Robustness and Plasticity of Epithelial Cell State in Development and Malignancy

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie vorgelegt der Philosophischen-Naturwissenschaftlichen Fakultät der Universität Basel

Von

Yoana Dimitrova,

Aus Sofia, Bulgaria

Basel, 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Mihaela Zavolan (Dissertationsleitung)

Prof. Dr. Gerhard M. Christofori (Korreferat)

Basel, den 15.09.2015

Prof. Dr. Jörg Schibler (Dekan)

Table of contents

Acknowledgements

I would first like to thank my thesis advisor Prof. Dr. Mihaela Zavolan at the University of Basel for guiding me, advising me and supporting me as a PhD student in her group. I also thank the members of the jury: Prof. Dr. Gerhard M. Christofori for his helpful insights and comments on my research project, Prof. em. Walter Keller, who readily accepted to be the president of the Jury for being so friendly and optimistic, and Dr. Alexander Feuerborn whose remarks were particularly useful. Similarly, I am grateful to my former university research advisors, Prof. Laurent Paquereau, Prof. Anne-Catherine Pratz and Dr. Jefferey Shaw whose examples, as both scientists and mentors, influenced my decision to pursue a career in science.

I am very thankful to Beatriz Dimitriades, William Aaron Grandy, Arnau Vilaseca-Vina, Andreas Gruber, Nitish Mittal, Daniel Mathow, Souvik Ghosh, Georges Martin and Xiaomo Wu for their help for the successful completion of this work. Especially, Aaron, Daniel and Xiaomo were implicated beyond their respective professional engagements and I am particularly appreciative of their support. A special thanks to Bea for our pleasant and motivating talks, and to George for sharing his experience and the incredible knowledge he has with the rest of us. Aaron is a great leader and biologist whose encouragement was very important for my survival as a graduate student. Likewise, Xiaomo's dedication and passion have always been an important example to me. I had the chance to learn a lot from her, and our lunch conversations have always been full of content. Andreas Gruber's help in proofreading and organizing data on top of his robust scientific input and funny jokes was critical. Thanks to Nitish for sharing great dinners, a beautiful wedding and most importantly his wide knowledge. Dominique Jedlinski and Afzal Pasha were both great colleagues and I am happy we could share our PhD time. I would like to thank the rest of my colleagues, to name some of them, Biter Bilen, Christoph Rodak, Jean Hausser, Jérémie Breda, Alexander Kanitz, Foivos Gypas, Ralf Schmidt, Rafal Gummienny, Anneke Brümmer, Joao C. Guimaraes, and Andrzej Jerzy Rzepiela as they have been actively involved in transforming our lab into a friendly and stimulating environment. I also thank the people who were not in Zavolan lab but helped me out. From fear of missing out I will not name them separately but among them are members of the D-BSSE sequencing facility, the Biozentrum's microscopy and proteomics facility, and the chemistry department. I keenly enjoyed the time, conversations and social activities shared with members of the Heinrich Reichert and Erick Van Nimwengen Groups, and in fact with all of my colleagues from the 6th floor of the Biozentrum.

I also show gratitude for my friends' and family's support. I particularly thank Kate for being always there no matter the distance and time. I thank my mom, dad, my sister and my uncle. I have the greatest mother and sister and love them deeply. My family gave me unconditional love and care and I am thankful that I can count on them in tough times. I still miss my father a lot and wish that he could have been here to share those precious moments with me.

I especially thank my husband Mitko. Without his love, patience and understanding this work would have never happened. His precious advices and thoughtful comments, combined with extensive proofreading, profound suggestions and great ideas that he shared are only a negligible part of his everlasting support. There have been plenty of difficult moments and he stayed next to me unconditionally. Thank you!

Last but not least, I want to thank my daughter, Kassia for being the cheerful and great kid she is. Although she joined this adventure at the end, I feel she was a key element to my success. Together with her father they are the most important people in my life. I love you!

Abstract

A better understanding of the molecular mechanisms that control pluripotency, differentiation and epithelial phenotypical plasticity is crucial for the development of the current knowledge in many general processes such as cell identity maintenance and cell fate decision-making.

Embryonic stem cells (ESC) pluripotency maintenance and differentiation are of key importance to the embryonic development, as well as to the progress in stem cells technologies. The role of miR-290-295 cluster members in preserving the pluripotent state and differentiation potential of mouse ESC is well established. Nevertheless, the precise list of targets translating the microRNAs functionality is incomplete. In our study we, firstly, identified and validated miR-290 targets with high confidence. We further confirmed the expression variation of IRF2 in response to miRNAs' depletion in ESC. Moreover, we revisited the involvement of nuclear factor kappa-B (NF-kB) pathway in the miRNA-dependent regulation in mESCs. Hence, our results provided new understanding of the role and mechanistic of miR-290-295 microRNAs involvement in ESC pluripotency and differentiation.

In a similar fashion to ESC pluripotency and differentiation mechanisms, a global analysis-approach that compares and combines data from different epithelial to mesenchymal transition (EMT) models enabled us to construct a more detailed network of regulatory entities implicated in epithelial plasticity. The maintenance and plasticity of the epithelial cell phenotype are important events not only during normal embryonic development, but also to cancer progression and metastasis formation. Comparing this network between mouse and human, we identified a new transcription factor (TF) motif TFAP2A/C that is consistently involved in EMT. When applying the NMuMG cellular model of TGFβ-induced EMT, we found that the predicted activity of the TFAP2A/C is inversely correlated to the *Tfap2a* mRNA expression during the process. We have confirmed that TFAP2A directly binds to the promoter of *Zeb2*, a TF central to EMT. Thus, it regulates the expression of this gene. Furthermore, the TFAP2A overexpression in NMuMG cells modulates the cells' epithelial phenotype and induces changes in cell adhesion and morphology. This

overexpression was followed by increased mRNA levels of EMT master regulator TFs, together with an elevated expression of genes involved in cellular adhesion. Therefore, we identified a potentially new role of TFAP2A transcription factor, which suggests that elements of its regulatory function during neural crest development might operate in mechanisms controlling epithelial plasticity in normal breast and tumor tissues.

Overall, we characterized another facet of microRNAs' function in pluripotency and differentiation in ESC, as well as a new aspect of the implication of TFAP2A in epithelial cell state integrity and plasticity. Therefore we contributed to expanding our insight of how are regulated at molecular level the cell identity homeostasis and the unfolding of cellular phenotypical plasticity.

Introduction

Molecular Basis of Cellular Specialization

The development of vertebrate organisms follows a strictly defined program that gives rise to a multitude of cell types and tissues from a single cell. A complex network of signaling cues, gene expression regulators and epigenetic factors define the fate of individual cells within the developing organism. Thus, even though all cells in an organism carry the same genetic information, they assume cell fates with little overlap in functionality**.** Terminally differentiated cells generally maintain their identity across various conditions and stimuli, but exogenously-driven changes in gene expression can reverse, or even drastically change cell fate. In the late eighties a pioneering study by Walter Gehring introduced the concept of a selector transcription factor, which governs a particular developmental decision (Schneuwly, Klemenz et al. 1987). In an ingenious experiment, he showed that the exogenous expression of a transcription factor, Antennapedia, promotes the development of legs at the place where antennae would normally develop in the fly *Drosophila melanogaster*. In another pilot study the overexpression of the transcription factor Myoblast determination protein 1 (MYOD1) in fibroblasts resulted in their transdifferentiation into myoblasts. These studies strengthen the idea that single genes, also called “master regulators”, are at the top of ”regulatory hierarchies” that define precise cellular states (Davis, Weintraub et al. 1987). Although the mechanisms that induce the differentiation of particular cell types are generally well studied, how cell identity is maintained in response to perturbations is not entirely understood. For example, in the case of postmitotic neuronal cell types the factors that determine the fate of a neuronal cell type are the same responsible for its maintenance: in the absence of inductive signal, autoregulatory feedback processes that involve maintenance-dedicated factors preserve the stability of the differentiated state (Deneris and Hobert 2014). That the maintenance of cell identity is important for organism function is undisputed. Among the various pathologies that are associated with loss of defined

differentiated phenotypes, cancer is perhaps the prototype. Many parallels have been drawn between cancer and metastasis and pluripotency (Goding, Pei et al. 2014). The aim of the project described here was to identify transcriptional and post-transcriptional regulators that can best explain gene expression changes that take place during two paradigmatic processes: embryonic development and epithelial-to-mesenchymal transition. A better understanding of the molecular circuits that underlie the homeostasis and plasticity of cell identity in these circumstances will benefit the general understanding of the differentiation programs that operate both in development and during pathological conditions.

Chapter I. Functions of miR-290 microRNAs in embryonic stem cells

Embryonic Stem Cells

Discovery of Embryonic Stem Cells

After fertilization, the ovum starts traveling through the female reproductive tract, taking approximately six days until it reaches the uterus (Figure 1) (Clift and Schuh 2013). During this time a series of mitotic divisions that do not change the size of the embryo take place. In the end of this process, towards the sixteen-cell stage, the embryo has a berry-like shape and it is called morula, from the Latin mora, meaning mulberry (Alberts 2002). Up to the 8-cell stage, the cells are totipotent, meaning that they can divide and give rise to any of the differentiated cells in the entire organism; they are identical to each other and mutually replaceable, meaning that when single cells are removed, the remaining ones will compensate (Alberts 2002).

Between the third and the forth cleavage (from the 8-cell to 16-cells stage), the previously poorly organized embryo will engage into a compaction process that will result in the separation of the cells to an outer and inner set.

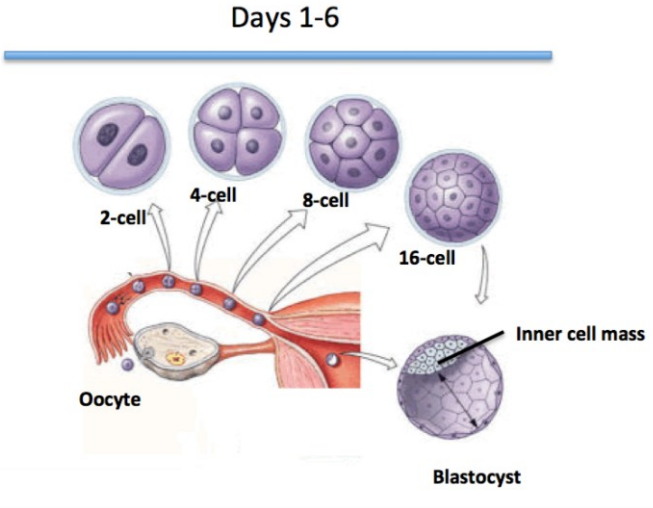

Figure 1. First six days of human embryogenesis. After fertilization of the oocyte, in approximately five to six days, the embryo divides, migrates and forms the blastocyst. Adapted from Fundamentals of Anatomy & Physiology, 7e By Frederic H.Martini, Copyright Pearson Education, Published by Benjamin Cummings, 2005.

After the compaction, the inner and outer cells will commit to different fates. The cells that constitute the wall of the sphere will give rise to extra-embryonic tissues, namely the trophectoderm. The cells from the inner cell mass will give rise to all tissues and organs of the adult organism (Alberts 2002). An internal fluid-filled cavity is created and the embryo is called blastocyst (Figure 1). Following the compaction is the gastrulation phase, in which the single layered blastula transforms into a trilaminar gastrula, which is composed of three germ layers: ectoderm, mesoderm and endoderm. Cells derived from the inner cell mass, also called embryonic stem cells (ESCs), can be explanted from the embryo and cultured *in vitro* (Evans and Kaufman 1981; Martin 1981). Although murine ESCs (mESCs) are extensively studied currently, the path to obtain these cells was quite long.

In 1954, Stevens and Little described a spontaneous testicular teratoma in mice, a complex tumor formation that contains a range of differentiated cells and tissues (Stevens and Little 1954). They argued that teratomas are composed of both undifferentiated pluripotent embryonic cells, as well as different cells of various types. They further determined that the pluripotent cells, which they called embryonal carcinoma (EC) cells, are able to give rise both to differentiated cells as well as self-renew (Stevens and Little 1954). Similarly, when early mouse embryos were grafted in adult animals, they generated teratomas (Solter, Skreb et al. 1970). Following these initial studies, much effort has been put into optimizing the growth conditions for embryonic cells in culture so that their properties and their differentiation *in vitro* can be studied (Evans 2011). In 1975, Minz and Illmensee generated a chimeric mouse by injecting EC cells in the mouse blastocyst. However, due to karyotypic abnormalities often present in the EC cells, the chimerism was never observed at the germ cells level (Mintz and Illmensee 1975). EC cells were found to have highly similar properties to normal non-cancerous cells from the inner cell mass of the blastocyst, for instance they are able to form embryoid bodies (EB) *in vitro* (Martin and Evans 1975). This finding paved the way to the isolation and culturing of ESCs (Martin and Evans 1975; Evans 2011). The first ESCs from mouse were cultured in the beginning of the eighties. (Evans and Kaufman 1981; Martin 1981). However, it took much longer until human ESCs were obtained in 1998 (Thomson, Itskovitz-Eldor et al. 1998).

ESCs are pluripotent, meaning that they are able to form any of the tissues and organs of the entire organism, except those forming the placenta and certain parts of the

embryo (Figure 3). Furthermore they are able to differentiate into any of the three germ layers. In addition, as any stem cell, they are capable of self-renewal, and thus they can also produce additional stem cells (Evans 2011). In contrast to EC cells, when injected in the mouse blastocyst, mouse ESCs give rise to chimeric mice, including their germ-lines (Robertson, Bradley et al. 1986). As a consequence, the *in vitro* modified genetic material of ESCs can be used to generate fully mutant animal by germ-line transmission. This in turns allows the study of target genes functions (Evans 2011). Stem cells form a topic of great current interest. They enable basic research on understanding the development and function of human cells and are expected to have a prospectively important role for testing drugs safety and efficacy. In addition they are highly relevant for the future of regenerative medicine, as they represent a promising source of tissue and cells in replacement therapies (Evans 2011; van Berlo and Molkentin 2014)

Pluripotency Maintenance in Cell Culture

In the years following their discovery, mouse ESC were maintained in culture together with a feeder layer of embryonic cells that were treated such that could not divide anymore (Evans and Kaufman 1981; Martin 1981). Later studies found that the factor that feeder cells provided and was important for maintaining mESC pluripotency *in vitro* is LIF (Leukemia inhibitory factor). When combined with fetal calf serum (FCS), LIF bypasses the requirement for feeder cells (Martin and Evans 1975; Smith, Heath et al. 1988; Williams, Hilton et al. 1988). When LIF is withdrawn, mESC still proliferate but their differentiation is induced, suggesting that LIF presence in culture media supports mESC's self-renewal capacity (Smith 2001). LIF is a cytokine, member of the interleukin 6 (IL-6) family, and it interacts with its corresponding transmembrane receptor Leukemia inhibitory factor receptor (LIFR). Upon ligand binding, LIFR dimerizes with Interleukin 6 signal transducer IL6ST/gp130 receptor, which further transduces the signal. The effect of LIF on mESC pluripotency is subsequently mediated via Janus kinases (JAKs)-dependent activation of the transcription factor Signal transducer and activator of transcription 3 (STAT3) (Niwa, Burdon et al. 1998; Smith 2001) (Figure 2). Furthermore STAT3 activation alone is sufficient to sustain mESC pluripotency (Matsuda, Nakamura et al.

1999). Despite the critical role of LIF-gp130 in maintaining cell cultures of mESC, the developing embryo is not dependent on this signaling path prior to gastrulation. Instead, this pathway is important in a process named diapause, in which a lactating mouse female is fecundated and the embryo development is blocked at the blastocyst stage before implantation, until the mother's hormone levels are reestablished (Nichols, Chambers et al. 2001; Smith 2001). To avoid the use of Fetal Calf Serum (FCS) which is heterogeneous in composition and therefore an important source of variability, the bone morphogenetic protein 4 (BMP4) can be added to the cell culture and thus allow the growth of mESC in chemically defined medium (Ying, Nichols et al. 2003).

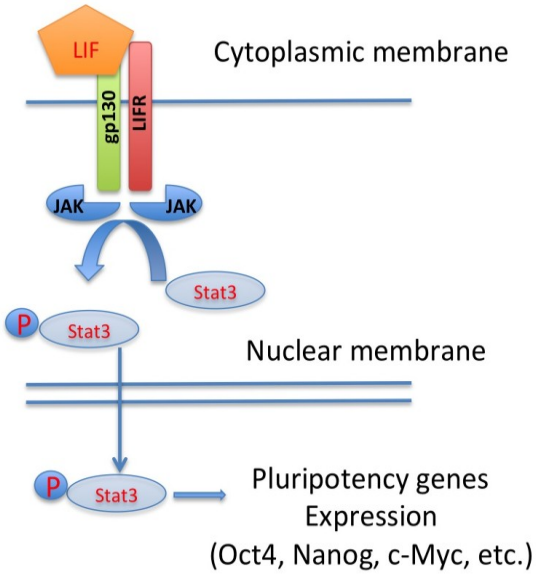

Figure 2. LIF signaling maintains the expression of pluripotency genes. Binding of LIF to its cellular receptor, which is a homodimer of LIFR and gp130, triggers STAT3 phosphorylation and concomitant signal transduction to the nucleus, where the expression of pluripotency genes is activated. Adapted from (Arabadjiev 2012)

The optimal culture conditions for human ESC (hESC) are surprisingly different than those for mESCs, and human ESCs differ from mouse ESC in their molecular profile, morphology and differentiation potential (Nichols and Smith 2009). Neither LIF addition, nor STAT3 activation, via gp130 receptor signaling, are sufficient for preserving hESC pluripotency in the absence of a feeder layer of mouse embryonic fibroblasts (MEF) (Humphrey, Beattie et al. 2004). MEFs can be replaced with

Matrigel® or similar secreted gelatinous protein mixtures, in the presence of MEF conditioned media (Xu, Inokuma et al. 2001). However, as Matrigel® is produced by Engelbreth-Holm-Swarm (EHS) mouse sarcoma cells, it has limitations and cannot be used when hESC are employed in clinical applications. Matrigel® like substances exhibit extensive lot-to-lot variability and can lead to xenogenic contamination (Villa-Diaz, Ross et al. 2013). Therefore extensive effort is currently being directed towards development of feeder-free, chemically defined conditions for the establishment and expansion of human pluripotent stem cells (hPSC) (Ludwig, Levenstein et al. 2006; Chen, Gulbranson et al. 2011; Rodin, Antonsson et al. 2014). Furthermore, adhesion independent suspension culture methods are of high interest for large scale derivation and propagation of hESC (Steiner, Khaner et al. 2010).

Pluripotency in Different Embryonic States

The embryonic stem cells that are derived from the cells of the inner cell mass (ICM) and are the progenitors of the epiblast, define the naïve pluripotent state (Evans and Kaufman 1981; Martin 1981). Shortly after the blastocyst stage the ICM will separate into two cell types: the epiblast, from which the embryo proper is formed, and the primitive endoderm, which gives rise to extra-embryonic tissues (Najm, Chenoweth et al. 2011). In the pre-implantation epiblast of female embryos both X chromosomes are active. This property is specific to the naïve pluripotent state. Upon implantation, the epiblast is subject to a series of developmental signals that will result in its conversion into a layer of epithelium, in parallel with random inactivation in one of the X chromosomes in XX epiblasts. Subsequently, the cells of this epithelium are subject to location-driven specification. The cells that constitute the post-implantation epiblast maintain a high degree of plasticity and their fate can be reoriented at this stage (Nichols and Smith 2009). However, in contrast to cells originating from the inner cell mass, post-implantation epiblast cells cannot give rise to chimeras when injected into blastocysts (Rossant 2008). Mouse pluripotent cells from the post-implantation epiblast, EpiSCs, have been already isolated and can be cultured in the presence of Fibroblast growth factor (FGF) and Activin instead of LIF (Brons, Smithers et al. 2007; Tesar, Chenoweth et al. 2007). Consistent with their pluripotency, EpiSC are efficient in teratoma formation (Tesar, Chenoweth et al.

2007). Furthermore, as expected from their origin, double X EpiSCs have one inactivated copy of X chromosome. EpiSCs represent the so-called primed pluripotent state (Nichols and Smith 2009). Through the exogenous expression of a single transcription factor, namely Kruppel-like factor 4 (KFL4), these cells can be reprogrammed to the naïve pluripotent state. The transition from mESC to EpiSC is achieved with growth factors cues (Guo, Yang et al. 2009). For ethical reasons hESC cannot be tested for their ability to form chimeras (Nichols and Smith 2009). However they are shown to be able to engraft into mouse blastocyst, and in certain cases they can undergo gastrulation and form human/mouse embryonic chimeras (James, Noggle et al. 2006). Although hESC have similar embryological origin as mESCs, they reassemble in many aspects EpiSC and are considered to be in primed rather than in ground pluripotent state (Mascetti and Pedersen 2014). It was recently demonstrated that human blastocyst inner cell mass derived cells, when kept in NHSM (naïve human stem cell medium), which contains LIF, a combination of other growth factors and small molecule inhibitors of core signaling pathways, display more similarities to mESC and are thought to preserve their ground state pluripotency. Furthermore, ICM-like hESCs significantly outperform the previously derived hESC cell lines in their ability to generate interspecies chimeras (Gafni, Weinberger et al. 2013).

ESC Differentiation

When placed in relevant growth conditions, ESCs can give rise to cells of any of the three germ layers (Figure 3). In contrast to mESCs, hESCs can also give rise to a population of cells that shares many characteristics with trophoblasts, when stimulated with BMP4 (Xu, Chen et al. 2002).

Different methods exist to promote the differentiation of ESC. The most widely used method is aggregation of ESC in suspension that results in the formation of a three-dimensional (3D) structure known as embryoid body (EB). This strategy was initially developed for the culture of EC cells and can also be applied to ESC (Martin and Evans 1975). The differentiation of EBs resembles in many aspects the developmental program that the ICM cells of the embryo undergo. However, a major difference is that the EBs lack a correct axial organization and body plan, and do not have

appropriate organization of the three germ layers. A recent improvement of this technique combines the growth of the embryoid colony in a 3D fibrin gel with a consequent step of cell anchorage to a collagen coated two-dimensional (2D) support, which promotes the proper germ layer organization of EBs (Poh, Chen et al. 2014). Other approaches consist in co-culturing the ESC on stromal cells that will stimulate their differentiation or using a layer of extracellular matrix proteins (Keller 2005). A series of cell types originating from any of the three germ layers: the mesoderm, the endoderm and the ectoderm can be produced from ESC. The differentiation of mesodermal cells gives rise to hematopoietic, vascular, cardiac, skeletal muscle, osteogenic, adipogenic, and chondrogenic lineages (Keller 2005; Salani, Donadoni et al. 2012; Slukvin 2013; Barad, Schick et al. 2014). With respect to the endoderm, pluripotent stem cells (PCS) were used to obtain various cell types from the gastrointestinal and respiratory tract, as well as hepatocytes, pancreatic cells and thyroid follicular cells (Kadzik and Morrisey 2012; Cheng, Tiyaboonchai et al. 2013; Sewell and Lin 2014; Sinagoga and Wells 2015). Concerning ectoderm-derived lineages, protocols that establish neuroectoderm and epidermis commitment are well defined. The neural differentiation leads to the three major cell types present in the central nervous system: neurons, astrocytes and oligodendrocytes. Furthermore, engendering of specialized neuronal sub-types such as dopaminergic, cholinergic and glutaminergic neurons is also possible (Keller 2005). An exciting recent development is the generation of organ-like tissues, named “organoids” through 3D cell cultures methods (Shamir and Ewald 2014). Organoids are structurally similar to the model organs, can be composed of cells derived from different germ layers and are formed from various cell lineages. Eyecup, gut, brain, kidney, liver and lung are amongst the successfully produced organoids. These experimental models are instrumental for better understanding of organ development and function in healthy, or pathological conditions (Shamir and Ewald 2014; Dye, Hill et al. 2015).

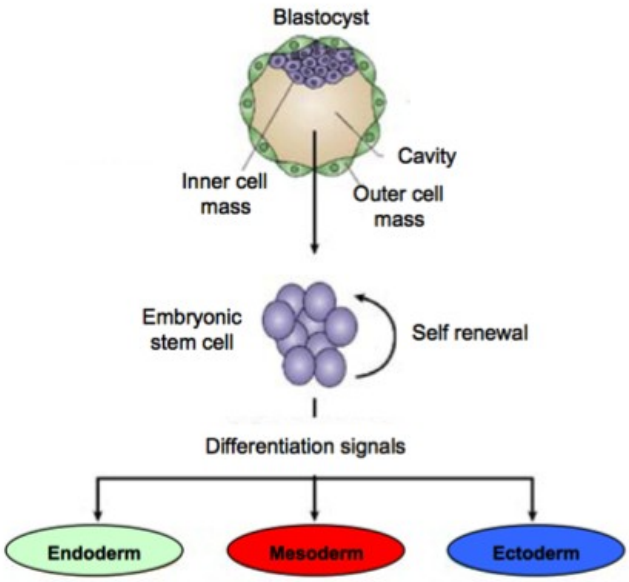

Figure 3. Embryonic stem cells differentiation potential. Adapted by permission of Macmillian Publishers Ltd Nature Reviews Genetics (O'Connor and Crystal 2006), copyright 2006. Embryonic stem cells are explanted form the blastocyst. They can be maintained in cell culture for indefinitely long periods of time or can be differentiated into any of the three germ layer-derived cell types.

Induced Pluripotency

In 2006, a crucial discovery changed the landscape of molecular and cellular biology. By expressing a cocktail of four transcription factors: Octamer-binding protein 4/OCT4, homeobox protein NANOG, transcription factor SOX2 and Myc proto-oncogenic protein/MYC (OSKM), in mouse fibroblasts Takahashi and Yamanaka obtained cells that were in many aspects similar to embryonic stem cells (Takahashi and Yamanaka 2006). This finding opened new avenues in stem cell research. The rapidly developing technology of generating induced Pluripotent Stem Cells (iPSC) promises to offer an alternative solution for disease modeling, drug discovery and regenerative medicine (Yamanaka 2012). In 2007 the successful reprogramming of human fibroblasts to iPSCs allowed the generation of cells that reassembled but

circumvented the ethical issues associated with hESCs. Moreover iPSCs have the important advantage of a matched genetic background with the patient from which the fibroblasts were initially isolated. iPSCs provide a unique opportunity in disease modeling and drug discovery, as they allow for the generation of virtually any cell type from a given subject and the use of large number of genetically variable cell lines and tissues in drug screening assays (Park, Arora et al. 2008).

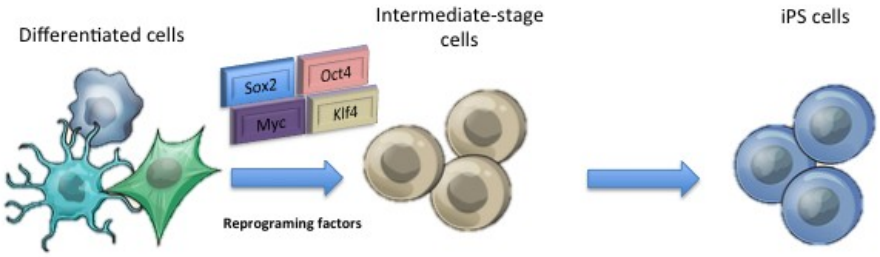

Figure 4. Induced pluripotency stem cells technology. Terminally differentiated cells can be reprogrammed, using a combination of four reprogramming transcription factors. Adapted by permission of Macmillian Publishers Ltd Nature (Loh and Lim 2013), copyrights 2013.

Indeed, many disease-modeling studies followed. IPSCs derived from people suffering of amyotrophic lateral sclerosis (ALS) were used to better understand the mechanism of the disease and to test for potential drugs (Dimos, Rodolfa et al. 2008). ALS is a devastating neurodegenerative disorder and more than 100 mutations in dozens of genes are known to be in its origin. Motor neurons derived from iPSCs of patients with different forms of ALS presenting different set of mutations allowed the identification of potential general mechanism of the disease. With the application of the same methodology an anti-epileptic drug showed promising results and will be further studied in clinic (Dimos, Rodolfa et al. 2008; Kiskinis, Sandoe et al. 2014; Wainger, Kiskinis et al. 2014). IPSC-derived cardiac myocytes and hepatocytes can be used as an alternative to test drug toxic effects (Yamanaka 2009). Another major future application of the stem cell technology is in regenerative medicine. In 2007 autologous iPSCs were used for the first time successfully in the treatment of sickle cell anemia in mice (Hanna, Wernig et al. 2007). Currently research is being directed towards making the iPSC technology useful in the treatment of macular degeneration,

spinal cord injuries, Parkinson's disease, and platelet deficiency (Takayama, Nishimura et al. 2010; Kriks, Shim et al. 2011; Nori, Okada et al. 2011; Okamoto and Takahashi 2011). The research on regeneration of cardiac tissue progressed very quickly in the last decade. Myocardial infraction (MI) or advanced heart failure, leads to the destruction of cardiac tissue and important loss of cardiomyocytes, which are leading cause of death. ES and iPS cells based methods currently show the highest regenerative potential for therapeutic cardiac regeneration (van Berlo and Molkentin 2014). Both strategies were tested in animal models and were demonstrated to be efficient (Kawamura, Miyagawa et al. 2012; Chong, Yang et al. 2014). However, prior to any of these approaches advancing towards the clinic, important safety questions related to the technology, namely the long term genetic stability of iPSC need to be addressed. Furthermore, proving that every cell in the treatment suspension is differentiated enough not to form cancer or teratoma remains challenging (van Berlo and Molkentin 2014).

Regulation of pluripotency

The stable pluripotent state results from a balance of signals promoting stemness and inhibiting differentiation (Smith 2001). Extracellular and intrinsic signaling are integrated by a network of molecules that involves complex interactions between transcription factors, RNA binding proteins, small and long non-coding RNAs and other regulators of gene expression (Young 2011). This tangled molecular circuitry is responsible for maintaining the epigenetic and transcriptional landscape of pluripotent stem cells (PSCs) in a ground state.

Transcriptional Control of Pluripotency

Transcription factors are proteins that interact directly or indirectly with DNA and thus activate or suppress the transcription of different genes. They can bind to elements that are either proximal to the promoter or distal, 100s of kb away (Young 2011).

Core Pluripotency Factors

Three transcription factors are known as the core pluripotency transcription factors that are responsible for maintaining stemness of PSC: OCT4, SOX2 and NANOG (Young 2011; Theunissen and Jaenisch 2014).

Murine OCT4 is encoded by *Pou5f1* gene and it belongs to the POU family of homeodomain transcription factors. Its expression during mouse development is restricted to the blastomere before compaction, the pre- and post- implantation epiblast and the primordial germ cells (Young 2011). In OCT4-deficient mouse embryos, the blastocyst forms, but the ICM cells are not pluripotent and rather committed to the trophoblast lineage (Nichols, Zevnik et al. 1998). The OCT4 levels regulate ESC's fate. Increased OCT4 levels induce ESC differentiation towards primitive endoderm and mesoderm, while its repression leads to trophectoderm specification (Niwa, Miyazaki et al. 2000). OCT4 acts in concert with SOX2 to maintain pluripotency and induce mesendoderm determination. SOX2 is a member of Sox (SRY-related HMG (High Mobility Group) box) and it is considered as transcriptional partner of OCT4. Consistently, the SOX2 expression pattern in the early development is highly similar to that of OCT4. The SOX2 protein and mRNA are found in the epiblast and primordial germ cells. In contrast, however, to OCT4, SOX2 is equally expressed in the post-implantation extra-embryonic ectoderm that further develops to mature placenta (Avilion, Nicolis et al. 2003). A further confirmation of the synergy between SOX2 and OCT4 is the similarity in the phenotype of their knockout in mouse embryo, which in both cases results in a failure to establish a pluripotent ICM population (Avilion, Nicolis et al. 2003; Yeo and Ng 2013). In mouse embryos, NANOG is expressed only in the parietal germ cells and the pluripotent cells populations that arise after compaction. *Nanog* is rapidly silenced upon the specification of these lineages. Homozygous *Nanog*-knockout mice lack defined epiblast cells population upon implantation (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003). In contrast *Nanog*-null homozygous mESC can self-renew indefinitely, although with lower efficiency, and without committing into EpiSC (Chambers, Silva et al. 2007). Despite the fact that NANOG is dispensable for

mESC pluripotency, it stabilizes their undifferentiated state(Chambers, Silva et al. 2007).

The core transcription factors exert their control over pluripotency in a cooperative manner. They act in concert to regulate the expression of their own genes and thus establish a positive auto-regulatory feedback loop. In addition, they co-regulate the genes involved in preserving pluripotency and at the same time contribute to the repression of those promoting differentiation (Young 2011).

Extended Network of Transcription Factors.

Though less important, other transcription factors were found to complement OCT4/SOX2/NANOG in the regulation of pluripotency. KFL4 and c-Myc/MYC are part of the Yamanaka's cocktail of factors that reprogram fibroblasts to pluripotent stem cells similar to ESC and have important roles in ESC pluripotency maintenance (Takahashi and Yamanaka 2006). In addition, studies aiming to understand the regulators of pluripotency identified more transcription factors to be tightly involved in governing this state together with the core transcription factors. A non- exhaustive list includes REX1, TCF3, SMAD1, STAT3, ESRRB, ZFX, Ronin/THAP11, KLF2, KLF5, SALL4, PRDM14, TCL1, DAX1, NAC1, ZFP281 and others (Chia, Chan et al. 2010; Kim, Woo et al. 2010; Young 2011) (Table 1). Some of them, like PRDM14, are specific for hESC and are dispensable for mESC pluripotency (Chia, Chan et al. 2010).

A recent systemic study of protein-protein interactions, as well as protein-DNA interaction of some of the pluripotency related factors in ESC, concluded that in their functional network, three regulatory cores exist (Kim, Woo et al. 2010). The first one is composed of the core pluripotency transcription factors and a number of other transcription factors, and it is responsible for activation of pluripotency related genes. The second one is focused around Polycomb Repressive Complex (PRC) and the genes that are part of its regulatory module are repressed in ESC. Finally, the third one is centered on MYC and forms a module of transcription factors that act together with it to positively regulate a subset of genes involved in PSC self-renewal and maintenance (Kim, Woo et al. 2010).

Transcription factors or cofactors	Gene function	Reference
OCT4	Core circuitry	(Nichols, Zevnik et al. 1998)
SOX2	Core circuitry	(Chambers and Smith 2004; Masui, Nakatake et al. 2007)(Avilion, Nicolis et al. 2003)
NANOG	Core circuitry	(Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003)
TCF3	WNT signaling to core circuitry	(Cole, Johnstone et al. 2008; Marson, Foreman et al. 2008)
STAT3	LIF signaling to core circuitry	(Niwa, Burdon et al. 1998)
SMAD1	BMP signaling to core circuitry	(Ying, Nichols et al. 2003)
SMAD2/3	TGFβ/Activin/Nodal signaling	(Beattie, Lopez et al. 2005; James, Levine et al. 2005)
MYC	Proliferation	(Cartwright, McLean et al. 2005)
ESRRB	Steroid hormone receptor	(Ivanova, Dobrin et al. 2006)
SALL4	Embryonic regulator	(Zhang, Tam et al. 2006)
TBX3	Mediates LIF signaling	(Ivanova, Dobrin et al. 2006)
ZFX	Self-renewal	(Galan-Caridad, Harel et al. 2007)

Ronin	Metabolism	(Beattie, Lopez et al. 2005)
KFL4	LIF signaling	(Jiang, Chan et al. 2008)
PRDM14	ESC identity	(Chia, Chan et al. 2010)
Mediator	Core circuitry	(Hu, Kim et al. 2009; Kagey, Newman et al. 2010)
Cohesin	Core circuitry	(Hu, Kim et al. 2009; Kagey, Newman et al. 2010)
PAF1 complex	Couples transcription with histone modification	(Ding, Paszkowski-Rogacz et al. 2009)
DAX1	OCT4 inhibitor	(Niakan, Davis et al. 2006; Sun, Nakatake et al. 2009)
CNOT3	MYC/ZFX cofactor	(Hu, Kim et al. 2009)
TRIM28	MYC/ZFX cofactor	(Fazzio, Huff et al. 2008; Hu, Kim et al. 2009)

Table 1. Transcription factors and cofactors implicated in the regulation of pluripotency in ESC. (Young 2011).

Function of MYC in Pluripotency Regulation

Myc transcription factors belong to the family of basic helix-loop-helix (bHLH) is composed of three members, namely MYC/c-Myc, MYCN/n-Myc and MYCL/L-Myc. They are well known oncogenes and are deregulated in many cancers (Luscher and Vervoorts 2012). Myc proteins bind to E-box elements in DNA and

heterodimerize with a protein named MAX. The interaction with MAX is critical for most MYC functions. MAX can also bind to some of the MAX dimerization proteins, Mxd/MAD, such as MXD1 MNT, and MGA, and thus MAD proteins antagonize MYC (Luscher and Vervoorts 2012). Different mechanisms are suggested to explain how MYC can mechanistically activate gene expression. In particular, those include promoter activation by recruitment of histone acetyl transferases (HATs) such as TIP60/KAT5 and GCN5/KAT2A, RNA polymerase II pause release mediated by p-TEFb, and a recent model, in which MYC is rather amplifying the already existing expression of genes than initiating *de novo* transcription (Cole and Nikiforov 2006; Rahl, Lin et al. 2010; Lin, Loven et al. 2012; Nie, Hu et al. 2012).
In mouse embryonic development, as well as in ESC, *Myc* deletion did not affect pluripotency (Sawai, Shimono et al. 1991; Stanton, Perkins et al 1992; Davis, Wims et al. 1993). However, when *Myc* and *Mycn* are simultaneously knocked out in ESC, the cells switch towards a differentiated state (Varlakhanova, Cotterman et al. 2010). Additionally, the overexpression of MYC in mESC replaces the need of LIF addition in culture media, which further underlines the importance of these transcription factors in supporting self-renewal and pluripotency (Cartwright, McLean et al. 2005).

Studies of transcription factors-DNA binding in ESC determined that gene targets occupied by core transcription factors differ from those regulated by MYC. Therefore it is suggested that MYC, together with other transcription factors, regulates gene expression in a regulatory module in the transcription factor network of pluripotency (Kim, Woo et al. 2010).

Epigenetic Landscape of Pluripotency

In the nucleus, the DNA is embedded together with proteins and RNA molecules in a structure called chromatin. The DNA is tightly packed in nucleosomes containing 147 nucleotides wrapped around a core of histone proteins. The nucleosomes are further stacked in complex 3D organisations (Woodcock and Ghosh 2010). Post-translational modifications in histones recruit different factors, affect the way nucleosomes are

compacted, and can further influence higher-order chromatin structures (Tee and Reinberg 2014).

Epigenetic regulation encompasses various mechanisms that act on the structure and accessibility of DNA to modify gene expression. They include DNA methylation, histone modification and chromatin remodelling (Tee and Reinberg 2014). Generally, in PSCs genes that are involved in commitment decisions are maintained in a silenced state, while still responsive to the appropriate signals (Han and Yoon 2012).

Chromatin Regulators	Gene function	Reference
Polycomb	Silencing of lineage-specific regulators	(Boyer, Plath et al. 2006; Lee, Jenner et al. 2006)
SETDB1 (ESET)	Silencing of lineage-specific regulators	(Bilodeau, Kagey et al. 2009)
esBAF	Nucleosome mobilization	(Ho, Ronan et al. 2009)
CHD1	Nucleosome mobilization	(Gaspar-Maia, Alajem et al. 2009)
CHD7	Nucleosome mobilization	(Schnetz, Handoko et al. 2010)
TIP60-p400	Histone acetylation	(Fazzio, Huff et al. 2008)

Table 2. Chromatin regulators, implicated in the regulation of pluripotency in ESC. (Young 2011)

The input from different signalling pathways can be integrated by changes in chromatin structure and in certain cases the signalling molecules directly interact with

the chromatin. In mESC JAK kinases, also involved in signalling downstream of LIF, can phosphorylate the tyrosine residue 41 of histone H3, and thus impede Heterochromatin protein 1α (HP-1α)/CBX5 interplay with chromatin to consequently affect the core pluripotency factors expression (Griffiths, Li et al. 2011; Ye and Blelloch 2014). In another study, a JNK effector of MAPK pathway was shown to phosphorylate histone H3 at serine (Ser3), in the course of ESC differentiation into neurons (Tiwari, Stadler et al. 2012). Moreover, a SWI/SNF-like ATP dependent chromatin remodelling complex, named esBAF, is essential for mESC maintenance and cooperates with LIF signalling by promoting genome-wide STAT3 binding (Ho, Ronan et al. 2009; Ho, Miller et al. 2011). A number of other chromatin regulators were found to be essential for ESC functionality and among those are cohesion/condensin complexes (Young 2011). In addition, histone modifying regulators, such as polycomb group protein (PcG) complexes, TIP60-p400 and SETDB1, are critical for pluripotency regulation (Table 2) (Young 2011).

RNA Binding Proteins

RNA binding proteins (RBP) participate in a large number of functions related to RNA processing, including splicing, poly-adenylation, nuclear export, translation, modification of RNA molecules, and degradation. Thus, it is nor surprising that RBPs are also associated with the pluripotency network, either promoting the undifferentiated state or differentiation along various lineages (Ye and Blelloch 2014). The role of different RBPs in pluripotency is reviewed in detail in (Ye and Blelloch 2014). An interesting example is that of Mettl3/MTA70 and Mettl14/MET14, two mammalian methyltransferases, which transfer the methyl group of *S*-adenosyl-*L*-methionine to produce N^6-methyladenosylated RNA (m^6A RNA) (Liu, Yue et al. 2014). When any of these two RBPs is depleted, mESC exhibit a reduction in m^6A RNA methylation and impaired self-renewal (Wang, Li et al. 2014). Transcripts enriched in this modification correspond to developmental regulators with particular chromatin state, where m^6A incorporation has a destabilizing effect on the RNA. MBNL proteins were shown to block ES-cell-specific alternative splicing and reprogramming and were shown to act in synergy with RBFOX2 to create a splicing program into iPSCs differentiation (Han, Irimia et

al. 2013; Venables, Lapasset et al. 2013). The Epithelial splicing regulatory protein 1, ESRP regulated the expression of pluripotency-associated regulators (Fagoonee, Bearzi et al. 2013). These results suggest that RBPs and importantly splicing factors are relevant for cell identity determination (Ye and Blelloch 2014).

microRNAs

MicroRNAs are small non-coding RNA of approximately 22 nucleotides in length that play important roles in vertebrate development, as well as in diverse physiological and cellular processes (He and Hannon 2004; Bartel 2009; Mencia, Modamio-Hoybjor et al. 2009; de Pontual, Yao et al. 2011). In mammals, microRNAs are loaded into Ago proteins and in most of the cases the microRNA-guided Ago protein will bind to a region in the mRNA 3'UTR that can have as few as 6 nucleotides complementarity to the 5` bases 2-7 of the microRNA, which is the so-called "microRNA seed region". Once the ribonucleoprotein RNA-induced silencing complex (RISC) complex, containing the microRNA loaded Ago, is localized on its mRNA target, the corresponding mRNA is destabilized, and the expression of the corresponding protein decreases (Lingel and Izaurralde 2004; Filipowicz, Bhattacharyya et al. 2008; Chekulaeva, Mathys et al. 2011) (Fabian, Sonenberg et al. 2010).

