAGGGGCTTCTA
HSPD1	CAGGGCCTTTGTGGATGTGTAAT CCTGGCACCATGATTTTTACCC
FIS1	GGCTCCTGGGAGTTGTAGTGTCC CTCCAGGCCCGTAGTCTGAGGT
PDE4DIP	AGTTTGAGACCAGCCTGGGAAAC CCACAGCCACACACCACTATGTC
RBM23	CACTGTCAGACCAATTCGGGAAG GTGCCCAGCTAAGTGATCAGACC

ACTB	GTGGACATCTCTTGGGCACTGAG ATGCCCTCCAAGAGCTCCTTCT
INTERGENIC	CTACGTTCTCTATGGGGGTGTGC TGGGTTAAGAATTTGAGGGTAAATGAA

Table S2. **Mouse siRNA**

Zfp827	GCCAGGACAUCUCGGUUAA AGAAUAUGUUUGAGCGCCA GGUAAGUGGUUGCCGAACU GGAAGAAACACCCGUAUUA

Human siRNA

ZNF827	AGAACAUGUUUGAGCGUCA GGACCAUGAUGACGUCUAA CCAGAGAAAGGGUUCGAAU GCACUUUAAUCGGGCGAGA

4. Combined discussion and future perspectives

Discussion

Individual tissues of a multicellular organism are a collection of different cell types, which have a specialized gene regulatory network critical for their functional integrity. For example, a brain tissue usually consist of hundreds of different cell types including neurons, astrocytes, oligodendrocytes or pericytes each having a specialized function. All these diverse cell types are generated by cellular differentiation processes characterized by sequential changes in the cellular fate specified by precise spatio-temporal transcriptional regulation. Such phenotypic remodeling and its accompanied transcriptome changes are regulated by a strong combinatorial action of epigenetic regulators and transcription factors. Similar mechanisms are essential throughout mammalian development; therefore, it is imperative to understand the dynamics of transcriptional and epigenetic changes that drive the process of cellular differentiation. My thesis provides advancement in our understanding of gene regulatory programs underlying neurogenesis and Epithelial to Mesenchymal Transition (EMT) via employing a highly multidisciplinary approach engaging epigenetics, genomics and developmental biology in combination with computational biology tools.

In the first project, I focused on the molecular characterization of novel transcription factors implicated in neurogenesis. The human genomic locus for the HMG-box class transcription factor TOX3 has been implicated in susceptibility to restless legs syndrome in genome-wide association studies, but the physiological role and function of TOX3 during brain development remains largely unknown. Our data demonstrate that Tox3 is highly expressed in the ventricular zone of the developing mouse brain and is necessary for the proliferation of neural stem cells [126]. Our findings indicate that similar to founding member of the TOX family transcription factor, Tox, lack of Tox3 reduces the proliferative potency of neural progenitor cells (NPCs) in the ventricular zone that leads to reduction of post-mitotic neurons in the upper layer of the neocortex [112]. Tox3 binds and activates the Nestin promoter in NPCs, but whether it recognizes DNA in either a sequence dependent or sequence-independent manner requires further

investigation. These findings of gene regulatory role of Tox3 provides mechanistic insights for the neurological disease such as restless legs syndrome associated with this gene locus [358]. Apart from studying Tox3 function in neurological disorders, the physiological role of Tox3 during embryonic neurogenesis will be an interesting aspect for further studies.

During my second project, I investigated genome-wide changes of the epigenetic landscape with respect to accessibility and H3K27ac pattern changes during embryonic corticogenesis, as well as upon changes in neuronal activity. Despite progress, we lack a comprehensive understanding of how specific genes are precisely turned on and off during neuronal development and how regulatory sequences including promoters as well as distal sites contribute to this process. Recent evidence has indicated a crucial role for distal regulatory elements in defining a cell type-specific gene expression program during development [359-361]. Using *in vitro* neuronal differentiation system of mouse embryonic stem cell (mESC) and FAIRE-seq assay, we have showed genome-wide remodeling of chromatin accessibility during neurogenesis. We observed that dividing neuronal progenitors display a more open chromatin landscape than post-mitotic neurons, reflecting genome compaction upon differentiation. This result also suggests that a large number of accessible sites in differentiated neurons may already pre-exist in neuronal progenitors, arguing for 'epigenetic priming' wherein these regions later facilitating timely gain transcriptional competence upon the induction of neurogenesis [126]. Our data reveal that the H3K27ac marked distal regulatory regions are uniquely gained at each stage of differentiation, which recruit distinct transcription factors known to be critical for neurogenesis of respective stages to drive expression of genes that define the respective cell identity. Furthermore activity-dependent plasticity of vertebrate neurons allows the brain to respond to external stimuli and rewire the synapses to assist in memory formation. Our comprehensive transcriptome analysis of neuronal activation via NMDA receptor stimulation shows that the downregulated genes included many neuron-specific genes while those induced contained extracellular matrix genes in addition to the classical neuronal activity genes [74]. Interestingly, such neuronal activity

driven transcriptome resembled closer to neuronal progenitors than to neurons and showed induction of classical neuronal progenitor genes (e.g. *Pax6, Sox2*) and downregulation of hallmark neuronal genes (e.g. *Mapt, Rbfox3*). Furthermore, in addition to known neuronal survival genes (e.g. *Btg2, Bdnf, Atf1*) this study also discovered a number of novel survival genes that were transcriptionally induced upon neuronal activity [74,75,169,170]. Basal neuronal activity or long-term potentiation (LTP) results in the formation of synapses and thus creates memories. On the other hand, after prolonged neuronal activity, long-term depression (LTD) is required to rewire old synapses to allow the acquisition of new learning. Current literature suggests that while LTP results in a gain of H3K27ac at distal sites, LTD event may involve the erasure of memory stored on chromatin during LTP and assists memory formation [74,87,139,169,170,174,175]. We observed that prolonged NMDA activity (LTD) results in a massive loss of chromatin accessibility and the H3K27ac mark at potential distal regulatory regions, whereas only a limited set of potential distal regulatory regions gained those features. The genes in close proximity to these lost distal regulatory regions were dominantly enriched in neuronal activity related genes. Taken together, our findings reveal the dynamics of distal regulatory elements and a vast trove of potential distal regulatory regions that function in determining transcriptome underlying stage-specific cellular identity during neurogenesis and in response to NMDA-induced neuronal activity [362]. Future work should aim to unravel the mechanistic details of the collaborative partnership between transcription factors and epigenetic mechanisms at distal regulatory regions and how this partnership contributes to the gene expression program that underlies neuronal development and activity.

From the third project onwards I switched gears towards analyzing molecular mechanisms underlying cell-fate changes during EMT. In the third project I focused on the role of JNK signaling in the progression of EMT. EMT is a multi-step process during which static epithelial cells become migratory mesenchymal cells. It is utilized during gastrulation, neural crest migration, tissue regeneration and wound healing while its aberrant activation is associated with organ fibrosis and cancer metastasis[188,190,191,211,289]. TGF-β-induced

EMT has been shown to involve activation of canonical Smad signaling and a number of non-canonical pathways including MAP kinases such as ERK, p38 and JNK. However the induction kinetics of these pathways as well as their crosstalk with canonical Smad signaling during EMT has not been explored. Our genome-wide expression profiling in combination with morphological analysis following JNK inhibition at various stages of EMT revealed that while Smad signaling is required for the onset of EMT, JNK pathway rather induced at later time points and is critical for the progression and maintenance of phenotypic and cellular changes associated with EMT. We further established the clinical relevance of these findings by showing that invasive tumors exhibit higher JNK signaling and lower Smad activity as compared to non-invasive tumors. Our findings have uncovered a previously unknown function of JNK signaling in modulating the epigenetic state and chromatin packaging at target gene promoters to mediate transcriptional reprogramming underlying EMT. Furthermore, we have also discovered several novel transcription factors (PRKRIR, PCBP4 and TSC22D1) that were critically required for the phenotypic and molecular changes that accompany EMT in a JNK dependent manner. Overall, these comprehensive findings extend beyond existing notions on the function of JNK activity during EMT, providing detailed kinetics of its phase of action during the process and unraveling the entire gene regulatory network downstream of this important signaling cascade through which it functions to determine mesenchymal fate [257]. Future work should aim for a finer mapping of JNK dependent gene regulatory network as well as its crosstalk with other non-canonical signaling pathways involved during EMT.

In the fourth and fifth project I aimed to identify and molecularly characterize novel factors critical for EMT. Our high resolution transcriptome analysis during EMT in primary mammary epithelial cells, identified strong induction of F-box protein, FBXO32 [197,257] and a C2H2 Zinc finger protein ZNF827 [328]. FBXO32 generally acts as a substrate-recognition subunit of the SKP1–Cullin1–F-box protein E3 ligase complex for ubiquitination and was shown to have a role in muscle homeostasis [269,294]. We found that FBXO32 depletion results in a phenotypic and molecular blockage of hallmark characteristics of EMT in

various established human and mouse cellular model systems as well as during cortical neurogenesis [198,295]. FBXO32 is strongly amplified in metastatic cancers, and its depletion in a NSG mouse xenograft model significantly inhibited tumor growth and metastasis. Our IP-mass spectrometry approach identified many FBXO32 interactors and substrates, among which CtBP1 was strongly enriched. Further comprehensive investigation revealed that FBXO32-dependent K63-mediated ubiquitination of CtBP1 is essential for its nuclear retention. Consequently, such nuclear retention of -CtBP1 is essential for its binding to promoters of key EMT genes and subsequent epigenetic remodeling including histone modifications and chromatin accessibility, thereby generating the gene expression program that drives essential microenvironment critical for EMT. Interestingly, these data also revealed that MMPs (e.g. *MMP1*, *MMP9*, and *MMP10*), which degrade and modify the ECM and play a crucial role in EMT, require FBXO32 for their transcriptional induction during this cascade of phenotypic remodeling[211,298]. Notably, inflammatory cytokines such as *CXCL1/2/3/5*, *IL-1A/1B/32/6*, and many of their receptors, such as *IL1R1, CXCR1* and *TGFBR1*, were also strongly induced by FBXO32. Further computational and experimental analysis discovered that the genes activated by FBXO32 via CtBP1 are also targeted by ATF3, which is known to enhance cell motility [299]. Altogether, these evidences emphasize the essential function of FBXO32 in establishing the proper microenvironment required to initiate EMT. This findings provide new insights for the crosstalk of ubiqutination pathway and epigenetic regulation during EMT [363].