MicroRNAs are derived from either independently regulated genes that are transcribed by RNA polymerase II or from processing of introns of genes that produce protein-coding or non-coding RNAs (Lee, Kim et al. 2004; Borchert, Lanier et al. 2006; Bortolin-Cavaille, Dance et al. 2009). Different type-III RNases are involved at different steps of miRNA biogenesis. Primary microRNA transcripts can be processed by Drosha-DGCR8 ribonuclease complex in the nucleus, or in the case of mirtons by the splicing machinery, and result in precursor microRNA (pre-microRNA) hairpins that are translocated into the cytoplasm by Exportin5 (Yi, Qin et al. 2003; Bohnsack, Czaplinski et al. 2004; Han, Lee et al. 2004; Ruby, Jan et al. 2007). In the cytoplasm, pre-microRNAs are further processed by the nuclease Dicer to give rise to double-stranded RNA molecules with 5' monophosphate and 3' hydroxyl groups (Hutvagner, McLachlan et al. 2001). One of the two strands, named guide strand, is incorporated

into the RISC and targets it to complementary messenger RNA (mRNA) (Kai and Pasquinelli 2010).

MicroRNAs (miRNAs) are critical for mammalian embryonic development, and this is exemplified by the lethality of Dicer and DGCR8 knockouts (Bernstein, Kim et al. 2003; Wang, Medvid et al. 2007). In PSC, a set of microRNAs is specifically associated with the undifferentiated cellular state. The list of microRNAs specifically enriched in the pluripotent state includes the clusters of miR-17-92, miR-302-367, hsa-miR-371-373, and its mouse orthologue mmu-mir-290-295, the miR-200 family, as well as the miR-106 and miR-195 miRNAs. A particular case is the C19MC cluster of embryonic microRNAs, encoding the miR-520 family that is only represented in human and primates (Suh, Lee et al. 2004; Bar, Wyman et al. 2008; Laurent, Chen et al. 2008; Morin, O'Connor et al. 2008). Interestingly, individual microRNAs or clusters, such as miR-17, miR-106, hsa-miR-372/mmu-miR-290, miR-302, and miR-520, have the same or very similar seed sequence that might differ in one nucleotide despite the fact that some of them have different evolutionary origin. As the seed sequence is critical for target recognition this observation suggests that there is important number of mRNA that are regulated by all of the above-mentioned microRNAs (Leonardo, Schultheisz et al. 2012). Some of the embryonic microRNAs, including miR-17-92, miR-106a-25, mmu-miR-290/ hsa-miR-372, miR-302-367, and miR-200, promote reprogramming, when expressed together with reprogramming factors (Leonardo, Schultheisz et al. 2012; Wang, Guo et al. 2013). Moreover, it has been reported that lentiviral expression of mmu-miR-302-367 or transient transfection of all three hsa-miR-200c, hsa-miR-302, and hsa-miR-363 are sufficient to reprogram respectively mouse or human fibroblast to iPSC (Anokye-Danso, Trivedi et al. 2011; Miyoshi, Ishii et al. 2011). Consistently, it has been argued that core pluripotency and OSKM reprogramming transcription factors promote the undifferentiated state partly by inducing the expression of microRNAs. For instance, OCT4, SOX2 and NANOG bind to the promoter of miR-106a-363, mmu-miR-290 and miR-302-367 in mESC (Marson, Levine et al. 2008). Myc is equally shown to transactivate the expression of miR-17-92 (O'Donnell, Wentzel et al. 2005).

Another set of miRNAs that belong to the let-7 family are related to various differentiation pathways and can negatively affect pluripotency. For instance, they inhibit reprogramming by stimulating the expression of prodifferentiation factors (Worringer, Rand et al. 2014).

Embryonic specific miRNAs are known to functionally enforce ESC cell cycle and glucose metabolism, to regulate mesenchymal to epithelial transition (MET) during reprogramming, to control DNA methylation in ESC, to influence m^6A modification of RNA and to affect PSC apoptosis (Miyoshi, Ishii et al. 2011; Leonardo, Schultheisz et al. 2012; Cao, Guo et al. 2015).

ESCs have a particular cell cycle, with a shortened G1 phase. ESC depleted in components of miRNA biogenesis pathway, and therefore, deficient in miRNAs, exhibit a cell cycle defect, which can be partially reversed by the transfection of mmu-miR-290 family members in mouse or hsa-miR-372 and hsa-miR-195 in hESC (Wang, Baskerville et al. 2008; Qi, Yu et al. 2009). MiRNAs from miR-290 family were shown to regulate self-renewal of mESC by regulating MYC, Lin28 and Sall4, while the let-7 family members controlled the same genes in an opposite manner (Melton, Judson et al. 2010). Surprisingly, a recent study of DGCR8-null mESC found that the mutant embryonic cells resemble ground state pluripotent stem cells cultured in 2i +LIF, a medium in presence of inhibitors of Extracellular regulated kinases ERK/MAPK and Glycogen synthase kinases GSK3, suggesting that microRNAs might not be critical for the ground state pluripotency (Kumar, Cahan et al. 2014). Furthermore, wild type ground state pluripotent stem cells express let-7 family members together with miR-290-295, suggesting that the two miRNA families might act synergistically to maintain this state. However, miRNA depletion in mECS, led to higher heterogeneity in the stem cell population, in agreement with previous studies that suggest a role of miRNA in controlling the noise in gene expression (Kumar, Cahan et al. 2014).

In mESCs, the miR-290-295 cluster miRNAs are among the most abundantly expressed (Houbaviy, Murray et al. 2003; Leung, Young et al. 2011). Moreover, homozygous deletion of this cluster in mice results in partially penetrant embryonic lethality and compromised fertility in females. Nevertheless, surviving male miR-290-$295^{-/-}$ mice are phenotypically normal (Medeiros, Dennis et al. 2011). This observation can probably be explained by the expression of microRNAs with same or similar seed that might compensate in certain cases the absence of miR-290-295. Alternatively, the fact that microRNAs deficient DGCR8-null mESCs do not impede the ground state pluripotent stem cells state might also be related to the partial penetrance of the miR-290-295 knock-out phenotype (Kumar, Cahan et al. 2014).

The important role of miRNAs in maintenance and differentiation of PSC is well accepted (Leonardo, Schultheisz et al. 2012). However, the targets and the mechanism of miR-290-295 cluster function in ESCs are not well understood (Leonardo, Schultheisz et al. 2012). To fill this gap, we have carried out an extensive analysis of data sets derived from mESCs that either expressed or were deficient in expression of miR-290-295 cluster miRNAs. We aimed to identify reproducible, high confidence and direct transcription factor targets of the miRNAs that propagate and perhaps amplify the effects of these miRNAs in the pluripotency network. Indeed, an initial computational analysis of these data carried by another PhD student in the group pinpointed a number of transcription factors that appeared to be involved in differentiation processes and to be directly regulated by the miR-290-295 cluster. We have followed up these results, constructing and testing luciferase reporters in a mouse cell line. In collaboration with other group members we have confirmed the expression variation of IRF2 in response to miRNAs depletion in ESC and we have validated the involvement of nuclear factor kappa-B (NF-kB) pathway in the miRNA-dependent regulation in mESCs. Overall, this study complements the current knowledge on the manner miR-290-295 regulates pluripotency, and proposes a new insight into its involvement in cell cycle, innate immune response, and chromatin modification in mESC.

Results

Manuscript published under the following title:

"Embryonic stem cell-specific microRNAs contribute to pluripotency by inhibiting regulators of multiple differentiation pathways."

Gruber AJ, Grandy WA, Balwierz PJ, Dimitrova YA, Pachkov M, Ciaudo C, Nimwegen Ev, Zavolan M.

Nucleic Acids Res. 2014 Aug;42(14):9313-26. doi: 10.1093/nar/gku544.
Epub 2014 Jul 16

Published online 16 July 2014

Nucleic Acids Research, 2014, Vol. 42, No. 14 **9313–9326**
doi: 10.1093/nar/gku544

Embryonic stem cell-specific microRNAs contribute to pluripotency by inhibiting regulators of multiple differentiation pathways

Andreas J. Gruber[1], William A. Grandy[1], Piotr J. Balwierz[1], Yoana A. Dimitrova[1], Mikhail Pachkov[1], Constance Ciaudo[2], Erik van Nimwegen[1] and Mihaela Zavolan[1,*]

[1]Biozentrum, University of Basel, Klingelberstrasse 50-70, CH-4056 Basel, Switzerland and [2]ETH Zürich, Otto-Stern-Weg 7, CH-8093 Zürich, Switzerland

Received April 17, 2013; Accepted June 5, 2014

ABSTRACT

The findings that microRNAs (miRNAs) are essential for early development in many species and that embryonic miRNAs can reprogram somatic cells into induced pluripotent stem cells suggest that these miRNAs act directly on transcriptional and chromatin regulators of pluripotency. To elucidate the transcription regulatory networks immediately downstream of embryonic miRNAs, we extended the motif activity response analysis approach that infers the regulatory impact of both transcription factors (TFs) and miRNAs from genome-wide expression states. Applying this approach to multiple experimental data sets generated from mouse embryonic stem cells (ESCs) that did or did not express miRNAs of the ESC-specific miR-290-295 cluster, we identified multiple TFs that are direct miRNA targets, some of which are known to be active during cell differentiation. Our results provide new insights into the transcription regulatory network downstream of ESC-specific miRNAs, indicating that these miRNAs act on cell cycle and chromatin regulators at several levels and downregulate TFs that are involved in the innate immune response.

INTRODUCTION

Embryonic stem cells (ESCs) originate from the inner cell mass of mammalian blastocysts. Due to their ability to self-renew as well as differentiate into various specialized cell types, they hold the promise of medical applications, such as stem cell therapy and tissue engineering. Therefore, the regulatory mechanisms behind pluripotency, stem cell fate and renewal are of great interest.

MiRNAs are short (∼22 nt long), single-stranded RNAs that post-transcriptionally regulate the expression of target genes (1). Computational and high-throughput studies suggest that a single miRNA can regulate hundreds of target genes (2,3) and that the majority of human mRNAs are regulated by miRNAs (4). Several studies found that the expression of ESC-specific miRNAs is required for initiation of stem cell differentiation and normal embryonic development (5–7). The ESC-specific miR-290-295 cluster accounts for ∼50% of the miRNA population of mouse ESCs (8–11) and its expression is downregulated relatively rapidly during differentiation (9,12). Interestingly, three of the seven miRNAs that are co-expressed from the miR-290-295 cluster, namely, miR-291a-3p, miR-294 and miR-295, are sufficient to force a G1→S transition (13) and promote induced pluripotency (14). All of these miRNAs, as well as those of another ESC-specific miRNA cluster, miR-302-367 (12,15), have the same sequence 'AAGUGCU' at positions 2-8 (also called the 'seed') which defines a family of miRNAs with related targets (4).

In contrast to the miR-290-295 cluster, miR-302-367 is also present in human and has been used to reprogram fibroblasts into induced pluripotent stem cells (iPSCs) (16). The reprogramming of differentiated cells into pluripotent stem cells by the ESC-specific miRNAs entails large gene expression and phenotypic changes that are likely to be due to regulatory cascades that involve several regulators. To identify *transcriptional regulators* that are immediate targets of the AAGUGCU seed family miRNAs, we analyzed data obtained in several previous studies that aimed to uncover the function of the miR-290-295 cluster.

These data consist of microarray-based measurements of mRNA expression in ESCs that were either deficient in miRNAs or expressed subsets of ESC-specific miRNAs (Supplementary Table S1). Sinkkonen *et al.* (17) analyzed mRNA expression of ESCs that express miRNAs ($Dicer^{+/-}$), ESCs that do not express miRNAs ($Dicer^{-/-}$) as well as $Dicer^{-/-}$ ESCs transfected with the miR-290-295 cluster miRNAs (miR-290, miR-291a-3p, miR-292-3p, miR-293, miR-294 and miR-295 mimics). The study

*To whom correspondence should be addressed. Tel: +41 61 267 1577; Fax: +41 61 267 1584; Email: mihaela.zavolan@unibas.ch

showed that the expression profile of ESCs can be restored to a large extent in Dicer$^{-/-}$ ESCs through transfection of miR-290-295 cluster miRNAs, and that these miRNAs are important for appropriate *de novo* DNA methylation in differentiating ESCs. Hanina *et al.* (18) profiled mRNA expression in Dicer$^{-/-}$ ESCs as well as in Dicer$^{-/-}$ ESCs transfected with miR-294. Combining these expression data with a biochemical approach to isolate Argonaute 2 (Ago2)-bound mRNAs, the study identified miR-294 targets in ESCs. It further concluded that miR-294 regulates a subset of genes that are also targeted by the Myc transcriptional regulator and that some of the effects of miR-294 expression may be due to the indirect upregulation of pluripotency factors, such as Lin28. Employing mRNA expression profiling of Dgcr8$^{-/-}$ ESCs, as well as miR-294-transfected Dgcr8$^{-/-}$ ESCs, Melton *et al.* (19) showed that self-renewal and differentiation of ESCs is regulated in an antagonistic manner by miR-294 and let-7. Finally, Zheng *et al.* (11) profiled mRNA expression of miRNA expressing ESCs and Dicer$^{-/-}$ ESCs and uncovered a pro-survival, anti-apoptotic function of the miR-290-295 cluster of miRNAs.

Altogether, these studies provide five separate experimental data sets that can be used to investigate the function of AAGUGCU seed family miRNAs in ESCs. They all determined mRNA expression profiles of ESCs with impaired miRNA expression (due to knockout of either Dgcr8 or Dicer components of the miRNA biogenesis pathway), as well as of ESCs that expressed miRNAs of the AAGUGCU seed family. The latter were either ES cells which expressed the full complement of miRNAs, or miRNA-deficient ESCs that were transfected with either miRNAs of the miR-290-295 cluster, or only miR-294. Although it has been observed that these studies resulted in sets of miRNA targets that are only partially overlapping (10), a meta-analysis that combines these data sets to identify the pathways that are most reproducibly targeted by the AAGUGCU miRNAs has not been performed.

In our study, we aimed to infer transcriptional regulators that are directly and consistently targeted by the AAGUGCU family of miRNAs, the pathways that these regulators control and the interactions that they have with each other. Toward this end, we modeled genome-wide mRNA expression in terms of computationally predicted target sites of both transcription factors (TFs) and miRNAs. This approach allowed us to identify a number of transcriptional regulators whose activity is consistently altered by miRNAs of the AAGUGCU seed family and that could contribute to the maintenance of pluripotency. Through reporter assays we validated these regulators as targets of AAGUGCU seed family miRNAs. Employing Dicer$^{-/-}$ mouse ES cells we showed that the expression of the IRF2 TF is strongly upregulated in the absence of miRNAs and that the nuclear concentration of the RelA component of the nuclear factor kappa-B (NF-κB) pathway upon stimulation with tumor necrosis factor α (TNF-α) is also increased. Our results give new insights into the functions of miRNAs in the regulatory circuitry of ESCs.

MATERIALS AND METHODS

Experimental data sets

Supplementary Table S1 summarizes the data sets that we obtained from the Gene Expression Omnibus (GEO) database of the National Center for Biotechnology Information (NCBI) and that we have used in our study. Each data set covers at least two distinct experimental conditions, with three replicates per condition. The first condition of every data set corresponds to an ESC line deficient in mature miRNAs due to Dicer- or Dgcr8-knockout. The second condition corresponds to either an ESC line expressing the entire complement of embryonically expressed miRNAs or the knockout cell line transfected with miR-294 or with mimics of the miR-290 cluster miRNAs (mir-290, mir-291a-3p, mir-292-3p, mir-293, mir-294 and mir-295).

Microarray analysis

Computational analysis of Illumina MouseWG-6 v2.0 Expression BeadChips from Hanina et al. (2010). We downloaded the processed data from the GEO database of NCBI (accession no. GSE20048). Probe-to-gene associations were made by mapping the probe sequences (provided by the authors) to the set of mouse transcript sequences (downloaded 2011-02-19 from the UCSC Genome Bioinformatics web site).

We computed average gene expression levels as weighted averages of the signals of all probes that perfectly matched to at least one transcript of the gene. Whenever a probe mapped to multiple genes, a weight of $1/n$ was assigned to each of the n genes that the probe matched. For a given replicate experiment, the $\log_2$ expression fold change of each gene was then determined by subtracting the $\log_2$-average expression of the gene in the first condition (control) from the $\log_2$-average expression in the second condition (treatment).

Computational analysis of Affymetrix Mouse Genome 430 2.0 chips from Sinkkonen et al. (2008) and Zheng et al. (2011). We downloaded the data from the GEO database (accessions GSE8503, GSE7141 and GSE30012) and analyzed the CEL files with the R software (http://www.R-project.org) using the BioConductor affy package (20). We used the GCRMA algorithm (21) for background correction and the MClust R package (22) to fit a two-component Gaussian mixture model to the $\log_2$-probe intensities and classify probes as expressed or not expressed. A probe was considered for further analysis if it was consistently classified as expressed in all three replicates of at least one of the two experimental conditions. The remaining probes were quantile normalized across all conditions and replicates of a particular experiment. Probe-to-gene associations were made by mapping probe sequences (provided on the Affymetrix web site, http://www.affymetrix.com) to mouse transcript sequences (as used by motif activity response analysis (MARA), downloaded from UCSC Genome Bioinformatics web site as described above). We then computed $\log_2$-gene expression fold changes as described for Illumina Expression BeadChips (see above).

Computational analysis of Affymetrix Mouse Gene 1.0 ST chips from Melton et al. (2010). We downloaded the data from the GEO database (accession no. GSE18840) and analyzed the CEL files with the R Bioconductor oligo package (23). We used the Robust Multi-array Average (RMA) algorithm (24) for background adjustment. The rest of the analysis, including the classification of probes into expressed/not expressed, the quantile normalization, and the calculation of $\log_2$ gene expression fold changes, was carried out as described above.

Proportions of AAGUGCU miRNA seed family targets among genes that are consistently downregulated in multiple experiments. For each gene and each experiment, we calculated the standard error in its $\log_2$ fold change across the replicates. A gene was considered significantly downregulated when it was down-regulated more than 1.96 standard-errors. We then determined the intersection set of significantly downregulated genes for every possible subset of the experiments S={*MeltonDGCR8KOVs294, SinkkonenDicerKOVs290, SinkkonenDicerKOVsWT*}. Subsequently, for every obtained intersection set, the proportion of AAGUGCU miRNA seed family targets (TargetScan aggregate P_{CT} score predictions (4)) was determined and plotted against the size of the corresponding intersection set.

Combined MARA of TFs and miRNAs. We carried out the MARA (25) separately for each experimental data set. MARA relates the expression level E driven by individual promoters (measured by microarrays) to the number of binding sites N that various regulators have in the promoters using a simple linear model

$$E_{ps} = \tilde{c}_s + c_p + \sum_m N_{pm} A_{ms}, \tag{1}$$

where c_p is a term reflecting the basal expression of promoter p, $\tilde{c}_s$ reflects the mean expression in sample s, and A_{ms} is the (unknown) *activity* of binding motif m in sample s (where with 'sample' we refer to any individual replicate of any condition of a data set, see section 'Experimental data sets' above). That is, using the predicted site-counts N_{pm} and the measured expression levels E_{ps} we used an approximation (1) to infer the activities A_{ms} of all motifs across all samples by ridge regression. In our analyses, we considered a curated set of 189 TF binding motifs (for detailed information about the motifs and the corresponding TFs see Supplementary Table S7). Furthermore, we included the binding sites in the 3'UTRs of mRNAs of 85 miRNA families by incorporating aggregate P_{CT} scores as provided by TargetScan (4) (predictions downloaded on the 27th of March 2012 from the TargetScan web site, http://www.targetscan.org). miRNAs are grouped into families by their seed sequences and in particular the *AAGUGCU* seed family corresponds to the following miRNAs: *mmu-miR-291a-3p*, *mmu-miR-294*, *mmu-miR-295*, *mmu-miR-302a*, *mmu-miR-302b* and *mmu-miR-302d*. An aggregate P_{CT} score was assigned to a promoter by averaging the aggregate P_{CT} scores of transcripts associated with this promoter.

For a given motif m, MARA provides for each sample s motif activities A^*_{ms} and associated errors σ_{ms}. More specifically, marginalizing over all other motifs, the likelihood $P(D|A_{ms})$ of the expression data D given the activity of a given motif is proportional to a Gaussian

$$P(D|A_{ms}) \propto \exp\left[-\frac{1}{2}\frac{(A_{ms} - A^*_{ms})^2}{\sigma^2_{ms}}\right]. \tag{2}$$

Given that all analysed experiments were performed in multiple replicates we were interested in averaging motif activities across replicates and we used the following Bayesian approach. For each motif m separately, we assumed that the activities across a group g of replicates belonging to a specific condition of an experiment (see section 'Experimental data sets' above) are normally distributed around some (unknown) mean $\bar{A}_{mg}$ with (unknown) variance σ^2_{mg}

$$P(A_{ms}|\bar{A}_{mg}, \sigma_{mg}) = \frac{1}{\sqrt{2\pi}\sigma_{mg}} \exp\left[-\frac{1}{2}\frac{(A_{ms} - \bar{A}_{mg})^2}{\sigma^2_{mg}}\right]. \tag{3}$$

By combining the prior from Equation (3) with the likelihood from Equation (2) for each replicate sample $s \in g$ and integrating out the (unobserved) true activities A_{ms} in each of the replicates, we obtained the probability of the form

$$P(D|\bar{A}_{mg}, \sigma_{mg}) = \prod_{s\in g} \frac{1}{\sqrt{2\pi(\sigma^2_{mg} + \sigma^2_{ms})}} \exp\left[-\frac{(A^*_{ms} - \bar{A}_{mg})^2}{2(\sigma^2_{mg} + \sigma^2_{ms})}\right]. \tag{4}$$

Formally, we would next integrate out the unknown standard deviation of activities in the group σ_{mg} of this likelihood. Unfortunately, this integral cannot be performed analytically. We thus approximated the integral by the value of the integrand at its maximum, *i.e.* we numerically found the value of σ_{mg} that maximizes expression (4). Assuming an uniform prior over mean activity $\bar{A}_{mg}$, we find that $P(\bar{A}_{mg}|D)$ is again a Gaussian with mean

$$\bar{A}^*_{mg} = \frac{\sum_{s\in g} \frac{A^*_{ms}}{(\sigma^*_{mg})^2 + (\sigma_{ms})^2}}{\sum_{s\in g} \frac{1}{(\sigma^*_{mg})^2 + (\sigma_{ms})^2}}, \tag{5}$$

and error

$$\bar{\sigma}^*_{mg} = \sqrt{\frac{1}{\sum_{s\in g} \frac{1}{(\sigma^*_{mg})^2 + (\sigma_{ms})^2}}}. \tag{6}$$

where σ^*_{mg} is the maximum likelihood estimate of Expression (4). We call the quantities defined in (5) and (6) averaged activities and averaged errors, respectively.

To identify motifs that consistently change in their activities across experiments, we wanted to further average motif activities across these experiments. However, because of the inherent differences in the *scale* of expression variation in the different experiments, the motif activities also varied in scale across the experiments. Thus, before averaging we first standardized the motif activities across the two conditions a and b. That is, for a given experiment we defined a scale L

$$L = \sqrt{\frac{\left(\bar{A}^{*b}_{mg}\right)^2 + \left(\bar{A}^{*a}_{mg}\right)^2}{2}}, \tag{7}$$

and rescaled the activities

$$\tilde{A}^*_{mg} = \frac{\bar{A}^*_{mg}}{L} \quad (8)$$

and their errors

$$\tilde{\sigma}^*_{mg} = \frac{\bar{\sigma}^*_{mg}}{L}. \quad (9)$$

These condition-specific, averaged and rescaled activities ($\tilde{A}^*_{mg}$) and errors ($\tilde{\sigma}^*_{mg}$) from the different experiments were then combined into two groups, *i.e.* the group of a conditions and the group of b conditions, and for each group we again averaged the activities exactly as described above for the replicates.

To rank the activity changes between two different experimental conditions (presence/absence of miRNAs) we determined a z-value for every motif m by dividing the change in averaged activities between the two different conditions a and b by the averaged errors as follows

$$z = \frac{\tilde{A}^{*b}_{mg} - \tilde{A}^{*a}_{mg}}{\sqrt{\left(\sigma^{*b}_{mg}\right)^2 + \left(\sigma^{*a}_{mg}\right)^2}}. \quad (10)$$

Consequently, from the results of Equation (10) we obtained a global z-value-based ranking of the motifs.

Motif–motif interaction network. To uncover which TFs were targeted by a particular motif m, we focused only on those TF genes, whose promoters were consistently (in all experiments) predicted by MARA to be targets of motif m. MARA computes a target score S for each potential target promoter of motif m. S corresponds to the log-likelihood ratio of the data D assuming the promoter is indeed a target, and assuming the promoter is independent of the regulator, i.e.

$$S = \log\left[\frac{P(D|\text{target})}{P(D|\text{nottarget})}\right]. \quad (11)$$

Assuming a uniform prior of 1/2 that the promoter is indeed a target, the posterior probability p that the promoter is a target given the data is

$$p = \frac{1}{1 + \frac{1}{e^s}}. \quad (12)$$

To obtain a combined probability p_c that a gene is a target of a particular motif across N different experiments the probability product was calculated by multiplying the probabilities p_n obtained in individual experiments n, *i.e.*

$$p_c = \prod_{n=1}^{N} p_n. \quad (13)$$

Evaluating miR-294 targets with luciferase assays

Cloning, cell culture and luciferase assay. We polymerase chain reaction (PCR)-amplified 3′UTR fragments of the putative target genes from Normal Murine Mammary Gland (NMuMG) genomic DNA and cloned them into pGEM-T Easy vector (Promega; A1360). We used site-directed mutagenesis and the QuickChange II kit (Stratagene; 200524-5) to generate deletion mutant constructs that differed in a few nucleotides in the miR-294 seed-matching region from the wild-type construct. All constructs, wild-type and mutated, were verified by sequencing and then subcloned into the empty psiCHECK-2 vector (Promega; C8021) at XhoI - NotI restriction sites. The sequences of the primers used for cloning and mutagenesis can be found in Supplementary Tables S9 and S10, respectively. NMuMG cells were reverse-transfected with Lipofectamine2000 reagent (Invitrogen; 11668019), and the corresponding psiCHECK-2 constructs in the presence of 50nM Syn-mmu-miR-294-3p mimic (QIAGEN; MSY0000372), or 50 nM of non-targeting negative control siRNA (Microsynth). Between 36 and 48 h post-transfection cells were collected and both *Renilla* and firefly luciferase activities were measured using Dual Glo Luciferase Assay System (Promega; E2940).

For each gene, expression was measured for both constructs in 3 separate experiments, and each experiment contained 3 technical replicates.

Analysis of the luciferase data. We denote by w_{ir} the logarithm (base 2) of the expression level of the luciferase construct containing the wild-type 3′UTR in experiment i replicate r and by m_{ir} the analogous expression for the mutant construct. For each gene the data thus consist of 9 values w and 9 values m. We took into account two sources of variability, namely, true expression variability across experiments and 'measurement noise' between replicates. We first describe the measurement noise. Assuming the true expression of the wild type was w_i, we assumed that the probability to measure expression level w_{ir} (in a given replicate r) follows a Gaussian distribution with a certain variance τ_i

$$P(w_{ir}|w_i, \tau_i) = \frac{1}{\tau_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{w_{ir} - w_i}{\tau_i}\right)^2\right], \quad (14)$$

thus allowing for the possibility that each experiment i has a *different* level of noise τ_i between replicates. The probability of the wild-type data of experiment i, assuming that τ_i is given, is simply the product of expressions $P(w_{ir}|w_i, \tau_i)$ over the three replicates $r = 1$ through 3. Using $\langle w_i\rangle$ and $\text{var}(w_i)$ to denote the mean and variance of the measurement across the replicates, we can rewrite this as

$$P(\{w_{ir}\}|w_i, \tau_i) \propto \frac{1}{\tau_i^3} \exp\left[-\frac{3}{2}\left(\frac{w_i - \langle w_i\rangle}{\tau_i}\right)^2 - \frac{3}{2}\frac{\text{var}(w_i)}{\tau_i^2}\right]. \quad (15)$$

Integrating over the unknown variable τ_i from 0 to infinity with a scale prior $P(\tau_i) \propto 1/\tau_i$ we obtain

$$P(\{w_{ir}\}|w_i) \propto \left(1 + \frac{(w_i - \langle w_i\rangle)^2}{\text{var}(w_i)}\right)^{3/2}. \quad (16)$$

Approximating this Student's t distribution by a Gaussian, that is, approximating the probability of the data in experiment i by a Gaussian with mean $\langle w_i\rangle$ and variance $\text{var}(w_i)$,

we have

$$P(\{w_{ir}\}|w_i) \approx \sqrt{\frac{3}{\mathrm{var}(w_i)}} \exp\left[-\frac{3(w_i - \langle w_i\rangle)^2}{2\mathrm{var}(w_i)}\right]. \quad (17)$$

Since the variability between replicates is much smaller than the variability across experiments, this approximation will have a negligible effect on the final outcome.

For the true variability between experiments, we denote by w the 'true' average expression of the wild-type construct. We assume that the deviation of the level w_i in experiment i from the mean w follows a Gaussian distribution with variance σ. We thus have

$$P(w_i|w,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{w_i - w}{\sigma}\right)^2\right]. \quad (18)$$

To obtain the probability of the data given w we multiply $P(\{w_{ir}\}|w_i)$ by $P(w_i|w,\sigma)$ and integrate over the unknown expression level w_i. We then obtain

$$P(\{w_{ir}\}|w,\sigma) \propto \frac{1}{\sqrt{\sigma^2 + \mathrm{var}(w_i)/3}} \exp\left[-\frac{(\langle w_i\rangle - w)^2}{2(\sigma^2 + \mathrm{var}(w_i)/3)}\right]. \quad (19)$$

The interpretation of this formula is straightforward. The deviation between the mean $\langle w_i\rangle$ of the observations in experiment i, and the average level w is Gaussian-distributed with a variance that is the sum of the variability σ^2 across experiments, and the variability $\mathrm{var}(w_i)/3$ associated with estimating w_i from the 3 replicate measurements due to measurement noise.

For the measurements of the mutant construct in experiment i we obtain an analogous equation

$$P(\{m_{ir}\}|m,\tilde{\sigma}) \propto \frac{1}{\sqrt{\tilde{\sigma}^2 + \mathrm{var}(m_i)/3}} \exp\left[-\frac{(\langle m_i\rangle - m)^2}{2(\tilde{\sigma}^2 + \mathrm{var}(m_i)/3)}\right], \quad (20)$$

where we have introduced the variability $\tilde{\sigma}$ of the true expression of the mutant construct across replicates. What we are interested in is the *difference* $w - m$ in log-expression of the wild-type and mutant construct. To this end, we define $\mu = w - m$ and $y = (m + w)/2$ and integrate over y. We then obtain

$$P(\{w_{ir}\},\{m_{ir}\}|\mu,\sigma,\tilde{\sigma}) \propto \frac{1}{\sqrt{\sigma^2 + \tilde{\sigma}^2 + \mathrm{var}(w_i)/3 + \mathrm{var}(m_i)/3}} \exp\left[-\frac{(\langle w_i\rangle - \langle m_i\rangle - \mu)^2}{2(\sigma^2 + \tilde{\sigma}^2 + \mathrm{var}(w_i)/3 + \mathrm{var}(m_i)/3)}\right]. \quad (21)$$

This is again a Gaussian with mean $\langle w_i\rangle - \langle m_i\rangle$ and a variance that is the sum of all variances σ^2, $\tilde{\sigma}^2$, $\mathrm{var}(w_i)/3$ and $\mathrm{var}(m_i)/3$.

Clearly, although both σ^2 and $\tilde{\sigma}^2$ are unknown, the only variable that enters in our equations is their sum. We thus simplify the notation by defining this sum as

$$\gamma^2 = \sigma^2 + \tilde{\sigma}^2. \quad (22)$$

Similarly, we redefine the variance associated with measurement noise as

$$t_i^2 = \mathrm{var}(w_i)/3 + \mathrm{var}(m_i)/3, \quad (23)$$

which leads to

$$P(\{w_{ir}\},\{m_{ir}\}|\mu,\gamma) \propto \frac{1}{\sqrt{\gamma^2 + t_i^2}} \exp\left[-\frac{(\langle w_i\rangle - \langle m_i\rangle - \mu)^2}{2(\gamma^2 + t_i^2)}\right]. \quad (24)$$

We now combine the data from the different experiments and remove the final unknown variable γ. The probability of all data given the variable of interest μ and unknown variability parameter γ is simply the product

$$P(D|\mu,\gamma) = \prod_{i=1}^{3} P(\{w_{ir}\},\{m_{ir}\}|\mu,\gamma). \quad (25)$$

To obtain the probability of the data D given μ we multiply this expression with a scale prior for γ, *i.e.* $P(\gamma) = 1/\gamma$, and integrate over γ

$$P(D|\mu) = \int_0^\infty P(D|\mu,\gamma)\frac{d\gamma}{\gamma}. \quad (26)$$

We performed the integration numerically with Mathematica to obtain $P(D|\mu)$, and used Bayes' theorem to compute the posterior distribution of the parameter μ, $P(\mu|D)$ as $P(D|\mu)/\int_{-\infty}^{\infty} P(D|\mu)d\mu$. Finally, we determined the 5 percentile, the 25 percentile, the median, the 75 percentile and the 95 percentile of this distribution again with the Mathematica software.

Mouse ESC (mESC) culture

The generation of Dicer(DCR)$^{flox/flox}$ and DCR$^{-/-}$ mouse ES cell lines has been described elsewhere (26). The cells were routinely screened for both pluripotency and differentiation markers (see Supplementary Figure S4). Both mES cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) (Gibco; 41966-029) supplemented with 15% of a special batch of fetal bovine serum tested for optimal growth of mESCs. In addition, the DMEM contained 1000 U/ml of a homegrown recombinant LIF (a kind gift of Thomas Grentzinger), 0.1mM 2β-mercaptoethanol (Millipore; ES-007-E), 1x L-Glutamine (Gibco; 25030-024), 1x Sodium Pyruvate (Gibco; 11360) and 1x Minimum Essential Medium, Non-Essential Amino Acids (MEM, NEAA) (Gibco; 11140-35). The cells were grown on gelatin-coated (Sigma; G1393) dishes. The medium was changed daily, and the cells were subcultured every 2–3 days. To induce NF-κB signaling, mESCs were treated with 20 ng/ml TNF-α (Cell Signaling Technology; 5178) for 24 h.

Quantitative reverse transcriptase-PCR (qRT-PCR)

Total RNA was extracted from mESCs using Tri Reagent (Sigma; T9424) following the supplier's protocol. Contaminating DNA was removed using the RQ1 RNase-Free DNase kit (Promega; M6101). The resulting DNA-free RNA was then purified using the RNeasy MinElute

Cleanup kit (Qiagen; 74204) and quantified using Nanodrop. Superscript III (Invitrogen; 18080) was then used to create cDNA following the manufacturer's recommendations. The cDNA was finally purified using QIAquick PCR Purification kit (Qiagen; 74204), quantified using Nanodrop and diluted to 8 ng/μl. Each qRT-PCR reaction was run using 2 μl of the purified cDNA in triplicate ($n = 3$) using Power SYBR Green PCR Master Mix (Applied Biosystems; 4367659) on a StepOne Plus RT-PCR System (Applied Biosystems). The following primer pairs were used in this study:

- Mouse IRF2 Fwd: 5′-CTG GGC GAT CCA TAC AGG AAA-3′
- Mouse IRF2 Rev: 5′-CTC AAT GTC GGG CAG GGA AT-3′
- Mouse E2F5 Fwd: 5′-GTT GTG GCT ACA GCA AAG CA-3′
- Mouse E2F5 Rev: 5′-GGC CAA CAG TGT ATC ACC ATG A-3′
- Mouse c-Myc Fwd: 5′-GTT GGA AAC CCC GCA GAC AG-3′
- Mouse c-Myc Rev: 5′-ATA GGG CTG TAC GGA GTC GT-3′
- Mouse GAPDH Fwd: 5′-CAT CAC TGC CAC CCA GAA GAC TG-3′
- Mouse GAPDH Rev: 5′-ATG CCA GTG AGC TTC CCG TTC AG-3′

qRT-PCR data were normalized using glyceraldehyde 3-phosphate dehydrogenase (GAPDH) expression and evaluated using the $2^{-\Delta\Delta Ct}$ method (27). Significant changes in gene expression were identified based on Student's *t*-test.

Western blots

To extract total proteins from mESCs, radioimmunoprecipitation assay buffer supplemented with 1x Complete, ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail (Roche; 11873580001) was used to lyze cell pellets. Cytosolic and nuclear protein fractions were enriched using a series of lysis buffers as follows:

- Lysis Buffer 1 (LB1): 50 mM Hepes-KOH, pH 7.5; 140 mM NaCl; 1 mM EDTA, pH 8.0; 10% v/v Glycerol; 0.5% v/v NP-40; 0.25% v/v Triton X-100.
- Lysis Buffer 2 (LB2): 10 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1mM EDTA, pH 8.0; 0.5 mM EGTA, pH 8.0.
- Lysis Buffer 3 (LB3): 10mM Tris-HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA, pH 8.0; 0.5 mM EGTA, pH 8.0; 0.1% v/v Na-Deoxycholate; 30% v/v N-Lauroylsarcosine.

All lysis buffers were supplemented with the protease inhibitor cocktail immediately before use. The cytosolic fraction was extracted by lyzing the cell pellets in LB1 that leaves the nuclear membrane intact. The nuclei were then pelleted (1,350 x g; 4°C; 5 min), washed with LB2, pelleted once more and finally lyzed with LB3 to release the nuclear contents. All protein lysates were quantified using the BCA Protein Assay kit (Pierce; 23227). The following antibodies (dilution 1:1000) were used in this study:

- Anti-IRF2 (Center) rabbit IgG (Abgent; AP11225c)
- Anti-NF-κB p65 (D14E12) XP rabbit IgG (Cell Signaling Technology; 8242)
- Anti-GAPDH (6C5) mouse IgG (Santa Cruz Biotechnology; sc-32233)
- Anti-Histone H3 (C-16) goat IgG (Santa Cruz Biotechnology; sc-8654)
- HRP-conjugated Polyclonal swine Anti-Rabbit (Dako; P0217)
- HRP-conjugated Polyclonal rabbit Anti-Mouse (Dako; P0260)
- HRP-conjugated Polyclonal rabbit Anti-Goat (Dako; P0449)

Western blot signals were visualized with the enhanced chemiluminescence blotting detection reagents (GE Healthcare; RPN2106). Cytosolic enrichment was confirmed via a postive GAPDH signal, while nuclear enrichment was confirmed by Histone H3. Western blot quantifications were performed using the ImageJ software by quantifying the pixels of each band and normalizing against a housekeeper, such as Histone H3.

RESULTS

General relationship between data sets

A common, though perhaps naive expectation is that combining data from experiments that have been independently performed in different labs, with different experimental procedures, allows one to identify essential properties of the system that are invariant with respect to details of the experimental approach. In our case, in any given experiment, confounding effects may have led to some genes being spuriously identified as targets of AAGUGCU miRNAs (false positives), and true targets of AAGUGCU miRNAs being missed (false negatives). For example, because it is unclear whether the miRNA processing enzymes solely function in this pathway, it is important to analyze data from ESCs in which the miRNA biogenesis has been impaired at different levels (Dicer in the studies of Sinkkonen *et al.* (17) and Hanina *et al.* (18) and Dgcr8 in the study of Melton *et al.* (19)). Furthermore, although ESCs expressing the full complement of miRNAs provide the most physiological reference point for the function of the miR-290-295 cluster miRNAs in normal, unstressed cells, the effect of these miRNAs in these cells is confounded by the effects of other co-expressed miRNAs. Similarly, if the profiled cell population was heterogeneous with respect to the pluripotency/differentiation status, the let-7 miRNAs may have masked the effect of miR-294, because these miRNAs have antagonistic effects (19).

Requiring targets to show consistent downregulation across multiple data sets can reduce the number of false positive miR-294 targets. On the other hand, requiring perfect consistency across a large number of experiments is likely to lead to too many false negatives, simply because different experiments have different levels of accuracy or confounding effects. Thus, we first investigated the relationship of gene-level expression changes between ESCs that did or did not express embryonic miRNAs in all pairs of

(a)

Publication	Abbr.	GEO Id	First Condition	Second Condition
Melton et al. (2010)	A	GSE18840	Dgcr8$^{-/-}$	miR-294 transfected Dgcr8$^{-/-}$
Sinkkonen et al. (2008)	B	GSE8503	Dicer$^{-/-}$	miR-290 cluster transfected Dicer$^{-/-}$
Sinkkonen et al. (2008)	C	GSE7141	Dicer$^{-/-}$	Dicer$^{+/-}$

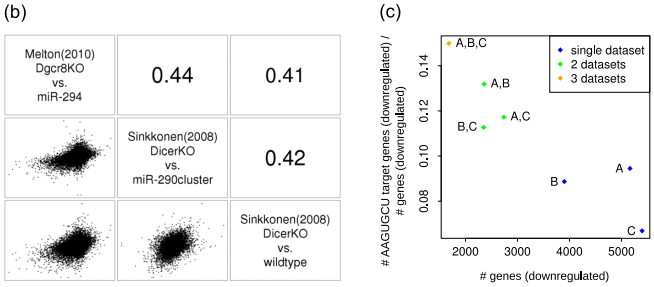

Figure 1. Overview of the mRNA expression data sets—**(a)** Data sources. **(b)** Matrix of scatter plots (below diagonal) and Pearson's correlation coefficients (above diagonal) of per-gene $\log_2$ fold changes in pairs of experiments. The names of the individual data sets are shown on the diagonal. **(c)** Proportion of predicted targets of the AAGUGCU seed family of miRNAs (TargetScan aggregate P_{CT} score based predictions (4)) among genes that are consistently downregulated in all three (orange), pairs (green) or individual data sets (blue) (indicated by the labels, key given in the 'Abbr.' column of the table in panel (a)), plotted against the number of genes that are consistently downregulated in all of the considered data sets.

experiments. Although pairwise Pearson's correlation coefficients were as low as 0.11 (Supplementary Figure S1), three of the five experimental data sets (Figure 1a), covering all described conditions (expression of miR-294, miR-290-295 cluster miRNAs or the entire complement of embryonically expressed miRNAs in a miRNA-deficient background) gave reasonably high pairwise correlation coefficients (Figure 1b). We therefore focused our discussion on these data sets, and for completeness, we present the results of a similar analysis of all five data sets in the Supplementary material (Supplementary Figure S2 and Tables S5 and S6). Of the ∼4000–5000 genes that were downregulated in a single experiment, a little less than 2000 genes were downregulated in all three experiments. Importantly, the proportion of predicted AAGUGCU seed family targets among downregulated genes increased when intersecting an increasing number of data sets (Figure 1c), indicating that the approach of a combined analysis of these data sets does have the potential to reveal important regulators that are immediately downstream of the AAGUGCU family of miRNAs. 252 of the genes downregulated in all three experiments were predicted AAGUGCU seed family targets (4) (Supplementary Table S2).

The transcriptional network regulated by the miRNAs of the AAGUGCU seed family in ESCs

As mentioned in the Introduction, the main aim of our study was to identify *transcriptional regulators* that are targeted by the AAGUGCU seed family and at the same time can account for the largest fraction of gene expression changes that are observed in cells that do or do not express the miRNAs. We therefore built on the MARA approach (28) that we recently made available in the form of an easy-to-use web application (25). In contrast to standard transcriptome analyses that strive to find genes (including transcription regulators) whose expression changes significantly between conditions, MARA aims to infer changes of the *regulatory impact* (also referred to as 'activity') of binding motifs. This is achieved by modeling gene expression as a linear function of the number of regulatory motif binding sites occurring in the promoter (for TFs) and 3′UTR (for miRNAs) of the gene and the unknown activity of each motif. The change in activity of a specific binding motif (*e.g.* of the Irf2 TF) in a specific condition (*e.g.* transfection of miR-294) is inferred from the expression changes of all (predicted) *targets* of this motif (determined by transcriptome profiling), taking into account the occurrences of sites for other regulators in these targets. For example, a decrease in Irf2 activity is inferred when the predicted Irf2 targets consistently show a decrease in expression that cannot be explained by the occurrence of binding sites for other regulatory motifs in the promoters or 3′UTRs of these targets. This means that MARA can uncover gene expression changes that are due not only to changes in the mRNA expression level of a regulator, but also to changes in the *active form* (*e.g.* for TFs through post-translational modifications, such as phosphorylation) of the regulator. MARA was initially developed for the characterization of transcription regulatory networks (28), and we have recently extended it to also model miRNA-dependent changes in mRNA stability (25). For this study we further extended the MARA approach to identify regulators whose activity not only changes most significantly between samples but also reproducibly across multiple data sets. Our approach is described in detail in the Materials and Methods section.

To verify that MARA can indeed uncover the key regulator in these experiments, namely, the miRNAs of the AAGUGCU seed family, we first applied MARA taking into account all TFs and miRNA seed families (see Supplementary Table S4). In subsequent analyses, however, we performed the MARA analysis with only the AAGUGCU seed family motif added to the full complement of TF motifs. This was because when all miRNAs are included in the analysis, MARA will also infer non-zero activities for other miRNAs, *e.g.* those with significantly overlapping sets of targets (29).

MARA quantifies the extent to which the activity of each motif varies across conditions by a *z*-statistic, that roughly corresponds to the ratio between the average deviation of the motif activity from zero and the standard deviation of the motif activity (see Materials and Methods). Supplementary Table S3 shows all motifs ranked by their absolute *z*-values.

MARA also predicts which promoters or 3′UTRs are targeted by each motif, quantifying the confidence in each predicted motif-target interaction by a posterior probability (see Materials and Methods). We used these probabilities to construct a regulatory network of motif–motif interactions (Figure 2) that provides a synthetic view of the regulatory impact of the AAGUGCU seed family of miRNAs on the transcriptional network of pluripotent stem cells. An arrow was drawn from motif *A* to motif *B* whenever motif *A* was predicted by MARA to regulate a TF *b* whose binding specificity is represented by motif *B*. Only motif-TF inter-

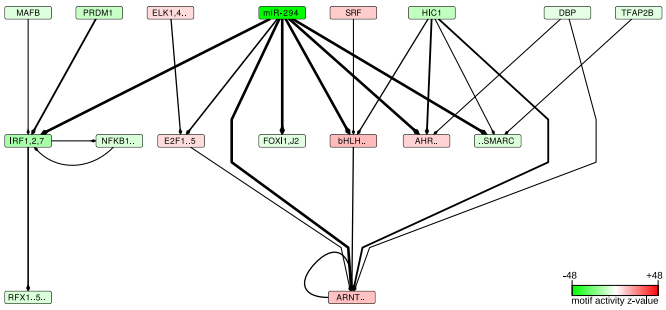

Figure 2. The transcriptional network inferred to be affected by the miRNAs of the AAGUGCU seed family (represented by miR-294)—A directed edge was drawn from a motif *A* to a motif *B* if *A* was consistently (across data sets) predicted to regulate a TF *b* whose sequence specificity is represented by motif *B*. The thickness of the edge is proportional to the product of the probabilities that *A* targets *b*. For the clarity of the figure, only motifs with absolute *z*-values >5 and only edges with a target probability product >0.3 are shown. The intensity of the color of a box representing a motif is proportional to the significance of the motif (the corresponding *z*-values can be found in Supplementary Table S3). Red indicates an increase and green a decrease in activity, corresponding to increased and decreased expression, respectively, of the tagets of the motif when the miRNAs are expressed. The full motif names as well as the corresponding TFs are listed in Supplementary Table S7.

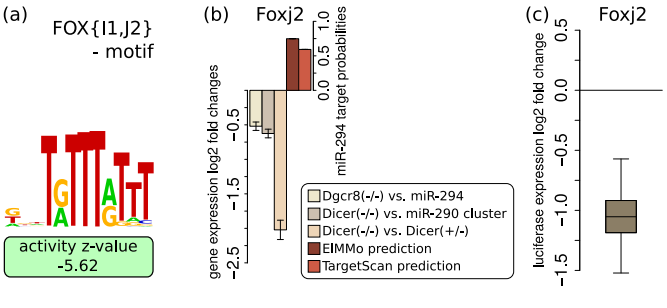

Figure 3. Foxj2 is a direct target of miR-294—**(a)** The *'FOX{I1,J2}'* motif shows a negative change in activity in the presence of miR-294. **(b)** Foxj2 mRNA $\log_2$ fold changes (±1.96*SEM; $n = 3$) in the Melton *et al.* Dgcr8$^{-/-}$ versus miR-294 transfection (yellow), Sinkkonen *et al.* Dicer$^{-/-}$ versus miR-290-295 cluster transfection (dark brown) and Dicer$^{-/-}$ versus Dicer$^{+/-}$ (light brown) data sets, as well as the prediction scores for these genes as targets of miR-294 as given by ElMMo (30) (dark red) and TargetScan (aggregate P_{CT}) (4) (light red). **(c)** A luciferase reporter construct carrying the 3′UTR of Foxj2 is downregulated upon co-transfection with miR-294 relative to a construct carrying the Foxj2 3′UTR but with a mutated miR-294 target site ($n = 9$).

actions that were predicted in all data sets and that involved motifs with high significance (|*z*-value| > 5) are shown.

The motif corresponding to the AAGUGCU seed family (represented by the dark green 'miR-294' motif in Figure 2) is by far the most significantly changing motif (see also Supplementary Table S3). Its negative change in activity upon miRNA expression is consistent with the destabilizing effect of the miRNA on its targets.

The motif with the second most significant change in activity, 'IRF1,2,7', is bound by the interferon regulatory factors. MARA predicts that this motif is directly targeted by miR-294, in line with previous suggestions that the interferon regulatory factors are targets of the miR-290 cluster miRNAs (18). We present a more detailed analysis of this motif in the next section.

A second motif whose activity decreases significantly upon miRNA expression is 'FOX{I1,J2}' (Figure 3a). Of the TFs associated with this motif, Foxj2 is predicted within all data sets to be directly regulated by miR-294 (Figure 2). Consistently, Foxj2 is downregulated upon miRNA expression on the mRNA level (Figure 3b). In order to validate that Foxj2 is a direct target of the miRNAs, as predicted by both ElMMo (30) and TargetScan (Figure 3b), we cloned the 3′UTR of Foxj2 downstream of a luciferase reporter and co-transfected this construct together with miR-294 in the murine mammary gland cell line NMuMG. For comparison, we generated a construct in which the presumed miRNA-294 target site was mutated and we performed similar co-transfection experiments. The results of this experiment clearly show that Foxj2 is indeed a functional target of miR-294 (Figure 3c). We carried out similar transfection experiments with control siRNAs, that do not target the reporter, and a standard analysis of these data is presented in Supplementary Figure S3. Little is known about the function of Foxj2 in cell fate. It appears to be expressed very early in development (31), but its overexpression has a negative effect on embryogenesis (32). Our results suggest that the AAGUGCU seed family of miRNAs contributes to the maintenance of an adequate expression of Foxj2 in pluripotent stem cells. The third most significant changing motif, basic-helix-loop-helix (referred to as 'bHLH..' in Figure 2), can be bound by many TFs (reviewed in (33)), some of which are predicted direct targets of miR-294.

To further elucidate the transcription regulatory network downstream of the AAGUGCU seed family of miRNAs, we analyzed in-depth the TFs whose associated motif had the most significant activity change (|*z*-value| > 5) and that were consistently predicted by MARA to be direct targets of the miR-294 seed family miRNAs across the multiple data sets (Table 1).

We found that the majority of these direct target TFs fall into three categories that have previously been associated with pluripotency: NF-κB-related interferon response factors that control NF-κB signalling, cell cycle regulators and epigenetic regulators.

AAGUGCU seed family miRNAs modulate Irf2-dependent transcription

The 'IRF1,2,7' motif shows the second strongest activity change upon changes in miR-294 expression (Figure 4a and Supplementary Table S3). Of the individual factors associated with this motif, Irf2 is the one that was consistently predicted by our analysis to be a direct target of the AAGUGCU seed family miRNAs across data sets (Table 1), consistent with the predictions of both ElMMo and TargetScan (Figure 4b). Irf2 was downregulated at the mRNA level across all analyzed data sets (Figure 4b). Consistently, we found that Irf2 is strongly downregulated in DCR$^{flox/flox}$ compared to DCR$^{-/-}$ ESCs, both at the mRNA level (Figure 4c) as well as at the protein level (Figure 4d). To validate Irf2 as a direct target of miR-294, we conducted luciferase assays as described above for Foxj2. Our results demonstrate that Irf2 is indeed targeted by miR-294 (Figure 4e). Although relatively little is known about the function of this factor in ESCs, a recent study showed

Table 1 TFs consistently predicted by MARA to be direct targets of miR-294 and whose absolute motif activity z-value is >5

Name	Motif	Motif Abbreviation	Activity z-value
Irf2	IRF1,2,7.p3	IRF1,2,7	-16.29
Mxd3	bHLH_family.p2	bHLH..	13.00
Clock	bHLH_family.p2	bHLH..	13.00
Arnt2	ARNT_ARNT2_BHLHB2_MAX_MYC_USF1.p2	ARNT..	11.60
Arnt2	AHR_ARNT_ARNT2.p2	AHR..	8.39
BAF170	DMAP1_NCOR{1,2}_SMARC.p2	..SMARC	-6.98
E2f5	E2F1..5.p2	E2F1..5	6.62
Foxj2	FOX{I1,J2}.p2	FOXI1,J2	-5.62

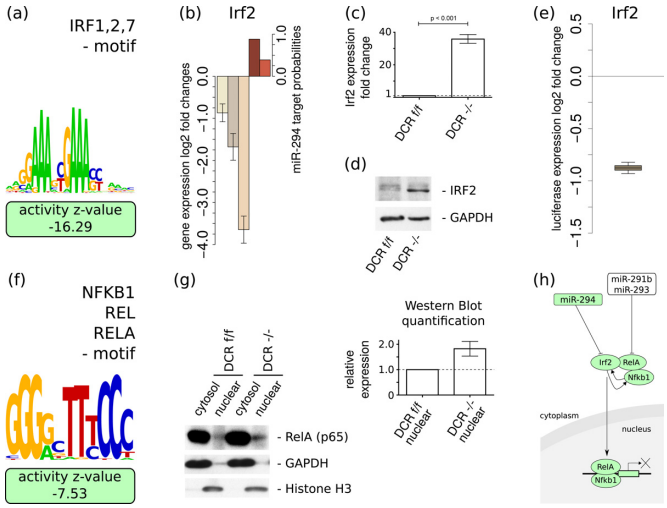

Figure 4. miR-294 targets the Irf2 TF and modulates 'IRF1,2,7' and 'NFKB1_REL_RELA' activities—**(a)** The activity of the 'IRF1,2,7' motif is strongly decreased in the presence of miR-294. **(b)** The expression of Irf2 is downregulated within all analysed data sets (±1.96*SEM; $n = 3$) and Irf2 is predicted by ElMMo and TargetScan to be a direct target of miR-294 (color scheme as in Figure 3). Low levels of Irf2 mRNA **(c)** and protein **(d)** in DCR$^{flox/flox}$ ES cells compared to miRNA deficient DCR$^{-/-}$ ESCs are observed with qRT-PCR and western blot, respectively. qRT-PCR experiments were run in triplicate (± SEM; $n = 3$). **(e)** The luciferase reporter construct carrying the Irf2 3′UTR shows a strong response to miR-294 co-transfection compared to a similar construct but with a mutated Irf2 target site ($n = 9$). **(f)** Sequence logo of the 'NFKB1_REL_RELA' motif that is associated with the canonical NF-κB pathway and that exhibits a significant decrease in activity in the presence of miR-294. **(g)** Western blots of RelA, GAPDH and Histone H3 in nuclear and cytoplasmic fractions in ESCs that do and do not express miRNAs. The densitometric quantification indicates an increased level of nuclear RelA in the DCR$^{-/-}$ ESCs compared to DCR$^{flox/flox}$ ESCs (± SEM; $n = 3$). **(h)** Proposed model of the inhibitory effect of miR-290-295 cluster miRNAs on the canonical NF-κB pathway in pluripotent stem cells. Regulatory motifs are denoted by colored rectangles and individual genes by ovals. See text for the evidence of individual interactions.

that Irf2 overexpression causes differentiation of ESCs (34). The strong impact of AAGUGCU miRNAs on Irf2 levels and the relatively large impact of the 'IRF1,2,7' motif on gene expression suggest that this regulatory connection plays an important role in maintaining ESC pluripotency.

Like the 'IRF1,2,7' motif, the 'NFKB1_REL_RELA' motif also exhibits a significantly lower activity when the embryonic miRNAs are expressed (Figure 4f). Western blot confirms that after stimulation with TNF-α, DCR$^{flox/flox}$ ESCs have lower levels of nuclear NF-κB pathway-associated marker RelA compared with miRNA-deficient DCR$^{-/-}$ ES cells (Figure 4g). This observation is consistent with a decreased activity of the canonical NF-κB signalling pathway in the presence of the miRNAs, which has been shown to be important for maintaining ESCs in a pluripotent state yet poised to undergo differentiation (35,36). Indeed, the Nanog pluripotency factor directly interacts with components of the NF-κB complex, inhibiting its transcriptional activity (35). Combining our results with recent reports that link the expression of the miR-290-295 cluster to signalling through the canonical NF-κB pathway and the latter to Irf2, the following model of the involvement of the miR-290-295 cluster in the regulation of NF-κB signalling emerges. Expression of the RelA component of the NF-κB complex is repressed post-transcriptionally by the miR-290-295 cluster members miR-291b-5p and miR-293 both of which do not belong to the AAGUGCU seed family of miRNAs (36). In humans, RelA recruitment to the nucleus, which is a pre-requisite for NF-κB complex-dependent transcription, appears to depend on IRF2 (37), whose knockdown interferes with transcriptional activation via NF-κB (37). Here we found that in mouse, IRF2 expression is also repressed by other members of the miR-290-295 cluster, namely, the AAGUGCU family of miRNAs. Thus, the miRNAs of the miR-290-295 cluster may act in concert to inhibit the canonical NF-κB signalling in ESCs (Figure 4h).

miRNAs of the AAGUGCU seed family impact the cell cycle at multiple levels

AAGUGCU seed family members of the miR-290-295 cluster were previously shown to accelerate the G1→S transition and promote proliferation of ESCs by targeting the cyclin E-Cdk2 regulatory pathway (13). Consistently, we found that these miRNAs increase the activity of transcription regulatory motifs associated with activation of the cell cycle (Figure 5a), in particular, the 'ARNT_ARNT2_BHLHB2_MAX_MYC_USF1' motif that is bound by Myc. This TF was previously found to increase upon miR-294 transfection (19). How the miRNAs, with intrinsically repressive function, increase the Myc activity on its targets is unknown. Our analysis suggests a few hypotheses.

Specifically, luciferase assays show that three cell cycle-associated TFs, namely, Mxd3 (also known as Mad3), E2f5 and Arnt2 are not only predicted but also experimentally confirmed direct targets of the AAGUGCU seed family

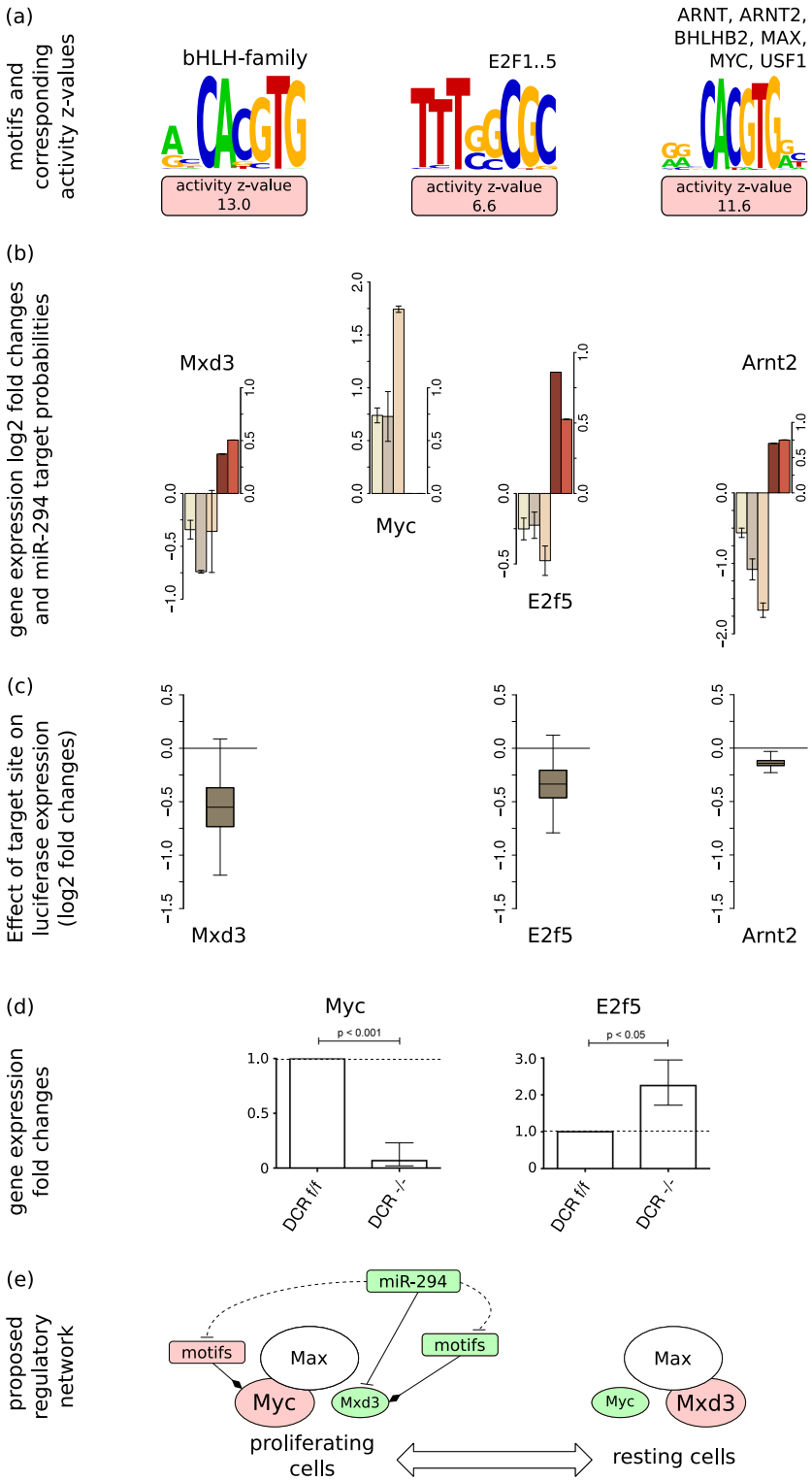

Figure 5. miR-294 impacts cell cycle regulation at multiple levels—**(a)** MARA analysis reveals that miR-294 induces positive activity changes of multiple motifs involved in cell cycle regulation. Shown are the sequence logos of these motifs: the Myc- and Arnt2-associated motif 'ARNT_ARNT2_BHLHB2_MAX_MYC_USF1', the putative Myc-regulating 'E2F1..5' motif and the Mxd3-associated 'bHLH-family' motif. **(b)** $\log_2$ mRNA fold changes (±1.96*SEM; $n = 3$) of Myc, Arnt2, E2f5 and Mxd3 (color scheme as in Figure 3) in the analyzed data sets. **(c)** Luciferase constructs carrying the 3′UTR of Arnt2, E2f5 or Mxd3, respectively, are downregulated upon co-transfection with miR-294 relative to constructs carrying the same 3′UTRs but with mutated miR-294 binding sites ($n = 9$). **(d)** qRT-PCR shows decreased expression of Myc and increased expression of E2f5 in $DCR^{-/-}$ ESCs relative to $DCR^{flox/flox}$ ESCs. qRT-PCR experiments were run in triplicate (±SEM; $n = 3$). **(e)** Proposed model of miR-294-dependent regulation of the Myc-Max/Mxd-Max network. Shapes scheme is as in Figure 4. Green or red shapes represent negative or positive changes (in motif activities or gene expression fold changes), respectively. Dashed lines indicate indirect and solid lines direct regulatory links between motifs/genes.

miRNAs (Figure 5b and c and Table 1). Mxd3 is one of the so-called 'Mad' partners of the Max protein (reviewed in (38)). In contrast to Myc, which forms a heterodimeric complex with Max in proliferating cells (39), the Mad factors Mad1, Mad3 (*i.e.* Mxd3) and Mad4 are primarily expressed and form complexes with Max in differentiating, growth-arrested cells (40). Mxd3 was further shown to specifically regulate the S-phase (41).

Second, we found that E2f5, one of the TFs associated with the 'E2F1..5′ motif, was consistently downregulated at the mRNA level in all analyzed data sets (Figure 5b) and luciferase assays further confirm that E2f5 is a target of miR-294 (Figure 5c), albeit with a small response to the miRNA. Consistently, E2f5 expression is increased in DCR$^{-/-}$ ESCs compared to DCR$^{flox/flox}$ ESCs (Figure 5d). The positive activity change of the E2F1..5 motif in the presence of the miRNAs (Figure 5a) suggests that this TF acts predominantly as repressor (as proposed before, reviewed in (42)). Notably, Myc is among the predicted targets of E2F1..5, providing an indirect path to the upregulation of Myc upon the presence of the miRNAs (Figure 5b and d).

Finally, Arnt2, a TF associated with the 'ARNT_ARNT2_BHLHB2_MAX_MYC_USF1' motif, but also with the 'AHR_ARNT_ARNT2' motif that corresponds to the complex of Arnt2 and Ahr, is also a predicted direct target of the AAGUGCU seed family which we validated in a luciferase assay (Figure 5c). This TF forms heterodimers with the aryl-hydrocarbon receptor (AHR) (43) and appears to be involved in the differentiation of ESCs into endothelial cells under hypoxic conditions (44), but otherwise little is known about its function. Given that Arnt2 and Myc (45) share the same binding motif, an interesting hypothesis is that Arnt2 competes with Myc for binding to targets and that its downregulation by AAGUGCU miRNAs allows Myc to act at promoters which would otherwise be bound by Arnt2. This hypothesis is again consistent with a positive Myc activity in ESCs, in which these miRNAs are expressed.

The model that we propose based on these results is that miRNAs of the AAGUGCU family regulate the cell cycle and the G→S transition through multiple pathways that come together in the increased expression of the crucial Myc regulator (Figure 5e). The miRNAs are able to downregulate the Mxd3 antagonist of Myc, the E2f5 repressor which would in turn result in the increased expression of E2f5 targets including Myc, and can downregulate Arnt2 which may compete with Myc for binding to regulatory sites.

miRNAs of the AAGUGCU seed family control multiple epigenetic regulators

As TFs, epigenetic regulators are also enriched among the targets of miRNAs (46). A role for the miR-290-295 cluster in epigenetic regulation was already proposed by Sinkkonen *et al.* (17), who found that expression of retinoblastoma-like 2 (Rbl-2) protein, a known repressor of the *de novo* methyltransferases, is controlled by these miRNAs. Through our analysis we found that the AAGUGCU miRNAs directly target the epigenetic regulator BAF170 (Smarcc2), a component of ATP-dependent, BAF (BRG1-associated factor) complexes (also known as SWI/SNF complexes) that remodel the nucleosome structure and thereby regulate gene expression (reviewed in (47)). The activity of the BAF170 motif changed significantly upon AAGUGCU miRNA expression in miRNA-deficient ESCs (Figure 6a, Table 1), accompanied by consistent downregulation of BAF170 mRNA (Figure 6b). Comparing constructs with and without the putative miR-294 binding site in the BAF170 3′UTR in a luciferase assay we found that BAF170 is significantly downregulated by miR-294 (Figure 6c), indicating that BAF170 is indeed a direct target of miR-294.

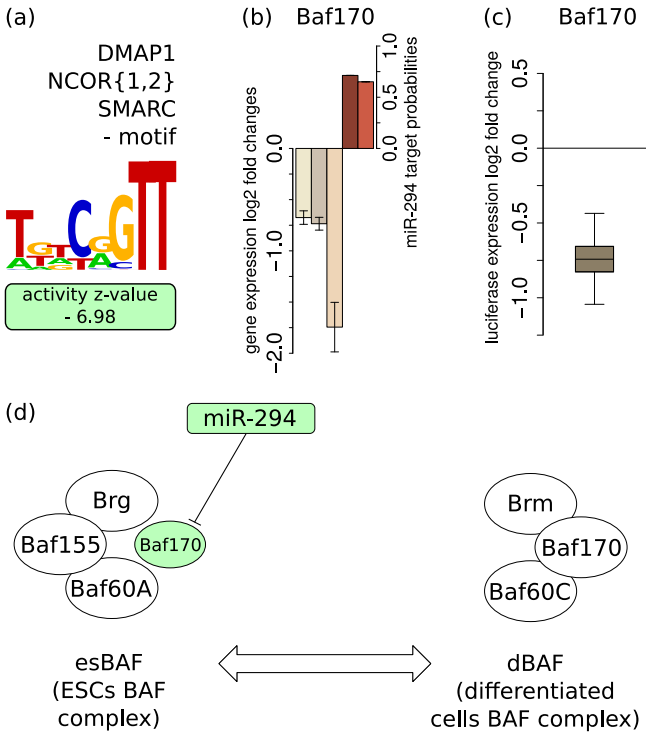

Figure 6. The BAF170 (Smarcc2) component of the dBAF chromatin remodeling complex is a direct target of miR-294—**(a)** MARA analysis reveals a negative activity change of the 'DMAP1_NCOR{1,2}_SMARC' motif in the presence of miR-294. **(b)** Expression of BAF170 (Smarcc2) is consistently downregulated in the presence of miR-294 in all considered experimental data sets (±1.96*SEM; $n = 3$; color scheme as in Figure 3). **(c)** A luciferase construct carrying the BAF170 3′UTR is downregulated upon co-transfection with miR-294 relative to a construct carrying a mutated 3′UTR ($n = 9$). **(d)** Model of the possible involvement of miR-294 in the maintenance of the ESC-specific chromatin remodeling complex esBAF. The miRNA-induced reduction in BAF170 levels may contribute to the maintenance of appropriate levels of esBAF complexes in ESCs thereby maintaining self-renewal and proliferation (48). Color, shapes and lines scheme is as in Figure 5.

Recently, it was shown that BAF170 is downregulated during miR-302-367-based reprogramming and that BAF170 knockdown increases the number of iPSC colonies in somatic cell reprogramming (49). As miRNAs of the miR-302-367 cluster share the seed sequence with miR-294, it is likely that miR-294 has similar effects on BAF170 expression and pluripotency.

The model that emerges from these studies is that the AAGUGCU family of miRNAs may play a role in the remodeling of BAF complexes. In ESCs, the BAF complex (esBAF), which contains a BAF155 subunit, shares a large

proportion of target genes with the pluripotency-associated TFs Oct4, Sox2 and Nanog (50) and is required for the self-renewal and maintenance of pluripotency in mESCs (48). Consistently, overexpression of esBAF components was found to promote reprogramming (51). In differentiated cells, however, the so-called differentiated cell BAF complex (dBAF) (52), contains the BAF170 and not the BAF155 subunit (48). The fact that induced BAF170 expression in ESCs decreases the level of BAF155 protein suggested that BAF170 can displace BAF155 from esBAF, thereby increasing its degradation rate (48). By preventing expression of BAF components that are specific to differentiated cells and that antagonize embryonic state-specific BAF (Figure 6d), the AAGUGCU family of miRNAs may promote an ESC-specific epigenetic state.

DISCUSSION

It has been established that ESC-specific miRNAs that share an AAGUGCU seed region are among the regulatory factors that are necessary to maintain a pluripotent ESC state. Strikingly, overexpression of a cluster of ESC-specific miRNAs was found sufficient for inducing reprogramming of differentiated cells into iPSCs. This suggests that the miRNAs can set into motion an entire regulatory cascade that leads to cell reprogramming. Several studies determined the gene expression profiles of ESCs that did and did not express AAGUGCU family miRNAs. An insight emerging from these studies was that miR-290-295 miRNAs regulate the cell cycle and apoptosis, either directly or indirectly.

To better understand how the direct regulatory factor targets of these miRNAs contribute to pluripotency, we made use of a recently developed method, called MARA, that models gene expression in terms of computationally predicted regulatory sites. The approach originates in regression models that were first proposed by Bussemaker *et al.* (53) for inferring regulatory elements from gene expression data. However, MARA's goal is different. It uses predicted regulatory sites in combination with a linear model to infer from gene expression data the activities of transcriptional regulators. The first application of MARA (28) to the reconstruction of the core transcriptional regulatory network of a differentiating human cell line, demonstrated that the method can successfully infer key regulatory interactions *ab initio*. Notably, it was found that MARA accurately infers the activities of the key regulatory motifs, in spite of computational predictions of regulatory sites being error-prone, and of gene expression likely being a much more complex function of the regulatory sites. The power of the method stems from the fact that motif activities are inferred from the *statistics* of expression of hundreds to thousands of putative target genes of each regulatory motif. Here we have used an extended version of the MARA model, which also includes predicted miRNA binding sites, to infer both transcriptional and post-transcriptional regulators of mRNA expression levels. A similar approach was recently applied by Setty *et al.* (54) to reconstruct the regulatory networks in glioblastoma.

The TF targets of the AAGUGCU miRNAs that we identified with the extended MARA model had the following properties:

(i) The activity of their corresponding motif changed significantly upon expression of the AAGUGCU miRNAs, meaning that the predicted targets of these regulators showed, on average, consistent expression changes.
(ii) Their expression was consistently downregulated at the mRNA level upon expression of the AAGUGCU miRNAs.
(iii) They were predicted as direct targets of the AAGUGCU family of miRNAs by miRNA target prediction programs.
(iv) They were consistently (*i.e.* within every analyzed data set) predicted by MARA to be directly regulated by the AAGUGCU seed family of miRNAs on the basis of the dependence of their expression changes on the presence of the miRNA binding sites in their 3′UTRs.
(v) They could be confirmed as AAGUGCU miRNA targets with luciferase assays.

Altogether, these lines of evidence firmly establish these transcriptional regulators as direct targets of the AAGUGCU seed family miRNAs, forming the first layer downstream of this miRNAs in the regulatory network of pluripotency.

First, our analysis suggests that AAGUGCU miRNAs target the cell cycle, and in particular the G1→S transition, through multiple pathways. By targeting the repressive cell cycle regulator E2f5, the miRNAs might directly promote the G1→S transition. In addition, the miRNAs seem to increase the activity of the proliferation-associated TF Myc through multiple indirect routes, including shifting the balance between Myc and its antagonist Mxd3 within transcription regulatory complexes that act on Myc target genes. Second, we found that the AAGUGCU miRNAs may affect the balance between chromatin remodeling complexes that are active in ESCs and in differentiated cells, a function probably important for keeping specific genomic regions from being silenced through heterochromatin formation. Third, we found that the AAGUGCU miRNAs directly target the interferon regulatory factor Irf2, whose expression is strongly increased in $DCR^{-/-}$ cells, consistent with a significant change in the regulatory impact that we inferred for this factor. Finally, our analysis uncovers a few transriptional regulators that have previously not been connected to the transcriptional network of pluripotent stem cells, including Foxj2, whose expression is strongly affected by the miRNAs and the Clock (circadian locomotor output cycles kaput) TF. Interestingly, circadian oscillations are not present in mouse ES cells, but are switched on during differentiation, and then disappear again upon reprogramming of differentiated cells into iPSCs (55). It is thus tempting to speculate that circadian oscillations in ESCs may be actively suppressed by the AAGUGCU miRNAs and that downregulation of these miRNAs during development may be necessary for the establishment of circadian rhythms. However, the response of the 3′UTR of Clock in luciferase assays was

very variable in our hands, and we were not able to unambiguously validate it as a direct target of miR-294.

As mentioned before, the AAGUGCU seed motif is not unique to miRNAs of the mouse-specific miR-290-295 cluster. It also occurs in the miR-302 family of miRNAs that is present in human and in a shifted version (at positions 3–9 instead of 2–8) it occurs in the miR-17/20a miRNAs of the oncogenic miR-17-92 cluster. Although miR-19 has been reported to be the key oncogenic component of this cluster (56), the strong effects that AAGUGCU miRNAs exert on the cell cycle raise the question of whether miR-17 and miR-20a may not play a role similar to miR-294 in malignant cells.

In summary, our analysis demonstrates that combining accurate predictions of regulatory elements with analysis of transcriptome-wide mRNA expression changes in response to specific manipulations is a general and powerful approach to uncovering key regulators within gene expression networks. In the future, incorporation of measurements of miRNA expression as well as of predictions of TF binding sites in miRNA genes will enable identification of feedback loops between miRNAs and TFs that are known to operate in many systems.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank the members of the Zavolan group for feedback on the manuscript.

FUNDING

Swiss National Science [#31003A_127307]; European Research Council Starting Grant [to M.Z.]. Werner Siemens fellowship at the Biozentrum [to A.J.G.]. Source of open access funding: Biozentrum, University of Basel.
Conflict of interest statement. None declared.

REFERENCES

1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
2. Lim,L.P., Lau,N.C., Garrett-Engele,P., Grimson,A., Schelter,J.M., Castle,J., Bartel,D.P., Linsley,P.S. and Johnson,J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.
3. Selbach,M., Schwanhäusser,B., Thierfelder,N., Fang,Z., Khanin,R. and Rajewsky,N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
4. Friedman,R.C., Farh,K.K.-H., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
5. Kanellopoulou,C., Muljo,S.A., Kung,A.L., Ganesan,S., Drapkin,R., Jenuwein,T., Livingston,D.M. and Rajewsky,K.C.P. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.*, **19**, 489–501 .
6. Murchison,E.P., Partridge,J.F., Tam,O.H., Cheloufi,S. and Hannon,G.J. (2005) Characterization of Dicer-deficient murine embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12135–12140.
7. Wang,Y., Medvid,R., Melton,C., Jaenisch,R. and Blelloch,R. (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat. Genet.*, **39**, 380–385.
8. Babiarz,J.E., Ruby,J.G., Wang,Y., Bartel,D.P. and Blelloch,R. (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.*, **22**, 2773–2785.
9. Ciaudo,C., Servant,N., Cognat,V., Sarazin,A., Kieffer,E., Viville,S., Colot,V., Barillot,E., Heard,E. and Voinnet,O. (2009) Highly dynamic and sex-specific expression of microRNAs during early ES cell differentiation. *PLoS Genet.*, **5**, e1000620.
10. Leung,A.K., Young,A.G., Bhutkar,A., Zheng,G.X., Bosson,A.D., Nielsen,C.B. and Sharp,P.A.C.P. (2011) Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat. Struct. Mol. Biol.*, **18**, 237–244 .
11. Zheng,G.X., Ravi,A., Calabrese,J.M., Medeiros,L.A., Kirak,O., Dennis,L.M., Jaenisch,R., Burge,C.B. and Sharp,P.A.C.P. (2011) A latent pro-survival function for the mir-290-295 cluster in mouse embryonic stem cells. *PLoS Genet.*, **7**, e1002054 .
12. Houbaviy,H.B., Murray,M.F. and Sharp,P.A. (2003) Embryonic stem cell-specific MicroRNAs. *Dev. Cell*, **5**, 351–358.
13. Wang,Y., Baskerville,S., Shenoy,A., Babiarz,J.E., Baehner,L. and Blelloch,R. (2008) Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.*, **40**, 1478–1483.
14. Judson,R.L., Babiarz,J.E., Venere,M. and Blelloch,R. (2009) Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat. Biotechnol.*, **27**, 459–461.
15. Kuo,C.-H., Deng,J.H., Deng,Q. and Ying,S.-Y. (2012) A novel role of miR-302/367 in reprogramming. *Biochem. Biophys. Res. Commun.*, **417**, 11–16.
16. Anokye-Danso,F., Trivedi,C.M., Juhr,D., Gupta,M., Cui,Z., Tian,Y., Zhang,Y., Yang,W., Gruber,P.J., Epstein,J.A. and Morrisey,E.E. (2011) Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*, **8**, 376–388.
17. Sinkkonen,L., Hugenschmidt,T., Berninger,P.,,Gaidatzis, D., Mohn,F., Artus-Revel,C.G., Zavolan,M., Svoboda,P. and Filipowicz,W. (2008) MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.*, **15**, 259–267 .
18. Hanina,S.A., Mifsud,W., Down,T.A., Hayashi,K., O'Carroll,D., Lao,K., Miska,E.A. and Surani,M.A. (2010) Genome-wide identification of targets and function of individual MicroRNAs in mouse embryonic stem cells. *PLoS Genet.*, **6**, e1001163.
19. Melton,C., Judson,R.L. and Blelloch,R. (2010) Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*, **463**, 621–626.
20. Gautier,L., Cope,L., Bolstad,B.M. and Irizarry,R.A. (2004) affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)*, **20**, 307–315.
21. Wu,Z., Irizarry,R.A., Gentleman,R., Martinez-Murillo,F. and Spencer,F. (2004) A model-based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.*, **99**, 909–917.
22. Fraley,C. and Raftery,A.E. (2009) mclust: Model-based clustering/normal mixture modeling. R package version 3.
23. Carvalho,B.S. and Irizarry,R.A. (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics (Oxford, England)*, **26**, 2363–2367.
24. Irizarry,R., Hobbs,B., Collin,F., D Beazer-Barclay,Y., J Antonellis,K., Scherf,U. and P Speed,T. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics Oxford England*, **4**, 249–264.
25. Balwierz,P.J., Pachkov,M., Arnold,P., Gruber,A.J., Zavolan,M. and van Nimwegen,E. (2014) ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res.*, **24**, 869–884.
26. Ciaudo,C., Jay,F., Okamoto,I., Chen,C.-J., Sarazin,A., Servant,N., Barillot,E., Heard,E. and Voinnet,O. (2013) RNAi-dependent and independent control of LINE1 accumulation and mobility in mouse embryonic stem cells. *PLoS Genet.*, **9**, e1003791.
27. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, **25**, 402–408.
28. Suzuki,H., Forrest,A.R.R., van Nimwegen,E., Daub,C.O., Balwierz,P.J., Irvine,K.M., Lassmann,T., Ravasi,T., Hasegawa,Y., de Hoon,M.J.L. *et al.* (2009) The transcriptional network that controls

growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.*, **41**, 553–562.
29. Tsang,J.S., Ebert,M.S. and van Oudenaarden,A. (2010) Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures. *Mol. Cell*, **38**, 140–153.
30. Gaidatzis,D., van Nimwegen,E., Hausser,J. and Zavolan,M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformat.*, **8**, 69–92.
31. Granadino,B., Arias-de-la Fuente,C., Pérez-Sánchez,C., Párraga,M., López-Fernández,L.A., del Mazo,J. and Rey-Campos,J. (2000) Fhx (Foxj2) expression is activated during spermatogenesis and very early in embryonic development. *Mech. Dev.*, **97**, 157–160.
32. Martín-de Lara,F., Sánchez-Aparicio,P., Arias de la Fuente,C. and Rey-Campos,J. (2008) Biological effects of FoxJ2 over-expression. *Transgenic Res.*, **17**, 1131–1141.
33. Skinner,M.K., Rawls,A., Wilson-Rawls,J. and Roalson,E.H. (2010) Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature. *Differentiation; Res. Biol. Diversity*, **80**, 1–8.
34. Yamamizu,K., Piao,Y., Sharov,A.A., Zsiros,V., Yu,H., Nakazawa,K., Schlessinger,D. and Ko,M. S.H. (2013) Identification of transcription factors for lineage-specific ESC differentiation. *Stem Cell Rep.*, **1**, 545–559.
35. Torres,J. and Watt,F.M. (2008) Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3. *Nat. Cell Biol.*, **10**, 194–201.
36. Lüningschrör,P., Stöcker,B., Kaltschmidt,B. and Kaltschmidt,C. (2012) miR-290 cluster modulates pluripotency by repressing canonical NF-κB signaling. *Stem Cells (Dayton, Ohio)*, **30**, 655–664.
37. Chae,M., Kim,K., Park,S.-M., Jang,I.-S., Seo,T., Kim,D.-M., Kim,I.-C., Lee,J.-H. and Park,J. (2008) IRF-2 regulates NF-kappaB activity by modulating the subcellular localization of NF-kappaB. *Biochem. Biophys. Res. Commun.*, **370**, 519–524.
38. Hurlin,P.J. and Huang,J. (2006) The MAX-interacting transcription factor network. *Seminars Cancer Biol.*, **16**, 265–274.
39. Ayer,D.E. and Eisenman,R.N. (1993) A switch from Myc:Max to Mad:Max heterocomplexes accompanies monocyte/macrophage differentiation. *Genes Dev.*, **7**, 2110–2119.
40. Hurlin,P.J., Quéva,C., Koskinen,P.J., Steingrímsson,E., Ayer,D.E., Copeland,N.G., Jenkins,N.A. and Eisenman,R.N. (1995) Mad3 and Mad4: novel Max-interacting transcriptional repressors that suppress c-myc dependent transformation and are expressed during neural and epidermal differentiation. *EMBO J.*, **14**, 5646–5659.
41. Quéva,C., McArthur,G.A., Iritani,B.M. and Eisenman,R.N. (2001) Targeted deletion of the S-phase-specific Myc antagonist Mad3 sensitizes neuronal and lymphoid cells to radiation-induced apoptosis. *Mol. Cell. Biol.*, **21**, 703–712.
42. Cam,H. and Dynlacht,B.D. (2003) Emerging roles for E2F: beyond the G1/S transition and DNA replication. *Cancer Cell*, **3**, 311–316.
43. Sekine,H., Mimura,J., Yamamoto,M. and Fujii-Kuriyama,Y. (2006) Unique and overlapping transcriptional roles of arylhydrocarbon receptor nuclear translocator (Arnt) and Arnt2 in xenobiotic and hypoxic responses. *J. Biol. Chem.*, **281**, 37507–37516.
44. Han,Y., Kuang,S.-Z., Gomer,A. and Ramirez-Bergeron,D.L. (2010) Hypoxia influences the vascular expansion and differentiation of embryonic stem cell cultures through the temporal expression of vascular endothelial growth factor receptors in an ARNT-dependent manner. *Stem Cells (Dayton, Ohio)*, **28**, 799–809.
45. Smith,K. and Dalton,S. (2010) Myc transcription factors: key regulators behind establishment and maintenance of pluripotency. *Regenerative Med.*, **5**, 947–959.
46. Gruber,A.J. and Zavolan,M. (2013) Modulation of epigenetic regulators and cell fate decisions by miRNAs. *Epigenomics*, **5**, 671–683.
47. Wilson,B.G. and Roberts,C.W.M. (2011) SWI/SNF nucleosome remodellers and cancer. *Nat. Rev. Cancer*, **11**, 481–492.
48. Ho,L., Ronan,J.L., Wu,J., Staahl,B.T., Chen,L., Kuo,A., Lessard,J., Nesvizhskii,A.I., Ranish,J. and Crabtree,G.R. (2009) An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 5181–5186.
49. Subramanyam,D., Lamouille,S., Judson,R.L., Liu,J.Y., Bucay,N., Derynck,R. and Blelloch,R. (2011) Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat. Biotechnol.*, **29**, 443–448.
50. Ho,L., Jothi,R., Ronan,J.L., Cui,K., Zhao,K. and Crabtree,G.R. (2009) An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 5187–5191.
51. Singhal,N., Graumann,J., Wu,G., Araúzo-Bravo,M.J., Han,D.W., Greber,B., Gentile,L., Mann,M. and Schöler,H.R. (2010) Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell*, **141**, 943–955.
52. He,L., Liu,H. and Tang,L. (2012) SWI/SNF chromatin remodeling complex: a new cofactor in reprogramming. *Stem Cell Rev.*, **8**, 128–136.
53. Bussemaker,H.J., Li,H. and Siggia,E.D. (2001) Regulatory element detection using correlation with expression. *Nat. Genet.*, **27**, 167–171.
54. Setty,M., Helmy,K., Khan,A.A., Silber,J., Arvey,A., Neezen,F., Agius,P., Huse,J.T., Holland,E.C. and Leslie,C.S. (2012) Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma. *Mol. Syst. Biol.*, **8**, 605–620.
55. Yagita,K., Horie,K., Koinuma,S., Nakamura,W., Yamanaka,I., Urasaki,A., Shigeyoshi,Y., Kawakami,K., Shimada,S., Takeda,J. *et al.* (2010) Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 3846–3851.
56. Mu,P., Han,Y.-C., Betel,D., Yao,E., Squatrito,M., Ogrodowski,P., de Stanchina,E., D'Andrea,A., Sander,C. and Ventura,A. (2009) Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. *Genes Dev.*, **23**, 2806–2811.