Similar to FBXO32, ZNF827 loss-of-function strongly inhibits EMT progression *in vitro* and *in vivo.* Further molecular analysis revealed these effects to be linked to ZNF827-dependent regulation of a major set of alternative splicing events that accompany EMT. Our comprehensive splicing analysis using transcriptome profiling during EMT identified a large number of such events (n = 3415) that affected 1134 genes. Among those genes, a substantial portion of alternative splicing events (n = 1586 events in 924 genes) were reverted to that in a normal epithelial state upon ZNF827 depletion, suggesting an important role for ZNF827 in the establishment of the splicing program required for acquisition of mesenchymal

identity[213,337,342]. Epigenetic regulation and transcriptional kinetics play critical roles in regulating alternative splicing of nascent transcripts via fine-tuning the recruitment kinetics of splicing components and in turn modulation of polymerase elongation rate [338]. We show that ZNF827 mediates these responses by directly targeting the underlying genomic loci. It further interacts with epigenetic (e.g. HDAC1) and splicing (e.g. U2AF1/2) machinery and assists in their recruitment to target chromatin and RNA sites respectively [337]. Among the genes regulated at the alternative splicing level by ZNF827 during EMT, we found strong enrichment for protein and nucleic acid biosynthetic process-relevant genes. This observation is highly relevant given that several recent studies have implied a critical role of metabolites and the underlying regulatory networks during EMT [335,336,346-348]. Our project to discover novel regulators of EMT including FBXO32 and ZNF827 showed the potential of the employed multidisciplinary approaches to decipher new molecules whose function was unknown previously, but is critical for the establishment of EMT process. These exciting findings further warrant an investigation into the function of these protein in various contexts of development and disease.

Altogether, our findings delineate diverse gene regulatory circuitries that cells utilize to orchestrate cellular differentiation during development and disease. This study provides new insights in epigenomic regulation during neurogenesis, neuronal activation, as well as EMT. These findings also suggested physiological implications of disease associated loci which will be interesting to follow up with respect to their potential for therapeutic intervention. Furthermore our studies provide a comprehensive resource of global gene expression profiles and epigenomic landscape during neurogenesis and EMT, which can further be explored by researchers in the field to gain new insight into these biological processes.

Concluding remarks

One of the fundamentals to understand the development of multicellular organism is to delineate the process of cellular differentiation which generates the huge variety of cell types within the tissue by precise cascades of spatio-temporal alterations that involve signaling pathways, epigenetic, transcriptional, post transcriptional and post translational machineries. In this thesis, I have attempted to understand and further enhance the knowledge of these changes during the process of neuronal differentiation and epithelial to mesenchymal transition (EMT). The findings described in this thesis have shed further light on how a particular transcription factor can control the cell-fate restriction by activation of lineage-specific marker to achieve uni-directional differentiation. Mapping of accessible chromatin sites along with distal regulation of transcriptional changes during neuronal development and neuronal activity have further enhanced our understanding of how distal regulatory sequences function to precisely control gene expression not only during development but also in mature neuronal function. During the process of EMT, the genome-wide transcriptomic and epigenomic datasets we have generated will serve as a useful resource to the scientific community to further study gene expression programs. Also, it will lay foundation for further functional investigation of novel coding and long noncoding RNAs transcriptionally modulated during distinct phases of this phenotypic transition. Furthermore, our comprehensive findings extend beyond existing notions on the function of JNK activity during EMT, providing detailed kinetics of its mode of action during the process and set a new dimension of how external signaling can modulate epigenome to drive changes in cell-fate. From the FBXO32 analysis, our findings identify a crosstalk between ubiqutination pathway and epigenetic machinery to facilitate transcriptional reprograming critical for EMT microenvironment. Furthermore ubiqutination dependent nuclear localization of transcriptional regulator such as CTBP1, adds another layer of complexity to such gene regulatory processes. Finally, our analysis reveals a novel and major deterministic role of one of the C2H2 Zinc finger proteins ZNF827 in regulating the repertoire of alternative splicing via modulation of RNA polymerase

kinetics that underlies EMT during development and disease. Overall, this thesis delineates diverse gene regulatory circuitries that cells utilize to orchestrate cell fate changes during development and disease.

References

1 Touchon, M. & Rocha, E. P. Coevolution of the Organization and Structure of Prokaryotic Genomes. *Cold Spring Harb Perspect Biol* **8**, a018168, doi:10.1101/cshperspect.a018168 (2016).

2 Fuerst, J. A. Intracellular compartmentation in planctomycetes. *Annu Rev Microbiol* **59**, 299-328, doi:10.1146/annurev.micro.59.030804.121258 (2005).

3 Funes, S. *et al.* A green algal apicoplast ancestor. *Science* **298**, 2155, doi:10.1126/science.1076003 (2002).

4 Rout, M. P. & Field, M. C. The Evolution of Organellar Coat Complexes and Organization of the Eukaryotic Cell. *Annual review of biochemistry* **86**, 637-657, doi:10.1146/annurev-biochem-061516-044643 (2017).

5 Rokas, A. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet* **42**, 235-251, doi:10.1146/annurev.genet.42.110807.091513 (2008).

6 Dacks, J. B. *et al.* The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *Journal of cell science* **129**, 3695-3703, doi:10.1242/jcs.178566 (2016).

7 Medvedev, S. P., Shevchenko, A. I. & Zakian, S. M. Molecular basis of Mammalian embryonic stem cell pluripotency and self-renewal. *Acta Naturae* **2**, 30-46 (2010).

8 Leslie, M. P., Shelton, D. E. & Michod, R. E. Generation time and fitness tradeoffs during the evolution of multicellularity. *J Theor Biol* **430**, 92-102, doi:10.1016/j.jtbi.2017.07.007 (2017).

9 Arendt, D. *et al.* The origin and evolution of cell types. *Nature reviews. Genetics* **17**, 744-757, doi:10.1038/nrg.2016.127 (2016).

10 Dahm, R. Friedrich Miescher and the discovery of DNA. *Developmental biology* **278**, 274-288, doi:10.1016/j.ydbio.2004.11.028 (2005).

11 Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine* **79**, 137-158 (1944).

12 Hershey, A. D. & Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* **36**, 39-56 (1952).

13 Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).

14 Meselson, M. & Stahl, F. W. The Replication of DNA in Escherichia Coli. *Proceedings of the National Academy of Sciences of the United States of America* **44**, 671-682 (1958).

15 Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nature reviews. Genetics* **17**, 47-62, doi:10.1038/nrg.2015.10 (2016).

16 Qi, P. *et al.* Reciprocal repression between TUSC7 and miR-23b in gastric cancer. *International journal of cancer* **137**, 1269-1278, doi:10.1002/ijc.29516 (2015).

17 Badeaux, A. I. & Shi, Y. Emerging roles for chromatin as a signal integration and storage platform. *Nature reviews. Molecular cell biology* **14**, 211-224, doi:10.1038/nrm3545 (2013).

18 Johnson, D. G. & Dent, S. Y. Chromatin: receiver and quarterback for cellular signals. *Cell* **152**, 685-689, doi:10.1016/j.cell.2013.01.017 (2013).

19 Smith, E. & Shilatifard, A. Transcriptional elongation checkpoint control in development and disease. *Genes & development* **27**, 1079-1088, doi:10.1101/gad.215137.113 (2013).

20 Razin, A. & Riggs, A. D. DNA methylation and gene function. *Science* **210**, 604-610 (1980).

21 Baylin, S. B. & Jones, P. A. Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol* **8**, doi:10.1101/cshperspect.a019505 (2016).

22 Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol* **6**, a019133, doi:10.1101/cshperspect.a019133 (2014).
23 Zechiedrich, E. L. & Cozzarelli, N. R. Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in Escherichia coli. *Genes & development* **9**, 2859-2869 (1995).
24 Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature reviews. Genetics* **18**, 517-534, doi:10.1038/nrg.2017.33 (2017).
25 Moran, U., Phillips, R. & Milo, R. SnapShot: key numbers in biology. *Cell* **141**, 1262-1262 e1261, doi:10.1016/j.cell.2010.06.019 (2010).
26 Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974).
27 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).
28 Woodcock, C. L. & Ghosh, R. P. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol* **2**, a000596, doi:10.1101/cshperspect.a000596 (2010).
29 van Steensel, B. Chromatin: constructing the big picture. *The EMBO journal* **30**, 1885-1895, doi:10.1038/emboj.2011.135 (2011).
30 Cheema, M. S. & Ausio, J. The Structural Determinants behind the Epigenetic Role of Histone Variants. *Genes (Basel)* **6**, 685-713, doi:10.3390/genes6030685 (2015).
31 Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D. & Zhao, Y. SnapShot: histone modifications. *Cell* **159**, 458-458 e451, doi:10.1016/j.cell.2014.09.037 (2014).
32 Marmorstein, R. & Zhou, M. M. Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb Perspect Biol* **6**, a018762, doi:10.1101/cshperspect.a018762 (2014).
33 Seto, E. & Yoshida, M. Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol* **6**, a018713, doi:10.1101/cshperspect.a018713 (2014).
34 Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nature reviews. Molecular cell biology* **6**, 838-849, doi:10.1038/nrm1761 (2005).
35 Lanouette, S., Mongeon, V., Figeys, D. & Couture, J. F. The functional diversity of protein lysine methylation. *Mol Syst Biol* **10**, 724, doi:10.1002/msb.134974 (2014).
36 Wang, S. & Wang, Y. Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis. *Biochimica et biophysica acta* **1829**, 1126-1135, doi:10.1016/j.bbagrm.2013.07.003 (2013).
37 Shi, Y. G. & Tsukada, Y. The discovery of histone demethylases. *Cold Spring Harb Perspect Biol* **5**, doi:10.1101/cshperspect.a017947 (2013).
38 Fiszbein, A. & Kornblihtt, A. R. Histone methylation, alternative splicing and neuronal differentiation. *Neurogenesis* **3**, e1204844, doi:10.1080/23262133.2016.1204844 (2016).
39 Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Molecular cell* **48**, 491-507, doi:10.1016/j.molcel.2012.11.006 (2012).
40 Kim, T. K. & Shiekhattar, R. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* **162**, 948-959, doi:10.1016/j.cell.2015.08.008 (2015).
41 Schaffner, W. Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem* **396**, 311-327, doi:10.1515/hsz-2014-0303 (2015).
42 Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature reviews. Genetics* **13**, 59-69, doi:10.1038/nrg3095 (2011).
43 Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187, doi:10.1016/j.cell.2016.09.018 (2016).

44 Robinson, P. J., Fairall, L., Huynh, V. A. & Rhodes, D. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6506-6511, doi:10.1073/pnas.0601212103 (2006).
45 Song, F. *et al.* Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380, doi:10.1126/science.1251413 (2014).
46 Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature reviews. Molecular cell biology* **13**, 436-447, doi:10.1038/nrm3382 (2012).
47 Sebe-Pedros, A. *et al.* High-Throughput Proteomics Reveals the Unicellular Roots of Animal Phosphosignaling and Cell Differentiation. *Developmental cell* **39**, 186-197, doi:10.1016/j.devcel.2016.09.019 (2016).
48 Sebe-Pedros, A. *et al.* The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. *Cell* **165**, 1224-1237, doi:10.1016/j.cell.2016.03.034 (2016).
49 Lynch, M. & Force, A. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459-473 (2000).
50 Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651-654, doi:10.1038/35036627 (2000).
51 Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41-42, doi:10.1038/35075138 (2001).
52 Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature reviews. Genetics* **5**, 101-113, doi:10.1038/nrg1272 (2004).
53 Hobert, O. Regulation of terminal differentiation programs in the nervous system. *Annu Rev Cell Dev Biol* **27**, 681-696, doi:10.1146/annurev-cellbio-092910-154226 (2011).
54 Tosches, M. A. Developmental and genetic mechanisms of neural circuit evolution. *Developmental biology*, doi:10.1016/j.ydbio.2017.06.016 (2017).
55 Namba, T. & Huttner, W. B. Neural progenitor cells and their role in the development and evolutionary expansion of the neocortex. *Wiley Interdiscip Rev Dev Biol* **6**, doi:10.1002/wdev.256 (2017).
56 Berdasco, M. & Esteller, M. DNA methylation in stem cell renewal and multipotency. *Stem Cell Res Ther* **2**, 42, doi:10.1186/scrt83 (2011).
57 Lee, H. K., Lee, H. S. & Moody, S. A. Neural transcription factors: from embryos to neural stem cells. *Mol Cells* **37**, 705-712, doi:10.14348/molcells.2014.0227 (2014).
58 Araujo, G. L., Araujo, J. A., Schroeder, T., Tort, A. B. & Costa, M. R. Sonic hedgehog signaling regulates mode of cell division of early cerebral cortex progenitors and increases astrogliogenesis. *Front Cell Neurosci* **8**, 77, doi:10.3389/fncel.2014.00077 (2014).
59 Mizutani, K., Yoon, K., Dang, L., Tokunaga, A. & Gaiano, N. Differential Notch signalling distinguishes neural stem cells from intermediate progenitors. *Nature* **449**, 351-355, doi:10.1038/nature06090 (2007).
60 Palma, V. & Ruiz i Altaba, A. Hedgehog-GLI signaling regulates the behavior of cells with stem cell properties in the developing neocortex. *Development* **131**, 337-345, doi:10.1242/dev.00930 (2004).
61 Zechner, D. *et al.* beta-Catenin signals regulate cell growth and the balance between progenitor cell expansion and differentiation in the nervous system. *Developmental biology* **258**, 406-418 (2003).
62 Brault, V. *et al.* Inactivation of the beta-catenin gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. *Development* **128**, 1253-1264 (2001).

63 Kalani, M. Y. *et al.* Wnt-mediated self-renewal of neural stem/progenitor cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16970-16975, doi:10.1073/pnas.0808616105 (2008).
64 Fujita, S. The discovery of the matrix cell, the identification of the multipotent neural stem cell and the development of the central nervous system. *Cell Struct Funct* **28**, 205-228 (2003).
65 Ahmed, S. *et al.* Transcription factors and neural stem cell self-renewal, growth and differentiation. *Cell Adh Migr* **3**, 412-424 (2009).
66 Bylund, M., Andersson, E., Novitch, B. G. & Muhr, J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci* **6**, 1162-1168, doi:10.1038/nn1131 (2003).
67 Collignon, J., Varlet, I. & Robertson, E. J. Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature* **381**, 155-158, doi:10.1038/381155a0 (1996).
68 Fuccillo, M., Joyner, A. L. & Fishell, G. Morphogen to mitogen: the multiple roles of hedgehog signalling in vertebrate neural development. *Nat Rev Neurosci* **7**, 772-783, doi:10.1038/nrn1990 (2006).
69 Custo Greig, L. F., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat Rev Neurosci* **14**, 755-769, doi:10.1038/nrn3586 (2013).
70 Woodworth, M. B., Custo Greig, L., Kriegstein, A. R. & Macklis, J. D. SnapShot: cortical development. *Cell* **151**, 918-918 e911, doi:10.1016/j.cell.2012.10.004 (2012).
71 Adrover, M. F. *et al.* Characterization of specific cDNA background synthesis introduced by reverse transcription in RT-PCR assays. *Biochimie* **92**, 1839-1846, doi:10.1016/j.biochi.2010.07.019 (2010).
72 Milner, B., Squire, L. R. & Kandel, E. R. Cognitive neuroscience and the study of memory. *Neuron* **20**, 445-468 (1998).
73 Koenig, O. [Perception and memory: illustration of an approach of the study of visual mechanisms in cognitive neuroscience]. *Revue neurologique* **151**, 451-456 (1995).
74 Chen, Y., Wang, Y., Modrusan, Z., Sheng, M. & Kaminker, J. S. Regulation of neuronal gene expression and survival by basal NMDA receptor activity: a role for histone deacetylase 4. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 15327-15339, doi:10.1523/JNEUROSCI.0569-14.2014 (2014).
75 Zhang, S. J. *et al.* Decoding NMDA receptor signaling: identification of genomic programs specifying neuronal survival and death. *Neuron* **53**, 549-562, doi:10.1016/j.neuron.2007.01.025 (2007).
76 Nestler, E. J. & Landsman, D. Learning about addiction from the genome. *Nature* **409**, 834-835, doi:10.1038/35057015 (2001).
77 Kandel, E. R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030-1038, doi:10.1126/science.1067020 (2001).
78 Huh, G. S. *et al.* Functional requirement for class I MHC in CNS development and plasticity. *Science* **290**, 2155-2159 (2000).
79 Kuipers, S. D. *et al.* BDNF-induced LTP is associated with rapid Arc/Arg3.1-dependent enhancement in adult hippocampal neurogenesis. *Sci Rep* **6**, 21222, doi:10.1038/srep21222 (2016).
80 Connor, S. A. & Wang, Y. T. A Place at the Table: LTD as a Mediator of Memory Genesis. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, doi:10.1177/1073858415588498 (2015).
81 Massey, P. V. & Bashir, Z. I. Long-term depression: multiple forms and implications for brain function. *Trends in neurosciences* **30**, 176-184, doi:10.1016/j.tins.2007.02.005 (2007).

82 Sanchez-Perez, A., Llansola, M., Cauli, O. & Felipo, V. Modulation of NMDA receptors in the cerebellum. II. Signaling pathways and physiological modulators regulating NMDA receptor function. *Cerebellum* **4**, 162-170, doi:10.1080/14734220510008003 (2005).
83 Pokorska, A. *et al.* Synaptic activity induces signalling to CREB without increasing global levels of cAMP in hippocampal neurons. *Journal of neurochemistry* **84**, 447-452 (2003).
84 Shaywitz, A. J. & Greenberg, M. E. CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annual review of biochemistry* **68**, 821-861, doi:10.1146/annurev.biochem.68.1.821 (1999).
85 Tao, X., West, A. E., Chen, W. G., Corfas, G. & Greenberg, M. E. A calcium-responsive transcription factor, CaRF, that regulates neuronal activity-dependent expression of BDNF. *Neuron* **33**, 383-395 (2002).
86 West, A. E. *et al.* Calcium regulation of neuronal gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 11024-11031, doi:10.1073/pnas.191352298 (2001).
87 Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187, doi:10.1038/nature09033 (2010).
88 Malik, A. N. *et al.* Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nature neuroscience* **17**, 1330-1339, doi:10.1038/nn.3808 (2014).
89 Bossy-Wetzel, E., Schwarzenbacher, R. & Lipton, S. A. Molecular pathways to neurodegeneration. *Nature medicine* **10 Suppl**, S2-9, doi:10.1038/nm1067 (2004).
90 Choi, D. W. Excitotoxic cell death. *Journal of neurobiology* **23**, 1261-1276, doi:10.1002/neu.480230915 (1992).
91 Hardingham, G. E. & Bading, H. The Yin and Yang of NMDA receptor signalling. *Trends in neurosciences* **26**, 81-89, doi:10.1016/S0166-2236(02)00040-1 (2003).
92 Lipton, S. A. & Rosenberg, P. A. Excitatory amino acids as a final common pathway for neurologic disorders. *The New England journal of medicine* **330**, 613-622, doi:10.1056/NEJM199403033300907 (1994).
93 Kalia, L. V., Kalia, S. K. & Salter, M. W. NMDA receptors in clinical neurology: excitatory times ahead. *The Lancet. Neurology* **7**, 742-755, doi:10.1016/S1474-4422(08)70165-0 (2008).
94 Choi, D. W. Glutamate neurotoxicity and diseases of the nervous system. *Neuron* **1**, 623-634 (1988).
95 Mehta, A., Prabhakar, M., Kumar, P., Deshmukh, R. & Sharma, P. L. Excitotoxicity: bridge to various triggers in neurodegenerative disorders. *European journal of pharmacology* **698**, 6-18, doi:10.1016/j.ejphar.2012.10.032 (2013).
96 Sgambato-Faure, V. & Cenci, M. A. Glutamatergic mechanisms in the dyskinesias induced by pharmacological dopamine replacement and deep brain stimulation for the treatment of Parkinson's disease. *Progress in neurobiology* **96**, 69-86, doi:10.1016/j.pneurobio.2011.10.005 (2012).
97 Ahmed, I. *et al.* Glutamate NMDA receptor dysregulation in Parkinson's disease with dyskinesias. *Brain : a journal of neurology* **134**, 979-986, doi:10.1093/brain/awr028 (2011).
98 Snyder, E. M. *et al.* Regulation of NMDA receptor trafficking by amyloid-beta. *Nature neuroscience* **8**, 1051-1058, doi:10.1038/nn1503 (2005).
99 Blanchet, P. J., Papa, S. M., Metman, L. V., Mouradian, M. M. & Chase, T. N. Modulation of levodopa-induced motor response complications by NMDA antagonists in Parkinson's disease. *Neuroscience and biobehavioral reviews* **21**, 447-453 (1997).
100 Ikonomidou, C., Qin Qin, Y., Labruyere, J. & Olney, J. W. Motor neuron degeneration induced by excitotoxin agonists has features in common with those seen in the SOD-1 transgenic mouse

model of amyotrophic lateral sclerosis. *Journal of neuropathology and experimental neurology* **55**, 211-224 (1996).