Chapter II. Implication of TFAP2A in Epithelial Plasticity in Breast Cancer and Development

Epithelial Plasticity

Epithelial to mesenchymal transitions (EMT) describes a transition that epithelial cells undergo, progressively losing their epithelial characteristics, and acquiring a new mesenchymal phenotype. Elizabeth Hay and colleagues observed and documented such changes for the first time, in the primitive streak formation of chicken embryo (Trelstad, Hay et al. 1967). The process is reversible, the term mesenchymal to epithelial transition (MET) being used to describe the opposite process (Lamouille, Xu et al. 2014). Incomplete EMT may also exist (Bryant and Mostov 2008).

The epithelial organization is evolutionary older than Metazoa. Simple, non-cadherin-based, and polarized epithelial structures are already found in the fruiting body formed by the unicellular *Dictyostelium discoideum* (Dickinson, Nelson et al. 2011). In multicellular organisms the epithelial tissue fulfills the function of creating a barrier between two different media, frequently segregating internal from external environments (Rodriguez-Boulan and Macara 2014). Characteristic of epithelial organization is the close contact in between adjacent cells, but also their apico-basal polarity, stabilized by adherens junctions, desmosomes and tight junctions (Bryant and Mostov 2008). The epithelium is separated from neighboring tissues by basal lamina, an assembly of extracellular proteins and glycoproteins (Thiery, Acloque et al. 2009). In contrast, the connective tissue surrounding epithelia is formed from unconfined mesenchymal or stromal cells, embedded in a 3D extracellular matrix (Thiery, Acloque et al. 2009).

A number of specific events occur during EMT, including the disassembly of epithelial junctions, loss of apico-basal cell polarity, cell morphological changes, cytoskeleton rearrangements, and increased cell motility (Spano, Heck et al. 2012). EMT is a critical process during embryonic development; it also participates in normal and pathological mechanisms such as wound healing, fibrosis and tumor metastasis. Moreover, ECS differentiation and reprogramming of somatic cells to

iPSC, involve respectively EMT and MET steps (Goding, Pei et al. 2014). In addition, EMT is also suggested to associate to the origin of cancer stem cells (Mani, Guo et al. 2008)

All of these processes, which involve complete transition or partial modification of epithelial phenotype, are collectively referred to as epithelial plasticity.

Epithelial Plasticity During Embryonic Development

EMT is fundamental process that is part of the embryonic development in vertebrates. The majority of adult organs and tissues are formed in result of a sequence of successive EMT and MET transitions. The mesenchymal cells formed from the primary EMT can undergo the reverse process that is MET, and thus form transient epithelial structures. For instance, the heart is formed after three consecutive rounds of EMT and MET (Thiery, Acloque et al. 2009). The successive cycles of conversion in between the epithelial and mesenchymal cell state, involved in embryonic cells differentiation and organogenesis, are named primary, secondary and tertiary EMT, respectively (Thiery, Acloque et al. 2009).

The first EMT event described in early embryogenesis is the formation of the primitive endoderm from ICM cells that later contributes to extra-embryonic tissues (Figure 5) (Thiery, Acloque et al. 2009). Another example of profoundly investigated EMT process is in the course of gastrulation, which leads to the segregation of three germ layers (Figure 5) (Nakaya and Sheng 2008). The EMT conversion is involved in the formation of a structure of cells that arises from the epiblast, and constitutes the primitive streak. Cells from the epiblast migrate towards the internal part of the embryo along the primitive streak to form the definitive endoderm and later the mesoderm layers of the embryo (Arnold and Robertson 2009; Thiery, Acloque et al. 2009).

The neural crest cells (NCC) population arises from the ectoderm, surrounded by the neural plate and prospective epidermis. NCCs gain extensive migratory capacity and separate from the neuroepithelium through a complete or partial EMT, the process is referred as delamination (Figure 5). Afterwards they migrate in the embryo and due to their multipotent nature are at the origin of a variety of cell types, including certain neurons, glial cells, constituents of the peripheral nervous system, cardiac structures,

endocrine cells, smooth muscle and others (Thiery, Acloque et al. 2009; Theveneau and Mayor 2012).

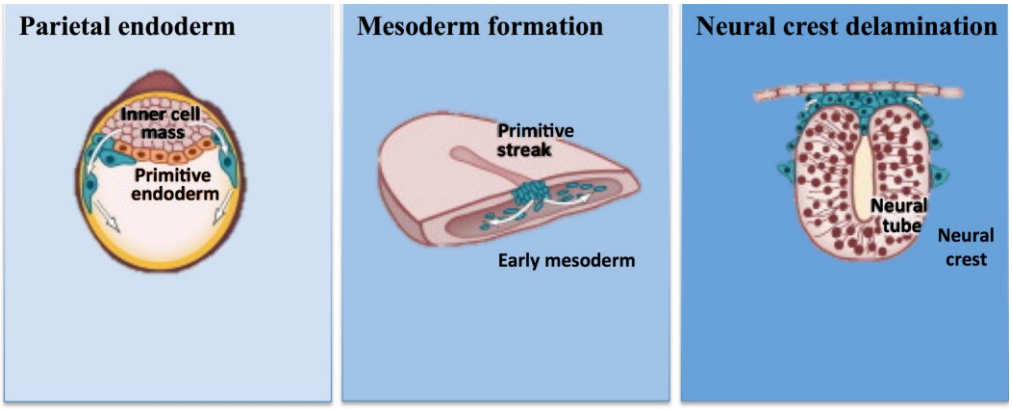

Figure 5. Primary EMT during Embryonic Development. Primary EMT starts before implantation of the embryo with the formation of the parietal endoderm. Next, after implantation, the mesodermal progenitors also undergo EMT during gastrulation. The neural crest delamination is a later event, following the embryogenesis. Reprinted from (Thiery, Acloque et al. 2009) with permission from Elsevier.

Epithelial Plasticity During Cancer Progression

The activation of invasion and metastasis is considered a hallmark in human tumors and it is a major contributor to cancer related mortality (Hanahan and Weinberg 2011; Slattum and Rosenblatt 2014) (Bill and Christofori 2015). The metastasis formation is considered as a multistep mechanism. It starts with tumor cells, invading the surrounding tissues, followed by their extravasation into neighboring blood and lymph vessels. After transit in the circulatory system, the cells undergo extravasation towards the parenchyma of distant tissues, which then they colonize (Fidler 2003; Hanahan and Weinberg 2011; Bill and Christofori 2015). EMT is related to increased migration and invasiveness, and it is thought to be involved in cancer malignancies, where epithelial tumor cells escape the initial tumor site, by invading nearby stroma and reaching the circulatory system (Bill and Christofori 2015). Furthermore, cooperation between cancer cells and platelets in blood vessels might contribute to lung metastasis by induction of EMT in the tumor cells (Labelle, Begum et al. 2011).

On the other hand MET is suggested to be important for the colonization step, which involves the metastatic outgrowth in the secondary cancer site (Brabletz 2012). Moreover, it has been proposed that EMT may play a role beyond metastatic formation, regulating processes like tumor initiation and drug resistance, by promoting the cancer stem cells phenotype (Mani, Guo et al. 2008; De Craene and Berx 2013; Plaks, Kong et al. 2015).

It is well established that certain cancer cell lines undergo EMT *in vitro* upon exogenous expression of EMT master regulators or stimulation with different factors, such as TGFβ. However, the clinical relevance of EMT in tumor formation and progression is still debated mainly because the evidence of EMT in the tumor are to certain extent speculative (De Craene and Berx 2013; Bill and Christofori 2015). In support of the involvement of EMT, the loss or down regulation of E-cadherin (CDH1), an epithelial adherens junction protein and hallmark of the transition, in carcinoma cells is frequently related to a malignancy progression (Cavallaro and Christofori 2004; Yilmaz and Christofori 2010). Nevertheless, the link between EMT and metastasis remains controversial in the lobular breast carcinoma, which is due to an inactivating mutation of *CDH1*, but has also well delineated epithelial features (Lombaerts, van Wezel et al. 2006). Aggressive breast cancer cell lines and tumors subtypes, such as basal B and claudin-low, were related with an EMT transcriptomic signature (Neve, Chin et al. 2006; Herschkowitz, Simin et al. 2007; Hennessy, Gonzalez-Angulo et al. 2009). Despite the fact that an EMT transcriptomic signature does not correlate with survival and it is not informative of poor outcome, specific regulators of EMT such as SOX4, PRRX1 and LXH2 do show such relationships (Ocana, Corcoles et al. 2012; Tiwari, Tiwari et al. 2013; Kuzmanov, Hopfer et al. 2014; Bill and Christofori 2015). It is not clear, however, if those tumors undergo EMT, at what stage they do and if this is essential for metastasis formation (Bill and Christofori 2015).

Gene Regulatory Networks Involved in EMT

A number of transcription factors were shown to be master regulators of EMT and to be able to induce it in different contexts. Among those are snail family zinc finger 1 and 2 (SNAI1 and 2), twist family bHLH transcription factor 1 and 2 (TWIST1 and

2), zinc-finger E-box binding homeobox 1 and 2 (ZEB1 and 2) (Table 3) (Lamouille, Xu et al. 2014). Beyond these transcription master regulators, epithelial plasticity is regulated at a number of distinct levels, including splicing and miRNA-dependent silencing, by interconnected regulatory networks (Figure 6) (De Craene and Berx 2013).

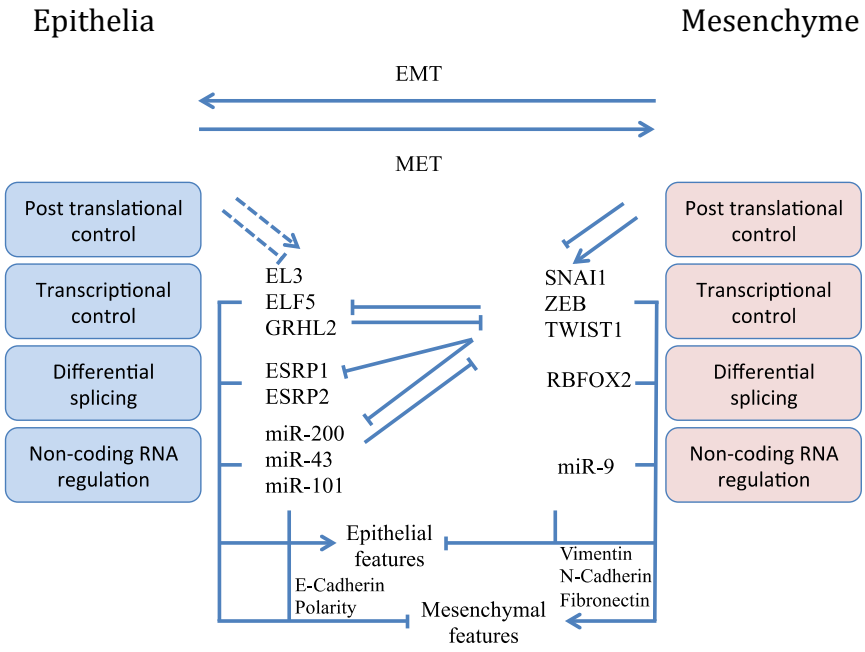

Figure 6. EMT is controlled by interconnected regulatory networks. ZEB, SNAI1 and TWIST1 are key EMT-inducing transcription factors and have a central role in the regulatory network. miRNAs form negative feedback loops with these transcription factors. Adapted by permission of Macmillian Publishers Ltd Nature Reviews Cancer (De Craene and Berx 2013), copyright 2013

EMT is associated with important changes in genes expression and large number of genes is affected (Lamouille, Xu et al. 2014). Among the epithelium-specific genes, changes are observed in those encoding epithelial tight junction and desmosome complexes, including claudins, occludin, desmoplakin and plakophilin, respectively (Lamouille, Xu et al. 2014). E-cadherin deserves a particular attention as its down-regulation is considered a hallmark of the transition (Yilmaz and Christofori 2010). Furthermore, the down-regulation of E-cadherin is often coupled with the up-regulation of N-cadherin, and therefore, mesenchymal adherens junctions replace the epithelial adherens junctions, in a phenomenon known as "cadherin switch" (Wheelock, Shintani et al. 2008; Yilmaz and Christofori 2010). Other adhesion

molecules are unregulated upon EMT, for instance neural cell adhesion molecule (NCAM) expression increases during the transition and it participates in focal adhesion assembly and cell migration (Lehembre, Yilmaz et al. 2008). In additions, the expression of genes involved in the formation of cytoskeleton, such as cytokeratin and vimentin is modified (Lamouille, Xu et al. 2014).

EMT is also coupled with changes in the way cells interact with components of the extracellular matrix (ECM) (Lamouille, Xu et al. 2014). Fibronectin (FN1) is found in the stromal ECM of breast tumors (Christensen 1992). Its expression is upregulated during the transition and its interactions with mammary epithelial cells promote EMT (Park and Schwarzbauer 2014). In the course of the transition cells down-regulate epithelial integrins and up-regulate those that enforce EMT and invasion (Yilmaz and Christofori 2009). For example, in Madin-Darby canine kidney (MDCK) cells the expression of SNAI1 induced up-regulation of αvβ3 integrin, recognized for its role in cancer invasion (Haraguchi, Okubo et al. 2008; Yilmaz and Christofori 2009).

Transcription Level Regulation of Epithelial Plasticity

Firstly, the SNAI1 transcription factor was identified to regulate EMT by interacting with CDH1 promoter and thus repressing its expression (Batlle, Sancho et al. 2000; Cano, Perez-Moreno et al. 2000). Consequently, a number of other transcription factors including SNAI2, ZEB1, ZEB2, E47, Kruppel-like factor 8 (KLF8) and Brachyury were reported to directly repress E-cadherin expression together with other epithelial junction proteins and thus promote an EMT phenotype. Another continuously growing set of transcription factors, such as SOX4, paired mesoderm homeobox 1 (PRRX1), TWIST1, some of the Forkhead box protein, for instance C2 (FOXC2), high mobility group A2 (HMGA2), goosecoid, E2-2 (TCF4), certain GATA proteins and SIX1 were equally described to control the process without directly regulating E-cadherin promoter (Tiwari, Gheldof et al. 2012; De Craene and Berx 2013; Lamouille, Xu et al. 2014). A non-exhaustive list of transcription factors implicated in EMT is presented in Table 3. It is important to note that despite the multitude of transcription factors controlling the transition like ZEB, SNAIL and TWIST, have been described to be consistently involved in EMT (Peinado, Olmeda et al. 2007; De Craene and Berx 2013). Their expression or activity is activated in the

early steps of the process. Furthermore, the EMT inducing transcription factors exert their action on same or related pathways and are frequently part of interconnected regulatory network (Taube, Herschkowitz et al. 2010; De Craene and Berx 2013). They repress epithelial genes, while also activating the mesenchymal ones. In many cases they act in synergy and they regulate mutually their expression (De Craene and Berx 2013). Often transcription factors outside ZEB, SNAIL and TWIST would rather ease EMT and/or feed to those core factors (Table 3) (De Craene and Berx 2013).

Transcription Factor	Function in EMT	Reference
SNAI1 and 2	Zinc-finger protein, transcriptional repressor	(Batlle, Sancho et al. 2000; Cano, Perez-Moreno et al. 2000; De Craene, van Roy et al. 2005)
ZEB1 and 2	Zinc-finger protein, transcriptional repressor	(Tiwari, Gheldof et al. 2012)
KLF8	Zinc-finger protein, transcriptional repressor and activator	(Wang, Zheng et al. 2007; Lahiri and Zhao 2012)
Brachyury	Transcriptional activator	(Fernando, Litzinger et al. 2010)
TWIST1 and 2	bHLH factors	(Peinado, Olmeda et al. 2007)
FOXD3	Neural crest specifier	(Dottori, Gross et al. 2001)
FOXQ1	Repressed expression of the core EMT regulator e-cadherin	(Zhang, Meng et al. 2011)
FOXO3A	Notch regulator	Gopinath et al.(2014

FOXC2	Transcriptional activator	(Mani, Yang et al. 2007)
Goosecoid	Homeobox protein	(Hartwell, Muir et al. 2006)
E2-2/TCF4	Class I bHLH factor	(Peinado, Olmeda et al. 2007)
SIX1	Homeobox protein	(McCoy, Iwanaga et al. 2009; Micalizzi, Christensen et al. 2009)
PRRX1	Homeobox protein	(Ocana, Corcoles et al. 2012)
GATA4 and 6	Downregulation of junctional dE-Cadherin	(Campbell, Whissell et al. 2011)
HMGA2	Co-regulates SNAI1 expression	(Thuault, Tan et al. 2008)
SOX9	Co-regulates SNAI2 expression	(Sakai, Suzuki et al. 2006)
CBFA–KAP1	Transcriptional activator	(Venkov, Link et al. 2007)
ZNF703 /Zeppo1	Represses E-Cadherin	(Slorach, Chou et al. 2011)
PRX1	Regulates BMP2 and TGFβ	(Makrodouli, Oikonomou et al. 2011)
SOX4	Controls EZH2 expression	(Tiwari, Tiwari et al. 2013)
E47/TCF3	bHLH factor	(Peinado, Olmeda et al. 2007)

Table 3. Transcription factors and cofactors implicated in the regulation of EMT. Adapted by permission of Macmillian Publishers Ltd Nature Reviews Cancer (De Craene and Berx 2013), copyright 2013

SNAIL Transcription Factors

In mammals two out of three SNAIL proteins, namely SNAI1 and SNAI2, are involved in EMT. Mouse *Snai1* gene is expressed in the prospective mesoderm as well as at the edge of the neural plate (Cano, Perez-Moreno et al. 2000). During embryonic development SNAIL gene family are essential for gastrulation, left-right patterning and affect neural crest delamination (Alberga, Boulay et al. 1991; Nieto, Sargent et al. 1994; Carver, Jiang et al. 2001; Murray and Gridley 2006). In the adult, SNAI1 presence is restrained to mesenchymal cells of lung, dermis and cartilage as well to wound-healing activated fibroblasts and mesenchymal stem cells (Franci, Takkunen et al. 2006; Batlle, Alba-Castellon et al. 2013). SNAIL is also expressed in human and mouse invasive tumors and it represses E-cadherin expression. (Batlle, Sancho et al. 2000; Cano, Perez-Moreno et al. 2000).

SNAIL transcription factors are conserved in metazoans and possess a particular domain organization (Barrallo-Gimeno and Nieto 2005). At their C-terminus they contain from two to six zinc-fingers, which confers them the capacity to specifically recognize the DNA E Box element (CTGGTG) (Nieto 2002). The N-terminus sequence of SNAIL proteins is more divergent (Nieto 2002). Traditionally SNAIL transcription factors are considered as transcriptional repressors, but they have been also demonstrated to activate certain target genes (Barrallo-Gimeno and Nieto 2005; Lamouille, Xu et al. 2014). In order to repress the expression of E-cadherin, SNAI1 interacts with distinct proteins and complexes that control histone modifications (Lamouille, Xu et al. 2014). Among those are Polycomb repressive complex (PRC2), SIN3A and histone deacetylases HDAC1, 2 and 3, and lysine specific demethylase 1 (LSD1) (Lamouille, Xu et al. 2014). It was also demonstrated that SNAI1 cooperates with ETS1 and SMAD3-SMAD4 complexes to govern the expression of EMT associated genes. SNAI2, on its side, recruits different repressor complex together with HDAC1/3 and C-terminal binding protein (CTBP) to promote EMT in a similar manner (Tiwari, Gheldof et al. 2012).

A variety of signaling pathways regulate SNAIL expression. The list of those that can activate SNAI1 activity and expression includes receptor tyrosine kinases (RTK), tumor growth factor (TGF), WNT, Notch, integrins, phosphatidylinositol 3 kinase (PI3K)-AKT, mitogen activated protein kinase (MAPK) and nuclear factor kB (NFkB) signaling (De Craene, van Roy et al. 2005; Tiwari, Gheldof et al. 2012; Lamouille, Xu et al. 2014). Most of those pathways are known to regulate SNAI1 at transcription level (De Craene, van Roy et al. 2005). However, the activity and stability of the transcription factor are also controlled by post-translational modifications. Glycogen synthase kinase 3b (GSK3β) can phosphorylate SNAI1 at Ser97 or Ser101 and protein kinase D1 (PKD1) at Ser11, which in both cases promotes its translocation to the cytoplasm (Zhou, Deng et al. 2004; Yook, Li et al. 2006; Du, Zhang et al. 2010). Additionally, GSK3β phosphorylates SNAI1 at Ser108, Ser112, Ser116 and Ser120 in its turn can stimulate the SNAI1 ubiquitin-mediated degradation (Zhou, Deng et al. 2004; Yook, Li et al. 2006). WNT, PI3K-AKT, NFkB and Notch signaling interfere with GSK3β phosphorylation of SNAI1 and thus lead to increased stability and activity of the transcription factor (Yook, Li et al. 2006; Lamouille, Xu et al. 2014). On the other hand, phosphorylation of SNAI1 at Ser246 and Thr203, respectively by PAK1 or large tumor suppressor 2 (LATS2), can stimulate the nuclear retention of the protein and therefore its activity (Yang, Rayala et al. 2005; Zhang, Rodriguez-Aznar et al. 2012). Finally Lox and lysyl oxidase-like 2 and 3 (LOXL2/3) stabilize SNAI1 protein (Tiwari, Gheldof et al. 2012).

ZEB Transcription Factors

The ZEB family consists of two transcription factors ZEB1 and ZEB2, which have overlapping functions, but not always overlapping expression patterns (Peinado, Olmeda et al. 2007). Both ZEB proteins induce EMT and cell migration (Tiwari, Gheldof et al. 2012). They are expressed in human fetal and adult tissues, including central nervous systems, heart, skeletal muscle and hematopoietic cells (Funahashi, Sekido et al. 1993; Genetta, Ruezinsky et al. 1994; Sekido, Murai et al. 1994; Postigo and Dean 2000). In addition, knockout of mouse *Zeb1* resulted in severe T-cell developmental impairment and various skeletal abnormalities, while knockout of *Zeb2* is embryonically lethal with delamination arrest and lack of TFAP2A positive

migrating neural crest cells (Higashi, Moribe et al. 1997; Takagi, Moribe et al. 1998; Van de Putte, Maruhashi et al. 2003).

ZEB1 and ZEB2 share important structural similarities (Postigo and Dean 2000). They possess two clusters of zinc-finger domains at both the C-terminal and N-terminal ends, and a homeodomain in the central part (Peinado, Olmeda et al. 2007). They recognize bipartite E Box DNA elements (CACCT and CACCTG) via their zinc-finger clusters and present similar sequence specificities (Postigo and Dean 2000; Peinado, Olmeda et al. 2007).

ZEB1 and ZEB2 form repressive complexes with SMAD proteins, CTBP or via BRG1 protein with Switch/Sucrose nonfermentable (SWI/SNF) complex (Tiwari, Gheldof et al. 2012; Lamouille, Xu et al. 2014). ZEB1 can also function as an activator by interacting with p300/CBP via PCAF and p300, and is potentially involved in histone demethylation via binding of LSD1 (Lamouille, Xu et al. 2014).

Signaling cascades downstream of TGFβ, WNT and RAS-MAPK pathways induce the expression of ZEB proteins (Peinado, Olmeda et al. 2007; Tiwari, Gheldof et al. 2012; Lamouille, Xu et al. 2014). Furthermore SNAI1 and TWIST1 were demonstrated to directly activate ZEB1 promoter (Dave, Guaita-Esteruelas et al. 2011). Both of the ZEB factors participate in a regulatory loop with miR-200 family of microRNAs, which antagonizes TGFβ-induced EMT (Gregory, Bert et al. 2008; Korpal, Lee et al. 2008; Park, Gaur et al. 2008). PRC2 inhibits the activity of ZEB2 at post-translational level by sumoylation (Tiwari, Gheldof et al. 2012).

bHLH Transcription Factors

Helix-loop-helix (HLH) proteins have a common structural organization consists of two amphipatic α-helices and middle loop linker involved in dimerization (Peinado, Olmeda et al. 2007). In certain HLH proteins supplementary basic domain exists therefore named bHLH, while others such as the Id proteins have no additional structural elements (Massari and Murre 2000). They can form hetero- or homo-dimers and in the case of bHLH recognize an E Box DNA consensus sequence (CANNTG) (Massari and Murre 2000). With respect to EMT, a number of HLH proteins are implicated, including TCF3, TCF4, TCF12, TWIST1,2 and the Id proteins (Peinado, Olmeda et al. 2007). TCF3, with its two isoforms E12 and E47, and TWIST 1 and 2

directly regulate E-cadherin expression and are essential inducers of EMT (Tiwari, Gheldof et al. 2012). The Id proteins are devoided from DNA binding domain but can still dimerize with other HLH factors. In this way they inhibit their action and therefore EMT (Kondo, Cubillo et al. 2004). Upon TGFβ1 induced EMT, Id1 protein is downregulated and the decrease of its expression correlates with that of E-cadherin, while expression of Id2 and Id3 can inhibit the transition (Tiwari, Gheldof et al. 2012).

microRNA Regulation of EMT

Non-coding RNAs and in particular microRNAs are critical modulators of developmental processes (Stefani and Slack 2008). Their abnormal expression is related to pathological conditions such as cancer (Nicoloso, Spizzo et al. 2009). In the recent years a large number of microRNAs that are involved in EMT, cancer progression and metastasis formation were identified (Nicoloso, Spizzo et al. 2009). For instance, miR-200 family members form a double-negative feedback loop with ZEB transcription factors (Figure 7). Five different microRNAs (miR-200a, miR-200b, miR-429, miR-200c, miR-141), encoded in two separate clusters are part of miR-200 family (Altuvia, Landgraf et al. 2005; Bracken, Gregory et al. 2008). The microRNA seed is critical for target recognition, and the members of the microRNA family comprise highly similar and conserved seed sequences. Therefore, this suggests that they also share important number of target genes (Lewis, Shih et al. 2003; Lewis, Burge et al. 2005). In the case of miR-200 family two seed sequences with a single nucleotide difference do exist: AAUACU, contained in miR-200bc/429, and AACACU found in miR-200a/141 (Feng, Wang et al. 2014). In the 3'UTR of *ZEB1* and *ZEB2* miR-200 family members have between 5 and 8 binding sites that allow a tight control over the transcription factors (Gregory, Bert et al. 2008; Park, Gaur et al. 2008). On the other hand ZEB represses the expression of the microRNA clusters by directly interacting with their promoters (Bracken, Gregory et al. 2008; Burk, Schubert et al. 2008). In this manner elevated miR-200 levels safeguard the epithelial state, whereas upon the transition ZEB increases and blocks the expression of miR-200 (Gregory, Bert et al. 2008). Ectopic expression of miR-200 in the mesenchymal cell state is sufficient to revert the process (Gregory, Bert et al. 2008).

Furthermore low levels of miR-200 family members are associated with aggressive tumor progression, metastasis formation, cancer stem cell stemness potential and chemoresistance (Feng, Wang et al. 2014). Additional targets of miR-200 family, such as *SUZ12* and *BMI*, also contribute to the microRNAs anti-metastatic potential and their role in promoting the epithelial cellular state (Wellner, Schubert et al. 2009; Iliopoulos, Lindahl-Allen et al. 2010). Other microRNAs-TF interactions have also been described in the context of EMT. For example, several microRNAs, including miR-29b, miR-30a and miR-34 family members directly target *SNAI1*, while miR-1 and miR-200b repress *SNAI2* (De Craene and Berx 2013; Lamouille, Xu et al. 2014). SNAIL transcription factors might control the expression of miR-200, miR-1 and miR-34 thus establishing a double-negative feedback regulation similar to the one described between ZEB TFs and miR-200 family members (Figure 7) (De Craene and Berx 2013). Downregulation of miR-335 is a hallmark of EMT and tumor progression, and it is mechanistically explained by the microRNA control over *SOX4* and tenascin-C mRNAs (Tavazoie, Alarcon et al. 2008; Tiwari, Gheldof et al. 2012).

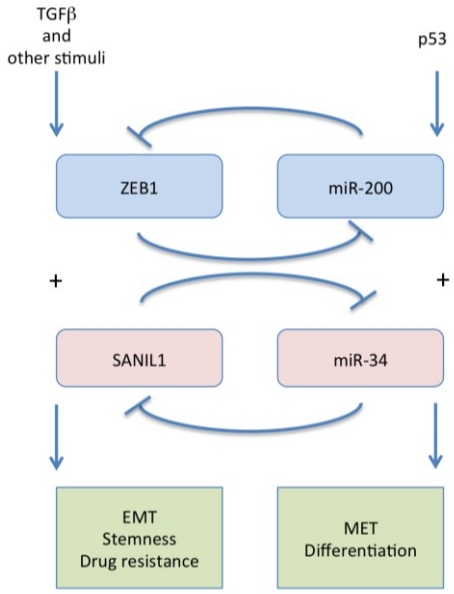

Figure 7. MicroRNAs-transcription factors feedback loops in EMT. TGFβ and other growth factors or extracellular stimuli induce EMT promoting transcription factors from the ZEB and SNAIL families, which on their turn activate EMT phenotypical changes, stemness and drug resistance. miR-200 and miR-34 family members form a double negative feedback loop with the EMT promoting TFs. ZEB and SNAIL block the transcription of the microRNAs, while miR-200 and miR-34

negatively regulate the expression of the TFs. In addition, p53 can induce the expression of both miRNA families, thus promoting the epithelial state. Adapted by permission of Macmillian Publishers Ltd Nature Reviews Cancer (Brabletz 2012), copyright 2012.

Apart from regulating the expression of master TFs microRNAs can also modulate EMT by controlling genes characteristic either of the epithelial or mesenchymal cell state (Lamouille, Xu et al. 2014). In mammary carcinomas miR-9 levels can be elevated and confer mesenchymal and invasive features to the tumors by targeting E-cadherin gene (Ma, Young et al. 2010). Another interesting example of a microRNA that promotes EMT and metastasis is miR-22. In mouse xenograft models it triggers EMT and stimulates metastasis formation, whereas in patients it correlates with poor clinical outcome. It is suggested that it indirectly contributes to the silencing of tumor suppressor miR-200 by targeting an enzyme, Ten eleven translocation (TET), involved in the miR-200 promoter demethylation (Song, Ito et al. 2013).

Splicing Factors in EMT

Alternative splicing is a regulatory mechanism of gene expression that leads to the formation of different proteins from a single gene and it is involved in cancer formation and EMT (Tiwari, Gheldof et al. 2012). The RNA binding proteins, ESRP1 and ESRP2 promote the epithelial state by inducing the splicing of epithelial-specific isoforms of certain EMT associated genes such as CD44, FGFR2 and CTNND1 (Warzecha, Sato et al. 2009; Warzecha, Jiang et al. 2010; Brown, Reinke et al. 2011). On the contrary RBFOX2 can regulate both epithelial and mesenchymal splicing and stimulate tissues invasiveness (Braeutigam, Rago et al. 2014). CELF, MBNL and hnRNP splicing factors were also associated with EMT (Shapiro, Cheng et al. 2011).

TFAP2A Transcription Factor

Transcription factor TFAP2A (also named AP-2α) belongs to the AP-2 family of transcription factors(Hilger-Eversheim, Moser et al. 2000; Eckert, Buhl et al. 2005). In 1987, Mitchell *et al.* were the first to discover and further designate TFAP2A as binding partner of the SV40 enhancer elements in HeLa cells (Mitchell, Wang et al.

1987). In human and mice there are five AP-2 transcription factors that share significant sequence similarity between 56 and 78 %, which are TFAP2A, TFAP2B, TFAP2C, TFAP2D and TFAP2E, or AP-2α, AP-2β, AP-2γ, AP-2δ and AP-2ε, respectively (Eckert, Buhl et al. 2005). They present a specific domain structural organization featuring an N-terminal prolin and glutamine rich trans-activation domains, a central basic region and very well conserved helix-loop-helix C-terminal region, involved in DNA binding and dimerization.(Williams and Tjian 1991) The protein can function as a hetero-or homo- dimer(Eckert, Buhl et al. 2005). In mice, *Tfap2a* gene is composed from 7 exons and it is expressed as four different isoforms; the first three (1a,1b and 1c) differ in the first exon and the forth isoform lacks the 2nd exon due to alternative splicing event(Meier, Koedood et al. 1995; Eckert, Buhl et al. 2005).

SELEX (Systemic Evolution of Ligand by EXponential Evolution) based, *in vitro* binding site enrichment assays, have determined that TFAP2A binds to a palindromic motif $GCCN_3GGC$ and its variations $GCCN_4GGC$, $GCCN_{3/4}GGG$ (Mohibullah, Donner et al. 1999). In more recent chip-seq experiments it was shown that human AP-2γ and AP-2α have consensus binding site of respectively SCCTSRGGS and SCCYSRGGS (S = G or C, R = A or G and Y = C or T)(Woodfield, Chen et al. 2010; Rada-Iglesias, Bajpai et al. 2012) (Bogachek, Chen et al. 2014) .

Mode of Action and Control

AP-2 family of proteins canonically functions as transcriptional regulators in the nucleus, where it regulates the expression of certain targets (Eckert, Buhl et al. 2005). They were shown to activate genes involved in apoptosis, cell growth, proliferation and differentiation (*CDKN1A*, *TGFA*, *ESR1*, *ERBB2*/HER-2/neu, *FOXA1*)(Bosher, Williams et al. 1995; Wang, Shin et al. 1997; Zeng, Somasundaram et al. 1997; Woodfield, Chen et al. 2010). In addition, it was shown that they can also repress the expression of many genes, including *MYC*, cyclin-D2, c/EBP- α /*CEBPA*, *MCAM*/MUC18.(Gaubatz, Imhof et al. 1995; Jean, Gershenwald et al. 1998; Jiang, Tang et al. 1998; Yu, Hitchler et al. 2009) In the promoter region of cyclin-D2, the TFAP2A binding site is found in close vicinity to MYC responsive E-Box element, therefore creating a mutually exclusive interaction (Hilger-Eversheim, Moser et al.

2000). Apart from its function as transcription factor, it has been demonstrated that TFAP2A participates in WNT signaling pathway by forming complex with adenomatous polyposis coli protein (APC) and β-catenin, disrupting the interaction between β-catenin and the transcription factors TCF and thereby blocking its effector function (Aqeilan, Palamarchuk et al. 2004). The activity of TFAP2A can be regulated by protein-protein interactions or by post-translational modifications (PTMs) (Eckert, Buhl et al. 2005). For instance TFAP2A binding to the PC4 transcription factor reduces its transcriptional self-interference and stimulates its transactivation potential (Zhong, Wang et al. 2003). TFAP2A can be sumoylated, inducing its degradation and suppressing its activity (Eloranta and Hurst 2002; Berlato, Chan et al. 2011). Equivalently, a phosphorylation by PKA at Ser239 of TFAP2A regulates its transactivation potential and oxidation of conserved cystein residues in TFAP2A DNA binding domain shifts its capacity to interact with DNA (Huang and Domann 1998; Garcia, Campillos et al. 1999; Grether-Beck, Felsner et al. 2003). It has been also proposed that TFAP2C sub-cellular localization can be regulated by an interaction of with WWOX that sequesters the transcription factor in the cytoplasm (Aqeilan, Palamarchuk et al. 2004). In addition, the different isoforms of TFAP2A were shown to differ in activity suggesting that TFAP2A activity can be further regulated at gene level by alternative splicing and/or alternative promoter usage (Buettner, Kannan et al. 1993; Berlato, Chan et al. 2011).

AP-2 Transcription Factors in Development and Cancer

In the developing embryo, AP-2 transcription factors are implicated in trophectoderm development, neural crest formation, as well as in the differentiation of numerous tissues and cell types (Eckert, Buhl et al. 2005). TFAP2A expression in the mouse embryo becomes apparent at day 8 in the lateral head mesenchyme and shortly after that in the neural tube and the primitive mesenchyme. The expression of α-, β- and γ-variants follows the same pattern between day 8-10. After day 11 on and in later stages they start to be expressed in different tissues(Moser, Ruschoff et al. 1997). The knockout of *Tfap2a* in mice results in a range of cranio-facial deformations originating from neural tube closure defect, as well as sensory organs and cranial ganglions abnormalities (Schorle, Meier et al. 1996; Zhang, Hagopian-Donaldson et

al. 1996). In humans, mutations in *TFAP2A* are linked to another developmental defect, namely the Branchio-Oculo-Facial Syndrome (BOFS) (Milunsky, Maher et al. 2008). Within the normal adult mammary tissue, TFAP2A is expressed in virgin and pregnant mice mammary gland. It is detected at the terminal end buds and also in the ductal epithelium, predominantly in the luminal cell population (Zhang, Brewer et al. 2003). Targeted overexpression of TFAP2A and TFAP2C in the mouse mammary gland results in lactation deficiency, increased proliferation and apoptosis, reduced alveolar budding and differentiation (Jager, Werling et al. 2003; Zhang, Brewer et al. 2003). Knockout of TFAP2C, a TFAP2A paralogue, in mouse mammary luminal cells, results in an increased number of terminal end buds with reduced distal migration (Cyr, Kulak et al. 2014).

Aberrant expression of TFAP2A has been observed in various cancers. In human nasopharyngeal carcinoma it is overexpressed and it is involved in tumorigenesis by targeting the HIF-1α/VEGF/PEDF pathway (Shi, Xie et al. 2014). On the contrary, reduced TFAP2A expression was reported to be associated with poor prognosis in gastric adenocarcinoma (Wang, Lv et al. 2011). The loss of TFAP2A is connected with the acquisition of the malignant phenotype in melanoma through regulation of cell adhesion molecules (ALCAM) (Melnikova and Bar-Eli 2008). In breast cancer, TFAP2A expression was found to be less organized than in normal mammary gland and it is associated with HER2/ ErbB-2 and ERα expression (Pellikainen, Naukkarinen et al. 2004). More recently TFAP2A and TFAP2C activation by loss of sumoylation in breast cancer was associated with the luminal breast cancer phenotype and it was suggested to interfere with EMT (Bogachek, Chen et al. 2014; Cyr, Kulak et al. 2014).

Results

Manuscript published under the following title:

"TFAP2A is a component of the Zeb1/2 network that regulates TGFβ1-induced epithelial to mesenchymal transition "

Yoana Dimitrova[1], Andreas Gruber[1], Nitish Mittal[1], Souvik Ghosh[1] , Beatrice Dimitriades[1] , Daniel Mathow[3], William Aaron Grandy[1], Gerhard Christofori[2], Mihaela Zavolan[1,4]
[1]Biozentrum, University of Basel, Klingelbergstrasse 50-70, CH-4056 Basel, Switzerland.
[2]Department of Biomedicine, University of Basel, Matenstrasse 28, CH-4058Basel, Switzerland
[3]Department of Cellular and Molecular Pathology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
[4]Correspondence should be addressed to Mihaela Zavolan Biozentrum, University of Basel, Klingelberstrasse 50-70, CH-4056 Basel, Switzerland
mihaela.zavolan@unibas.ch.

Biology Direct (2017) 12:8 DOI 10.1186/s13062-017-0180-7

Dimitrova et al. Biology Direct (2017) 12:8
DOI 10.1186/s13062-017-0180-7

Biology Direct

RESEARCH Open Access

TFAP2A is a component of the ZEB1/2 network that regulates TGFB1-induced epithelial to mesenchymal transition

Yoana Dimitrova[1], Andreas J. Gruber[1], Nitish Mittal[1], Souvik Ghosh[1], Beatrice Dimitriades[1], Daniel Mathow[3], William Aaron Grandy[1], Gerhard Christofori[2] and Mihaela Zavolan[1*]

Abstract

Background: The transition between epithelial and mesenchymal phenotypes (EMT) occurs in a variety of contexts. It is critical for mammalian development and it is also involved in tumor initiation and progression. Master transcription factor (TF) regulators of this process are conserved between mouse and human.