101 Ulas, J. *et al.* Selective increase of NMDA-sensitive glutamate binding in the striatum of Parkinson's disease, Alzheimer's disease, and mixed Parkinson's disease/Alzheimer's disease patients: an autoradiographic study. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **14**, 6317-6324 (1994).

102 Young, A. B. *et al.* NMDA receptor losses in putamen from patients with Huntington's disease. *Science* **241**, 981-983 (1988).

103 Woolf, C. J. & Mannion, R. J. Neuropathic pain: aetiology, symptoms, mechanisms, and management. *Lancet* **353**, 1959-1964, doi:10.1016/S0140-6736(99)01307-0 (1999).

104 Rogawski, M. A. The NMDA receptor, NMDA antagonists and epilepsy therapy. A status report. *Drugs* **44**, 279-292 (1992).

105 Cortes-Mendoza, J., Diaz de Leon-Guerrero, S., Pedraza-Alva, G. & Perez-Martinez, L. Shaping synaptic plasticity: the role of activity-mediated epigenetic regulation on gene transcription. *Int J Dev Neurosci* **31**, 359-369, doi:10.1016/j.ijdevneu.2013.04.003 (2013).

106 Ghasemi, M. & Schachter, S. C. The NMDA receptor complex as a therapeutic target in epilepsy: a review. *Epilepsy Behav* **22**, 617-640, doi:10.1016/j.yebeh.2011.07.024 (2011).

107 Martynoga, B., Drechsel, D. & Guillemot, F. Molecular control of neurogenesis: a view from the mammalian cerebral cortex. *Cold Spring Harb Perspect Biol* **4**, doi:10.1101/cshperspect.a008359 (2012).

108 Yuan, S. H., Qiu, Z. & Ghosh, A. TOX3 regulates calcium-dependent transcription in neurons. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2909-2914, doi:10.1073/pnas.0805555106 (2009).

109 Dittmer, S. *et al.* TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. *Journal of cell science* **124**, 252-260, doi:10.1242/jcs.068759 (2011).

110 Pertz, O., Hodgson, L., Klemke, R. L. & Hahn, K. M. Spatiotemporal dynamics of RhoA activity in migrating cells. *Nature* **440**, 1069-1072, doi:10.1038/nature04665 (2006).

111 Bastien, D. *et al.* IL-1alpha Gene Deletion Protects Oligodendrocytes after Spinal Cord Injury through Upregulation of the Survival Factor Tox3. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **35**, 10715-10730, doi:10.1523/JNEUROSCI.0498-15.2015 (2015).

112 Artegiani, B. *et al.* Tox: a multifunctional transcription factor and novel regulator of mammalian corticogenesis. *The EMBO journal* **34**, 896-910, doi:10.15252/embj.201490061 (2015).

113 Chen, M. B. *et al.* Association between polymorphisms of trinucleotide repeat containing 9 gene and breast cancer risk: evidence from 62,005 subjects. *Breast Cancer Res Treat* **126**, 177-183, doi:10.1007/s10549-010-1114-6 (2011).

114 Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-1093, doi:10.1038/nature05887 (2007).

115 Winkelmann, J. *et al.* Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS genetics* **7**, e1002171, doi:10.1371/journal.pgen.1002171 (2011).

116 Spieler, D. *et al.* Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. *Genome research* **24**, 592-603, doi:10.1101/gr.166751.113 (2014).

117 Xiong, L. *et al.* MEIS1 intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. *Human molecular genetics* **18**, 1065-1074, doi:10.1093/hmg/ddn443 (2009).

118 Bibel, M., Richter, J., Lacroix, E. & Barde, Y. A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc* **2**, 1034-1043, doi:10.1038/nprot.2007.147 (2007).
119 Johansson, C. B. *et al.* Identification of a neural stem cell in the adult mammalian central nervous system. *Cell* **96**, 25-34 (1999).
120 Zimmerman, L. *et al.* Independent regulatory elements in the nestin gene direct transgene expression to neural stem cells or muscle precursors. *Neuron* **12**, 11-24 (1994).
121 Nye, J. S., Kopan, R. & Axel, R. An activated Notch suppresses neurogenesis and myogenesis but not gliogenesis in mammalian cells. *Development* **120**, 2421-2430 (1994).
122 Shih, A. H. & Holland, E. C. Notch signaling enhances nestin expression in gliomas. *Neoplasia* **8**, 1072-1082, doi:10.1593/neo.06526 (2006).
123 Gunther, H. S. *et al.* Glioblastoma-derived stem cell-enriched cultures form distinct subgroups according to molecular and phenotypic criteria. *Oncogene* **27**, 2897-2909, doi:10.1038/sj.onc.1210949 (2008).
124 Codega, P. *et al.* Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche. *Neuron* **82**, 545-559, doi:10.1016/j.neuron.2014.02.039 (2014).
125 Tox-Albumin Diphtheria. *Science* **19**, 198-199, doi:10.1126/science.ns-19.479.198 (1892).
126 Ziller, M. J. *et al.* Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature* **518**, 355-359, doi:10.1038/nature13990 (2015).
127 O'Flaherty, E. & Kaye, J. TOX defines a conserved subfamily of HMG-box proteins. *BMC Genomics* **4**, 13 (2003).
128 Borghese, L. *et al.* Inhibition of notch signaling in human embryonic stem cell-derived neural stem cells delays G1/S phase transition and accelerates neuronal differentiation in vitro and in vivo. *Stem Cells* **28**, 955-964, doi:10.1002/stem.408 (2010).
129 Kageyama, R., Ohtsuka, T., Hatakeyama, J. & Ohsawa, R. Roles of bHLH genes in neural stem cell differentiation. *Experimental cell research* **306**, 343-348, doi:10.1016/j.yexcr.2005.03.015 (2005).
130 Stros, M., Launholt, D. & Grasser, K. D. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci* **64**, 2590-2606, doi:10.1007/s00018-007-7162-3 (2007).
131 Johannessen, M., Delghandi, M. P. & Moens, U. What turns CREB on? *Cell Signal* **16**, 1211-1227, doi:10.1016/j.cellsig.2004.05.001 (2004).
132 Lee, H. J., Mignacca, R. C. & Sakamoto, K. M. Transcriptional activation of egr-1 by granulocyte-macrophage colony-stimulating factor but not interleukin 3 requires phosphorylation of cAMP response element-binding protein (CREB) on serine 133. *The Journal of biological chemistry* **270**, 15979-15983 (1995).
133 Jin, Z. *et al.* Different transcription factors regulate nestin gene expression during P19 cell neural differentiation and central nervous system development. *The Journal of biological chemistry* **284**, 8160-8173, doi:10.1074/jbc.M805632200 (2009).
134 Tiwari, V. K. *et al.* Target genes of Topoisomerase IIbeta regulate neuronal survival and are defined by their chromatin state. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E934-943, doi:10.1073/pnas.1119798109 (2012).
135 Baumgart, J. & Grebe, N. C57BL/6-specific conditions for efficient in utero electroporation of the central nervous system. *J Neurosci Methods* **240**, 116-124, doi:10.1016/j.jneumeth.2014.11.004 (2015).
136 Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905, doi:10.1126/science.1237905 (2013).
137 Hu, X. L., Wang, Y. & Shen, Q. Epigenetic control on cell fate choice in neural stem cells. *Protein & cell* **3**, 278-290, doi:10.1007/s13238-012-2916-6 (2012).

138 Borrelli, E., Nestler, E. J., Allis, C. D. & Sassone-Corsi, P. Decoding the epigenetic language of neuronal plasticity. *Neuron* **60**, 961-974, doi:10.1016/j.neuron.2008.10.012 (2008).
139 Feng, J., Fouse, S. & Fan, G. Epigenetic regulation of neural gene expression and neuronal function. *Pediatric research* **61**, 58R-63R, doi:10.1203/pdr.0b013e3180457635 (2007).
140 Bell, O., Tiwari, V. K., Thoma, N. H. & Schubeler, D. Determinants and dynamics of genome accessibility. *Nature reviews. Genetics* **12**, 554-564, doi:10.1038/nrg3017 (2011).
141 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).
142 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).
143 Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature genetics* **44**, 148-156, doi:10.1038/ng.1064 (2012).
144 Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283, doi:10.1038/nature09692 (2011).
145 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
146 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
147 Pott, S. & Lieb, J. D. What are super-enhancers? *Nature genetics* **47**, 8-12, doi:10.1038/ng.3167 (2014).
148 Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research* **17**, 877-885, doi:10.1101/gr.5533506 (2007).
149 Koohy, H., Down, T. A. & Hubbard, T. J. Chromatin accessibility data sets show bias due to sequence specificity of the DNase I enzyme. *PloS one* **8**, e69853, doi:10.1371/journal.pone.0069853 (2013).
150 Simon, J. M., Giresi, P. G., Davis, I. J. & Lieb, J. D. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature protocols* **7**, 256-267, doi:10.1038/nprot.2011.444 (2012).
151 Song, L. *et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome research* **21**, 1757-1767, doi:10.1101/gr.121541.111 (2011).
152 Bibel, M. *et al.* Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat Neurosci* **7**, 1003-1009, doi:10.1038/nn1301 (2004).
153 Tiwari, V. K. *et al.* A chromatin-modifying function of JNK during stem cell differentiation. *Nature genetics* **44**, 94-100, doi:10.1038/ng.1036 (2012).
154 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).
155 Lienert, F. *et al.* Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS genetics* **7**, e1002090, doi:10.1371/journal.pgen.1002090 (2011).
156 Thakurela, S. *et al.* Gene regulation and priming by topoisomerase IIalpha in embryonic stem cells. *Nature communications* **4**, 2478, doi:10.1038/ncomms3478 (2013).
157 Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**, 755-766, doi:S1097-2765(08)00358-4 [pii]

10.1016/j.molcel.2008.05.007 (2008).

158 Doyle, B., Fudenberg, G., Imakaev, M. & Mirny, L. A. Chromatin loops as allosteric modulators of enhancer-promoter interactions. *PLoS computational biology* **10**, e1003867, doi:10.1371/journal.pcbi.1003867 (2014).
159 Samee, M. A. & Sinha, S. Quantitative modeling of a gene's expression from its intergenic sequence. *PLoS computational biology* **10**, e1003467, doi:10.1371/journal.pcbi.1003467 (2014).
160 Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461, doi:10.1038/nature12787 (2014).
161 Lee, S. *et al.* STAT3 promotes motor neuron differentiation by collaborating with motor neuron-specific LIM complex. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11445-11450, doi:10.1073/pnas.1302676110 (2013).
162 Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N. & Yagi, T. CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell reports* **2**, 345-357, doi:10.1016/j.celrep.2012.06.014 (2012).
163 Dekker, J. CTCF and cohesin help neurons raise their self-awareness. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8799-8800, doi:10.1073/pnas.1206195109 (2012).
164 Lee, S. *et al.* Fusion protein Isl1-Lhx3 specifies motor neuron fate by inducing motor neuron genes and concomitantly suppressing the interneuron programs. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 3383-3388, doi:10.1073/pnas.1114515109 (2012).
165 Tanaka, H. *et al.* Islet1 selectively promotes peripheral axon outgrowth in Rohon-Beard primary sensory neurons. *Developmental dynamics : an official publication of the American Association of Anatomists* **240**, 9-22, doi:10.1002/dvdy.22499 (2011).
166 Sharma, K. *et al.* LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. *Cell* **95**, 817-828 (1998).
167 Uwanogho, D. *et al.* Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. *Mechanisms of development* **49**, 23-36 (1995).
168 Yao, B. & Jin, P. Unlocking epigenetic codes in neurogenesis. *Genes & development* **28**, 1253-1271, doi:10.1101/gad.241547.114 (2014).
169 West, A. E., Griffith, E. C. & Greenberg, M. E. Regulation of transcription factors by neuronal activity. *Nature reviews. Neuroscience* **3**, 921-931, doi:10.1038/nrn987 (2002).
170 Zheng, S. *et al.* NMDA-induced neuronal survival is mediated through nuclear factor I-A in mice. *The Journal of clinical investigation* **120**, 2446-2456, doi:10.1172/JCI33144 (2010).
171 Coyle, J. T. & Tsai, G. NMDA receptor function, neuroplasticity, and the pathophysiology of schizophrenia. *International review of neurobiology* **59**, 491-515, doi:10.1016/S0074-7742(04)59019-0 (2004).
172 Nord, A. S. *et al.* Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521-1531, doi:10.1016/j.cell.2013.11.033 (2013).
173 Bowers, E. M. *et al.* Virtual ligand screening of the p300/CBP histone acetyltransferase: identification of a selective small molecule inhibitor. *Chemistry & biology* **17**, 471-482, doi:10.1016/j.chembiol.2010.03.006 (2010).
174 Deisseroth, K., Heist, E. K. & Tsien, R. W. Translocation of calmodulin to the nucleus supports CREB phosphorylation in hippocampal neurons. *Nature* **392**, 198-202, doi:10.1038/32448 (1998).
175 Lopez-Atalaya, J. P. & Barco, A. Can changes in histone acetylation contribute to memory formation? *Trends in genetics : TIG* **30**, 529-539, doi:10.1016/j.tig.2014.09.003 (2014).
176 Maeda, M., Johnson, K. R. & Wheelock, M. J. Cadherin switching: essential for behavioral but not morphological changes during an epithelium-to-mesenchyme transition. *Journal of cell science* **118**, 873-887, doi:10.1242/jcs.01634 (2005).

177 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).
178 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
179 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-515, doi:10.1038/nbt.1621 (2010).
180 Langmead, B. Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 11**, Unit 11 17, doi:10.1002/0471250953.bi1107s32 (2010).
181 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
182 Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. QuasR: quantification and annotation of short reads in R. *Bioinformatics*, doi:10.1093/bioinformatics/btu781 (2014).
183 Liu, C. *et al.* NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic acids research* **33**, D112-115, doi:10.1093/nar/gki041 (2005).
184 Flicek, P. *et al.* Ensembl 2014. *Nucleic acids research* **42**, D749-755, doi:10.1093/nar/gkt1196 (2014).
185 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
186 Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research* **37**, W305-311, doi:10.1093/nar/gkp427 (2009).
187 Hay, E. D. The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. *Developmental dynamics : an official publication of the American Association of Anatomists* **233**, 706-720, doi:10.1002/dvdy.20345 (2005).
188 Acloque, H., Adams, M. S., Fishwick, K., Bronner-Fraser, M. & Nieto, M. A. Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. *The Journal of clinical investigation* **119**, 1438-1449, doi:10.1172/JCI38019 (2009).
189 Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* **119**, 1420-1428, doi:10.1172/JCI39104 (2009).
190 Nieto, M. A. Epithelial-Mesenchymal Transitions in development and disease: old views and new perspectives. *The International journal of developmental biology* **53**, 1541-1547, doi:10.1387/ijdb.072410mn (2009).
191 Nieto, M. A. The ins and outs of the epithelial to mesenchymal transition in health and disease. *Annual review of cell and developmental biology* **27**, 347-376, doi:10.1146/annurev-cellbio-092910-154036 (2011).
192 Potenta, S., Zeisberg, E. & Kalluri, R. The role of endothelial-to-mesenchymal transition in cancer progression. *British journal of cancer* **99**, 1375-1379, doi:10.1038/sj.bjc.6604662 (2008).
193 Zeisberg, E. M. *et al.* Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nat Med* **13**, 952-961, doi:10.1038/nm1613 (2007).
194 Zeisberg, M. *et al.* Fibroblasts derive from hepatocytes in liver fibrosis via epithelial to mesenchymal transition. *The Journal of biological chemistry* **282**, 23337-23347, doi:10.1074/jbc.M700194200 (2007).
195 Thiery, J. P. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* **2**, 442-454, doi:10.1038/nrc822 (2002).

196 Tiwari, N., Gheldof, A., Tatari, M. & Christofori, G. EMT as the ultimate survival mechanism of cancer cells. *Semin Cancer Biol* **22**, 194-207, doi:10.1016/j.semcancer.2012.02.013 (2012).
197 Tiwari, N. *et al.* Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming. *Cancer Cell* **23**, 768-783, doi:10.1016/j.ccr.2013.04.020 (2013).
198 Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nature reviews. Molecular cell biology* **15**, 178-196, doi:10.1038/nrm3758 (2014).
199 Thompson, E. W. & Haviv, I. The social aspects of EMT-MET plasticity. *Nat Med* **17**, 1048-1049, doi:10.1038/nm.2437 (2011).
200 Thiery, J. P. & Sleeman, J. P. Complex networks orchestrate epithelial-mesenchymal transitions. *Nature reviews. Molecular cell biology* **7**, 131-142, doi:10.1038/nrm1835 (2006).
201 Nawshad, A., Lagamba, D., Polad, A. & Hay, E. D. Transforming growth factor-beta signaling during epithelial-mesenchymal transformation: implications for embryogenesis and tumor metastasis. *Cells, tissues, organs* **179**, 11-23, doi:10.1159/000084505 (2005).
202 Feng, X. H. & Derynck, R. Specificity and versatility in tgf-beta signaling through Smads. *Annu Rev Cell Dev Biol* **21**, 659-693, doi:10.1146/annurev.cellbio.21.022404.142018 (2005).
203 Xu, J., Lamouille, S. & Derynck, R. TGF-beta-induced epithelial to mesenchymal transition. *Cell research* **19**, 156-172, doi:10.1038/cr.2009.5 (2009).
204 Barrallo-Gimeno, A. & Nieto, M. A. The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* **132**, 3151-3161, doi:10.1242/dev.01907 (2005).
205 Postigo, A. A., Depp, J. L., Taylor, J. J. & Kroll, K. L. Regulation of Smad signaling through a differential recruitment of coactivators and corepressors by ZEB proteins. *The EMBO journal* **22**, 2453-2462, doi:10.1093/emboj/cdg226 (2003).
206 Kowanetz, M., Valcourt, U., Bergstrom, R., Heldin, C. H. & Moustakas, A. Id2 and Id3 define the potency of cell proliferation and differentiation responses to transforming growth factor beta and bone morphogenetic protein. *Mol Cell Biol* **24**, 4241-4254 (2004).
207 Chaudhury, A. & Howe, P. H. The tale of transforming growth factor-beta (TGFbeta) signaling: a soigne enigma. *IUBMB life* **61**, 929-939, doi:10.1002/iub.239 (2009).
208 Zhang, Y. E. Non-Smad pathways in TGF-beta signaling. *Cell research* **19**, 128-139, doi:10.1038/cr.2008.328 (2009).
209 Gui, T., Sun, Y., Shimokado, A. & Muragaki, Y. The Roles of Mitogen-Activated Protein Kinase Pathways in TGF-beta-Induced Epithelial-Mesenchymal Transition. *Journal of signal transduction* **2012**, 289243, doi:10.1155/2012/289243 (2012).
210 Akhurst, R. J. & Hata, A. Targeting the TGFbeta signalling pathway in disease. *Nat Rev Drug Discov* **11**, 790-811, doi:10.1038/nrd3810 (2012).
211 Radisky, E. S. & Radisky, D. C. Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer. *J Mammary Gland Biol Neoplasia* **15**, 201-212, doi:10.1007/s10911-010-9177-x (2010).
212 Skrypek, N., Goossens, S., De Smedt, E., Vandamme, N. & Berx, G. Epithelial-to-Mesenchymal Transition: Epigenetic Reprogramming Driving Cellular Plasticity. *Trends in genetics : TIG*, doi:10.1016/j.tig.2017.08.004 (2017).
213 Neumann, D. P., Goodall, G. J. & Gregory, P. A. Regulation of splicing and circularisation of RNA in epithelial mesenchymal plasticity. *Semin Cell Dev Biol*, doi:10.1016/j.semcdb.2017.08.008 (2017).
214 Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. *Cell* **166**, 21-45, doi:10.1016/j.cell.2016.06.028 (2016).
215 Pijnenborg, R., Dixon, G., Robertson, W. B. & Brosens, I. Trophoblastic invasion of human decidua from 8 to 18 weeks of pregnancy. *Placenta* **1**, 3-19 (1980).