Methods: From a computational analysis of a variety of high-throughput sequencing data sets we initially inferred that TFAP2A is connected to the core EMT network in both species. We then analysed publicly available human breast cancer data for TFAP2A expression and also studied the expression (by mRNA sequencing), activity (by monitoring the expression of its predicted targets), and binding (by electrophoretic mobility shift assay and chromatin immunoprecipitation) of this factor in a mouse mammary gland EMT model system (NMuMG) cell line.

Results: We found that upon induction of EMT, the activity of TFAP2A, reflected in the expression level of its predicted targets, is up-regulated in a variety of systems, both murine and human, while TFAP2A's expression is increased in more "stem-like" cancers. We provide strong evidence for the direct interaction between the TFAP2A TF and the ZEB2 promoter and we demonstrate that this interaction affects ZEB2 expression. Overexpression of TFAP2A from an exogenous construct perturbs EMT, however, in a manner similar to the downregulation of endogenous TFAP2A that takes place during EMT.

Conclusions: Our study reveals that TFAP2A is a conserved component of the core network that regulates EMT, acting as a repressor of many genes, including ZEB2.

Reviewers: This article has been reviewed by Dr. Martijn Huynen and Dr. Nicola Aceto.

Keywords: Epithelial-to-mesenchymal transition, EMT, Transcription regulatory network, TFAP2A, ZEB2, TGFb1, NMuMG

Background

The epithelial to mesenchymal transition (EMT) is defined as the process in which cells that display predominantly epithelial features transition to a state in which they exhibit mesenchymal characteristics. EMT has well-established and important roles in different stages of embryonic development: it is observed during gastrulation, in the generation of the primitive mesoderm, during neural crest (NC) formation, and in the development of many organs such as heart valves, skeletal muscle, and the palate [1]. EMT-like phenomena were also described in adult organisms, as part of normal developmental changes, as well as during pathological processes [2]. For example, during breast development, an EMT-like program referred to as epithelial plasticity is thought to be part of branching morphogenesis, which leads to the formation of the complex ductal tree [3]. Recent findings suggest that an EMT program may increase the "stemness" potential of epithelial cells [4].

The mammary gland epithelium is composed of an internal luminal layer, and an external, basal layer of myoepithelial cells. Recent studies suggest that these

* Correspondence: mihaela.zavolan@unibas.ch
[1]Biozentrum, University of Basel, Klingelbergstrasse 50-70, CH-4056 Basel, Switzerland
Full list of author information is available at the end of the article

different cell types derive from a common stem cell, through a process that involves epithelial plasticity [5, 6]. Whereas this process is very well coordinated in normal development, its dysregulation in cancer leads to outcomes that are difficult to predict [3]. While the majority of experimental results indicate that manipulating EMT also affects cancer metastasis, recent reports on cancer cells circulating in the blood stream or resulting from genetic lineage tracing have questioned a critical role of EMT in the formation of metastases, but have demonstrated a role in chemotherapy resistance [7–9]. In breast cancer, it is believed that EMT affects the basal epithelial phenotype and is responsible for an increased metastatic potential [10].

The TFAP2A transcription factor (TF) is expressed early in embryogenesis, where it contributes to cell fate determination in the formation of the neural crest and the epidermis. The knockout of Tfap2a in mouse is lethal due to neural crest formation defects [11]. In humans, mutations in *TFAP2A* have been linked to the developmental defects in the Branchio-Oculo-Facial Syndrome (BOFS) [12].

TFAP2A is a member of the AP-2 family of TFs, which in humans and mice is composed of five members, TFAP2A, TFAP2B, TFAP2C, TFAP2D and TFAP2E, or AP-2α, AP-2β, AP-2γ, AP-2δ and AP-2ε, respectively. These proteins share important sequence similarities and have a specific structural organization with a proline and glutamine-rich trans-activation domain located at the N-terminus, a central region with positively-charged amino acids, and a highly conserved helix-loop-helix region at the C-terminus. The last two domains are involved in DNA binding and dimerization, the proteins being able to form hetero- or homo-dimers [13]. The *TFAP2A* gene is composed of seven exons. In mice, four different isoforms have been described [14]. Systemic Evolution of Ligand by EXponential enrichment (SELEX)-based, in vitro assays, have determined that AP-2α binds to the palindromic motif $GCCN_3GGC$ and to some close variants, $GCCN_4GGC$, $GCCN_{3/4}GGG$ [15]. More recent ChIP-seq experiments inferred SCCTSRGGS and SCCYSRGGS (S = G or C, R = A or G and Y = C or T) as the consensus sites for human AP-2γ and AP-2α, respectively [16].

In the adult mammary gland, TFAP2A is expressed in virgin and pregnant mice. Its mRNA and protein are detected at the terminal end buds and also in the ductal epithelium, predominantly in the luminal cell population [17]. Targeted overexpression of TFAP2A and TFAP2C in the mouse mammary gland results in lactation deficiency, increased proliferation and apoptosis, reduced alveolar budding and differentiation [17, 18]. Knockout of the TFAP2C paralog of TFAP2A in mouse mammary luminal cells results in an increased number of terminal end buds with reduced distal migration [19].

Aberrant expression of TFAP2A has been observed in various cancers. It is overexpressed in human nasopharyngeal carcinoma and is involved in tumorigenesis by targeting the HIF-1α/VEGF/PEDF pathway [20]. In contrast, reduced AP-2α expression was reported to be associated with poor prognosis in gastric adenocarcinoma [21]. The loss of TFAP2A is connected with the acquisition of the malignant phenotype in melanoma through regulation of cell adhesion molecules (ALCAM) [22]. TFAP2A expression was found to be less organized in breast cancer compared to normal mammary gland and it is associated with HER2/ErbB-2 and ERα expression [23].

To define conserved EMT regulatory networks, we started by analyzing seven mouse and human datasets obtained from EMT systems, altogether containing thirty-six mRNA sequencing samples. We found that TFAP2A is one of the factors that contribute most significantly to mRNA-level expression changes that take place during embryonic stem cell (ESC) differentiation to mesoderm or to NC cells, during normal mammary gland development, and most importantly, in breast cancer models. To investigate TFAP2A's involvement in EMT we used mouse mammary gland epithelial cell line NMuMG, a well-known model of EMT [24]. We demonstrate, for the first time, that the expression and activity of *Tfap2a* are modulated during TGFβ1-induced transdifferentiation of these cells. We further show that TFAP2A directly binds to the *Zeb2* promoter, modulating its transcriptional output. TFAP2A overexpression in NMuMG cells results in increased levels of EMT-inducing TFs, and promotes an EMT-like phenotype. Our study sheds a new light on the role of TFAP2A in processes that involve EMT, including breast cancer, and it contributes to a deeper understanding of the molecular and cellular mechanism of cancer development and metastasis.

Methods

Expression vectors and constructs

Mouse TFAP2A cDNA was kindly provided by Prof. Qingjie [25]. The TFAP2A-FLAG fusion was subcloned into pDONR201 plasmid, using a Gateway® BP Clonase® II Enzyme mix (#11789-020, Life Technologies) and it was further subcloned into pCLX vector, using Gateway® LR Clonase® II Enzyme mix (#11791-020, Life Technologies).

Cell culture

We used a subclone of NMuMG cells that was generated as previously described (NMuMG/E9) [24]. Cells

were cultured in Dulbecco's modified Eagle's medium (DMEM #D5671, Sigma Aldrich) with high glucose and L-glutamine, supplemented with 10% fetal bovine serum (#f-7524, Sigma-Aldrich) and where indicated were treated with 2 ng/ml TGFβ1 (#240-B, R&D Systems). Transient transfection was done using Lipofectamine2000 (#11668-019, Life Technologies) according to the manufacturer's instructions. For time course experiments, cells were grown in six well plates for up to 14 days and treated with 2 ng/mL TGFβ1. In addition, NMuMG pCLX-TFAP2A or NMuMG pCLX-GFP cells induced with 2 μg/mL of doxycycline for 6 days, and further treated or not treated with TGFβ1 for 72 hours were used to study the effect of TFAP2A overexpression.

Lentiviral infection

Stable populations of NMuMG cells expressing the blasticidine-resistant marker together with TFAP2A-FLAG under a doxycycline-inducible promoter were obtained with the pCLX expression system [26]. Lentiviral particles were produced in HEK293-LV cells using the helper vectors pMDL, pREV and the envelope-encoding vector pVSV. For infection, viral supernatants were added to target cells in the presence of polybrene (#TR-1003-G, Millipore) (1 μg/ml). Cells were further incubated at 37 °C under 5% CO_2 in a tissue culture incubator for 72 h, prior to selection with blasticidine at 10 μg/ml (#15205-25 mg, Sigma-Aldrich).

Light microscopy and immunofluorescence

Cells were treated with doxycycline or TGFβ1 for the indicated times, and were grown on gelatin coated glass coverslips. Cells were fixed with 4% paraformaldehyde in 1x PBS for 15 min (Fig. 2a, b). They were later permeabilized and blocked for 30 min with 0.1% Triton X-100 (#T8787, Sigma-Aldrich), 10% goat serum (#16210072, Gibco®, Life Technologies), and 1% BSA (#A9647, Sigma-Aldrich) in PBS (#20012-019, Gibco®, Life Technologies). Afterwards, the coverslips were incubated with the indicated primary antibodies overnight at 4 °C, and then with Alexa Fluor 488,647 conjugated secondary antibodies, (Molecular Probes, Life Technologies), for one hour at room temperature. Where appropriate, Acti-stain™ 555 (#PHDH1, Cytoskeleton) diluted 1:200 was added together with secondary antibody stain. The coverslips were mounted with VECTASHIELD™ DAPI Mounting Media (Vector Laboratories) on microscope slides and imaged with a confocal microscope (Zeiss LSM 700 Inverted).

Quantitative real-time reverse transcription PCR

Total RNA was extracted with TRI Reagent® (#T9424, Sigma-Aldrich) and further purified with Direct-zol™ RNA MiniPrep kit (#R2050, Zymo Research). Reverse transcription was performed with SuperScript® III Reverse Transcriptase (#18080-044, Life Technologies) according to the manufacturer's instructions. For qPCR, 8 ng of cDNA was used in a reaction with Power SYBR® Green PCR Master Mix (#4367659, Applied Biosystems). Gene expression changes are normalized to the expression of the house-keeping genes *Gapdh* and *Rplp0*.

mRNA sequencing

For the mRNA-seq library preparation, a well of a 6-well plate of NMuMG cells was used, either treated with growth factor and/or doxycycline, or with control reagents for the indicated times. mRNA-seq libraries were prepared as already described [27].

Chromatin immunoprecipitation (ChIP), sequencing library preparation and data analysis

The ChIP protocol was adapted from [28]. Cells were crosslinked in fixing buffer (50 mM HEPES pH 7.5, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 100 mM NaCl, 1% formaldehyde) for 10 min with continuous rocking at room temperature (RT), and then quenched with 125 mM glycine for 5 min. Cells were washed three times with cold PBS and collected by scrapping. Nuclei were isolated, and lysed to obtain crosslinked chromatin. Simultaneously, the antibody was coupled with protein G magnetic beads (#88848, Pierce™) by incubating 100 μl of protein G beads with 10 μg of TFAP2A-specific antibody (Novus) and 10 μg of rabbit IgG (#PP64, Millipore) as a negative control, for minimum 1 h at RT with continuous rotation. A probe sonicator was then used in cold conditions to reduce heating, for six cycles of 30 s pulse-on at amplitude value of 60 and 1 min and 15 s pulse-off to obtain chromatin fragments of 100–500 bp followed by centrifugation at 20,000 g for 10 min at 4 °C to get rid of nuclear debris. Further, 3% chromatin was kept as input control from each sample and an equal amount (around 750–1000 μg) of chromatin was incubated with magnetic beads-coupled antibody at 4 °C overnight with continuous rotation. Immuno-complexes were washed with 1 mL of wash buffers as described in the original protocol. Samples of washed immuno-complexes along with the input were further treated with RNase and then with proteinase K followed by overnight reverse crosslinking at 65 °C with continuous shaking at 1400 rpm in a thermoblock with heating lid. DNA was purified using Agencourt AMPure XP (#A63880, Beckman Coulter) beads as detailed in the reference. The enrichment of specific target genes was quantified by qRT-PCR, comparing the TFAP2A-ChIP with the IgG negative control.

Libraries of ChIPed and input DNA were prepared according to the instruction manual of NEBNext® ChIP-Seq Library Prep Reagent Set for Illumina. In brief, end

repair of input and ChIPed DNA was done by incubating with T4 DNA Polymerase Klenow fragment and T4 PNK enzyme at 20 °C for 30 min. The reaction was purified using Ampure beads according to the instruction manual. An A nucleotide overhang at the 3' end was produced by treating the end repaired DNA with dATP and Klenow Fragment ($3' \rightarrow 5'$ exo$^-$) at 37 °C for 20 min followed by DNA purification. Double stranded DNA adapters were ligated to dA overhang DNA by T4 DNA ligase reaction at 37 °C for 30 min followed by DNA purification and size selection as described in the instruction manual. Size selected DNA was PCR-amplified for 16 cycles using NEBNext® High-Fidelity 2X PCR Master Mix with Illumina universal forward primer and indexed reverse primer, that enabled multiplexing of samples for sequencing. Amplified DNA was finally purified and sequenced on an Illumina Hiseq2500 instrument. The obtained sequencing reads were mapped to the genome and visualized within the clipz genome browser (www.clipz.unibas.ch).

Antibodies and reagents

We used primary antibodies against the following proteins: TFAP2A (#sc-12726, Santa Cruz Biotechnology) for Western Blot (WB) and TFAP2A (#NBP1-95386, Novus Biologicals, Bio-Techne) for immunofluorescence and immunoprecipitation, actin (#sc-1615, Santa Cruz Biotechnology), E-cadherin (#610181, BD Transduction Laboratories), N-cadherin (#610921, BD Transduction Laboratories), Fibronectin (#F3648, Sigma-Aldrich), GAPDH (#sc-32233, Santa Cruz Biotechnology), vimentin (#v2258, Sigma-Aldrich). Recombinant human TGFβ1 was obtained from R&D Systems.

Electrophoretic Mobility Shift Assay (EMSA)

TnT® T7 Quick Coupled Transcription/Translation System (#L1171, Promega) was used to express in vitro translated TFAP2A from the pcDNA3-TFAP2A construct. Double-stranded oligonucleotide probes were end-labeled with ^{32}P and purified on autoseq G-50 columns (#27-5340-01, Amersham). Binding reactions containing probe, TFAP2A protein, poly (dI-dC) (#81349, Sigma-Aldrich) non-specific competitor in gel retention buffer (25 mM HEPES pH 7.9, 1 mM EDTA, 5 mM DTT, 150 mM NaCl, 10% Glycerol) and electrophoresis were carried out as described previously [29].

Combined motif activity response analysis

The datasets used in the following analysis are listed in Additional file 1: Table S1. We applied the ISMARA tool to each dataset as previously described [30]. Briefly, the Motif Activity Response Analysis (MARA) infers the activity of regulatory motifs from the number of binding sites of each motif m in each promoter p ($N_{m,p}$) and the genome-wide expression driven by these promoters p in samples s ($E_{p,s}$):

$$E_{p,s} = \tilde{c}_s + c_p + \sum_m N_{m,p} A_{m,s}$$

where $\tilde{c}_s$ represents the mean expression in sample s, c_p is the basal expression of promoter p, and $A_{m,s}$ is the (unknown) activity of motif m in sample s. To identify motifs that consistently change in activity across datasets we used a computational strategy as previously described [31]. In brief, first we obtained the average activities over the replicates of each condition in every dataset. Next, because the range of gene expression levels and consequently the motif activities varied across datasets, we re-centered and then standardized the averaged motif activities $\overline{A}^*_{m,g}$ and corresponding errors $\overline{\sigma}^*_{m,g}$, belonging to a specific condition g. To standardize the activities in a given dataset with the epithelial-like condition labeled as a and the mesenchymal-like condition by b we defined a scaling factor $\mathrm{S} = \sqrt{\frac{\left(\overline{A}^{*b}_{m,g}\right)^2 + \left(\overline{A}^{*a}_{m,g}\right)^2}{2}}$, and then rescaled the activities $\tilde{A}^*_{m,g} = \frac{\overline{A}^*_{m,g}}{\mathrm{S}}$ and the corresponding errors $\tilde{\sigma}^*_{m,g} = \frac{\overline{\sigma}^*_{m,g}}{\mathrm{S}}$. Subsequently, we separated the condition-specific, averaged and rescaled activities ($\tilde{A}^*_{m,g}$) and errors ($\tilde{\sigma}^*_{m,g}$) obtained from different datasets into two groups, depending on whether they originated from epithelial-like cells (a) or mesenchymal-like cells (b). We averaged activities belonging to the same group as done for sample replicates before (see above and [31]). Finally, to rank motif activity changes during EMT we calculated for every motif m a z-score by dividing the change in averaged activities by the averaged errors:

$$\mathrm{z} = \frac{\overline{\mathrm{A}}^{*b}_{m,g} - \overline{\mathrm{A}}^{*a}_{m,g}}{\sqrt{\left(\bar{\sigma}^{*b}_{m,g}\right)^2 + \left(\bar{\sigma}^{*a}_{m,g}\right)^2}}$$

Constructing motif-motif interaction networks

ISMARA predicts potential targets for each motif m by calculating a target score R as the logarithm of the ratio of two likelihoods: the likelihood of the data D assuming that a promoter p is a target of the motif, and the likelihood of the data assuming that it is not:

$$R = log\left(\frac{P(D|target\ promoter)}{P(D|not\ target\ promoter)}\right)$$

The posterior probability p that a promoter is a target given the data and assuming a uniform prior of 0.5 is given by $p = \frac{1}{1+\frac{1}{e^R}}$. To construct motif-motif interactions, we focused on those transcription regulators, whose

regulatory regions were consistently (within all datasets) predicted by ISMARA to be targeted by motifs of other regulators. We obtained a combined probability p_{comb} that a regulator is a target of a particular motif m across I different datasets by calculating the probability product of the probabilities obtained from individual datasets:

$$p_{comb} = \prod_{i=1}^{I} p_i$$

GOBO analysis

The top 100 target genes of the TFAP2 {A,C}.p2 motif as derived by applying ISMARA to the Neve et al. data set [32] were analyzed with the Gene Expression-Based Outcome for Breast Cancer Online (GOBO) tool [33]. For each gene only the promoter with the highest ISMARA target score was considered for the analysis.

Estimating gene expression $\log_2$ fold changes from mRNA sequencing data

For each sample s the expression values driven by each promoter of a gene g (determined by ISMARA, see above) were summed up to estimate the expression of gene g in sample s. Log_2 gene expression fold changes were then calculated for TGFβ1-treated pCLX-GFP (pCLX-GFP + TGF-beta), pCLX-TFAP2A (pCLX-TFAP2A), and for TGFβ1-treated pCLX-TFAP2A (pCLX-TFAP2A + TGF-beta) cell lines relative to the pCLX-GFP (pCLX-GFP) control cells.

Results

TFAP2A/C motif activity increases upon EMT in both mouse and human systems

Aiming to identify major regulators of EMT and to further construct a conserved network of their interactions, we used the Motif Activity Response Analysis (MARA) approach, which combines high-throughput measurements of mRNA expression with computational prediction of regulatory elements [30]. The published ISMARA tool [30] allows not only the automated analysis of individual data sets, but also the inference of motifs that most generally explain gene expression changes across multiple experiments.

The results from the combined MARA analysis of different EMT mRNA expression datasets from breast epithelial cell lines of mouse and human, and from the differentiation of human pluripotent stem cells into NC cell and mesoderm (Additional file 1: Table S1) are shown in Fig. 1 [34–40]. How much a given motif contributes to the observed gene expression changes is quantified in terms of a combined z-score, which in our case represents the significance of the motif activity change between the epithelial and mesenchymal cell types (denoted by the intensity of the color in Fig. 1a and b and listed in Additional file 1: Tables S2 and S3). Based on the genome-wide computational prediction of binding sites for transcription regulators we can further infer motif interaction networks. In Fig. 1, an arrow is drawn between two motifs A and B when any of the regulators that recognizes motif B is a predicted target of motif A. The motif interaction networks derived from mouse and human EMT models suggest that only a small fraction of the TFs has a highly conserved and significant role in both species. The core transcriptional network of EMT, containing the TFs *Zeb1*, *Zeb2* and *Snai1*, is conserved, as expected. The motifs that correspond to these factors have negative activity changes during EMT (represented by the blue color on the scheme) which indicates that the expression of their targets decreases, as expected from their known repressive function during the process [41]. The TFAP2A/C motif is also a conserved component of both mouse and human EMT networks. Its target genes are upregulated during EMT (reflected by the red color in the figure) and thus the motif itself is predicted to have a highly significant positive change in activity. Furthermore, in both human and mouse systems, the TFAP2A/C motif is predicted to target both *Zeb1* and *Zeb2* TFs (Fig. 1a and b).

TFAP2A expression and activity changes in EMT and breast cancer

We made use of the murine mammary gland cell line NMuMG to further investigate the role of the AP-2 family members TFAP2A and TFAP2C in EMT. Upon induction with TGFβ1, NMuMG cells undergo EMT, which manifests itself through E-cadherin downregulation, formation of actin stress fibers and an elongated, mesenchymal-like cell shape (Fig. 2a, b and [36]). mRNA-seq revealed that of the five members of the AP-2 family, only *Tfap2a* is expressed in this system, with reads covering all its exons (Additional file 1: Figure S1). Immunofluorescence staining of endogenous TFAP2A demonstrated that the protein has a predominantly nuclear localization (Fig. 2a, b). 48 h after the TGFβ1 stimulation we observed that *Tfap2a* mRNA levels decreased moderately and further declined during the 14 days time course, while the common EMT markers such as E-cadherin, Fibronectin and Vimentin followed the expected trend (Fig. 2c).

We next generated mRNA-seq data from a 14 days time course of NMuMG cells stimulated with 2 ng/mL TGFβ1. Applying ISMARA to these data revealed the dynamics of TFAP2A activity during the entire length of the time course (Fig. 2d). As the paralogous TFAP2A and TFAP2C bind similar sequences, we therefore refer to their shared binding motif as TAFP2 {A,C}. In contrast to its mRNA expression (Fig. 2c), the TFAP2A

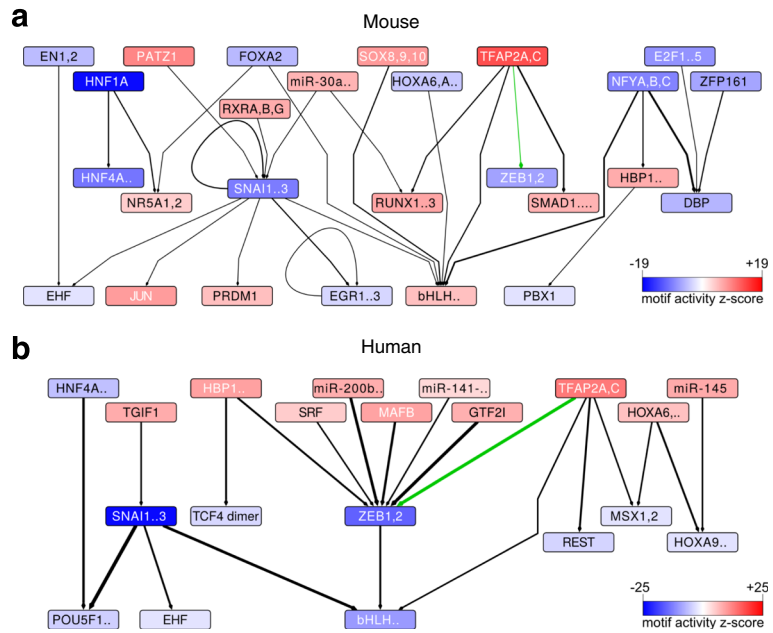

Fig. 1 The transcriptional networks inferred from different EMT systems. Motif–motif interaction networks derived from mouse (**a**) and human (**b**) datasets. An *arrow* was drawn from a motif A to a motif B if motif A was consistently (across datasets from the corresponding species) predicted to regulate a transcriptional regulator *b* that is known to bind motif B. The probability product that A targets *b* is reflected by the thickness of the line. For readability, only motifs with an absolute z-score > 2.0 and having at least one interaction with another such motif (with a target probability product > 0.35 for human and > 0.15 for mouse) are depicted. The color intensity of the nodes representing motifs is proportional to the significance of the motif given by its z-score. *Red* indicates increased and *green* indicates decreased activity upon EMT

transcriptional activity, reflected in the behavior of its targets, increases during EMT (Figs. 1 and 2d). This indicates that TFAP2A probably acts as a repressor in this context. Despite the fact that *Tfap2a* transcript levels and the TAFP2{A,C} motif activity exhibit a clear negative correlation, we observed the highest increase in activity in the first 6 h of treatment, while the changes in *Tfap2a* mRNA were delayed until a later time point. This may indicate that *Tfap2a* is regulated at the protein level. Considering that a rapid reduction of the active form of a regulator (here within 6 h) can only be achieved by post-translational mechanisms such as phosphorylation and/or targeted protein decay, the delayed response at the mRNA level appears coherent [42, 43]. Consistent with the changes observed at mRNA level, TFAP2A protein levels tend to decrease in the first 72 h after the TGFβ1 treatment (Fig. 2e and f).

To gain further insight into the relationship between TFAP2A expression and activity, we examined the mRNA expression data that was previously generated from human breast cancer cell lines [32]. The Neve et al. data set contained 51 samples that were separated in three categories according to their transcriptomic signature. Using the GOBO online tool we found that TFAP2A expression is reduced in the basal B breast cancer cell lines (Fig. 3a), which have a higher expression of the mesenchymal markers compared to the basal A type cell lines (Additional file 1: Figure S6). This is consistent with our observations in the mouse cell line [33]. We also analyzed the Neve et al. dataset [32] in ISMARA to identify the most significant TFAP2{A,C} targets, based on their ISMARA-provided z-score. Using the top 100 TFAP2{A,C} targets as input for the GOBO tool, we found that their expression is significantly increased in the basal B sub-type (Fig. 3b). Thus, we found a strikingly consistent negative correlation between TFAP2A mRNA and the expression of its transcriptional targets in the Neve et al. dataset, as well as in the data that we obtained in the NMuMG model. Remarkably, in a large panel of breast tumor datasets originating from more than 1500 patients, the expression of TFAP2A mRNA is also downregulated in the basal sub-type cancer category (Fig. 3c) [33]. More generally, using mRNA expression data from The Cancer Genome Atlas, we found that the expression of TFAP2A is positively correlated with that of epithelial markers and negatively correlated with that of mesenchymal markers, in normal breast tissue samples as well as in samples from breast tumors (Additional file 1: Figure S7).

TFAP2A binds directly to the Zeb2 promoter region

In addition to the significant activity change of the TFAP2{A,C} motif activity in human and mouse EMT systems (Fig. 1a and b), the interaction of the TFAP2{A,C} and ZEB1,2 motifs was also conserved in the

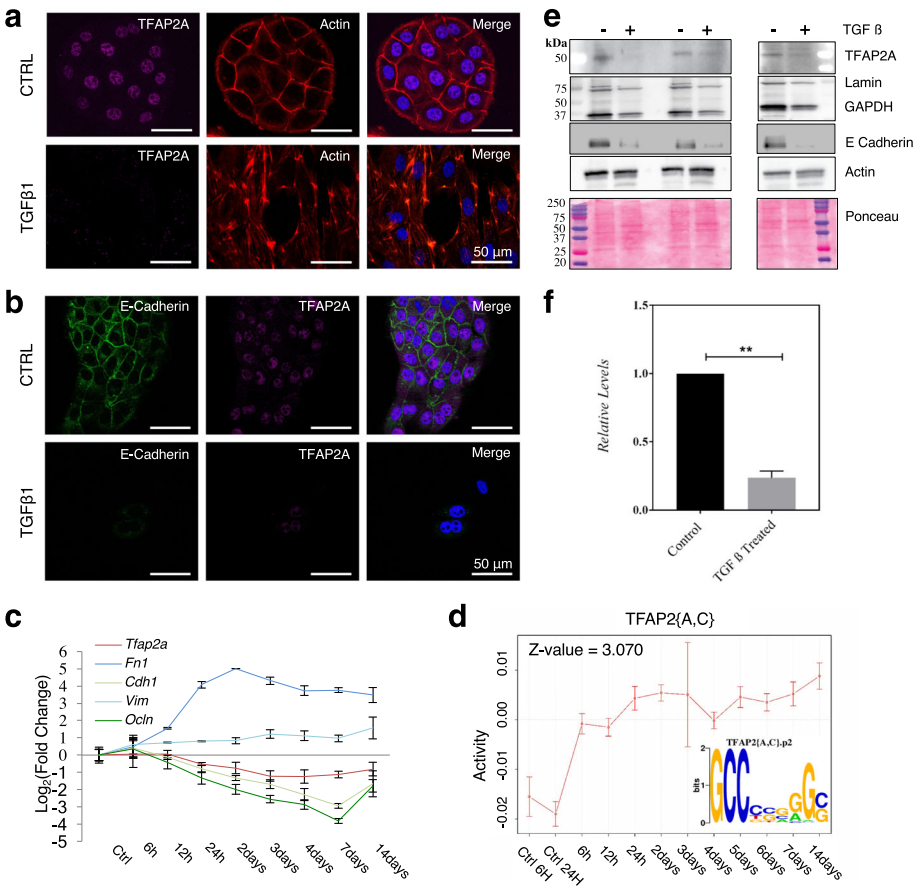

Fig. 2 TFAP2A expression and activity profile in the NMuMG EMT model. **a**-**b** NMuMG cells were treated with 2 ng/mL of TGFβ1 for 72 h and were stained for TFAP2A and F-Actin (**a**) and TFAP2A and E-cadherin (**b**). The merged panels represent colocalization of the imaged markers with the nucleus which was stained with DAPI and compared to controls. Scale bar represents 50 μm. **c** NMuMG cells were treated for 14 days with 2 ng/mL of TGFβ1. Quantitative RT-PCR of Tfap2a during the time course of this treatment indicates that Tfap2a mRNA levels are reduced upon EMT. The EMT markers E-cadherin (Cdh1), Fibronectin (Fn1), Occludin (Ocln), and Vimentin (Vim) follow the expected trend. **d** Two mRNA-seq samples from independent wells were prepared from a time course of NMuMG cells treated for 14 days with 2 ng/mL of TGFβ1, and the data was consequently analyzed with ISMARA [30]. The figure depicts the dynamics of TFAP2A/C transcriptional activity during the time course. The sequence logo of the TFAP2A/C binding motif is also indicated. **e**-**f** Lysates from NMuMG/E9 cells treated with 2 ng/mL of TGFβ1 for 72 h were probed for TFAP2A, GAPDH and Lamin B expression by WB and their levels compared with the expression levels of Actin and also to the Ponceau-stained membrane (**e**). The bar plot represents the densitometric quantification of the TFAP2A protein levels upon treatment compared to the control (**f**) ** indicates a *p*-value < 0.01 in the paired *t*-test ($P = 0.0014$)

EMT networks of both species. Our analysis predicted that TFAP2{A,C} controls the expression of *ZEB1* and *ZEB2* genes in both systems. The *Zeb2* target has a higher score than *Zeb1* in NMuMG cells (target scores from the initial ISMARA analysis were 0.7 for ZEB1 and 0.51 for ZEB2 in human, and 0.18 and 0.52, respectively in mouse). To validate the interaction between TFAP2A and the *Zeb2* promoter we performed an Electrophoretic Mobility Shift Assay (EMSA). From the SwissRegulon database of transcription factor binding sites that were predicted based on evolutionary conservation (www.swissregulon.ch), we found that the region around the second exon of the *Zeb2* gene, in which the ATG start codon resides, contains seven clusters of consensus binding sites for TFAP2{A,C} with a relatively high posterior probability. The corresponding region is represented in Fig. 4a. Two transcription start sites (TSS), annotated in the SwissRegulon, based on cap analysis of gene expression (CAGE) data [44], are in close proximity to the TFAP2{A,C} binding sites, in the intronic region between the first and the second exon (Fig. 4a) [44]. To confirm that the TFAP2A TF binds to the predicted sites, we carried out EMSA with radiolabeled oligonucleotides, each spanning one of the predicted binding sites (Fig. 4a and b). In the presence of the broad competitor poly-dI-dC, most of the probes give a shift upon addition of TFAP2A. The addition of an excessive amount of cold probes containing the same binding sites (Wt), results in a reduction of the shifted radiolabelled oligonucleotides, indicating competition for specific binding. This is further

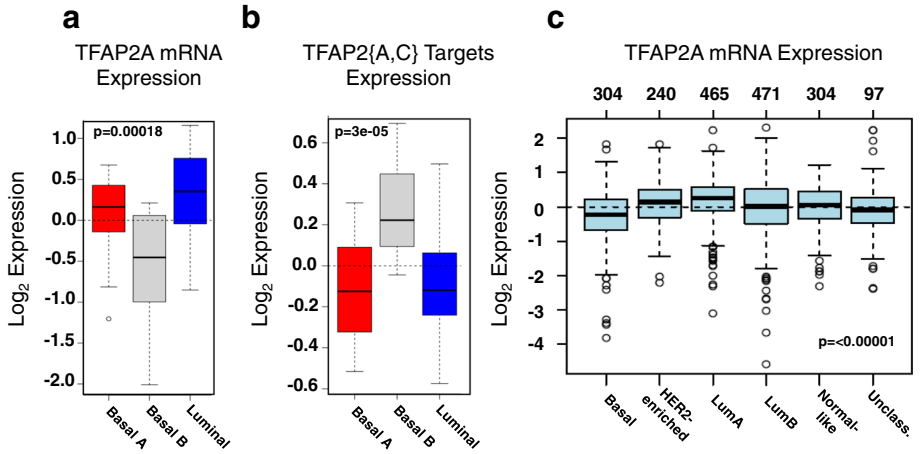

Fig. 3 TFAP2A expression and activity in breast cancers. Box plots of TFAP2A gene expression (**a**) and expression levels of the top 100 ISMARA-inferred TFAP2A targets (**b**) in a panel of breast cancer cell lines grouped in the basal A (*red*), basal B (*grey*) and luminal (*blue*) subgroups based on the annotation from Neve et al. [32]. **c** Box plot of TFAP2A gene expression for tumor samples stratified according to PAM50 subtypes [57]. All plots were generated with the GOBO online tool [33]

demonstrated by the fact that only few probes, indicated with red arrows, restored their shift in the presence of cold competitors that contained mutated versions of TFAP2A binding sites (M) (Fig. 4b).

To validate this regulatory interaction in NMuMG cells we have generated a stable cell line in which the overexpression of TFAP2A can be induced with doxycycline (see Methods; Additional file 1: Figure S2). As a control we established a similar cell line using an expression construct in which the TFAP2A coding region (CDR) was replaced by green fluorescent protein (GFP) CDR. Using an antibody that recognizes the endogenous TF we further confirmed that TFAP2A binds to the *Zeb2* promoter region by TFAP2A-chromatin immunoprecipitation (ChIP) followed by quantitative PCR: the *Zeb2* promoter was significantly enriched in the TFAP2A-ChIP from cell lines expressing either exogenously-encoded TFAP2A ($p = 0.005$). Cells expressing only endogenous TFAP2A also showed an enrichment of the the *Zeb2*, albeit not to the same level of significance ($p = 0.06$) (Fig. 4c).

Visualization of ChIP-seq data that we also obtained in this system, with the CLIPZ genome browser (www.clipz.unibas.ch) [45], confirms the presence of a peak in the predicted binding region that is only present in the TFAP2A-ChIP sample, but not in the Input controls (Fig. 4d) or the IgG (not shown). Overall, these results confirm that TFAP2A directly interacts with the *Zeb2* promoter, both in vitro as well as in the NMuMG cell line.

TFAP2A overexpression in NMuMG modulates epithelial plasticity

Finally, we used the above-mentioned cell lines to investigate the consequences of perturbed TFAP2A expression. Induced expression of TFAP2A, but not GFP, in untreated NMuMG cells led to morphological changes visible in phase contrast microscopy (Fig. 5a); compared to GFP-expressing cells, TFAP2A-expressing cells lose their epithelial polygonal cell shape and disperse on the plate. Consistently, qRT-PCR showed that adhesion-related genes were specifically deregulated upon TFAP2A induction (Additional file 1: Figure S3a and S3b). As expected, the treatment of GFP-expressing cells with TGFβ1 for 3 days leads to the induction of EMT markers *Snai1*, *Zeb2* and *Vim*. The expression of endogenous *Tfap2a* decreases upon the treatment of GFP-expressing NMuMG cells with TGFβ1. However, the induction of TFAP2A expression in the absence of TGFβ1 treatment appears to promote the expression of core EMT TFs such as *Snai1*, and *Zeb2* (Fig. 5b and Additional file 1: Figure S3c), without affecting the expression of E-cadherin at the mRNA level (Additional file 1: Figure S3a).

To better understand the effect of TFAP2A overexpression, we carried out transcriptional profiling of these four cell populations, namely untreated and TGFβ1-treated GFP-expressing cells, and untreated and TGFβ1-treated (for 72 h) TFAP2A overexpressing cells. The *Tfap2a* expression is increased upon doxycycline induction (Fig. 5b), but it decreases upon TGFβ1 treatment of GFP-expressing control cells (as we have observed before). Notably, the MARA analysis of these data reveals an increased activity of the TFAP2{A, C} motif in TGFβ1-induced, GFP-expressing cells, as we have initially observed in wild-type NMuMG cells, but also in TFAP2A-overexpressing cells treated with the growth factor when compared to GFP-expressing cells (Fig. 5c). The TGFβ1 treatment of TFAP2A-overexpressing cells further

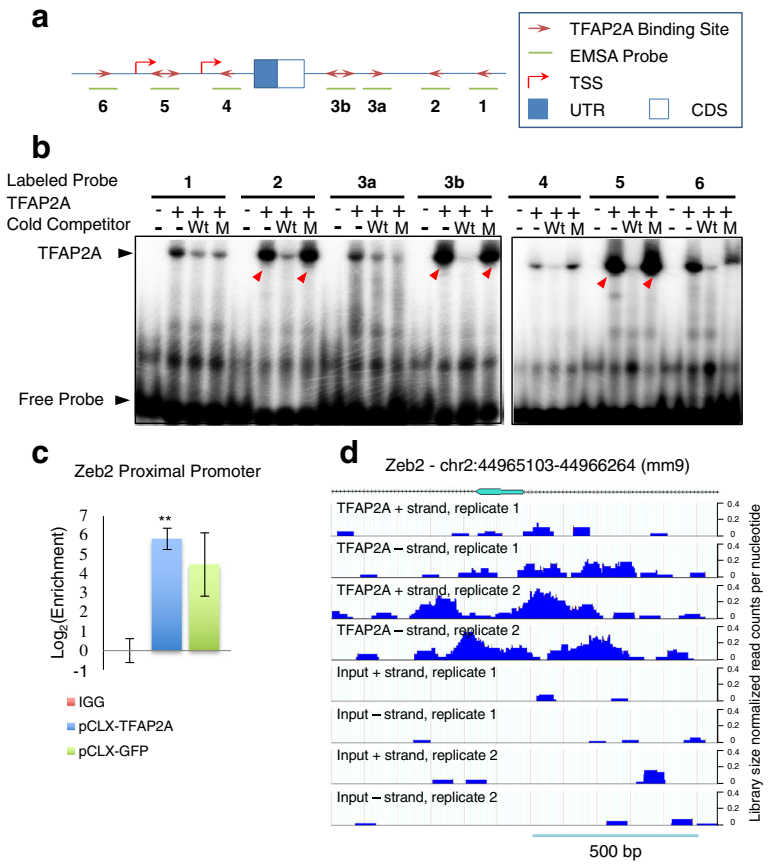

Fig. 4 TFAP2A binds directly to the Zeb2 promoter region. **a** Sketch of the region around the second exon of mouse *Zeb2*, showing the two transcription start sites found in SwissRegulon [44]. The *blue filled box* indicates the non-coding untranslated region (UTR) in exon 2, while the *white filled box* designates the start of the coding region (CDS). The predicted TFAP2A binding sites from SwissRegulon are marked with *red arrows*, and the probes that were used in (**b**) are indicated with *green lines* below the gene structure. Predicted transcription start sites (TSS) are also indicated. **b** Radiography of TFAP2A Electrophoretic Mobility Shift Assay (EMSA) with radiolabeled oligonucleotides, each spanning one of the predicted binding sites. The presence or absence of TFAP2A protein in the assay is indicated by a + or – sign, respectively. Cold competitors were used at 200-fold excess over the radiolabelled probes. Wt corresponds to unlabeled probe; M indicates a double-stranded oligonucleotide with a mutated TFAP2A binding site. *Red arrows* indicate the predicted TFAP2A binding probes that behave as expected from specific binding of TFAP2A. **c** TFAP2A ChIP was performed in NMuMG cells stably transduced with pCLX-TFAP2A (denoted as TFAP2A-OE (*blue*)) or with pCLX-GFP (denoted as TFAP2A-GFP (*green*)) viral vectors and further treated with 2 μg/mL doxycycline. Quantitative PCR data shows the enrichment of *Zeb2* promoter relative to a non-transcribed genomic region in TFAP2A-ChIP normalized to IgG control (*red*). Two independent experiments were performed for each condition and shown are means and standard deviations. The one-tail paired *t*-test indicates that TFAP2A is significantly enriched at the *Zeb2* (** for $p < 0.01$). **d** ChIP-seq libraries from TFAP2A ChIP or input chromatin were generated and the coverage of the genomic region spanning the second exon of *Zeb2* by reads is shown in a mouse genome browser (www.clipz.unibas.ch and [45]). The results of two independent experiments are presented. The TFAP2A ChIP-seq the *Zeb2* promoter region previously assessed by qPCR is enriched with respect to the input control sample. Mapping, annotation and visualization of deep-sequencing data was done with the ClipZ server [45]

increases the TFAP2A activity. Thus, the exogenously introduced TFAP2A has an opposite transcriptional activity relative to the endogenous form.

The activity of the SNAI1 motif decreases upon TGFβ1 treatment while its mRNA level increases, as expected from its known repressive activity in mesenchymal cells [41] (compare Fig. 5b and c). However, the >4-fold increase in *Snai1* mRNA that occurred upon TFAP2A overexpression was followed only by a small decrease in SNAI1 motif activity. Interestingly, the TGFβ1-induced decrease of SNAI1 activity is less pronounced when the TGFβ1 treatment is carried out in TFAP2A-overexpressing cells (Fig. 5b and c). These results indicate that overexpression of TFAP2A perturbs the course of TGFβ1-induced EMT in NMuMG cells.