216 Perez-Pomares, J. M. & Munoz-Chapuli, R. Epithelial-mesenchymal transitions: a mesodermal cell strategy for evolutive innovation in Metazoans. *The Anatomical record* **268**, 343-351, doi:10.1002/ar.10165 (2002).
217 Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871-890, doi:10.1016/j.cell.2009.11.007 (2009).
218 Yang, J., Dai, C. & Liu, Y. A novel mechanism by which hepatocyte growth factor blocks tubular epithelial to mesenchymal transition. *Journal of the American Society of Nephrology : JASN* **16**, 68-78, doi:10.1681/ASN.2003090795 (2005).
219 Willis, B. C. & Borok, Z. TGF-beta-induced EMT: mechanisms and implications for fibrotic lung disease. *American journal of physiology. Lung cellular and molecular physiology* **293**, L525-534, doi:10.1152/ajplung.00163.2007 (2007).
220 Hocevar, B. A., Brown, T. L. & Howe, P. H. TGF-beta induces fibronectin synthesis through a c-Jun N-terminal kinase-dependent, Smad4-independent pathway. *The EMBO journal* **18**, 1345-1356, doi:10.1093/emboj/18.5.1345 (1999).
221 Yu, L., Hebert, M. C. & Zhang, Y. E. TGF-beta receptor-activated p38 MAP kinase mediates Smad-independent TGF-beta responses. *The EMBO journal* **21**, 3749-3759, doi:10.1093/emboj/cdf366 (2002).
222 Santibanez, J. F. JNK mediates TGF-beta1-induced epithelial mesenchymal transdifferentiation of mouse transformed keratinocytes. *FEBS letters* **580**, 5385-5391, doi:10.1016/j.febslet.2006.09.003 (2006).
223 Alcorn, J. F. *et al.* Jun N-terminal kinase 1 regulates epithelial-to-mesenchymal transition induced by TGF-beta1. *Journal of cell science* **121**, 1036-1045, doi:10.1242/jcs.019455 (2008).
224 Liu, Q. *et al.* Transforming growth factor {beta}1 induces epithelial-mesenchymal transition by activating the JNK-Smad3 pathway in rat peritoneal mesothelial cells. *Peritoneal dialysis international : journal of the International Society for Peritoneal Dialysis* **28 Suppl 3**, S88-95 (2008).
225 Ying, L. *et al.* Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition. *Molecular bioSystems* **8**, 2289-2294, doi:10.1039/c2mb25070e (2012).
226 Santangelo, L. *et al.* The stable repression of mesenchymal program is required for hepatocyte identity: a novel role for hepatocyte nuclear factor 4alpha. *Hepatology* **53**, 2063-2074, doi:10.1002/hep.24280 (2011).
227 Spath, G. F. & Weiss, M. C. Hepatocyte nuclear factor 4 provokes expression of epithelial marker genes, acting as a morphogen in dedifferentiated hepatoma cells. *The Journal of cell biology* **140**, 935-946 (1998).
228 Valcourt, U., Thuault, S., Pardali, K., Heldin, C. H. & Moustakas, A. Functional role of Meox2 during the epithelial cytostatic response to TGF-beta. *Molecular oncology* **1**, 55-71, doi:10.1016/j.molonc.2007.02.002 (2007).
229 Ohshima, J. *et al.* Two candidate tumor suppressor genes, MEOX2 and SOSTDC1, identified in a 7p21 homozygous deletion region in a Wilms tumor. *Genes, chromosomes & cancer* **48**, 1037-1050, doi:10.1002/gcc.20705 (2009).
230 Rosenbauer, F. *et al.* Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. *Nature genetics* **38**, 27-37, doi:10.1038/ng1679 (2006).
231 Tanaka-Okamoto, M. *et al.* Increased susceptibility to spontaneous lung cancer in mice lacking LIM-domain only 7. *Cancer science* **100**, 608-616, doi:10.1111/j.1349-7006.2009.01091.x (2009).
232 Walter, M. J. *et al.* Reduced PU.1 expression causes myeloid progenitor expansion and increased leukemia penetrance in mice expressing PML-RARalpha. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12513-12518, doi:10.1073/pnas.0504247102 (2005).

233 Psichari, E., Balmain, A., Plows, D., Zoumpourlis, V. & Pintzas, A. High activity of serum response factor in the mesenchymal transition of epithelial tumor cells is regulated by RhoA signaling. *The Journal of biological chemistry* **277**, 29490-29495, doi:10.1074/jbc.M112368200 (2002).
234 Choi, H. N. *et al.* Serum response factor enhances liver metastasis of colorectal carcinoma via alteration of the E-cadherin/beta-catenin complex. *Oncology reports* **21**, 57-63 (2009).
235 Eastham, A. M. *et al.* Epithelial-mesenchymal transition events during human embryonic stem cell differentiation. *Cancer research* **67**, 11254-11262, doi:10.1158/0008-5472.CAN-07-2253 (2007).
236 Ullmann, U. *et al.* Epithelial-mesenchymal transition process in human embryonic stem cells cultured in feeder-free conditions. *Molecular human reproduction* **13**, 21-32, doi:10.1093/molehr/gal091 (2007).
237 Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell stem cell* **7**, 51-63, doi:10.1016/j.stem.2010.04.014 (2010).
238 Plachta, N., Bibel, M., Tucker, K. L. & Barde, Y. A. Developmental potential of defined neural progenitors derived from mouse embryonic stem cells. *Development* **131**, 5449-5456, doi:dev.01420 [pii]

10.1242/dev.01420 (2004).
239 Saito, T. In vivo electroporation in the embryonic mouse central nervous system. *Nat Protoc* **1**, 1552-1558, doi:10.1038/nprot.2006.276 (2006).
240 Alcorn, J. F. *et al.* c-Jun N-terminal kinase 1 is required for the development of pulmonary fibrosis. *American journal of respiratory cell and molecular biology* **40**, 422-432, doi:10.1165/rcmb.2008-0174OC (2009).
241 Chang, Y. & Wu, X. Y. JNK1/2 siRNA inhibits transforming-growth factor-beta1-induced connective tissue growth factor expression and fibrotic function in THSFs. *Molecular and cellular biochemistry* **335**, 83-89, doi:10.1007/s11010-009-0245-8 (2010).
242 McDonald, O. G., Wu, H., Timp, W., Doi, A. & Feinberg, A. P. Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nat Struct Mol Biol* **18**, 867-874 (2011).
243 Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153**, 38-55, doi:10.1016/j.cell.2013.03.008 (2013).
244 Oliva, A. A., Jr., Atkins, C. M., Copenagle, L. & Banker, G. A. Activated c-Jun N-terminal kinase is required for axon formation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26**, 9462-9470, doi:10.1523/JNEUROSCI.2625-06.2006 (2006).
245 Amura, C. R., Marek, L., Winn, R. A. & Heasley, L. E. Inhibited neurogenesis in JNK1-deficient embryonic stem cells. *Molecular and cellular biology* **25**, 10791-10802, doi:10.1128/MCB.25.24.10791-10802.2005 (2005).
246 Tognon, C. E. *et al.* ETV6-NTRK3-mediated breast epithelial cell transformation is blocked by targeting the IGF1R signaling pathway. *Cancer research* **71**, 1060-1070, doi:10.1158/0008-5472.CAN-10-3096 (2011).
247 Bae, J. S. *et al.* Serum response factor induces epithelial to mesenchymal transition with resistance to sorafenib in hepatocellular carcinoma. *International journal of oncology* **44**, 129-136, doi:10.3892/ijo.2013.2154 (2014).
248 Hayashida, T. *et al.* HOXB9, a gene overexpressed in breast cancer, promotes tumorigenicity and lung metastasis. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 1100-1105, doi:10.1073/pnas.0912710107 (2010).
249 Lee, C. C. *et al.* TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer. *The Journal of biological chemistry* **287**, 2798-2809, doi:10.1074/jbc.M111.258947 (2012).

250 Sanchez-Tillo, E. *et al.* ZEB1 Promotes invasiveness of colorectal carcinoma cells through the opposing regulation of uPA and PAI-1. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**, 1071-1082, doi:10.1158/1078-0432.CCR-12-2675 (2013).
251 Burk, U. *et al.* A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO reports* **9**, 582-589, doi:10.1038/embor.2008.74 (2008).
252 Gross, I. *et al.* The intestine-specific homeobox gene Cdx2 decreases mobility and antagonizes dissemination of colon cancer cells. *Oncogene* **27**, 107-115, doi:10.1038/sj.onc.1210601 (2008).
253 Ortega, F., Berninger, B. & Costa, M. R. Primary culture and live imaging of adult neural stem cells and their progeny. *Methods in molecular biology* **1052**, 1-11, doi:10.1007/7651_2013_22 (2013).
254 Papageorgis, P. TGFbeta Signaling in Tumor Initiation, Epithelial-to-Mesenchymal Transition, and Metastasis. *Journal of oncology* **2015**, 587193, doi:10.1155/2015/587193 (2015).
255 Lim, J. & Thiery, J. P. Epithelial-mesenchymal transitions: insights from development. *Development* **139**, 3471-3486, doi:10.1242/dev.071209 (2012).
256 Jung, H. Y., Fattet, L. & Yang, J. Molecular pathways: linking tumor microenvironment to epithelial-mesenchymal transition in metastasis. *Clinical cancer research : an official journal of the American Association for Cancer Research* **21**, 962-968, doi:10.1158/1078-0432.CCR-13-3173 (2015).
257 Sahu, S. K. *et al.* JNK-dependent gene regulatory circuitry governs mesenchymal fate. *The EMBO journal* **34**, 2162-2181, doi:10.15252/embj.201490693 (2015).
258 Stankiewicz, T. R., Gray, J. J., Winter, A. N. & Linseman, D. A. C-terminal binding proteins: central players in development and disease. *Biomolecular concepts* **5**, 489-511, doi:10.1515/bmc-2014-0027 (2014).
259 Byun, J. S. & Gardner, K. C-Terminal Binding Protein: A Molecular Link between Metabolic Imbalance and Epigenetic Regulation in Breast Cancer. *International journal of cell biology* **2013**, 647975, doi:10.1155/2013/647975 (2013).
260 Chinnadurai, G. The transcriptional corepressor CtBP: a foe of multiple tumor suppressors. *Cancer research* **69**, 731-734, doi:10.1158/0008-5472.CAN-08-3349 (2009).
261 Chinnadurai, G. Transcriptional regulation by C-terminal binding proteins. *The international journal of biochemistry & cell biology* **39**, 1593-1607, doi:10.1016/j.biocel.2007.01.025 (2007).
262 Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525-530, doi:10.1038/nature16064 (2015).
263 Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472-476, doi:10.1038/nature15748 (2015).
264 Snider, N. T. & Omary, M. B. Post-translational modifications of intermediate filament proteins: mechanisms and functions. *Nature reviews. Molecular cell biology* **15**, 163-177, doi:10.1038/nrm3753 (2014).
265 Wang, Z., Liu, P., Inuzuka, H. & Wei, W. Roles of F-box proteins in cancer. *Nat Rev Cancer* **14**, 233-247, doi:10.1038/nrc3700 (2014).
266 Frescas, D. & Pagano, M. Deregulated proteolysis by the F-box proteins SKP2 and beta-TrCP: tipping the scales of cancer. *Nat Rev Cancer* **8**, 438-449, doi:10.1038/nrc2396 (2008).
267 Bai, C. *et al.* SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**, 263-274 (1996).
268 Skaar, J. R., Pagan, J. K. & Pagano, M. Mechanisms and function of substrate recruitment by F-box proteins. *Nature reviews. Molecular cell biology* **14**, 369-381, doi:10.1038/nrm3582 (2013).
269 Bodine, S. C. *et al.* Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science* **294**, 1704-1708, doi:10.1126/science.1065874 (2001).

270 Li, H. H. *et al.* Atrogin-1 inhibits Akt-dependent cardiac hypertrophy in mice via ubiquitin-dependent coactivation of Forkhead proteins. *The Journal of clinical investigation* **117**, 3211-3223, doi:10.1172/JCI31757 (2007).
271 D'Hulst, G. *et al.* Effect of acute environmental hypoxia on protein metabolism in human skeletal muscle. *Acta physiologica* **208**, 251-264, doi:10.1111/apha.12086 (2013).
272 Chaillou, T. *et al.* Hypoxia transiently affects skeletal muscle hypertrophy in a functional overload model. *American journal of physiology. Regulatory, integrative and comparative physiology* **302**, R643-654, doi:10.1152/ajpregu.00262.2011 (2012).
273 Chou, J. L. *et al.* Promoter hypermethylation of FBXO32, a novel TGF-beta/SMAD4 target gene and tumor suppressor, is associated with poor prognosis in human ovarian cancer. *Lab Invest* **90**, 414-425, doi:10.1038/labinvest.2009.138 (2010).
274 Mei, Z. *et al.* FBXO32 Targets c-Myc for Proteasomal Degradation and Inhibits c-Myc Activity. *The Journal of biological chemistry* **290**, 16202-16214, doi:10.1074/jbc.M115.645978 (2015).
275 Tanaka, N. *et al.* Acquired platinum resistance involves epithelial to mesenchymal transition through ubiquitin ligase FBXO32 dysregulation. *JCI Insight* **1**, e83654, doi:10.1172/jci.insight.83654 (2016).
276 Guo, W. *et al.* FBXO32, a new TGF-beta/Smad signaling pathway target gene, is epigenetically inactivated in gastric cardia adenocarcinoma. *Neoplasma* **62**, 646-657, doi:10.4149/neo_2015_078 (2015).
277 Zhang, X. L. *et al.* CtBP1 is involved in epithelial-mesenchymal transition and is a potential therapeutic target for hepatocellular carcinoma. *Oncology reports* **30**, 809-814, doi:10.3892/or.2013.2537 (2013).
278 Subramanian, T. & Chinnadurai, G. Association of class I histone deacetylases with transcriptional corepressor CtBP. *FEBS letters* **540**, 255-258 (2003).
279 Yang, Y. *et al.* Inhibitors of ubiquitin-activating enzyme (E1), a new class of potential cancer therapeutics. *Cancer research* **67**, 9472-9481, doi:10.1158/0008-5472.CAN-07-0568 (2007).
280 Morel, A. P. *et al.* EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. *PLoS genetics* **8**, e1002723, doi:10.1371/journal.pgen.1002723 (2012).
281 Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-715, doi:10.1016/j.cell.2008.03.027 (2008).
282 Komander, D. & Rape, M. The ubiquitin code. *Annual review of biochemistry* **81**, 203-229, doi:10.1146/annurev-biochem-060310-170328 (2012).
283 Boxer, L. D., Barajas, B., Tao, S., Zhang, J. & Khavari, P. A. ZNF750 interacts with KLF4 and RCOR1, KDM1A, and CTBP1/2 chromatin regulators to repress epidermal progenitor genes and induce differentiation genes. *Genes & development* **28**, 2013-2026, doi:10.1101/gad.246579.114 (2014).
284 Przybylo, J. A. & Radisky, D. C. Matrix metalloproteinase-induced epithelial-mesenchymal transition: tumor progression at Snail's pace. *Int J Biochem Cell Biol* **39**, 1082-1088, doi:10.1016/j.biocel.2007.03.002 (2007).
285 Orlichenko, L. S. & Radisky, D. C. Matrix metalloproteinases stimulate epithelial-mesenchymal transition during tumor development. *Clinical & experimental metastasis* **25**, 593-600, doi:10.1007/s10585-008-9143-9 (2008).
286 Biswas, S. *et al.* CXCL13-CXCR5 co-expression regulates epithelial to mesenchymal transition of breast cancer cells during lymph node metastasis. *Breast cancer research and treatment* **143**, 265-276, doi:10.1007/s10549-013-2811-8 (2014).
287 Fietz, S. A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proceedings of the National Academy of*

Sciences of the United States of America **109**, 11836-11841, doi:10.1073/pnas.1209647109 (2012).

288 Busskamp, V. *et al.* Rapid neurogenesis through transcriptional activation in human stem cells. *Mol Syst Biol* **10**, 760, doi:10.15252/msb.20145508 (2014).

289 Radisky, D. C. *et al.* Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. *Nature* **436**, 123-127, doi:10.1038/nature03688 (2005).

290 Zheng, H. *et al.* PKD1 phosphorylation-dependent degradation of SNAIL by SCF-FBXO11 regulates epithelial-mesenchymal transition and metastasis. *Cancer cell* **26**, 358-373, doi:10.1016/j.ccr.2014.07.022 (2014).

291 Xu, M. *et al.* Atypical ubiquitin E3 ligase complex Skp1-Pam-Fbxo45 controls the core epithelial-to-mesenchymal transition-inducing transcription factors. *Oncotarget* **6**, 979-994 (2015).

292 Rieser, E., Cordier, S. M. & Walczak, H. Linear ubiquitination: a newly discovered regulator of cell signalling. *Trends in biochemical sciences* **38**, 94-102, doi:10.1016/j.tibs.2012.11.007 (2013).

293 Erpapazoglou, Z., Walker, O. & Haguenauer-Tsapis, R. Versatile roles of k63-linked ubiquitin chains in trafficking. *Cells* **3**, 1027-1088, doi:10.3390/cells3041027 (2014).

294 Gomes, M. D., Lecker, S. H., Jagoe, R. T., Navon, A. & Goldberg, A. L. Atrogin-1, a muscle-specific F-box protein highly expressed during muscle atrophy. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 14440-14445, doi:10.1073/pnas.251541198 (2001).

295 Singh, S. & Solecki, D. J. Polarity transitions during neurogenesis and germinal zone exit in the developing central nervous system. *Frontiers in cellular neuroscience* **9**, 62, doi:10.3389/fncel.2015.00062 (2015).

296 Cleveland, B. M. & Evenhuis, J. P. Molecular characterization of atrogin-1/F-box protein-32 (FBXO32) and F-box protein-25 (FBXO25) in rainbow trout (Oncorhynchus mykiss): Expression across tissues in response to feed deprivation. *Comp Biochem Physiol B Biochem Mol Biol* **157**, 248-257, doi:10.1016/j.cbpb.2010.06.010 (2010).

297 Barnes, C. J. *et al.* Functional inactivation of a transcriptional corepressor by a signaling kinase. *Nature structural biology* **10**, 622-628, doi:10.1038/nsb957 (2003).

298 Kuo, P. L., Shen, K. H., Hung, S. H. & Hsu, Y. L. CXCL1/GROalpha increases cell migration and invasion of prostate cancer by decreasing fibulin-1 expression through NF-kappaB/HDAC1 epigenetic regulation. *Carcinogenesis* **33**, 2477-2487, doi:10.1093/carcin/bgs299 (2012).

299 Yin, X. *et al.* ATF3, an adaptive-response gene, enhances TGF{beta} signaling and cancer-initiating cell features in breast cancer cells. *Journal of cell science* **123**, 3558-3565, doi:10.1242/jcs.064915 (2010).

300 Sanchez-Elsner, T. *et al.* Synergistic cooperation between hypoxia and transforming growth factor-beta pathways on human vascular endothelial growth factor gene expression. *The Journal of biological chemistry* **276**, 38527-38535, doi:10.1074/jbc.M104536200 (2001).

301 Lebrun, J. J. The Dual Role of TGFbeta in Human Cancer: From Tumor Suppression to Cancer Metastasis. *ISRN molecular biology* **2012**, 381428, doi:10.5402/2012/381428 (2012).

302 Michalski, A. *et al.* Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP* **10**, M111 011015, doi:10.1074/mcp.M111.011015 (2011).

303 Olsen, J. V. *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* **4**, 709-712, doi:10.1038/nmeth1060 (2007).

304 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).

305 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805, doi:10.1021/pr101065j (2011).

306 Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods* **4**, 207-214, doi:10.1038/nmeth1019 (2007).
307 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
308 Wrighton, K. H. Cell migration: EMT promotes contact inhibition of locomotion. *Nat Rev Mol Cell Biol* **16**, 518, doi:10.1038/nrm4045 (2015).
309 Heerboth, S. *et al.* EMT and tumor metastasis. *Clin Transl Med* **4**, 6, doi:10.1186/s40169-015-0048-3 (2015).
310 Li, M. *et al.* Epithelial-mesenchymal transition: An emerging target in tissue fibrosis. *Exp Biol Med (Maywood)*, doi:10.1177/1535370215597194 (2015).
311 Navandar, M. *et al.* ERK signalling modulates epigenome to drive epithelial to mesenchymal transition. *Oncotarget* **8**, 29269-29281, doi:10.18632/oncotarget.16493 (2017).
312 De Craene, B. & Berx, G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* **13**, 97-110, doi:10.1038/nrc3447 (2013).
313 Zhang, J., Lu, A., Beech, D., Jiang, B. & Lu, Y. Suppression of breast cancer metastasis through the inhibition of VEGF-mediated tumor angiogenesis. *Cancer therapy* **5**, 273-286 (2007).
314 Vuong, C. K., Black, D. L. & Zheng, S. The neurogenetics of alternative splicing. *Nature reviews. Neuroscience* **17**, 265-281, doi:10.1038/nrn.2016.27 (2016).
315 Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252-1269, doi:10.1016/j.cell.2013.02.034 (2013).
316 Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587-1593, doi:10.1126/science.1230612 (2012).
317 Irimia, M. & Blencowe, B. J. Alternative splicing: decoding an expansive regulatory layer. *Current opinion in cell biology* **24**, 323-332, doi:10.1016/j.ceb.2012.03.005 (2012).
318 Licatalosi, D. D. & Darnell, R. B. RNA processing and its regulation: global insights into biological networks. *Nature reviews. Genetics* **11**, 75-87, doi:10.1038/nrg2673 (2010).
319 Warzecha, C. C. *et al.* An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *The EMBO journal* **29**, 3286-3300, doi:10.1038/emboj.2010.195 (2010).
320 Shapiro, I. M. *et al.* An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS genetics* **7**, e1002218, doi:10.1371/journal.pgen.1002218 (2011).
321 Roca, H. *et al.* Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer. *PloS one* **8**, e76773, doi:10.1371/journal.pone.0076773 (2013).
322 Bonomi, S. *et al.* HnRNP A1 controls a splicing regulatory circuit promoting mesenchymal-to-epithelial transition. *Nucleic acids research* **41**, 8665-8679, doi:10.1093/nar/gkt579 (2013).
323 Braeutigam, C. *et al.* The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. *Oncogene* **33**, 1082-1092, doi:10.1038/onc.2013.50 (2014).
324 Weirauch, M. T. & Hughes, T. R. A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Sub-cellular biochemistry* **52**, 25-73, doi:10.1007/978-90-481-9069-0_3 (2011).
325 Ding, G., Lorenz, P., Kreutzer, M., Li, Y. & Thiesen, H. J. SysZNF: the C2H2 zinc finger gene database. *Nucleic acids research* **37**, D267-273, doi:10.1093/nar/gkn782 (2009).
326 Schmitges, F. W. *et al.* Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome research* **26**, 1742-1752, doi:10.1101/gr.209643.116 (2016).