Discussion

Metastasis is the leading cause of death among breast cancer patients and a deeper understanding of the process is necessary for the development of treatment strategies [46]. The development of malignancy has been related to epithelial plasticity, and unsurprisingly, regulatory modules

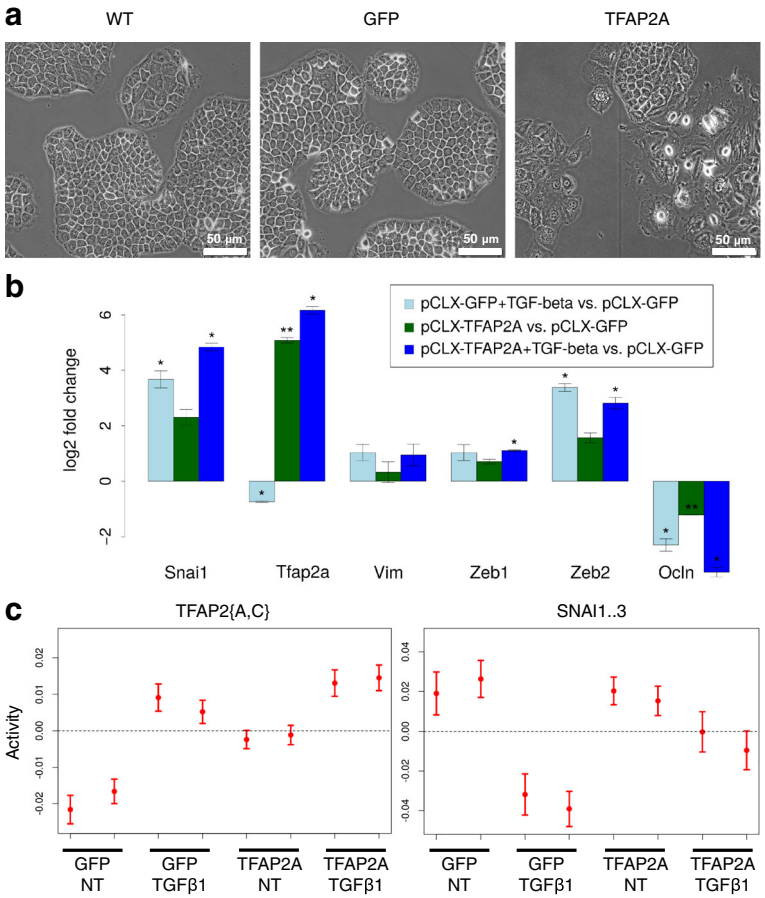

Fig. 5 TFAP2A overexpression in NMuMG modulates epithelial plasticity. **a** Expression of either GFP or TFAP2A was induced by 72 h doxycycline treatment in NMuMG cells stably transduced with either pCLX-GFP or pCLX-TFAP2A. Morphological changes and sparse cell arrangement are visible in phase contrast microscopy upon TFAP2A expression. Scale bar: 50 µm. **b** Gene expression $\log_2$ fold changes of EMT markers (TFs) were calculated from mRNA-seq samples of doxycycline-induced, TGFβ1-treated (72 h, 2 ng/mL) pCLX-GFP (pCLX-GFP + TGF-beta), doxycycline-induced pCLX-TFAP2A (pCLX-TFAP2A), as well as of doxycycline-induced, TGFβ1-treated (72 h, 2 ng/mL) pCLX-TFAP2A (pCLX-TFAP2A + TGF-beta) cell lines relative to doxycycline-induced pCLX-GFP (pCLX-GFP) cell line. Shown are the mean $\log_2$ fold changes (+/- 1 standard deviation) from two experiments. TFAP2A overexpression is apparent in both TFAP2A-induced samples (*dark green* and *dark blue*) but is not induced in cells treated with TGFβ1 alone (*light blue*). The EMT-inducing TFs have increased expression upon TFAP2A induction. * indicates a *p*-value ≤ 0.05 and ** a *p*-value ≤ 0.01 in a two-tailed *t*-test. **c** The transcriptional activities of TFAP2{A,C} and SNAI1..3 motifs in different conditions, as inferred with ISMARA from mRNA-seq data as described in (**b**). The two replicates from each condition are plotted next to each other

and networks that are involved in normal human development are hijacked during tumorigenic processes [41]. Although the regulatory network behind EMT has been intensely studied, by integrating data from multiple systems, recently developed computational methods can continue to provide new insights. In this study we have compared data from both developmental processes and cancer models of epithelial plasticity aiming to identify key regulators that are evolutionarily conserved. We found only a small number of motifs that have a significant activity change upon EMT in both human and mouse systems. Of these, SNAI1..3 and ZEB1..2 correspond to TFs that form the core EMT network [35]. We did not explicitly recover motifs for GSC, TWIST and FOXC2/SLUG. However, only the last factor has a specific motif represented in ISMARA. Motifs for miR-200 and the TGFβ1-related TGFI1 were only identified from the human samples. A novel insight derived from our analysis was that the motif corresponding to the TFAP2A and/or TFAP2C TFs also has a significant contribution to the expression changes that occur upon EMT in both species (Fig. 1a and b). The mechanistic link between TFAP2A/C and EMT was so far unknown, although TFAP2A was previously found important for neural crest formation and implicated in the activation of EMT inducing factors [47]. Furthermore, TFAP2A and TFAP2C have been implicated in mammary gland tumorigenesis and metastasis formation [16, 19]. Our data demonstrates that TFAP2A activity dynamically changes in the early time points of the TGFβ1 induced EMT in NMuMG cells, and thus suggests that TFAP2A regulates early steps in this

process (Fig. 2c). Although our analysis of the EMT time series indicated that the expression of *Tfap2a* is negatively correlated with the expression of its targets (reflected in the motif activity, Additional file 1: Figure S4), overexpression of TFAP2A induces changes that are similar to those occurring upon *Tfap2a* downregulation during EMT. This observation can have multiple causes. One is that TFAP2A activity is regulated post-translationally, similar to the core EMT TFs [41]. For instance, the SNAI1 protein has a rapid turn-over and its stability and activity are regulated by post-translational phosphorylation, lysine oxidation and ubiquitylation [41]. Indeed, it has been demonstrated that the sumoylation and phosphorylation of the TFAP2A protein can affect its transcription activation or DNA binding functions [48, 49]. Therefore, it is possible that during EMT, the activity of TFAP2A on its targets changes from repressive to activating and its mRNA levels may decrease due to a feedback regulatory mechanism. A regulatory step at the protein level is also suggested by the fact that the highest increase in TFAP2A activity is observed in the first 6 h of treatment whereas the changes in the *Tfap2a* mRNA are delayed to a later time point (Fig. 2c and d). Alternatively, TFAP2A may activate some of its targets and repress others, so that which effect dominates overall will depend on other factors or on TFAP2A expression levels. The dual transcription activity of TFAP2A has also been reported before [16]. Yet another possibility is that depending on its mode of expression and of post-translational modifications, TFAP2A may form distinct complexes with other factors to activate or repress its targets. Additional experiments will be necessary to address these possibilities. Nevertheless, our data provides evidence for a direct regulatory link between TFAP2A/C and the core EMT regulators ZEB1 and ZEB2 in both human and mouse. In mouse, we found that TFAP2A binds to the *Zeb2* promoter (Fig. 4), and that *Zeb2* levels increase when TFAP2A is overexpressed (Fig. 5b). These results indicate that TFAP2A regulates EMT-inducing factors transcriptionally. Although we have not investigated it in detail here, our TFAP2A-ChIP-seq data suggests that other critical regulators of EMT such as *Snai1*, *Sox4*, *Ezh2* and *Esrp2* may also be targets of TFAP2A (Additional file 1: Figure S5). This further strengthens the hypothesis that TFAP2A is part of a densely-connected network of genes that are essential for EMT [50–52]. Consistent with exogenous TFAP2A-induced activation of EMT markers, the NMuMG cells that overexpressed TFAP2A underwent phenotypical changes that were indicative of the acquisition of a mesenchymal phenotype (Fig. 5a). Furthermore, an EMT signature of positively regulated genes was significantly represented among genes that were up-regulated in TFAP2A-overexpressing NMuMG cells compared to control, GFP-expressing cells (Additional file 1: Table S4) [35]. Genes involved in cellular adhesion and glycosphingolipid metabolism, which has been recently suggested to regulate cellular adhesion via *St3gal5* and, more upstream, *Zeb1* [53], seems to also be affected by TFAP2A overexpression (Fig. 5b; Additional file 1: Figure S3b and S3c). Cell adhesion is concomitantly affected (Fig. 5a). Thus, our results support the link between TFAP2A and ZEB TFs, although overexpression of TFAP2A leads to cellular that are observed upon TGFβ1-induced down-regulation of endogenous TFAP2A. One cannot exclude that the observed induction of an EMT response upon TFAP2A overexpression is due to a phenomenon similar to the so-called 'squelching effect' [54]. The activity of TFAP2A does not appear to be sufficient for the induction of a complete EMT phenotype in the absence of TGFβ1 (Fig. 5a, c). Previously, ChIP-chip-based measurements of SMAD2/3 binding in human keratinocytes upon TGFβ stimulation indicated that SMAD2/3 binding sites co-occur with those for TFAP2A/C TFs, leading to the hypothesis that TFAP2A is involved in mediating the TGFβ signaling [55]. However, maintaining a high TFAP2A level in the context of TGFβ signaling may interfere with the activity of EMT TFs (Fig. 5c), consistent with our observation that EMT factors such as SNAI1 have less repressive activity when TFAP2A is overexpressed during TGFβ1-induced EMT. This in turn could be the rationale for the moderate downregulation in *Tfap2a* levels that we observed in the later phases of the TGFβ1-induced EMT time course (Fig. 2c). Consistent with previous studies that suggested that TFAP2A activation is connected with the luminal breast phenotype, thus promoting the epithelial state [16], here we found that endogenously-encoded TFAP2A is down-regulated upon TGFβ1-induced EMT. Interestingly, PRRX1, another TF that promotes EMT in a developmental context, was found to both induce the transition, and reduce the metastatic potential in tumors [56]. This suggests that the two processes are not always coupled and that a tumor suppressor can also activate EMT. This may be the case with TFAP2A as well; while it mediates the initiation of EMT, its sustained expression may interfere with EMT signaling. Our data thus connects TFAP2A to the core regulatory network that orchestrates the epithelium-to-mesenchyme transition in normal development as well as in cancers.

Conclusions

Applying recently developed computational methods to a set of epithelial plasticity datasets we have construct a conserved transcription factor motif interaction network that operates during the epithelium-to-mesenchyme transition. Our analysis recovered the known core EMT TFs and further linked the TFAP2A/C motif to this core network. Employing the NMuMG model cell line we provided further evidence that TFAP2A is involved in EMT, most likely in the early stages. We found that TFAP2A binds to the promoter of the Zeb2 master

regulator of EMT and that TFAP2A overexpression in NMuMG cells induces an increase in Zeb2 expression. Finally overexpression of TFAP2A in NMuMG cells promoted the expression of EMT markers and of cellular features related to the acquisition of a mesenchymal phenotype. Overall, our data links TFAP2A to the core TF network that is regulating EMT in normal development as well as in cancers.

Reviewers' comments

Reviewer's report 1: Dr. Martijn Huynen, Nijmegen Centre for Molecular Life Science, The Netherlands

Reviewer comments

The manuscript describes an elegant computational analysis of the regulatory motifs associated with the EMT transition, followed by the experimental validation that a new factor, TFAP2A, plays an important role in this process. In general I do find the first part of the paper very convincing, it computationally identifies the factor, confirms the results in independent data, and confirms binding of the factor to a predicted target. I do get a bit confused by the results of the overexpression of TFAP2A, and the arguments used to make these results consistent with the first part of the paper.

Author's response: *We thank the reviewer for the positive assessment of our computational analysis. Although we did find publicly available data that supports our conclusions about the involvement of TFAP2A in EMT, we nevertheless sought to validate its role ourselves. We tried to explain better the rationale and the results in the revision, even though some results remain paradoxical.*

Does Fig. 1 contain the complete set of motifs that are predicted to be "differentially active" in the transition? If so, is it a coincidence that they are all connected to each other?

Author's response: *We have described the selection of the motifs that we show in the legend of the Figure. Briefly, we only showed motifs with an absolute z-score > 2 and arrows that represent predictions with probabilities larger than a threshold (0.35 for human and 0.15 for mouse). For the readability of Fig.* 1*, only motifs that have at least a predicted interaction with another motif at the mentioned thresholds are considered. However, realizing that motifs with significant activity that are not connected to other motifs may also be of interest, we have now included the full tables of motif activity changes upon EMT as Additional file* 1*: Tables S2 and S3.*

I am surprised by the low level of conservation between the species. Are there some motifs from e.g. human that are just below a threshold? The authors argue "The motif interaction networks derived from mouse and human EMT models suggest that only a small fraction of the TFs has a highly conserved and significant role in both species." How reliable are those species-specific predictions, and how reliable is the absence of a signal in these analyses, with these data.

Author's response: *Although we selected sequencing data sets obtained from systems where EMT presumably occurs for both species, we unfortunately did not have matching systems available for human and mouse. So indeed, the precise scores of the different motifs depend on the data sets that we used and given sufficient data, other motifs may emerge as having similar behaviour in mouse and human EMT systems. Nevertheless, we found it reassuring that the core EMT factors that were extensively studied so far, such as SNAI and ZEB emerged from our analysis. That the TFAP2A,C motif also has a conserved function was unexpected and prompted us to study it further.*

If I understand the manuscript correctly, the downregulation of TFAP2A is associated with the epithelial to mesenchymal transition. Why then overexpress TFAP2A? Even is this has to do with technical limitations, I would like to see that mentioned explicitly to better understand the logic of the approach.

Author's response: *Our initial analysis indicated that the expression of TFAP2A is down-regulated during EMT (Fig.* 2*), while its motif activity increases, suggesting that TFAP2A may function as a repressor. Therefore, we overexpressed TFAP2A, reasoning that this should perturb the process of TGFb1-induced EMT. Indeed, this is also what we observe. However, analysis of the sequencing data obtained after TFAP2A overexpressionoverexpression also revealed some paradoxical results, which we addressed in our discussion.*

I find the discussion why "overexpression of TFAP2A induces changes that are similar to those occurring upon Tfap2a downregulation during EMT" lengthy and unconvincing. The authors first perform a very thorough quantitative analysis of gene expression and motif occurrence data, based on the simplifying but defendable assumptions of their linear model, confirm their findings in independent breast cancer data (Fig. 3). Then they use a large number of ad-hoc arguments to explain the inconsistencies in their results. They may all be true, but they are not convincing. Given the apparent contradictory results of the overexpression, I am surprised by the sentence "Finally, we confirm that overexpression of TFAP2A in NMuMG cells modulates epithelial plasticity and cell adhesion" in the abstract as those results do not confirm a specific hypothesis based on the results of the quantitative analysis.

Author's response: *We have revised the discussion to hopefully make it more streamlined. We agree with the reviewer that the initial computational analysis suggested a clear picture of TFAP2's involvement in EMT. However, as we tried to go deeper into the mouse model,*

the results that we obtained were more complex than we anticipated. We felt it was important to show the unexpected overexpression results, but in the revision we have included only the initial characterization of this cell line, without following it into the phenotypic analysis. We hope that our revised description of the results makes it clear what we have learned from the different systems about the behaviour and role of TFAP2A.

In Fig. 5c there is a line connecting the various constructs. I take it this is not meant to implicate some sort of continuity? I do fully support publication once these issues have been handled.

Author's response: *Thank you for pointing this out. We have removed the lines to prevent the illusion of continuity of the data points.*

editorial: The legend with Fig. 3 could use some work "ABasal" or "Basal A"?

Author's response: *We thank the reviewer for pointing this out. We have fixed this issue and made the labels easier to read.*

TFAP2A expression was found to be less organized in breast cancer compared to normal mammary gland. - > glands

Author's response: *We think that the original formulation is correct.*

what is "substantially expressed"

Author's response: *We have explained that only Tfap2a (and not the other family members) has read coverage in all exons.*

It would be nice to specify which TFs of the core EMT network of ref 33 are retrieved and which are not.

Author's response: *We have expanded the text accordingly.*

"transcriptional" can often be replaced by "transcription", e.g. in "transcriptional regulation" page 18,

Author's response: *We have changed the term in all places where we thought it makes sense.*

line 20 "the interactions of the TFAP2{A,C}" appears redundant.

Author's response: *We removed the redundancy.*

page 22. "in untreated NMuMG cells lead to morphological changes" – > "led"

Author's response: *Fixed.*

"an EMT signature of positively regulated genes were significantly represented" – > "was"

Author's response: *Fixed.*

Reviewer's report 1: Dr. Nicola Aceto, Department of Biomedicine, University of Basel, Switzerland

Reviewer comments

Dimitrova et al. present a manuscript in which they highlight the transcription factor TFAP2A as a novel EMT regulator. They suggest that TFAP2A target genes, such as ZEB2, are upregulated during EMT in the NMuMG mouse model. Further, they conclude that the interaction between TFAP2A and ZEB2 promoter affects ZEB2 expression, hence modulating the EMT process itself and providing evidence for a role of TFAP2A in cancer progression. Altogether, this is an interesting manuscript yet requiring a few modifications and clarifications to convincingly argue in favor of TFAP2A's role in cancer progression.

(1) Introduction: the authors write their introductory paragraph arguing that e.g. "cancer progression, metastasis and chemotherapy resistance have all been linked to EMT". However, the role of EMT for each of these processes is highly debated in the field, and I would suggest the authors to provide a more balanced introduction, where it is clearly stated (and referenced) that the role/requirement of EMT in all these processed has still to be fully understood, especially in clinically-relevant settings.

Author's response: *We have rephrased and provided additional references to make the introduction more balanced.*

(2) Fig. 2a: I remain unconvinced about the degree of EMT that is triggered by TGFb in NMuMG cells. For instance, why only a small fraction of control cells express E-cad (roughly 30%)? Looking at the TGFb-treated cells, this ratio appears to remain the same (3/9 cells, i.e. roughly 30%). TFAP2A-positive vs negative cells in control vs TGFb also do not seem to change much, and neither does actin. I would suggest the authors to provide more quantitative data here (% of positive cells for each marker, or signal intensity) that comprise several fields of view.

Author's response: *To answer the reviewer's questions, we have redone the experiment, and imaged the cells with higher magnification. The results in the revised Fig.* 2 *clearly show that TFAP2A is abundantly expressed and nuclearly localized in control cells, while this staining pattern is abrogated upon TGFb1 treatment. In almost all control cells, the expression of E-cadherin is clearly visible, as is its localization close to the plasma membrane, features which are also abrogated by the TGFb1 treatment. E-cadherin levels estimated by Western blot (Fig.* 2*e) also indicate down-regulation upon TGFb1 treatment.*

(3) Fig. 2c: how relevant is a Z-value of 3, with an activity range varying from -0.02 to 0.01? Looking at Fig. 1 (Z-values ranging from -19 to +19), can the authors convincingly state that TFAP2 target genes (and TFAP2 activity, respectively) significantly change upon TGFb treatment in NMuMG cells?

Author's response: *Please note that Fig.* 1 *was generated based on multiple data sets and that is why the z-scores cover a much larger range. Based on a standard normal distribution of z-scores we consider values larger than 2 (in absolute value) significant.*

(4) Fig. 2d: somehow related to the previous point. Changes in TFAP2A protein levels are not very impressive. Is the change statistically significant? Control does not seem to have any error bar, was it repeated more than once?

Author's response: *We have repeated this experiment as well, using three biological replicates, adding an additional control (actin, in addition to lamin and GAPDH) and also Ponceau staining (current Fig.* 2*e). Although the overall protein levels are similar between conditions, TFAP2A's expression decreases upon TGFb1 treatment (as apparent also from the immunofluorescence staining, Fig.* 2*a). The controls that we initially used, lamin and GAPDH, also decrease to some extent upon TGFb1 treatment, which is probably why the relative change in TFAP2A in our initial figure was not very impressive. However, relative to the total protein level as well as to actin, TFAP2A expression is clearly reduced by the TGFb1 treatment.*

(5) Fig. 3: The authors observe a correlation between low TFAP2A expression and basal type of breast cancer. Two questions arise here: (a) is basalB more EMT-like than basal-A?

Author's response: *In the original publication (Ringner* et al. *PLoS One, 6:e17911, 2011), the basal B type is considered "more stem like".*

(b) how are TFAP2A target genes behaving in the larger dataset with 1500 samples?

Author's response: *Unfortunately we could not carry out this analysis on the GOBO web server.*

(6) Fig. 5: could the authors elaborate more about their conclusion "TFAP2A perturbs the course of TGFb-induced EMT in NMuMG cells"? It seems here that TFAP2A mRNA expression and activity are somewhat disconnected here, yet in previous experiments they seem to be going along quite well (e.g. see Fig. 2b-c and Fig. 3).

Author's response: *The reviewer, as reviewer #1 as well, rightly points out that the TFAP2A that is expressed from the exogenous construct seems to behave differently than the endogenously-encoded gene. This is also apparent from the quantification of TFAP2A expression in TGFb1-treated control cells, that only express endogenously encoded TFAP2A (which is down-regulated by the treatment) and in TFAP2A overexpressionoverexpression (where the expression is up-regulated, as expected, Fig.* 5*b). We discuss possible causes for this discrepancy in our manuscript (*Discussion *section). Although we did not identify the precise cause for it, we felt that it was important to show these results.*

(7) Fig. 6: In some instances (i.e. in TGFb-treated samples), actin staining seems to extend to regions that do not display any Hoechst staining. For example, in TFAP2A + TGFb sample, actin staining shows cells on the lower right corner of the image, but those cells do not show up in the Hoechst staining.

Author's response: *We think that this had to do with the intensity of the signal. However, we removed this figure from the revised version of the manuscript.*

(8) Differences in the aggregation index are not very impressive, and when taken per se would not be a strong argument of the involvement of TFAP2A in EMT. Instead, what would be the effect -in terms of EMT genes expression- of depleting TFAP2A in NMuMG cells treated with TGFb?

Author's response: *Because endogenous TFAP2A is down-regulated upon TGFb1 treatment, we initially sought to perturb the course of EMT by overexpressing TFAP2A and we carried out most of the experiments with this construct. It turned out that the overexpression of TFAP2A leads to similar molecular signatures as the downregulation of endogenous TFAP2A that takes place upon TGFb1-induced EMT. We agree with the reviewer that presenting the results with this construct as well as with the siRNAs makes the interpretation very difficult. We therefore decided to remove this figure and close the study at the point where the exogenous construct showed paradoxical results.*

The authors show in Additional file 1: Figure S3 some EMT genes, but it seems that genes such as Vim and Ocln are missing.

Author's response: *We have regenerated panel b in Fig.* 5 *based on the mRNA-seq samples that we used to infer the motif activities shown in panel c of the figure and we have included also Ocln, aside from Vim, whose expression we also estimated by qPCR. Both of the markers behave as expected in EMT. The additional qPCR validations are now shown in Additional file* 1*: Figure S3c.*

Also, what is the TFAP2A knockdown level with the siRNAs?

Author's response: *As we explained above, because the results of perturbing TFAP2A expression were difficult to interpret, we decided to not pursue too far the perturbation experiments. Therefore, we removed Fig. 6 and we did not include the siRNA quantifications in the revised manuscript.*

(9) Generally, it would be great to show some functional assays related to EMT (e.g. Boyden chamber, etc.) to reinforce the involvement of TFAP2A in this process

Author's response: *We agree with the reviewer that it would be exciting to carry out these studies. However, as the reviewer probably appreciates, this regulatory network is very complex and the perturbation experiments did not turn out as we expected. We therefore decided to follow the suggestion of reviewer #1, concentrating on the comparative analysis of the different systems that yielded consistent results and not trying to resolve the specific mechanism of TFAP2A, which likely depends on the precise form of the protein that is expressed from the endogenous locus.*

Reviewer's report 2: Dr. Martijn Huynen, Nijmegen Centre for Molecular Life Science, The Netherlands

Reviewer comments

This reviewer provided no additional comments.

Reviewer's report 2: Dr. Nicola Aceto, Department of Biomedicine, University of Basel, Switzerland

Reviewer's comments

Dimitrova et al. present a revised version of the manuscript that addressed and discussed some of the initial concerns. While I find the manuscript worthy of publication, a few points are still worth mentioning: (1) In an answer to my previous question #5 (see 1st review) the authors argue that Basal B is considered more stem-like (therefore more mesenchymal) than Basal A. However, EMT and stem-like are two very different features of cancer cells as well as normal tissues, which may or may not overlap depending on a variety of factors. For instance, a number of tumor cell lines that are fully epithelial can display stem-like features (tumor initiation, self-renewal, differentiation). My original question was more whether by looking at gene expression data of Basal B, this tumor type expresses significantly more EMT markers than Basal A. This would reinforce their conclusions.

Author's response: *To answer the reviewer's question we have used the GOBO tool to compare the expression levels of various epithelial and mesenchymal markers in Basal A and Basal B tumor types. As shown in the new Additional file* 1*: Figure S6, epithelial markers have higher expression in Basal A tumors, whereas mesenchymal markers have higher expression in Basal B tumors. This is in line with the concept that Basal B tumors are more mesenchymal.*

(2) Regarding patient data it would be more convincing to check the expression of TFAP2 (as well as its target genes and EMT markers) in several independent datasets to reinforce the conclusions of the authors.

Author's response: *To answer the reviewer's second question, we have used yet another data set, namely expression profiles of tumors and normal tissue samples from The Cancer Genome Atlas, to further examine the relationship between the expression of TFAP2A and that of various epithelial and mesenchymal markers. These results, summarized in the new Additional file* 1*: Figure S7, show that the TFAP2A expression is positively correlated with that of epithelial markers and negatively correlated with that of mesenchymal markers. This is again consistent with the results we obtained in our experimental system (Fig.* 2*).*

Additional file

Additional file 1: Supplementary information. (DOCX 19403 kb)

Abbreviations

BOFS: Branchio-Oculo-Facial Syndro; CAGE: Cap analysis of gene expression; CDR: Coding region; ChIP: Chromatin immunoprecipitation; EMT: Epithelial to mesenchymal transition; ESC: Embryonic stem cells; GOBO: Gene expression-based outcome for breast cancer online; MARA: Motif-activity response analysis; NC: Neural crest; NCBI: National Center for Biotechnology Information; NMuMG: Mouse mammary gland epithelial cell line; SELEX: Systemic evolution of Ligand by EXponential enrichment; SRA: Sequence read archive; TF: Transcription factor; WB: Western blot

Acknowledgements

We thank Arnau Vina-Vilaseca for excellent technical assistance. The authors also thank Xiaomo Wu and Geoges Martin for advice on setting-up EMSA experiments.

Funding

This work was supported by the Swiss National Science Foundation grant # 31003A_147013 to MZ and by the SystemsX.ch initiative in systems biology, through the RTD project 51RT-0_126031. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The sequencing data can be accessed at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) with the SRA accession ID SRP067296. The datasets that were taken from other studies and analyzed for this study are listed in Additional file 1: Table S1.

Authors' contributions

YD, SG, NM, BD, and DM performed experiments; YD, WAG, MZ designed and GC and MZ supervised the study; AJG, YD and SG performed data analysis; YD, AJG, SG and MZ wrote the manuscript. All authors edited the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

All authors have read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Author details

[1]Biozentrum, University of Basel, Klingelbergstrasse 50-70, CH-4056 Basel, Switzerland. [2]Department of Biomedicine, University of Basel, Mattenstrasse 28, CH-4058 Basel, Switzerland. [3]Department of Cellular and Molecular Pathology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

Received: 17 November 2016 Accepted: 22 March 2017
Published online: 17 April 2017

References

1. Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. Cell. 2009;139(5):871–90.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74.
3. Micalizzi DS, Farabaugh SM, Ford HL. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. J Mammary Gland Biol Neoplasia. 2010;15(2):117–34.
4. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008;133(4):704–15.
5. Scheel C, Eaton EN, Li SH, Chaffer CL, Reinhardt F, Kah KJ, Bell G, Guo W, Rubin J, Richardson AL, et al. Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. Cell. 2011;145(6):926–40.

6. Prater MD, Petit V, Alasdair Russell I, Giraddi RR, Shehata M, Menon S, Schulte R, Kalajzic I, Rath N, Olson MF, et al. Mammary stem cells have myoepithelial cell properties. Nat Cell Biol. 2014;16(10):942–50.
7. Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, Wu CC, LeBleu VS, Kalluri R. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature. 2015;527(7579):525–30.
8. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. 2013;339(6119):580–4.
9. Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong ST, Choi H, El Rayes T, Ryu S, Troeger J, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature. 2015;527(7579):472–6.
10. Sarrio D, Rodriguez-Pinilla SM, Hardisson D, Cano A, Moreno-Bueno G, Palacios J. Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. Cancer Res. 2008;68(4):989–97.
11. Zhang J, Hagopian-Donaldson S, Serbedzija G, Elsemore J, Plehn-Dujowich D, McMahon AP, Flavell RA, Williams T. Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. Nature. 1996;381(6579):238–41.
12. Milunsky JM, Maher TA, Zhao G, Roberts AE, Stalker HJ, Zori RT, Burch MN, Clemens M, Mulliken JB, Smith R, et al. TFAP2A mutations result in branchio-oculo-facial syndrome. Am J Hum Genet. 2008;82(5):1171–7.
13. Williams T, Tjian R. Characterization of a dimerization motif in AP-2 and its function in heterologous DNA-binding proteins. Science. 1991;251(4997):1067–71.
14. Meier P, Koedood M, Philipp J, Fontana A, Mitchell PJ. Alternative mRNAs encode multiple isoforms of transcription factor AP-2 during murine embryogenesis. Dev Biol. 1995;169(1):1–14.
15. Mohibullah N, Donner A, Ippolito JA, Williams T. SELEX and missing phosphate contact analyses reveal flexibility within the AP-2[alpha] protein: DNA binding complex. Nucleic Acids Res. 1999;27(13):2760–9.
16. Bogachek MV, Chen Y, Kulak MV, Woodfield GW, Cyr AR, Park JM, Spanheimer PM, Li Y, Li T, Weigel RJ. Sumoylation pathway is required to maintain the basal breast cancer subtype. Cancer Cell. 2014;25(6):748–61.
17. Zhang J, Brewer S, Huang J, Williams T. Overexpression of transcription factor AP-2alpha suppresses mammary gland growth and morphogenesis. Dev Biol. 2003;256(1):127–45.
18. Jager R, Werling U, Rimpf S, Jacob A, Schorle H. Transcription factor AP-2gamma stimulates proliferation and apoptosis and impairs differentiation in a transgenic model. Mol Cancer Res. 2003;1(12):921–9.
19. Cyr AR, Kulak MV, Park JM, Bogachek MV, Spanheimer PM, Woodfield GW, White-Baer LS, O'Malley YQ, Sugg SL, Olivier AK, Zhang W, Domann FE, Weigel RJ. TFAP2C governs the luminal epithelial phenotype in mammary development and carcinogenesis. Oncogene. 2014;34(4):436–44.
20. Shi D, Xie F, Zhang Y, Tian Y, Chen W, Fu L, Wang J, Guo W, Kang T, Huang W, et al. TFAP2A regulates nasopharyngeal carcinoma growth and survival by targeting HIF-1alpha signaling pathway. Cancer Prev Res. 2014;7(2):266–77.
21. Wang W, Lv L, Pan K, Zhang Y, Zhao JJ, Chen JG, Chen YB, Li YQ, Wang QJ, He J, et al. Reduced expression of transcription factor AP-2alpha is associated with gastric adenocarcinoma prognosis. PLoS One. 2011;6(9), e24897.
22. Melnikova VO, Bar-Eli M. Transcriptional control of the melanoma malignant phenotype. Cancer Biol Ther. 2008;7(7):997–1003.
23. Pellikainen J, Naukkarinen A, Ropponen K, Rummukainen J, Kataja V, Kellokoski J, Eskelinen M, Kosma VM. Expression of HER2 and its association with AP-2 in breast cancer. Eur J Cancer. 2004;40(10):1485–95.
24. Maeda M, Johnson KR, Wheelock MJ. Cadherin switching: essential for behavioral but not morphological changes during an epithelium-to-mesenchyme transition. J Cell Sci. 2005;118(Pt 5):873–87.
25. Li Q, Luo C, Lohr CV, Dashwood RH. Activator protein-2alpha functions as a master regulator of multiple transcription factors in the mouse liver. Hepatol Res. 2011;41(8):776–83.
26. Giry-Laterriere M, Cherpin O, Kim YS, Jensen J, Salmon P. Polyswitch lentivectors: "all-in-one" lentiviral vectors for drug-inducible gene expression, live selection, and recombination cloning. Hum Gene Ther. 2011;22(10):1255–67.
27. Gruber AR, Martin G, Muller P, Schmidt A, Gruber AJ, Gumienny R, Mittal N, Jayachandran R, Pieters J, Keller W, et al. Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. Nat Commun. 2014;5:5465.
28. Blecher-Gonen R, Barnett-Itzhaki Z, Jaitin D, Amann-Zalcenstein D, Lara-Astiaso D, Amit I. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat Protoc. 2013;8(3):539–54.
29. Wu X, Gehring W. Cellular uptake of the Antennapedia homeodomain polypeptide by macropinocytosis. Biochem Biophys Res Commun. 2014; 443(4):1136–40.
30. Balwierz PJ, Pachkov M, Arnold P, Gruber AJ, Zavolan M, van Nimwegen E. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. Genome Res. 2014;24(5):869–84.
31. Gruber AJ, Grandy WA, Balwierz PJ, Dimitrova YA, Pachkov M, Ciaudo C, Nimwegen E, Zavolan M. Embryonic stem cell-specific microRNAs contribute to pluripotency by inhibiting regulators of multiple differentiation pathways. Nucleic Acids Res. 2014;42(14):9313–26.
32. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006;10(6):515–27.
33. Ringner M, Fredlund E, Hakkinen J, Borg A, Staaf J. GOBO: gene expression-based outcome for breast cancer online. PLoS One. 2011;6(3), e17911.
34. Evseenko D, Zhu Y, Schenke-Layland K, Kuo J, Latour B, Ge S, Scholes J, Dravid G, Li X, MacLellan WR, et al. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. Proc Natl Acad Sci U S A. 2010;107(31):13742–7.
35. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci U S A. 2010;107(35):15449–54.
36. Diepenbruck M, Waldmeier L, Ivanek R, Berninger P, Arnold P, van Nimwegen E, Christofori G. Tead2 expression levels control the subcellular distribution of Yap and Taz, zyxin expression and epithelial-mesenchymal transition. J Cell Sci. 2014;127(Pt 7):1523–36.
37. Brunskill EW, Potter AS, Distasio A, Dexheimer P, Plassard A, Aronow BJ, Potter SS. A gene expression atlas of early craniofacial development. Dev Biol. 2014;391(2):133–46.
38. Feuerborn A, Srivastava PK, Kuffer S, Grandy WA, Sijmonsma TP, Gretz N, Brors B, Grone HJ. The Forkhead factor FoxQ1 influences epithelial differentiation. J Cell Physiol. 2011;226(3):710–9.
39. Tiwari N, Meyer-Schaller N, Arnold P, Antoniadis H, Pachkov M, van Nimwegen E, Christofori G. Klf4 is a transcriptional regulator of genes critical for EMT, including Jnk1 (Mapk8). PLoS One. 2013;8(2), e57329.
40. Kreitzer FR, Salomonis N, Sheehan A, Huang M, Park JS, Spindler MJ, Lizarraga P, Weiss WA, So PL, Conklin BR. A robust method to derive functional neural crest cells from human pluripotent stem cells. American journal of stem cells. 2013;2(2):119–31.
41. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer. 2013;13(2):97–110.
42. Nardozzi JD, Lott K, Cingolani G. Phosphorylation meets nuclear import: a review. Cell Commun Signal. 2010;8:32.
43. Westermarck J. Regulation of transcription factor function by targeted protein degradation: an overview focusing on p53, c-Myc, and c-Jun. Methods Mol Biol. 2010;647:31–6.
44. Pachkov M, Erb I, Molina N, van Nimwegen E. SwissRegulon: a database of genome-wide annotations of regulatory sites. Nucleic Acids Res. 2007;35(Database issue):D127–31.
45. Khorshid M, Rodak C, Zavolan M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. Nucleic Acids Res. 2011;39(Database issue):D245–52.
46. Bill R, Christofori G. The relevance of EMT in breast cancer metastasis: Correlation or causality? FEBS Lett. 2015;589(14):1577–87.
47. Rada-Iglesias A, Bajpai R, Prescott S, Brugmann SA, Swigut T, Wysocka J. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell. 2012;11(5):633–48.
48. Berlato C, Chan KV, Price AM, Canosa M, Scibetta AG, Hurst HC. Alternative TFAP2A isoforms have distinct activities in breast cancer. Breast Cancer Res. 2011;13(2):R23.
49. Garcia MA, Campillos M, Marina A, Valdivieso F, Vazquez J. Transcription factor AP-2 activity is modulated by protein kinase A-mediated phosphorylation. FEBS Lett. 1999;444(1):27–31.
50. Tiwari N, Tiwari VK, Waldmeier L, Balwierz PJ, Arnold P, Pachkov M, Meyer-Schaller N, Schubeler D, van Nimwegen E, Christofori G. Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming. Cancer Cell. 2013;23(6):768–83.

51. Cano A, Perez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, Portillo F, Nieto MA. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol. 2000;2(2):76–83.
52. Horiguchi K, Sakamoto K, Koinuma D, Semba K, Inoue A, Inoue S, Fujii H, Yamaguchi A, Miyazawa K, Miyazono K, et al. TGF-beta drives epithelial-mesenchymal transition through deltaEF1-mediated downregulation of ESRP. Oncogene. 2012;31(26):3190–201.
53. Mathow D, Chessa F, Rabionet M, Kaden S, Jennemann R, Sandhoff R, Grone HJ, Feuerborn A. Zeb1 affects epithelial cell adhesion by diverting glycosphingolipid metabolism. EMBO Rep. 2015;16(3):321–31.
54. Heslot H, Gaillardin C. Molecular biology and genetic engineering of yeasts. Boca Raton: CRC Press; 1992.
55. Koinuma D, Tsutsumi S, Kamimura N, Taniguchi H, Miyazawa K, Sunamura M, Imamura T, Miyazono K, Aburatani H. Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TFAP2A in transforming growth factor beta signaling. Mol Cell Biol. 2009;29(1):172–86.
56. Ocana OH, Corcoles R, Fabra A, Moreno-Bueno G, Acloque H, Vega S, Barrallo-Gimeno A, Cano A, Nieto MA. Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. Cancer Cell. 2012;22(6):709–24.
57. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27(8):1160–7.

Supplementary Information

TFAP2A is a component of the ZEB1/2 network that regulates TGFB1-induced epithelial to mesenchymal transition

Yoana Dimitrova[1], Andreas J. Gruber[1], Nitish Mittal[1], Souvik Ghosh[1], Beatrice Dimitriades[1], Daniel Mathow[3], William Aaron Grandy[1], Gerhard Christofori[2], Mihaela Zavolan[1,4]

[1]Biozentrum, University of Basel, Klingelbergstrasse 50-70, CH-4056 Basel, Switzerland.
[2]Department of Biomedicine, University of Basel, Mattenstrasse 28, CH-4058 Basel, Switzerland
[3]Department of Cellular and Molecular Pathology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
[4]Correspondence should be addressed to Mihaela Zavolan Biozentrum, University of Basel, Klingelberstrasse 50-70, CH-4056 Basel, Switzerland mihaela.zavolan@unibas.ch.

Table-S1. Datasets used for the generation of Figure 1.

Accession	Species	Description	Reference
GSE44727	human	iPS cells vs iPS derived neural crest cells	1
GSE23833	mouse	NM18 cells transfected with 25nm scrambled siRNA for 48hrs vs transfected with 25nm scrambled siRNA for 48hrs and treated with TGFβ1 for 40hrs	2
GSE49151	mouse	NMuMG/E9 cells treated with Control siRNA 1 or Control siRNA 1 and TGFβ1	3
GSE21668	human	Undifferentiated embryonic stem cells, H9 vs mesodermal progenitor population	4
GSE9691	human	HMLE cells untreated vs TGFβ1 treated	5
GSE55711	mouse	Py2T untreated vs Py2T 5 days TGFβ	6
GSE55964	mouse	Neuroepithelim vs neural crest	7

Table-S2. Motif activity changes derived from human EMT models (sorted by absolute z-scores).

Motif	Z-Score
SNAI1..3.p2	-24.288498
ZEB1.p2	-15.499399
TFAP2{A,C}.p2	13.229919
bHLH_family.p2	-9.973254
ARNT_ARNT2_BHLHB2_MAX_MYC_USF1.p2	-9.771872
ATF6.p2	9.291562
HBP1_HMGB_SSRP1_UBTF.p2	9.226859
MAFB.p2	9.049806
UCCAGUU	8.346224
TGIF1.p2	8.153858
AAUACUG	8.008944
GTF2I.p2	7.767217
ZNF238.p2	7.406887
TFEB.p2	7.375766
GAGGUAG	6.920853
NR4A2.p2	-6.870112
IKZF2.p2	6.531350
ESRRA.p2	-6.402536
ZNF384.p2	6.350658
XBP1.p3	6.286392
FOX{I1,J2}.p2	6.271248
HNF4A_NR2F1,2.p2	-6.229455
ZNF423.p2	-6.162108
HOX{A6,A7,B6,B7}.p2	6.099096
SREBF1,2.p2	6.086621
CUCCCAA	-5.966596
KLF12.p2	5.946872
RFX1..5_RFXANK_RFXAP.p2	5.541792
CRX.p2	-5.209987
TEAD1.p2	5.085874
MYOD1.p2	-5.075488
SRF.p3	5.044129
GUAAACA	5.000653
UCACAGU	4.838087
RXR{A,B,G}.p2	4.822125
FOXN1.p2	4.817176
TFAP4.p2	4.809295
IKZF1.p2	-4.799092
NKX3-1.p2	-4.792145
POU5F1_SOX2{dimer}.p2	-4.723304
NR6A1.p2	-4.421826
UGUGCUU	-4.412173
DBP.p2	4.382192
NFE2L1.p2	4.332266
REST.p3	-4.301219
TFDP1.p2	4.187116
PAX5.p2	-4.168027
FEV.p2	4.160352
NFE2L2.p2	-4.136635
MTF1.p2	-4.071565
GAUUGUC	4.043714
PITX1..3.p2	-4.039921
HLF.p2	-4.017240

PBX1.p2	4.008166
MYBL2.p2	3.965618
HAND1,2.p2	-3.960443
TCF4_dimer.p2	-3.950222
GGAAUGU	3.914362
STAT2,4,6.p2	3.882966
AACACUG	3.866572
UUUUUGC	-3.861829
SPIB.p2	3.848118
LHX3,4.p2	-3.825488
GAGAUGA	3.706771
BPTF.p2	3.694402
AAGUGCU	3.625894
NHLH1,2.p2	3.563439
ESR1.p2	-3.563007
POU6F1.p2	3.556815
SRY.p2	-3.501360
FOX{D1,D2}.p2	3.460886
NR1H4.p2	-3.444054
AGUGGUU	3.428867
UUGGCAC	3.428561
POU1F1.p2	-3.427446
AHR_ARNT_ARNT2.p2	-3.402815
ACAGUAU	3.396875
SOX17.p2	3.388932
ZIC1..3.p2	3.380114
EN1,2.p2	-3.358870
PRRX1,2.p2	3.351237
GATA6.p2	3.313266
AGCACCA	3.258944
AGCAGCG	-3.188248
PAX4.p2	3.186388
SOX5.p2	-3.159007
STAT1,3.p3	3.158086
YY1.p2	-3.150770
NFKB1_REL_RELA.p2	3.142962
UUGGCAA	3.116745
CACAGUG	3.073683
SPZ1.p2	3.050127
NKX2-3_NKX2-5.p2	-3.014212
TFAP2B.p2	-2.951623
AGCAGCA	2.933069
GUAACAG	2.918261
HOXA9_MEIS1.p2	-2.883912
ZBTB16.p2	-2.880101
UCACAUU	-2.879964
CUUUGGU	2.874260
RBPJ.p2	-2.843359
UUGGUCC	2.837260
ETS1,2.p2	-2.824302
HES1.p2	-2.801785
MSX1,2.p2	-2.768325
NRF1.p2	2.764340
EHF.p2	-2.747095
GUAGUGU	2.740297
AAGGUGC	2.704792

TBP.p2	2.627495
UUGUUCG	2.612341
ACAUUCA	2.609181
HNF1A.p2	-2.574668
RORA.p2	-2.552431
CDC5L.p2	-2.543031
UCAAGUA	2.485514
EP300.p2	2.478716
TBX4,5.p2	2.452506
SOX{8,9,10}.p2	2.443552
CREB1.p2	-2.429477
JUN.p2	2.416194
CEBPA,B_DDIT3.p2	-2.399835
FOX{F1,F2,J1}.p2	-2.253559
HIF1A.p2	-2.245699
GCAGCAU	2.223872
RXRA_VDR{dimer}.p2	-2.208193
PATZ1.p2	-2.199593
GFI1.p2	2.060710
ZNF148.p2	-2.056085
MYB.p2	2.050658
UGCAUAG	2.024762
AGGUAGU	1.984808
POU3F1..4.p2	-1.981746
CDX1,2,4.p2	-1.969056
LEF1_TCF7_TCF7L1,2.p2	1.949589
ZBTB6.p2	-1.941322
FOXP3.p2	-1.914797
TFCP2.p2	1.888318
EOMES.p2	1.885703
MEF2{A,B,C,D}.p2	-1.881713
TAL1_TCF{3,4,12}.p2	-1.876640
AACAGUC	1.861035
CAGUGCA	1.830254
CCUUCAU	1.823326
NANOG{mouse}.p2	-1.809389
UCCCUUU	1.803226
GUGCAAA	1.799525
BACH2.p2	-1.778019
GGCUCAG	1.772541
CTCF.p2	1.734880
GGCAAGA	1.734210
EBF1.p2	-1.724006
GCUACAU	1.657170
ZFP161.p2	1.629917
NFY{A,B,C}.p2	1.620166
HMX1.p2	-1.605599
FOXL1.p2	1.586486
AUGGCUU	1.583830
CCAGCAU	-1.563745
NFIX.p2	-1.550922
T.p2	-1.545058
ONECUT1,2.p2	-1.463027
AGCUGCC	-1.446600
FOXD3.p2	-1.313682
NKX2-2,8.p2	1.311760

MAZ.p2	1.311673
PAX6.p2	-1.307375
AACCUGG	-1.306978
AUGGCAC	1.279688
ARID5B.p2	-1.251386
UAAGACG	1.223018
CCAGUGU	1.199163
AIRE.p2	1.197368
ZNF143.p2	-1.163473
ADNP_IRX_SIX_ZHX.p2	1.112200
CAGCAGG	-1.112018
AGUGCAA	1.106508
GCUGGUG	1.105037
AAUCUCU	-1.096328
DMAP1_NCOR{1,2}_SMARC.p2	-1.088310
FOXO1,3,4.p2	1.083276
UAAGACU	1.079116
HSF1,2.p2	1.069167
ELF1,2,4.p2	1.041648
NKX2-1,4.p2	-1.019541
RXRG_dimer.p3	0.998680
STAT5{A,B}.p2	0.989426
AR.p2	-0.987129
HOX{A4,D4}.p2	-0.963680
ACUGCAU	-0.961381
CGUGUCU	-0.915790
FOXA2.p3	-0.889308
UAAUGCU	0.882457
ATF4.p2	-0.880505
LMO2.p2	-0.852996
CUACAGU	-0.851979
TLX2.p2	0.838461
AUGACAC	-0.836765
NR3C1.p2	0.830968
GGAGUGU	-0.802679
ACCCUGU	-0.777607
NFIL3.p2	-0.769865
ACAGUAC	0.769100
PAX2.p2	0.761529
NFE2.p2	0.736988
ALX1.p2	-0.736058
FOXQ1.p2	-0.692554
GGAAGAC	0.671442
CGUACCG	0.656731
GAUCAGA	0.656140
PDX1.p2	0.650568
HIC1.p2	0.635679
EGR1..3.p2	0.629637
GGACGGA	-0.612295
ACUGGCC	-0.610809
POU2F1..3.p2	-0.610074
AAUGCCC	-0.584786
GATA1..3.p2	0.581988
NFATC1..3.p2	0.581690
SMAD1..7,9.p2	-0.561780
NKX3-2.p2	0.551091

NKX6-1,2.p2	0.543763
NR5A1,2.p2	-0.534370
ATF2.p2	-0.522030
FOX{C1,C2}.p2	0.519247
GZF1.p2	0.514649
AGCUUAU	0.509781
PPARG.p2	0.485195
IRF1,2,7.p3	-0.472821
VSX1,2.p2	0.470825
EWSR1-FLI1.p2	-0.452944
UAUUGCU	-0.435356
E2F1..5.p2	0.425908
UGAAAUG	-0.412954
UGCAUUG	-0.406212
AAUCUCA	-0.403747
AUGUGCC	0.398050
UGUGCGU	-0.380793
GFI1B.p2	-0.379048
AAAGUGC	-0.378342
SOX2.p2	-0.378160
ELK1,4_GABP{A,B1}.p3	0.340488
GAGAACU	0.334635
CCCUGAG	0.306717
GTF2A1,2.p2	-0.297972
AACGGAA	-0.286985
MZF1.p2	0.286773
GLI1..3.p2	0.268608
GGCAGUG	-0.258552
GUCAGUU	-0.257585
TP53.p2	-0.220073
PRDM1.p3	-0.216004
RREB1.p2	-0.206545
AACCGUU	-0.202728
AAGGCAC	0.200079
EVI1.p2	0.198591
KLF4.p3	-0.184265
HOX{A5,B5}.p2	0.176179
PAX8.p2	-0.161384
FOSL2.p2	0.138129
POU5F1.p2	0.137880
AUUGCAC	-0.119868
PAX3,7.p2	-0.115442
NANOG.p2	0.104343
ATF5_CREB3.p2	0.094706
GAUAUGU	-0.084700
SPI1.p2	-0.081698
TLX1..3_NFIC{dimer}.p2	0.080810
UGACCUA	0.076086
MYFfamily.p2	0.069070
FOS_FOS{B,L1}_JUN{B,D}.p2	0.068486
HMGA1,2.p2	0.055254
CUX2.p2	-0.036901
RUNX1..3.p2	-0.033572
SP1.p2	-0.023642
ACCCGUA	0.014366

Table-S3. Motif activity changes derived from mouse EMT models (sorted by absolute z-scores).

Motif	Z-Score
HNF1A.p2	-18.319605
TFAP2{A,C}.p2	13.139745
HNF4A_NR2F1,2.p2	-10.316133
SNAI1..3.p2	-9.802821
NFY{A,B,C}.p2	-9.260140
NFIL3.p2	-9.153861
PATZ1.p2	9.037959
GTF2A1,2.p2	8.129773
E2F1..5.p2	-8.107501
XBP1.p3	7.993492
SOX{8,9,10}.p2	7.578368
JUN.p2	7.391577
ESRRA.p2	-7.092741
ZEB1.p2	-6.970979
ZFP161.p2	-6.773266
RUNX1..3.p2	6.677799
GAGGUAG	6.640954
TFCP2.p2	6.502995
RXR{A,B,G}.p2	6.232599
HBP1_HMGB_SSRP1_UBTF.p2	6.211917
ZBTB16.p2	-6.189857
GATA1..3.p2	6.074720
GUGCAAA	5.968291
SMAD1..7,9.p2	5.715875
HES1.p2	-5.703150
GUAAACA	5.658502
DBP.p2	-5.514266
FOXA2.p3	-5.288262
FOSL2.p2	5.275215
EN1,2.p2	-5.001922
STAT2,4,6.p2	4.958305
YY1.p2	-4.954490
AGCACCA	4.928008
SOX2.p2	4.919780
PRDM1.p3	4.820454
TLX2.p2	4.812390
MYB.p2	-4.762636
ATF6.p2	4.710928
bHLH_family.p2	4.694504
ZNF238.p2	4.656929
PAX2.p2	-4.622154
RFX1..5_RFXANK_RFXAP.p2	-4.531748
UAUUGCU	4.389541
MYFfamily.p2	4.378652
KLF12.p2	4.358762
ZNF423.p2	4.227038
NR1H4.p2	4.225841
LMO2.p2	-4.188487
NFATC1..3.p2	4.178978
HOX{A6,A7,B6,B7}.p2	-4.123296

GLI1..3.p2	4.085050
AGCUGCC	-4.051388
ESR1.p2	3.971369
MEF2{A,B,C,D}.p2	3.903434
NKX2-2,8.p2	3.857453
IRF1,2,7.p3	3.845818
FOXL1.p2	3.817527
NR5A1,2.p2	3.787010
ACAGUAC	3.741890
NFE2L2.p2	3.715022
CEBPA,B_DDIT3.p2	3.686488
POU1F1.p2	-3.601421
EBF1.p2	3.576447
PAX5.p2	-3.448711
GFI1.p2	-3.441597
CCCUGAG	3.391263
ZBTB6.p2	-3.359916
ZIC1..3.p2	3.355679
AAGUGCU	-3.342459
AGGUAGU	-3.290403
AUGGCAC	3.285942
UUGUUCG	-3.273784
CCAGUGU	3.246929
FEV.p2	3.164132
UCACAUU	-3.154298
AGUGCAA	-3.125763
UUGGUCC	3.098766
EGR1..3.p2	-3.096638
AGCAGCA	3.094618
SPIB.p2	3.082550
FOX{I1,J2}.p2	-3.035529
UGUGCUU	2.945159
TFDP1.p2	2.916955
RORA.p2	-2.858638
GGCAGUG	2.851382
TLX1..3_NFIC{dimer}.p2	-2.774303
NANOG{mouse}.p2	-2.761615
TFAP2B.p2	2.759048
BPTF.p2	-2.699265
EP300.p2	2.671825
EOMES.p2	-2.609712
CTCF.p2	2.570247
NKX3-2.p2	-2.518638
IKZF1.p2	-2.497992
ARID5B.p2	2.487411
CREB1.p2	-2.472282
AUUGCAC	-2.444454
REST.p3	2.442849
GGAAUGU	2.441918
MTF1.p2	2.431261
SOX5.p2	-2.409517
FOX{D1,D2}.p2	2.387016
AAGGUGC	2.379286
NKX2-3_NKX2-5.p2	2.367255
ALX1.p2	-2.352391
UGCAUAG	2.323399

ZNF148.p2	-2.252424
UCACAGU	-2.232014
HIC1.p2	2.222399
PBX1.p2	-2.212457
ATF5_CREB3.p2	-2.152149
KLF4.p3	-2.127946
ACAGUAU	-2.127841
EHF.p2	-2.110093
IKZF2.p2	2.100513
GAUUGUC	-2.083073
DMAP1_NCOR{1,2}_SMARC.p2	-2.044785
SPZ1.p2	2.036622
MYOD1.p2	-2.020948
AAUACUG	2.010256
FOX{F1,F2,J1}.p2	1.951040
STAT5{A,B}.p2	1.944012
UGAAAUG	-1.941875
ELK1,4_GABP{A,B1}.p3	-1.931232
LHX3,4.p2	1.906492
SRY.p2	1.894688
ZNF384.p2	-1.824031
AGUGGUU	1.818739
MYBL2.p2	1.784815
GATA6.p2	-1.762058
TFAP4.p2	1.760603
TAL1_TCF{3,4,12}.p2	1.682352
AGCUUAU	-1.675293
ONECUT1,2.p2	-1.633590
UCAAGUA	1.596472
FOX{C1,C2}.p2	-1.549272
FOXD3.p2	1.515649
UUGGCAA	-1.504195
HLF.p2	1.501009
HMX1.p2	-1.489057
ADNP_IRX_SIX_ZHX.p2	-1.482936
UCCAGUU	-1.457275
SRF.p3	1.452882
PRRX1,2.p2	1.451130
NANOG.p2	1.449412
NKX6-1,2.p2	1.447207
NFIX.p2	-1.431007
AHR_ARNT_ARNT2.p2	-1.397611
ZNF143.p2	-1.383203
ACUGGCC	-1.362517
HSF1,2.p2	-1.360968
UCCCUUU	-1.357800
TBX4,5.p2	-1.348500
NFE2L1.p2	1.343626
GCAGCAU	1.331934
CAGUGCA	1.306031
STAT1,3.p3	1.298060
AACAGUC	1.292869
HAND1,2.p2	1.287578
UUUUUGC	-1.280302
GFI1B.p2	-1.267603
FOXQ1.p2	-1.246515

GZF1.p2	1.202089
ELF1,2,4.p2	1.198595
NRF1.p2	-1.197642
T.p2	1.190501
GGACGGA	1.116742
CUX2.p2	-1.071701
HOXA9_MEIS1.p2	-1.064329
LEF1_TCF7_TCF7L1,2.p2	1.056026
TGIF1.p2	-1.034514
UAAUGCU	-0.996918
HOX{A4,D4}.p2	-0.983296
POU5F1.p2	-0.980397
PPARG.p2	-0.965255
AIRE.p2	0.924218
GUAACAG	0.910750
NFKB1_REL_RELA.p2	-0.905472
NR6A1.p2	0.882506
CDC5L.p2	-0.862852
TEAD1.p2	0.852762
GGAGUGU	-0.843095
FOXP3.p2	-0.838785
AUGACAC	-0.812233
GCUGGUG	0.807405
CUCCCAA	0.798569
RXRA_VDR{dimer}.p2	0.780997
AR.p2	0.736254
BACH2.p2	-0.736097
AACCUGG	-0.734849
TP53.p2	-0.723473
UAAGACU	-0.715628
SPI1.p2	0.704271
GUAGUGU	-0.702330
RREB1.p2	-0.698246
HOX{A5,B5}.p2	0.693836
GTF2I.p2	0.678249
MSX1,2.p2	0.626156
POU5F1_SOX2{dimer}.p2	0.624742
AAUCUCU	-0.614284
UGCAUUG	-0.601594
ATF2.p2	-0.587577
MAZ.p2	-0.569641
GCUACAU	-0.562374
UAAGACG	-0.562314
AGCAGCG	0.559258
AUGGCUU	0.542322
CGUGUCU	0.538620
GUCAGUU	0.505634
FOS_FOS{B,L1}_JUN{B,D}.p2	0.498148
CRX.p2	-0.489037
ACUGCAU	-0.478429
POU6F1.p2	-0.463786
CCAGCAU	0.459199
TCF4_dimer.p2	0.445521
MAFB.p2	0.424878
AACACUG	-0.405680
TFEB.p2	-0.371070

CDX1,2,4.p2	-0.369270
TBP.p2	0.349531
CGUACCG	-0.336247
SP1.p2	0.333897
GAGAUGA	0.331195
PAX6.p2	-0.323456
EWSR1-FLI1.p2	-0.319026
GGCUCAG	0.318037
AACGGAA	0.309046
POU2F1..3.p2	0.299161
AAUCUCA	-0.293698
GAUCAGA	0.292939
CACAGUG	0.279905
PAX8.p2	-0.271819
ATF4.p2	-0.259694
GAUAUGU	0.255698
AAGGCAC	-0.242607
HIF1A.p2	-0.238108
PAX4.p2	0.236628
CUUUGGU	-0.231595
NKX2-1,4.p2	-0.226075
NKX3-1.p2	0.225196
UGACCUA	-0.218150
PDX1.p2	0.209241
GGCAAGA	-0.203007
PAX3,7.p2	-0.200453
POU3F1..4.p2	0.190423
MZF1.p2	0.186999
FOXO1,3,4.p2	-0.166946
GGAAGAC	0.161723
AACCGUU	-0.161281
GAGAACU	-0.154159
PITX1..3.p2	-0.139404
VSX1,2.p2	-0.132444
SOX17.p2	-0.122967
UGUGCGU	0.122439
NFE2.p2	0.119513
ARNT_ARNT2_BHLHB2_MAX_MYC_USF1.p2	-0.105121
ACCCUGU	-0.094109
UUGGCAC	0.088067
EVI1.p2	-0.087064
CUACAGU	0.079026
RXRG_dimer.p3	-0.061463
RBPJ.p2	-0.057384
AAAGUGC	-0.054632
ACCCGUA	-0.043866
ACAUUCA	0.043493
NHLH1,2.p2	-0.033172
HMGA1,2.p2	-0.032751
ETS1,2.p2	0.031012
AAUGCCC	-0.025423
CCUUCAU	0.019178
CAGCAGG	0.016615
SREBF1,2.p2	0.010177
FOXN1.p2	-0.009457
NR3C1.p2	-0.006974

Table-S4. Fisher test of EMT signature genes represented in differentially expressed genes in response to TFAP2A overexpression.

EMT Signature Set[5]/Tfap2a	EMT Sign. Genes: Changing	EMT Sign. Genes: Not Changing	All Genes: Changing	All Genes: Not Changing	Pval
Up/Up	18	42	2007	12798	0.0007559
Up/Down	6	55	1538	13267	0.6181231
Down/Down	19	87	1538	13267	0.0130989
Down/Up	27	79	2007	12798	0.0008046

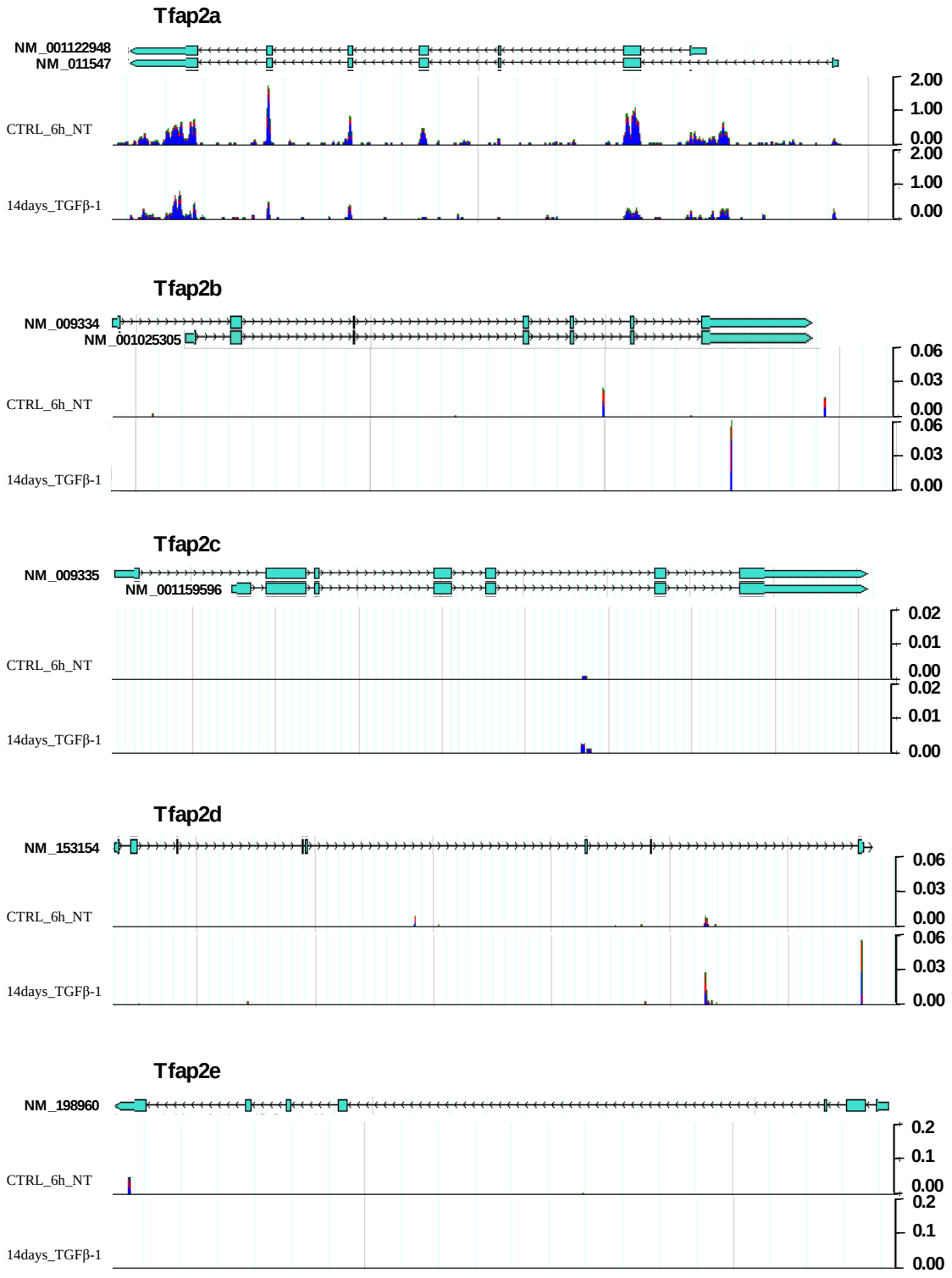

Figure-S1: Expression Profile of AP-2 family members. mRNAseq libraries from the following conditions NMuMG CTRL cells or cells treated for 14 days with growth factor, were generated and read coverage of the genomic region spanning the genes from the family of AP-2 transcription factors is shown in a mouse genome

browser (www. clipz.unibas.ch [8]). High densities of the reads (reads per million unit) are only present for the Tfap2a gene, while sporadic reads can be assigned to the other family members suggesting they have no or little expression. Mapping, annotation and visualization of deep-sequencing data was done with the CLIPZ server [8].

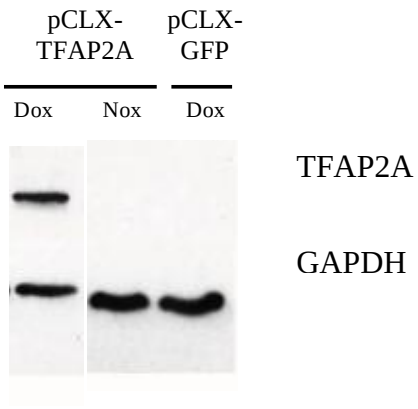

Figure-S2: Overexpression of TFAP2A protein is detected by Western Blot.

Lysates from NMuMG cells stably transduced with pCLX-TFAP2A or pCLX-GFP treated for 72 hours with 2 μg/mL doxycycline (Dox) or not (Nox) were probed for TFAP2A expression by WB. GAPDH is used as normalization control. Overexpression of TFAP2A is detectable only in the pCLX-TFAP2A doxycycline-induced cells.

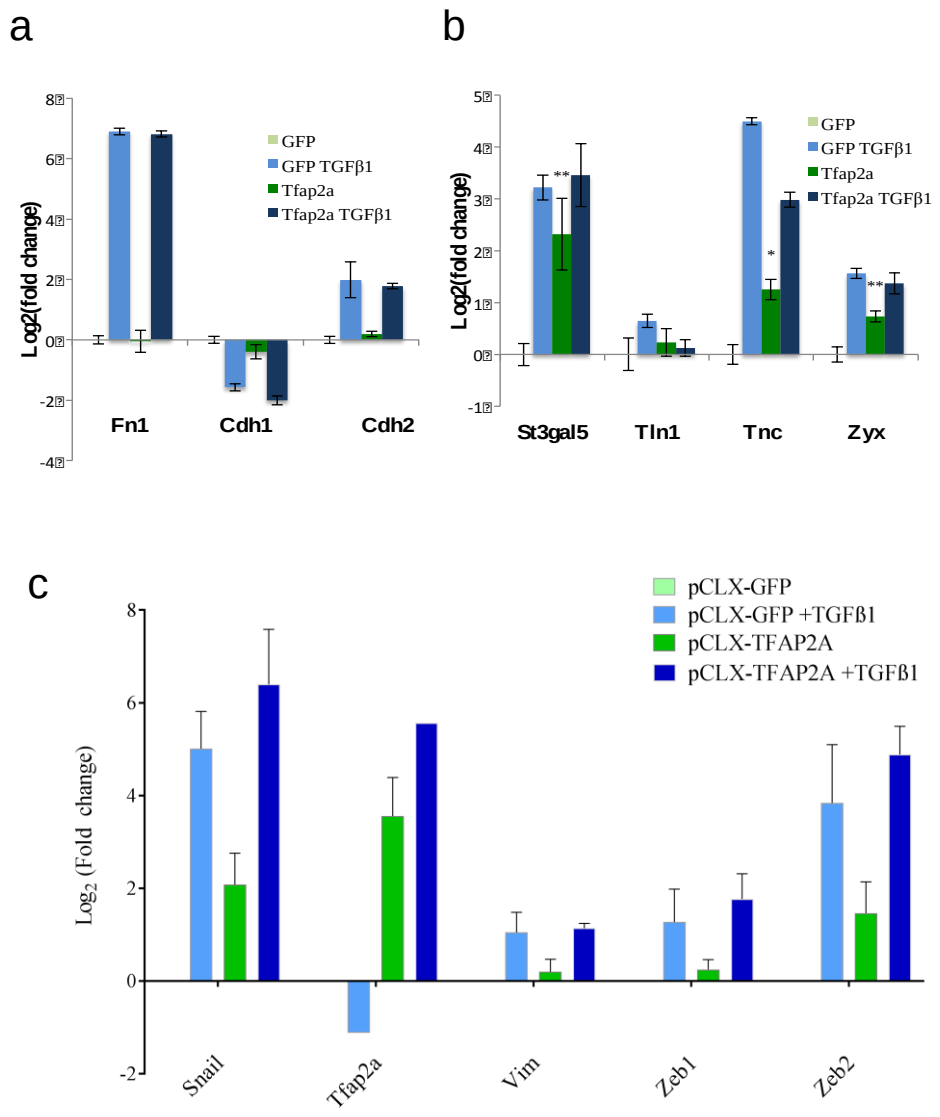

Figure-S3: TFAP2A overexpression in NMuMG modulates epithelial plasticity.

Quantitative RT-PCR of EMT markers on extracts from the doxycycline induced pCLX-TFAP2A or pCLX-GFP cell lines, treated or not with 2 ng/mL of TGFβ1 for 72 hours. The results represented in the figure are the mean values from three experiments for cells non-treated with growth factor (light and dark green) and two independent experiments in the case of TGFβ1-induced samples. a) EMT markers that do not show differential expression upon TFAP2A induction. b) Set of genes (Stg3gal5, Tln1, Tnc, Zyx) involved in focal adhesion that are unregulated in both

TGFβ1-induced EMT and upon TFAP2A induction. In addition, the St3gal5 gene that was recently involved in cell adhesion downstream of TGFβ1 signaling and Zeb1 is also significantly changing in both conditions [9]. The genes that displayed significant enrichment in the TFAP2A cell line versus the GFP cell line, as estimated from the performed one-tail paired *t*-test, are indicated with asterix (*) for $p<0.05$ and (**) for $p < 0.01$, respectively. (c) Quantitative RT-PCR of EMT markers (TFs) in extracts from the doxycycline-induced pCLX-GFP and pCLX-TFAP2A cell lines, treated or not with 2 ng/mL of TGFβ1 for 72 hours. Shown are the means from three experiments on cells not treated with growth factor (light and dark green) and two independent experiments in TGFβ1-induced cells. TFAP2A expression in TGFβ1-induced cells was only measured once. TFAP2A overexpression is apparent in both TFAP2A-induced samples (dark green and dark blue) but is not induced in cell treated with TGFβ1 alone (light blue). The EMT-inducing TFs have increased expression upon TFAP2A induction.

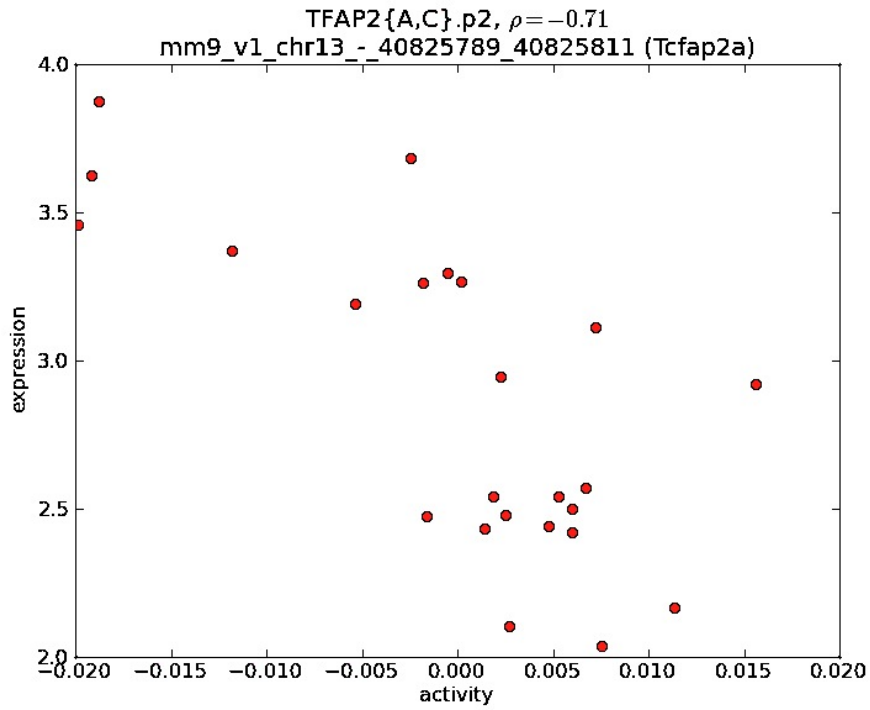

Figure-S4: Correlation between TFAP2A/C activity and mRNA expression levels during EMT time course. mRNA-seq samples in two replicates form independent wells were prepared from a time course of NMuMG cells treated for 14 days with 2 ng/mL of TGFβ1, and the data was consequently analyzed with MARA [10]. The figure emphasizes the correlation between TFAP2A/C transcriptional activity and mRNA expression levels during the time course.

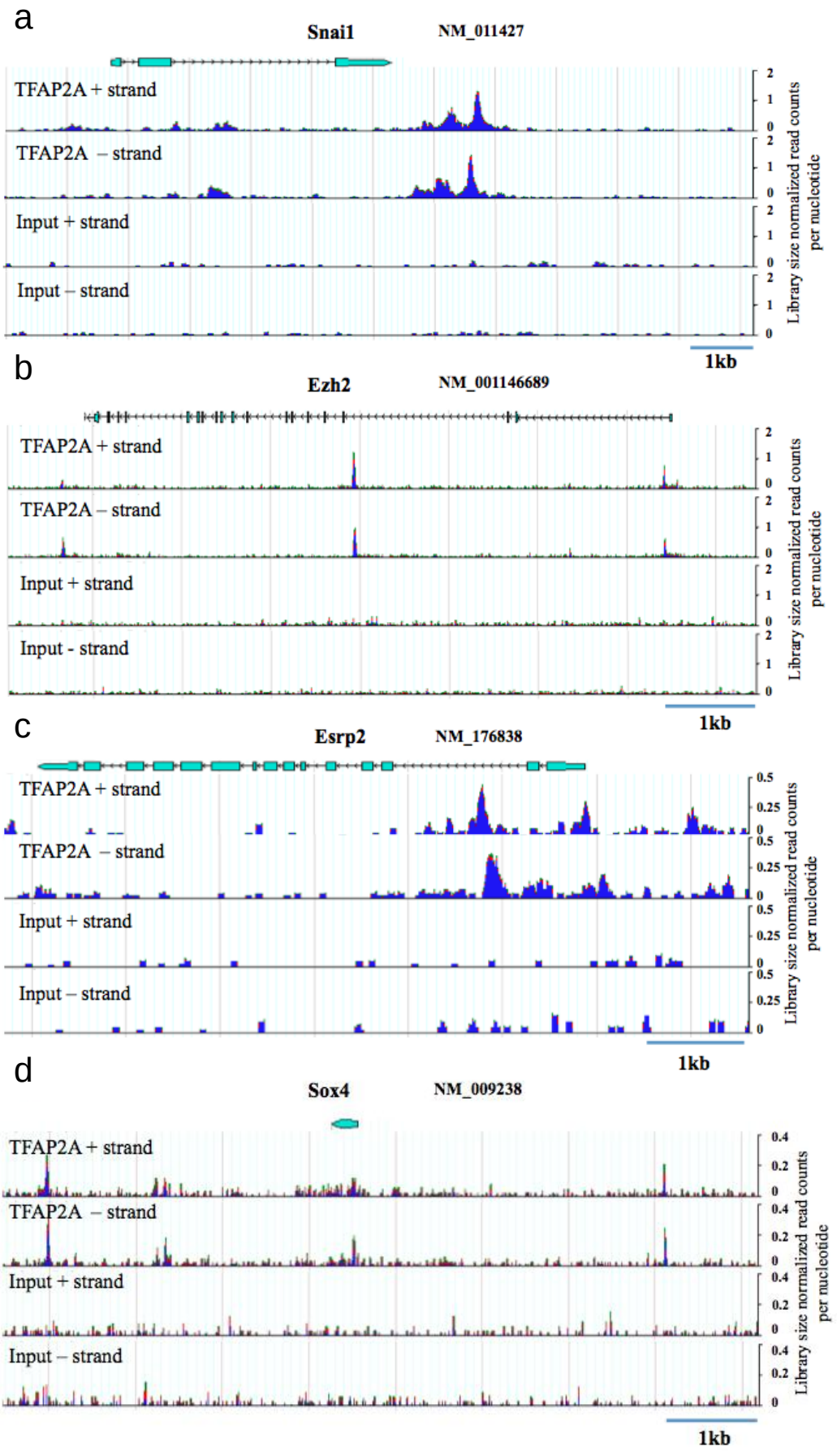
a
Snai1
NM_011427
TFAP2A + strand
TFAP2A – strand
Input + strand
Input – strand
Library size normalized read counts per nucleotide
1kb
b
Ezh2
NM_001146689
TFAP2A + strand
TFAP2A – strand
Input + strand
Input - strand
Library size normalized read counts per nucleotide
1kb
c
Esrp2
NM_176838
TFAP2A + strand
TFAP2A – strand
Input + strand
Input – strand
Library size normalized read counts per nucleotide
1kb
d
Sox4
NM_009238
TFAP2A + strand
TFAP2A – strand
Input + strand
Input – strand
Library size normalized read counts per nucleotide
1kb

Figure-S5: TFAP2A binds directly to crucial regulars of

EMT. ChIP-seq libraries from TFAP2A ChIP or Input were generated and the coverage of the genomic region spanning Snai1 (a), Ezh2 (b), Esrp2 (c), Sox4 (d) genes by reads is shown in a mouse genome browser (www.clipz.unibas.ch and [8]). The results of a representative experiment are presented. Mapping, annotation and visualization of deep-sequencing data was done with the ClipZ server [8].

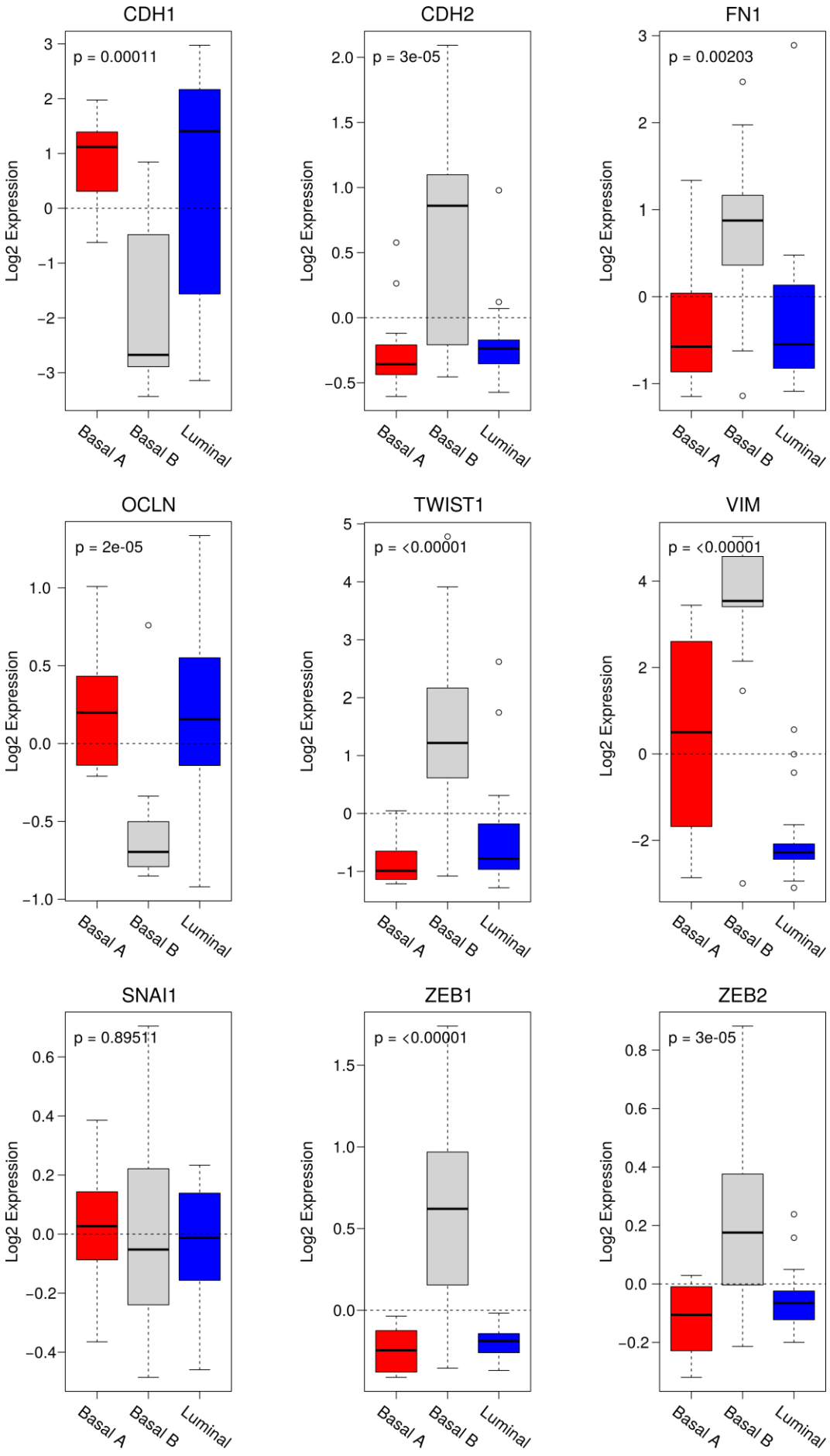
CDH1
p = 0.00011
CDH2
p = 3e-05
FN1
p = 0.00203
OCLN
p = 2e-05
TWIST1
p = <0.00001
VIM
p = <0.00001
SNAI1
p = 0.89511
ZEB1
p = <0.00001
ZEB2
p = 3e-05
Log2 Expression
Basal A
Basal B
Luminal

Figure-S6: EMT marker expression in breast cancers.

Box plots of marker gene expression in a panel of breast cancer cell lines grouped in the basal A (red), basal B (grey) and luminal (blue) subgroups based on the annotation from Neve et al. [11] All plots were generated with the GOBO online tool [12].

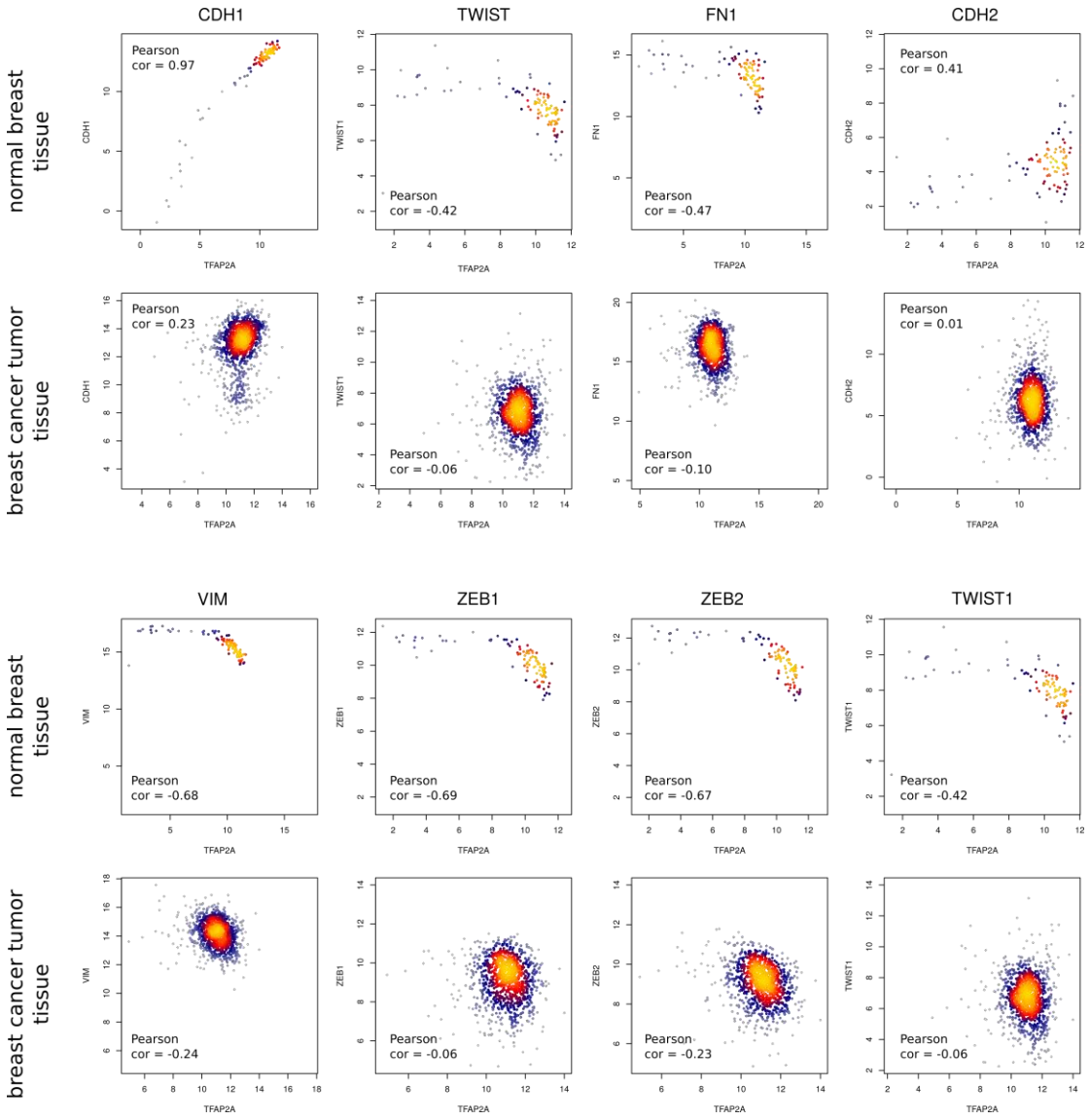

Figure-S7: Correlation of log2 expression values of TFAP2A and EMT marker genes in normal and tumor breast tissues.