327 Han, H. *et al.* Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Molecular cell* **65**, 539-553 e537, doi:10.1016/j.molcel.2017.01.011 (2017).
328 Conomos, D., Reddel, R. R. & Pickett, H. A. NuRD-ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. *Nature structural & molecular biology* **21**, 760-770, doi:10.1038/nsmb.2877 (2014).
329 Klijn, C. *et al.* A comprehensive transcriptional portrait of human cancer cell lines. *Nature biotechnology* **33**, 306-312, doi:10.1038/nbt.3080 (2015).
330 Saldi, T., Cortazar, M. A., Sheridan, R. M. & Bentley, D. L. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *Journal of molecular biology* **428**, 2623-2635, doi:10.1016/j.jmb.2016.04.017 (2016).
331 Banyard, J. & Bielenberg, D. R. The role of EMT and MET in cancer dissemination. *Connect Tissue Res*, 1-11, doi:10.3109/03008207.2015.1060970 (2015).
332 Iorns, E. *et al.* A new mouse model for the study of human breast cancer metastasis. *PloS one* **7**, e47995, doi:10.1371/journal.pone.0047995 (2012).
333 Simic, P. *et al.* SIRT1 suppresses the epithelial-to-mesenchymal transition in cancer metastasis and organ fibrosis. *Cell Rep* **3**, 1175-1186, doi:10.1016/j.celrep.2013.03.019 (2013).
334 Morel, A. P. *et al.* Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PloS one* **3**, e2888, doi:10.1371/journal.pone.0002888 (2008).
335 Carriere, P. P. *et al.* Glyceollin I Reverses Epithelial to Mesenchymal Transition in Letrozole Resistant Breast Cancer through ZEB1. *Int J Environ Res Public Health* **13**, ijerph13010010, doi:10.3390/ijerph13010010 (2016).
336 Cha, Y. H., Yook, J. I., Kim, H. S. & Kim, N. H. Catabolic metabolism during cancer EMT. *Archives of pharmacal research* **38**, 313-320, doi:10.1007/s12272-015-0567-x (2015).
337 Warns, J. A., Davie, J. R. & Dhasarathy, A. Connecting the dots: chromatin and alternative splicing in EMT. *Biochem Cell Biol*, 1-14, doi:10.1139/bcb-2015-0053 (2015).
338 Fong, N. *et al.* Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes & development* **28**, 2663-2676, doi:10.1101/gad.252106.114 (2014).
339 Schor, I. E., Rascovan, N., Pelisch, F., Allo, M. & Kornblihtt, A. R. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 4325-4330, doi:10.1073/pnas.0810666106 (2009).
340 Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annual review of biochemistry* **84**, 165-198, doi:10.1146/annurev-biochem-060614-034242 (2015).
341 Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. *Nature structural & molecular biology* **16**, 1128-1133, doi:10.1038/nsmb.1666 (2009).
342 Warzecha, C. C. & Carstens, R. P. Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT). *Semin Cancer Biol* **22**, 417-427, doi:10.1016/j.semcancer.2012.04.003 (2012).
343 Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457-463, doi:10.1038/nature08909 (2010).
344 Singh, R. & Valcarcel, J. Building specificity with nonspecific RNA-binding proteins. *Nature structural & molecular biology* **12**, 645-653, doi:10.1038/nsmb961 (2005).
345 Konieczny, P., Stepniak-Konieczna, E. & Sobczak, K. MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic acids research* **42**, 10873-10887, doi:10.1093/nar/gku767 (2014).
346 Shaul, Y. D. *et al.* Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell* **158**, 1094-1109, doi:10.1016/j.cell.2014.07.032 (2014).

347 Jiang, F., Ma, S., Xue, Y., Hou, J. & Zhang, Y. LDH-A promotes malignant progression via activation of epithelial-to-mesenchymal transition and conferring stemness in muscle-invasive bladder cancer. *Biochemical and biophysical research communications* **469**, 985-992, doi:10.1016/j.bbrc.2015.12.078 (2016).

348 Liu, M., Quek, L. E., Sultani, G. & Turner, N. Epithelial-mesenchymal transition induction is associated with augmented glucose uptake and lactate production in pancreatic ductal adenocarcinoma. *Cancer & metabolism* **4**, 19, doi:10.1186/s40170-016-0160-x (2016).

349 Althammer, S., Gonzalez-Vallinas, J., Ballare, C., Beato, M. & Eyras, E. Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* **27**, 3333-3340, doi:10.1093/bioinformatics/btr570 (2011).

350 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current protocols in bioinformatics* **47**, 11 12 11-34, doi:10.1002/0471250953.bi1112s47 (2014).

351 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

352 Dobin, A. & Gingeras, T. R. Optimizing RNA-Seq Mapping with STAR. *Methods in molecular biology* **1415**, 245-262, doi:10.1007/978-1-4939-3572-7_13 (2016).

353 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

354 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**, 417-419, doi:10.1038/nmeth.4197 (2017).

355 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

356 Guberman, J. M. *et al.* BioMart Central Portal: an open database network for the biological community. *Database : the journal of biological databases and curation* **2011**, bar041, doi:10.1093/database/bar041 (2011).

357 Yates, A. *et al.* Ensembl 2016. *Nucleic acids research* **44**, D710-716, doi:10.1093/nar/gkv1157 (2016).

358 Sahu, S. K. *et al.* TOX3 regulates neural progenitor identity. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1859**, 833-840, doi:https://doi.org/10.1016/j.bbagrm.2016.04.005 (2016).

359 Zentner, G. E. & Scacheri, P. C. The chromatin fingerprint of gene enhancer elements. *The Journal of biological chemistry* **287**, 30888-30896, doi:10.1074/jbc.R111.296491 (2012).

360 Orom, U. A. & Shiekhattar, R. Noncoding RNAs and enhancers: complications of a long-distance relationship. *Trends in genetics : TIG* **27**, 433-439, doi:10.1016/j.tig.2011.06.009 (2011).

361 Capdevila, J. & Izpisua Belmonte, J. C. Patterning mechanisms controlling vertebrate limb development. *Annu Rev Cell Dev Biol* **17**, 87-132, doi:10.1146/annurev.cellbio.17.1.87 (2001).

362 Thakurela, S., Sahu, S. K., Garding, A. & Tiwari, V. K. Dynamics and function of distal regulatory elements during neurogenesis and neuroplasticity. *Genome research* **25**, 1309-1324, doi:10.1101/gr.190926.115 (2015).

363 Sahu, S. K. *et al.* FBXO32 promotes microenvironment underlying epithelial-mesenchymal transition via CtBP1 during tumour metastasis and brain development. *Nature communications* **8**, 1523, doi:10.1038/s41467-017-01366-x (2017).

Publications

- FBXO32 employs CtBP1 to promote microenvironment underlying Epithelial-Mesenchymal Transition during tumour metastasis and brain development.
  Sahu SK, Tiwari N, Pataskar A,., Strand S, Beli P, Tiwari VK.
  Nature Communication. 2017 Nov
- Erk-dependent epigenetic reprogramming underlies Epithelial to Mesenchymal Transition.
  #Navandar M, #Garding A, **Sahu SK**, Pataskar A, Schick S, Tiwari VK.
  Oncotarget. Feb 2017
- TOX3 plays a role in neuronal progenitor proliferation.
  Sahu SK #, Fritz A #, Tiwari N #, Kovacs Z, ., Berninger B, Tiwari VK, Methner A.
  Biochim Biophys Acta. 2016 Apr, (**# Equal contributions**)
- Dynamics and function of distal regulatory elements during neurogenesis and neuroplasticity.
  Thakurela S #, **Sahu SK #**, Garding A, Tiwari VK.
  Genome Res. 2015 Sep, (**# Equal contributions**)
- JNK-dependent gene regulatory circuitry governs mesenchymal fate.
  Sahu SK, Garding A, Tiwari N,., Berninger B, Nitsch R, Schmidt M, Tiwari VK.
  EMBO Journal. 2015 Aug
- The repair of oxidized purines in the DNA of human lymphocytes requires an activation involving NF-YA-mediated upregulation of OGG1.
  von der Lippen C, **Sahu S**, Seifermann M, Tiwari VK, Epe B.
  DNA Repair. 2015 Jan
- Dynamics of chromatin accessibility and epigenetic state in response to UV damage.
  Schick S, Fournier D, Thakurela S, **Sahu SK**, Garding A, Tiwari VK.
  Journal of Cell Sci. 2015 Dec
- Regulation of constitutive & alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery.
  Bezzi M, Teo SX, Muller J, Mok WC, **Sahu SK**, Guccione E.
  Genes Dev. 2013 Sep
- Symmetric di-methylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance.
  Migliori V, Müller J, Phalke S, Low D, Bezzi M, Mok WC, **Sahu SK**, ., Mapelli M, Guccione E.
  Nat Struct Mol Biol. 2012 Jan

Under Review:

- ZNF827 dependent splicing dynamic is critical for EMT during development and disease.
  Sahu SK, Agirre E, Diken M, Strand S, Luco RF, Tiwari VK. **(Under preparation)**