The plots show the correlation of TFAP2A expression with the expression of the indicated epithelial and mesenchymal markers in normal breast tissue samples (n = 98) and of breast cancer samples (n = 1080). Normalized expression values of the indicated genes were obtained from The Cancer Genome Atlas project in Breast Invasive Carcinoma (TCGA-BRCA), more specifically from the Broad Institute TCGA Genome Data Analysis Center (2016): Analysis-ready

standardized TCGA data from Broad GDAC Firehose 2016_01_28 run. Broad Institute of MIT and Harvard. Dataset. https://doi.org/10.7908/C11G0KM9).

Supplementary methods

Analysis of mRNA-Sequencing data

Mapping and annotation of sequencing reads was done using the CLIPZ webserver [8]. Differential gene expression analysis was conducted using the Bioconductor DESeq package [13], whereat genes with an adjusted p-value < 0.05 were considered differentially expressed. Gene ontology (GO) analysis of differentially expressed genes was done with the topGO package [14] using the ”weight01” algorithm with a node size of one and ”Fisher” statistics.

Analysis of EMT signature genes enrichment

The differentially expressed genes between NMuMG cells that overexpress TFAP2A and NMuMG cells that overexpress GFP were compared to the EMT signature set of up- or downregulated genes [5]. Those with an adjusted p-value < 0.05 and following the defined direction of modulation (Up or Down) were considered as changing, whereas those with an adjusted p-value > 0.05 or not following the defined trend were defined as non-changing. In this manner four categories of genes sets comparing the signature gene set vs the genes differentially expressed upon TFAP2A overexpression were created Up/Up, Up/Down, Down/Down and Down/Up and for each category the number of genes Changing and Not-Changing is calculated. A Fisher statistics was used to calculate the enrichment of the genes in each of the four categories as

compared to all genes up- or downregulated upon TFAP2A overexpression (http://www.quantitativeskills.com/sisa/statistics/fisher.htm), the one sided p-value p(O>=E) is represented.

REFERENCES

1 Kreitzer, F. R. *et al.* A robust method to derive functional neural crest cells from human pluripotent stem cells. *Am J Stem Cells* **2**, 119-131 (2013).

2 Feuerborn, A. *et al.* The Forkhead factor FoxQ1 influences epithelial differentiation. *J Cell Physiol* **226**, 710-719, doi:10.1002/jcp.22385 (2011).

3 Tiwari, N. *et al.* Klf4 is a transcriptional regulator of genes critical for EMT, including Jnk1 (Mapk8). *PLoS One* **8**, e57329, doi:10.1371/journal.pone.0057329 (2013).

4 Evseenko, D. *et al.* Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proc Natl Acad Sci U S A* **107**, 13742-13747, doi:10.1073/pnas.1002077107 (2010).

5 Taube, J. H. *et al.* Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci U S A* **107**, 15449-15454, doi:10.1073/pnas.1004900107 (2010).

6 Diepenbruck, M. *et al.* Tead2 expression levels control the subcellular distribution of Yap and Taz, zyxin expression and epithelial-mesenchymal transition. *J Cell Sci* **127**, 1523-1536, doi:10.1242/jcs.139865 (2014).

7 Brunskill, E. W. *et al.* A gene expression atlas of early craniofacial development. *Dev Biol* **391**, 133-146, doi:10.1016/j.ydbio.2014.04.016 (2014).

8 Khorshid, M., Rodak, C. & Zavolan, M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res* **39**, D245-252, doi:10.1093/nar/gkq940 (2011).

9 Mathow, D. *et al.* Zeb1 affects epithelial cell adhesion by diverting glycosphingolipid metabolism. *EMBO Rep* **16**, 321-331, doi:10.15252/embr.201439333 (2015).

10 Balwierz, P. J. *et al.* ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res* **24**, 869-884, doi:10.1101/gr.169508.113 (2014).

11 Neve, R. M. *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-527, doi:10.1016/j.ccr.2006.10.008 (2006).

12 Ringner, M., Fredlund, E., Hakkinen, J., Borg, A. & Staaf, J. GOBO: gene expression-based outcome for breast cancer online. *PLoS One* **6**, e17911, doi:10.1371/journal.pone.0017911 (2011).

13 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
14 topGO: Enrichment analysis for Gene Ontology. (2010).

Discussion and Perspectives

During embryonic development cells with initially epithelial charachteristics gain mesenchymal and migratory capacities, due to phenotypical plasticity events (Thiery, Acloque et al. 2009). Epithelial to mesenchymal transition is involved in normal development; wound healing or pathological conditions such as tumor progression, metastasis and invasiveness, and fibrosis (Lamouille, Xu et al. 2014). A highly complex program at transcriptional, post-transcriptional and post-translational level ensures that cells expressing epithelial features will acquire mesenchymal phenotype during the transition (Tiwari, Gheldof et al. 2012; De Craene and Berx 2013).

With the aim to identify conserved regulatory modules that operate in both cancer and normal development, we analyzed and compared several transcriptomics datasets from mouse and human EMT models, including neural crest differentiation, mesoderm specification and breast cancer. We used MARA (Motif Activity Response Analysis), an online tool that models the predicted regulatory sites for transcription factors (motifs) that explain the measured genome-wide expression changes in the input datasets (https://ismara.unibas.ch/fcgi/mara) (Balwierz, Pachkov et al. 2014). Further, we have constructed motif-motif interaction networks of both mouse and human EMT models. Next, we identified that TFAP2A/C motif increased activity is conserved between mouse and human and it has a central place in both EMT motif–motif interaction networks. Moreover, the interaction between TFAP2A/C and ZEB motifs is also maintained in between the two mammalian networks. Consequently, to validate the predicted observations we used NMuMG cells; mouse breast cancer cell line that is widely used as an *in vitro* model of TGFβ induced EMT. Performing MARA analysis on transcriptomics data from this model we confirmed that TFAP2A/C motif changes are reproducible and in line with the previous observations. In addition, in the course of the transition the expression levels of *Tfap2a* mRNA are negatively correlated with the TF predicted activity change. What is more, we have demonstrated that TFAP2A directly interacts with *Zeb2* promoter *in vitro* as well as in NMuMG cells. The overexpression of TFAP2A in NMuMG cells, followed by its increased activity, is translated in altered cells epithelial phenotype. TFAP2A

induced: i) changes in cellular morphology and cytoskeleton rearrangements; ii) a decrease in cellular adhesion, iii) increased mRNA levels of EMT master regulator TFs and iv) elevated expression of genes implicated in focal adhesion. Overall our results support a role of TFAP2A as a general modulator of EMT, which promotes the transition and it is probably involved in the initial step of the process. Those results are in a contradiction with recently published reports suggesting that TFAP2A and its homologue TFAP2C both govern the luminal cell phenotype and hence the epithelial cell state in breast cancer. Both studies from Weigel laboratory demonstrate that TFAP2C and/or TFAP2A modulate the expression of certain luminal genes, such as *CD44* and ESRα (Bogachek, Chen et al. 2014; Cyr, Kulak et al. 2014). In addition, they show that TFAP2A activation by sumoylation inhibition abrogates the tumor formation capacities of breast cancer cell lines in xenografts, and leads to an increase in $CD24^{hi}/CD44^{low}$ cells population (Bogachek, Chen et al. 2014). However the $CD24^{low}/CD44^{high}$ phenotype and tumor formation capacity of breast cancer cell lines are not directly related with EMT (Ocana, Corcoles et al. 2012). A recent study identified that increased expression of PRRX1 is a potent EMT inducer in cancer, while in the same time leading to reduction of the cancer stem cells population ($CD24^{low}/CD44^{high}$) thus uncoupling cancer stem cells formation and EMT (Ocana, Corcoles et al. 2012). A loss of function/ gain of function study of the roles of TFAP2A and TFAP2C in cancer cell lines confirmed that both TFs reduced the proliferation, but instead increased cells migration and invasion properties (Orso, Penna et al. 2008). Furthermore, TFAP2A knockdown led to increased adhesion, while the TF overexpression had the inverse effect (Orso, Penna et al. 2008). Therefore the results described by Weigel and colleagues can be mostly explained by the TFAP2A and TFAP2C effect on proliferation and depletion in cancer stem cell population, rather than directly relating them to the EMT phenotype. Furthermore our observations are also in line with previously described role of TFAP2A in neural crest specification, where it controls the expression of EMT, inducing transcription factors such as SNAIL and ZEB (Rada-Iglesias, Bajpai et al. 2012).

In a perspective, TF luciferase reporters will be used to further confirm the predicted changes of TFAP2A activity. The most intriguing part of our results is the negative correlation between the mRNA levels of *Tfap2a* and its activity, suggesting that the TF either functions predominantly as repressor or that a negative auto-regulatory feed-back loop is operational. In both cases the luciferase reporters will be

instrumental to elucidate the precise mode of action of TFAP2A. In the case of auto-regulatory loop one can further hypothesize that post-transcriptional or post-translational mechanisms activate TFAP2A, to allow it directly or indirectly inhibit its own expression. Different splicing isoforms of TFAP2A were shown to have variations in transcriptional activity, as well as post-translational modifications, including ubiquitylation and phosphorylation, which are known to modulate its activity (Garcia, Campillos et al. 1999; Bogachek, Chen et al. 2014). For example, a K10R mutant of TFAP2A might be overexpressed in parallel together with wild-type TFAP2A in NMuMG cells and consequently observe for any phenotypical differences. Proteomic analysis by mass spectrometry can identify phosphorylated residues in the transcription factor and their relative proportion can be estimated with high precision using Single Reaction Monitoring/Multiple Reaction Monitoring (SRM/MRM) label-free method (Wolf-Yadlin, Hautaniemi et al. 2007). In addition, the link between TFAP2A expression levels and NMuMG adhesion properties needs to be studied in more detail, in order to confirm that the genes significantly affected at mRNA levels are also changing at protein level. Next, the siRNA-induced depletion of TFAP2A and the consequent effect on adhesion and EMT-TF in the presence or absence of TGFβ also needs to be addressed. Last but not least, results obtained in NMuMG cells needs to be validated in other EMT models ideally of both mouse and human and thus to confirm the general applicability of the model.

A similar computational approach was applied to elucidate the mechanism by which miR-290-295 cluster functions in maintenance and differentiation of ESCs (Leonardo, Schultheisz et al. 2012). To better define the direct targets of miR-290-295 microRNAs, we have carried out an extensive analysis of data sets derived from mESCs that either expressed or were deficient in expression of miR-290-295 cluster miRNAs. In this manner, we determined direct and reproducible transcription factor targets of the miRNAs that mediate the effects of these miRNAs in pluripotency. An initial computational analysis of these data predicted a set of miR-290 transcription factors that might be involved in the differentiation processes. The computationally predicted targets were validated with luciferase reporters in a mouse cell line. Finally, we demonstrated the expression variation of IRF2 in response to miRNAs depletion in ESC and, importantly, the involvement of nuclear factor kappa-B (NF-kB) pathway in the miRNA-dependent regulation in mESCs. This study provided a deeper understanding of the mechanism, by which miR-290 regulates pluripotency, and also

propose a extended vision of how microRNAs are implicated in cell cycle, innate immune response, and chromatin modification in mESC.

Pluripotency maintenance and embryonic stem cells differentiation, as well as epithelial tissues homeostasis and plasticity, are complex processes that are regulated by inter-connected networks at transcriptional, post-transcriptional and post-translational level. An intensive research effort was dedicated in understanding the molecular mechanisms that operate in both networks. Recent advances in computational biology and systemic approaches allow for a new more explicit and accurate models of the regulatory organization of those processes. Therefore the improvement and application of the newly emerging strategies will lead to identification of new entities that reproducibly and faithfully control critical circuitries in ESCs pluripotency maintenance or in EMT.

References

Alberga, A., J. L. Boulay, et al. (1991). "The snail gene required for mesoderm formation in Drosophila is expressed dynamically in derivatives of all three germ in layers." Development **111**(4): 983-992.

Alberts, B. (2002). Molecular biology of the cell. New York, Garland Science.

Altuvia, Y., P. Landgraf, et al. (2005). "Clustering and conservation patterns of human microRNAs." Nucleic acids research **33**(8): 2697-2706.

Anokye-Danso, F., C. M. Trivedi, et al. (2011). "Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency." Cell stem cell **8**(4): 376-388.

Aqeilan, R. I., A. Palamarchuk, et al. (2004). "Physical and functional interactions between the Wwox tumor suppressor protein and the AP-2gamma transcription factor." Cancer research **64**(22): 8256-8261.

Arabadjiev (2012). "Of mice and men-differential mechanisms of maintaining the undifferentiated state in mESC and hESC." Biodiscovery **3**(1).

Arnold, S. J. and E. J. Robertson (2009). "Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo." Nature reviews. Molecular cell biology **10**(2): 91-103.

Avilion, A. A., S. K. Nicolis, et al. (2003). "Multipotent cell lineages in early mouse development depend on SOX2 function." Genes & development **17**(1): 126-140.

Balwierz, P. J., M. Pachkov, et al. (2014). "ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs." Genome research **24**(5): 869-884.

Bar, M., S. K. Wyman, et al. (2008). "MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries." Stem cells **26**(10): 2496-2505.

Barad, L., R. Schick, et al. (2014). "Human embryonic stem cells vs human induced pluripotent stem cells for cardiac repair." The Canadian journal of cardiology **30**(11): 1279-1287.

Barrallo-Gimeno, A. and M. A. Nieto (2005). "The Snail genes as inducers of cell movement and survival: implications in development and cancer." Development **132**(14): 3151-3161.

Bartel, D. P. (2009). "MicroRNAs: Target Recognition and Regulatory Functions." Cell **136**(2): 215-233.

Batlle, E., E. Sancho, et al. (2000). "The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells." Nature cell biology **2**(2): 84-89.

Batlle, R., L. Alba-Castellon, et al. (2013). "Snail1 controls TGF-beta responsiveness and differentiation of mesenchymal stem cells." Oncogene **32**(28): 3381-3389.

Beattie, G. M., A. D. Lopez, et al. (2005). "Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers." Stem cells **23**(4): 489-495.

Berlato, C., K. V. Chan, et al. (2011). "Alternative TFAP2A isoforms have distinct activities in breast cancer." Breast cancer research : BCR **13**(2): R23.

Bernstein, E., S. Y. Kim, et al. (2003). "Dicer is essential for mouse development." Nature genetics **35**(3): 215-217.

Bill, R. and G. Christofori (2015). "The relevance of EMT in breast cancer metastasis: Correlation or causality?" FEBS letters **589**(14): 1577-1587.

Bilodeau, S., M. H. Kagey, et al. (2009). "SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state." Genes & development **23**(21): 2484-2489.

Bogachek, M. V., Y. Chen, et al. (2014). "Sumoylation pathway is required to maintain the basal breast cancer subtype." Cancer cell **25**(6): 748-761.
Bohnsack, M. T., K. Czaplinski, et al. (2004). "Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs." RNA **10**(2): 185-191.
Borchert, G. M., W. Lanier, et al. (2006). "RNA polymerase III transcribes human microRNAs." Nat Struct Mol Biol **13**(12): 1097-1101.
Bortolin-Cavaille, M. L., M. Dance, et al. (2009). "C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts." Nucleic Acids Res **37**(10): 3464-3473.
Bosher, J. M., T. Williams, et al. (1995). "The developmentally regulated transcription factor AP-2 is involved in c-erbB-2 overexpression in human mammary carcinoma." Proceedings of the National Academy of Sciences of the United States of America **92**(3): 744-747.
Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-353.
Brabletz, T. (2012). "EMT and MET in metastasis: where are the cancer stem cells?" Cancer cell **22**(6): 699-701.
Brabletz, T. (2012). "To differentiate or not--routes towards metastasis." Nature reviews. Cancer **12**(6): 425-436.
Bracken, C. P., P. A. Gregory, et al. (2008). "A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition." Cancer research **68**(19): 7846-7854.
Braeutigam, C., L. Rago, et al. (2014). "The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion." Oncogene **33**(9): 1082-1092.
Brons, I. G., L. E. Smithers, et al. (2007). "Derivation of pluripotent epiblast stem cells from mammalian embryos." Nature **448**(7150): 191-195.
Brown, R. L., L. M. Reinke, et al. (2011). "CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression." The Journal of clinical investigation **121**(3): 1064-1074.
Bryant, D. M. and K. E. Mostov (2008). "From cells to organs: building polarized tissue." Nature reviews. Molecular cell biology **9**(11): 887-901.
Buettner, R., P. Kannan, et al. (1993). "An alternatively spliced mRNA from the AP-2 gene encodes a negative regulator of transcriptional activation by AP-2." Molecular and cellular biology **13**(7): 4174-4185.
Burk, U., J. Schubert, et al. (2008). "A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells." EMBO reports **9**(6): 582-589.
Campbell, K., G. Whissell, et al. (2011). "Specific GATA factors act as conserved inducers of an endodermal-EMT." Developmental cell **21**(6): 1051-1061.
Cano, A., M. A. Perez-Moreno, et al. (2000). "The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression." Nature cell biology **2**(2): 76-83.
Cao, Y., W. T. Guo, et al. (2015). "miR-290/371-Mbd2-Myc circuit regulates glycolytic metabolism to promote pluripotency." The EMBO journal **34**(5): 609-623.
Cartwright, P., C. McLean, et al. (2005). "LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism." Development **132**(5): 885-896.
Carver, E. A., R. Jiang, et al. (2001). "The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition." Molecular and cellular biology **21**(23): 8184-8188.

Cavallaro, U. and G. Christofori (2004). "Cell adhesion and signalling by cadherins and Ig-CAMs in cancer." Nature reviews. Cancer **4**(2): 118-132.
Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-655.
Chambers, I., J. Silva, et al. (2007). "Nanog safeguards pluripotency and mediates germline development." Nature **450**(7173): 1230-1234.
Chambers, I. and A. Smith (2004). "Self-renewal of teratocarcinoma and embryonic stem cells." Oncogene **23**(43): 7150-7160.
Chekulaeva, M., H. Mathys, et al. (2011). "miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs." Nat Struct Mol Biol **18**(11): 1218-1226.
Chen, G., D. R. Gulbranson, et al. (2011). "Chemically defined conditions for human iPSC derivation and culture." Nature methods **8**(5): 424-429.
Cheng, X., A. Tiyaboonchai, et al. (2013). "Endodermal stem cell populations derived from pluripotent stem cells." Current opinion in cell biology **25**(2): 265-271.
Chia, N. Y., Y. S. Chan, et al. (2010). "A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity." Nature **468**(7321): 316-320.
Chong, J. J., X. Yang, et al. (2014). "Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts." Nature **510**(7504): 273-277.
Christensen, L. (1992). "The distribution of fibronectin, laminin and tetranectin in human breast cancer with special attention to the extracellular matrix." APMIS. Supplementum **26**: 1-39.
Clift, D. and M. Schuh (2013). "Restarting life: fertilization and the transition from meiosis to mitosis." Nature reviews. Molecular cell biology **14**(9): 549-562.
Cole, M. D. and M. A. Nikiforov (2006). "Transcriptional activation by the Myc oncoprotein." Current topics in microbiology and immunology **302**: 33-50.
Cole, M. F., S. E. Johnstone, et al. (2008). "Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells." Genes & development **22**(6): 746-755.
Cyr, A. R., M. V. Kulak, et al. (2014). "TFAP2C governs the luminal epithelial phenotype in mammary development and carcinogenesis." Oncogene.
Dave, N., S. Guaita-Esteruelas, et al. (2011). "Functional cooperation between Snail1 and twist in the regulation of ZEB1 expression during epithelial to mesenchymal transition." The Journal of biological chemistry **286**(14): 12024-12032.
Davis, A. C., M. Wims, et al. (1993). "A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice." Genes & development **7**(4): 671-682.
Davis, R. L., H. Weintraub, et al. (1987). "Expression of a single transfected cDNA converts fibroblasts to myoblasts." Cell **51**(6): 987-1000.
De Craene, B. and G. Berx (2013). "Regulatory networks defining EMT during cancer initiation and progression." Nature reviews. Cancer **13**(2): 97-110.
De Craene, B., F. van Roy, et al. (2005). "Unraveling signalling cascades for the Snail family of transcription factors." Cellular signalling **17**(5): 535-547.
de Pontual, L., E. Yao, et al. (2011). "Germline deletion of the miR-17[sim]92 cluster causes skeletal and growth defects in humans." Nat Genet **43**(10): 1026-1030.
Deneris, E. S. and O. Hobert (2014). "Maintenance of postmitotic neuronal cell identity." Nature neuroscience **17**(7): 899-907.
Dickinson, D. J., W. J. Nelson, et al. (2011). "A polarized epithelium organized by beta- and alpha-catenin predates cadherin and metazoan origins." Science **331**(6022): 1336-1339.

Dimos, J. T., K. T. Rodolfa, et al. (2008). "Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons." Science **321**(5893): 1218-1221.
Ding, L., M. Paszkowski-Rogacz, et al. (2009). "A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity." Cell stem cell **4**(5): 403-415.
Dottori, M., M. K. Gross, et al. (2001). "The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate." Development **128**(21): 4127-4138.
Du, C., C. Zhang, et al. (2010). "Protein kinase D1 suppresses epithelial-to-mesenchymal transition through phosphorylation of snail." Cancer research **70**(20): 7810-7819.
Dye, B. R., D. R. Hill, et al. (2015). "In vitro generation of human pluripotent stem cell derived lung organoids." eLife **4**.
Eckert, D., S. Buhl, et al. (2005). "The AP-2 family of transcription factors." Genome biology **6**(13): 246.
Eloranta, J. J. and H. C. Hurst (2002). "Transcription factor AP-2 interacts with the SUMO-conjugating enzyme UBC9 and is sumolated in vivo." The Journal of biological chemistry **277**(34): 30798-30804.
Evans, M. (2011). "Discovering pluripotency: 30 years of mouse embryonic stem cells." Nature reviews. Molecular cell biology **12**(10): 680-686.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-156.
Fabian, M. R., N. Sonenberg, et al. (2010). "Regulation of mRNA translation and stability by microRNAs." Annu Rev Biochem **79**: 351-379.
Fagoonee, S., C. Bearzi, et al. (2013). "The RNA binding protein ESRP1 fine-tunes the expression of pluripotency-related factors in mouse embryonic stem cells." PloS one **8**(8): e72300.
Fazzio, T. G., J. T. Huff, et al. (2008). "An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity." Cell **134**(1): 162-174.
Feng, X., Z. Wang, et al. (2014). "MiR-200, a new star miRNA in human cancer." Cancer letters **344**(2): 166-173.
Fernando, R. I., M. Litzinger, et al. (2010). "The T-box transcription factor Brachyury promotes epithelial-mesenchymal transition in human tumor cells." The Journal of clinical investigation **120**(2): 533-544.
Fidler, I. J. (2003). "The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited." Nature reviews. Cancer **3**(6): 453-458.
Filipowicz, W., S. N. Bhattacharyya, et al. (2008). "Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?" Nat Rev Genet **9**(2): 102-114.
Franci, C., M. Takkunen, et al. (2006). "Expression of Snail protein in tumor-stroma interface." Oncogene **25**(37): 5134-5144.
Funahashi, J., R. Sekido, et al. (1993). "Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis." Development **119**(2): 433-446.
Gafni, O., L. Weinberger, et al. (2013). "Derivation of novel human ground state naive pluripotent stem cells." Nature **504**(7479): 282-286.
Galan-Caridad, J. M., S. Harel, et al. (2007). "Zfx controls the self-renewal of embryonic and hematopoietic stem cells." Cell **129**(2): 345-357.
Garcia, M. A., M. Campillos, et al. (1999). "Transcription factor AP-2 activity is modulated by protein kinase A-mediated phosphorylation." FEBS letters **444**(1): 27-31.

Gaspar-Maia, A., A. Alajem, et al. (2009). "Chd1 regulates open chromatin and pluripotency of embryonic stem cells." Nature **460**(7257): 863-868.
Gaubatz, S., A. Imhof, et al. (1995). "Transcriptional activation by Myc is under negative control by the transcription factor AP-2." The EMBO journal **14**(7): 1508-1519.
Genetta, T., D. Ruezinsky, et al. (1994). "Displacement of an E-box-binding repressor by basic helix-loop-helix proteins: implications for B-cell specificity of the immunoglobulin heavy-chain enhancer." Molecular and cellular biology **14**(9): 6153-6163.
Goding, C. R., D. Pei, et al. (2014). "Cancer: pathological nuclear reprogramming?" Nature reviews. Cancer **14**(8): 568-573.
Gregory, P. A., A. G. Bert, et al. (2008). "The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1." Nature cell biology **10**(5): 593-601.
Grether-Beck, S., I. Felsner, et al. (2003). "Mitochondrial cytochrome c release mediates ceramide-induced activator protein 2 activation and gene expression in keratinocytes." The Journal of biological chemistry **278**(48): 47498-47507.
Griffiths, D. S., J. Li, et al. (2011). "LIF-independent JAK signalling to chromatin in embryonic stem cells uncovered from an adult stem cell disease." Nature cell biology **13**(1): 13-21.
Guo, G., J. Yang, et al. (2009). "Klf4 reverts developmentally programmed restriction of ground state pluripotency." Development **136**(7): 1063-1069.
Han, H., M. Irimia, et al. (2013). "MBNL proteins repress ES-cell-specific alternative splicing and reprogramming." Nature **498**(7453): 241-245.
Han, J., Y. Lee, et al. (2004). "The Drosha-DGCR8 complex in primary microRNA processing." Genes & development **18**(24): 3016-3027.
Han, J. W. and Y. S. Yoon (2012). "Epigenetic landscape of pluripotent stem cells." Antioxidants & redox signaling **17**(2): 205-223.
Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.
Hanna, J., M. Wernig, et al. (2007). "Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin." Science **318**(5858): 1920-1923.
Haraguchi, M., T. Okubo, et al. (2008). "Snail regulates cell-matrix adhesion by regulation of the expression of integrins and basement membrane proteins." The Journal of biological chemistry **283**(35): 23514-23523.
Hartwell, K. A., B. Muir, et al. (2006). "The Spemann organizer gene, Goosecoid, promotes tumor metastasis." Proceedings of the National Academy of Sciences of the United States of America **103**(50): 18969-18974.
He, L. and G. J. Hannon (2004). "MicroRNAs: small RNAs with a big role in gene regulation." Nat Rev Genet **5**(7): 522-531.
Hennessy, B. T., A. M. Gonzalez-Angulo, et al. (2009). "Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics." Cancer research **69**(10): 4116-4124.
Herschkowitz, J. I., K. Simin, et al. (2007). "Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors." Genome biology **8**(5): R76.
Higashi, Y., H. Moribe, et al. (1997). "Impairment of T cell development in deltaEF1 mutant mice." The Journal of experimental medicine **185**(8): 1467-1479.
Hilger-Eversheim, K., M. Moser, et al. (2000). "Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis and cell-cycle control." Gene **260**(1-2): 1-12.

Ho, L., E. L. Miller, et al. (2011). "esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function." Nature cell biology **13**(8): 903-913.
Ho, L., J. L. Ronan, et al. (2009). "An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency." Proceedings of the National Academy of Sciences of the United States of America **106**(13): 5181-5186.
Houbaviy, H. B., M. F. Murray, et al. (2003). "Embryonic stem cell-specific MicroRNAs." Developmental cell **5**(2): 351-358.
Hu, G., J. Kim, et al. (2009). "A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal." Genes & development **23**(7): 837-848.
Huang, Y. and F. E. Domann (1998). "Redox modulation of AP-2 DNA binding activity in vitro." Biochemical and biophysical research communications **249**(2): 307-312.
Humphrey, R. K., G. M. Beattie, et al. (2004). "Maintenance of pluripotency in human embryonic stem cells is STAT3 independent." Stem cells **22**(4): 522-530.
Hutvagner, G., J. McLachlan, et al. (2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." Science **293**(5531): 834-838.
Iliopoulos, D., M. Lindahl-Allen, et al. (2010). "Loss of miR-200 inhibition of Suz12 leads to polycomb-mediated repression required for the formation and maintenance of cancer stem cells." Molecular cell **39**(5): 761-772.
Ivanova, N., R. Dobrin, et al. (2006). "Dissecting self-renewal in stem cells with RNA interference." Nature **442**(7102): 533-538.
Jager, R., U. Werling, et al. (2003). "Transcription factor AP-2gamma stimulates proliferation and apoptosis and impairs differentiation in a transgenic model." Molecular cancer research : MCR **1**(12): 921-929.
James, D., A. J. Levine, et al. (2005). "TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells." Development **132**(6): 1273-1282.
James, D., S. A. Noggle, et al. (2006). "Contribution of human embryonic stem cells to mouse blastocysts." Developmental biology **295**(1): 90-102.
Jean, D., J. E. Gershenwald, et al. (1998). "Loss of AP-2 results in up-regulation of MCAM/MUC18 and an increase in tumor growth and metastasis of human melanoma cells." The Journal of biological chemistry **273**(26): 16501-16508.
Jiang, J., Y. S. Chan, et al. (2008). "A core Klf circuitry regulates self-renewal of embryonic stem cells." Nature cell biology **10**(3): 353-360.
Jiang, M. S., Q. Q. Tang, et al. (1998). "Derepression of the C/EBPalpha gene during adipogenesis: identification of AP-2alpha as a repressor." Proceedings of the National Academy of Sciences of the United States of America **95**(7): 3467-3471.
Kadzik, R. S. and E. E. Morrisey (2012). "Directing lung endoderm differentiation in pluripotent stem cells." Cell stem cell **10**(4): 355-361.
Kagey, M. H., J. J. Newman, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture." Nature **467**(7314): 430-435.
Kai, Z. S. and A. E. Pasquinelli (2010). "MicroRNA assassins: factors that regulate the disappearance of miRNAs." Nature structural & molecular biology **17**(1): 5-10.
Kawamura, M., S. Miyagawa, et al. (2012). "Feasibility, safety, and therapeutic efficacy of human induced pluripotent stem cell-derived cardiomyocyte sheets in a porcine ischemic cardiomyopathy model." Circulation **126**(11 Suppl 1): S29-37.
Keller, G. (2005). "Embryonic stem cell differentiation: emergence of a new era in biology and medicine." Genes & development **19**(10): 1129-1155.

Kim, J., A. J. Woo, et al. (2010). "A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs." Cell **143**(2): 313-324.
Kiskinis, E., J. Sandoe, et al. (2014). "Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1." Cell stem cell **14**(6): 781-795.
Kondo, M., E. Cubillo, et al. (2004). "A role for Id in the regulation of TGF-beta-induced epithelial-mesenchymal transdifferentiation." Cell death and differentiation **11**(10): 1092-1101.
Korpal, M., E. S. Lee, et al. (2008). "The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2." The Journal of biological chemistry **283**(22): 14910-14914.
Kriks, S., J. W. Shim, et al. (2011). "Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease." Nature **480**(7378): 547-551.
Kumar, R. M., P. Cahan, et al. (2014). "Deconstructing transcriptional heterogeneity in pluripotent stem cells." Nature **516**(7529): 56-61.
Kuzmanov, A., U. Hopfer, et al. (2014). "LIM-homeobox gene 2 promotes tumor growth and metastasis by inducing autocrine and paracrine PDGF-B signaling." Molecular oncology **8**(2): 401-416.
Labelle, M., S. Begum, et al. (2011). "Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis." Cancer cell **20**(5): 576-590.
Lahiri, S. K. and J. Zhao (2012). "Kruppel-like factor 8 emerges as an important regulator of cancer." American journal of translational research **4**(3): 357-363.
Lamouille, S., J. Xu, et al. (2014). "Molecular mechanisms of epithelial-mesenchymal transition." Nature reviews. Molecular cell biology **15**(3): 178-196.
Laurent, L. C., J. Chen, et al. (2008). "Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence." Stem cells **26**(6): 1506-1516.
Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-313.
Lee, Y., M. Kim, et al. (2004). "MicroRNA genes are transcribed by RNA polymerase II." EMBO J **23**(20): 4051-4060.
Lehembre, F., M. Yilmaz, et al. (2008). "NCAM-induced focal adhesion assembly: a functional switch upon loss of E-cadherin." The EMBO journal **27**(19): 2603-2615.
Leonardo, T. R., H. L. Schultheisz, et al. (2012). "The functions of microRNAs in pluripotency and reprogramming." Nature cell biology **14**(11): 1114-1121.
Leung, A. K., A. G. Young, et al. (2011). "Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs." Nature structural & molecular biology **18**(2): 237-244.
Lewis, B. P., C. B. Burge, et al. (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." Cell **120**(1): 15-20.
Lewis, B. P., I. H. Shih, et al. (2003). "Prediction of mammalian microRNA targets." Cell **115**(7): 787-798.
Lin, C. Y., J. Loven, et al. (2012). "Transcriptional amplification in tumor cells with elevated c-Myc." Cell **151**(1): 56-67.
Lingel, A. and E. Izaurralde (2004). "RNAi: finding the elusive endonuclease." RNA **10**(11): 1675-1679.

Liu, J., Y. Yue, et al. (2014). "A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation." Nature chemical biology **10**(2): 93-95.
Loh, K. M. and B. Lim (2013). "Stem cells: Close encounters with full potential." Nature **502**(7469): 41-42.
Lombaerts, M., T. van Wezel, et al. (2006). "E-cadherin transcriptional downregulation by promoter methylation but not mutation is related to epithelial-to-mesenchymal transition in breast cancer cell lines." British journal of cancer **94**(5): 661-671.
Ludwig, T. E., M. E. Levenstein, et al. (2006). "Derivation of human embryonic stem cells in defined conditions." Nature biotechnology **24**(2): 185-187.
Luscher, B. and J. Vervoorts (2012). "Regulation of gene transcription by the oncoprotein MYC." Gene **494**(2): 145-160.
Ma, L., J. Young, et al. (2010). "miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis." Nature cell biology **12**(3): 247-256.
Makrodouli, E., E. Oikonomou, et al. (2011). "BRAF and RAS oncogenes regulate Rho GTPase pathways to mediate migration and invasion properties in human colon cancer cells: a comparative study." Molecular cancer **10**: 118.
Mani, S. A., W. Guo, et al. (2008). "The epithelial-mesenchymal transition generates cells with properties of stem cells." Cell **133**(4): 704-715.
Mani, S. A., J. Yang, et al. (2007). "Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers." Proceedings of the National Academy of Sciences of the United States of America **104**(24): 10069-10074.
Marson, A., R. Foreman, et al. (2008). "Wnt signaling promotes reprogramming of somatic cells to pluripotency." Cell stem cell **3**(2): 132-135.
Marson, A., S. S. Levine, et al. (2008). "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells." Cell **134**(3): 521-533.
Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proceedings of the National Academy of Sciences of the United States of America **78**(12): 7634-7638.
Martin, G. R. and M. J. Evans (1975). "Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro." Proceedings of the National Academy of Sciences of the United States of America **72**(4): 1441-1445.
Mascetti, V. L. and R. A. Pedersen (2014). "Naivete of the human pluripotent stem cell." Nature biotechnology **32**(1): 68-70.
Massari, M. E. and C. Murre (2000). "Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms." Molecular and cellular biology **20**(2): 429-440.
Masui, S., Y. Nakatake, et al. (2007). "Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells." Nature cell biology **9**(6): 625-635.
Matsuda, T., T. Nakamura, et al. (1999). "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells." The EMBO journal **18**(15): 4261-4269.
McCoy, E. L., R. Iwanaga, et al. (2009). "Six1 expands the mouse mammary epithelial stem/progenitor cell pool and induces mammary tumors that undergo epithelial-mesenchymal transition." The Journal of clinical investigation **119**(9): 2663-2677.
Medeiros, L. A., L. M. Dennis, et al. (2011). "Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects." Proceedings of the National Academy of Sciences of the United States of America **108**(34): 14163-14168.

Meier, P., M. Koedood, et al. (1995). "Alternative mRNAs encode multiple isoforms of transcription factor AP-2 during murine embryogenesis." Developmental biology **169**(1): 1-14.
Melnikova, V. O. and M. Bar-Eli (2008). "Transcriptional control of the melanoma malignant phenotype." Cancer biology & therapy **7**(7): 997-1003.
Melton, C., R. L. Judson, et al. (2010). "Opposing microRNA families regulate self-renewal in mouse embryonic stem cells." Nature **463**(7281): 621-626.
Mencia, A., S. Modamio-Hoybjor, et al. (2009). "Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss." Nat Genet **41**(5): 609-613.
Micalizzi, D. S., K. L. Christensen, et al. (2009). "The Six1 homeoprotein induces human mammary carcinoma cells to undergo epithelial-mesenchymal transition and metastasis in mice through increasing TGF-beta signaling." The Journal of clinical investigation **119**(9): 2678-2690.
Milunsky, J. M., T. A. Maher, et al. (2008). "TFAP2A mutations result in branchio-oculo-facial syndrome." American journal of human genetics **82**(5): 1171-1177.
Mintz, B. and K. Illmensee (1975). "Normal genetically mosaic mice produced from malignant teratocarcinoma cells." Proceedings of the National Academy of Sciences of the United States of America **72**(9): 3585-3589.
Mitchell, P. J., C. Wang, et al. (1987). "Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen." Cell **50**(6): 847-861.
Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-642.
Miyoshi, N., H. Ishii, et al. (2011). "Reprogramming of mouse and human cells to pluripotency using mature microRNAs." Cell stem cell **8**(6): 633-638.
Mohibullah, N., A. Donner, et al. (1999). "SELEX and missing phosphate contact analyses reveal flexibility within the AP-2[alpha] protein: DNA binding complex." Nucleic acids research **27**(13): 2760-2769.
Morin, R. D., M. D. O'Connor, et al. (2008). "Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells." Genome research **18**(4): 610-621.
Moser, M., J. Ruschoff, et al. (1997). "Comparative analysis of AP-2 alpha and AP-2 beta gene expression during murine embryogenesis." Developmental dynamics : an official publication of the American Association of Anatomists **208**(1): 115-124.
Murray, S. A. and T. Gridley (2006). "Snail family genes are required for left-right asymmetry determination, but not neural crest formation, in mice." Proceedings of the National Academy of Sciences of the United States of America **103**(27): 10300-10304.
Najm, F. J., J. G. Chenoweth, et al. (2011). "Isolation of epiblast stem cells from preimplantation mouse embryos." Cell stem cell **8**(3): 318-325.
Nakaya, Y. and G. Sheng (2008). "Epithelial to mesenchymal transition during gastrulation: an embryological view." Development, growth & differentiation **50**(9): 755-766.
Neve, R. M., K. Chin, et al. (2006). "A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes." Cancer cell **10**(6): 515-527.
Niakan, K. K., E. C. Davis, et al. (2006). "Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis." Molecular genetics and metabolism **88**(3): 261-271.

Nichols, J., I. Chambers, et al. (2001). "Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines." Development **128**(12): 2333-2339.
Nichols, J. and A. Smith (2009). "Naive and primed pluripotent states." Cell stem cell **4**(6): 487-492.
Nichols, J., B. Zevnik, et al. (1998). "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4." Cell **95**(3): 379-391.
Nicoloso, M. S., R. Spizzo, et al. (2009). "MicroRNAs--the micro steering wheel of tumour metastases." Nature reviews. Cancer **9**(4): 293-302.
Nie, Z., G. Hu, et al. (2012). "c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells." Cell **151**(1): 68-79.
Nieto, M. A. (2002). "The snail superfamily of zinc-finger transcription factors." Nature reviews. Molecular cell biology **3**(3): 155-166.
Nieto, M. A., M. G. Sargent, et al. (1994). "Control of cell behavior during vertebrate development by Slug, a zinc finger gene." Science **264**(5160): 835-839.
Niwa, H., T. Burdon, et al. (1998). "Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3." Genes & development **12**(13): 2048-2060.
Niwa, H., J. Miyazaki, et al. (2000). "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." Nature genetics **24**(4): 372-376.
Nori, S., Y. Okada, et al. (2011). "Grafted human-induced pluripotent stem-cell-derived neurospheres promote motor functional recovery after spinal cord injury in mice." Proceedings of the National Academy of Sciences of the United States of America **108**(40): 16825-16830.
O'Connor, T. P. and R. G. Crystal (2006). "Genetic medicines: treatment strategies for hereditary disorders." Nature reviews. Genetics **7**(4): 261-276.
O'Donnell, K. A., E. A. Wentzel, et al. (2005). "c-Myc-regulated microRNAs modulate E2F1 expression." Nature **435**(7043): 839-843.
Ocana, O. H., R. Corcoles, et al. (2012). "Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1." Cancer cell **22**(6): 709-724.
Okamoto, S. and M. Takahashi (2011). "Induction of retinal pigment epithelial cells from monkey iPS cells." Investigative ophthalmology & visual science **52**(12): 8785-8790.
Orso, F., E. Penna, et al. (2008). "AP-2alpha and AP-2gamma regulate tumor progression via specific genetic programs." FASEB journal : official publication of the Federation of American Societies for Experimental Biology **22**(8): 2702-2714.
Park, I. H., N. Arora, et al. (2008). "Disease-specific induced pluripotent stem cells." Cell **134**(5): 877-886.
Park, J. and J. E. Schwarzbauer (2014). "Mammary epithelial cell interactions with fibronectin stimulate epithelial-mesenchymal transition." Oncogene **33**(13): 1649-1657.
Park, S. M., A. B. Gaur, et al. (2008). "The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2." Genes & development **22**(7): 894-907.
Peinado, H., D. Olmeda, et al. (2007). "Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype?" Nature reviews. Cancer **7**(6): 415-428.
Pellikainen, J., A. Naukkarinen, et al. (2004). "Expression of HER2 and its association with AP-2 in breast cancer." European journal of cancer **40**(10): 1485-1495.

Plaks, V., N. Kong, et al. (2015). "The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells?" Cell stem cell **16**(3): 225-238.
Poh, Y. C., J. Chen, et al. (2014). "Generation of organized germ layers from a single mouse embryonic stem cell." Nature communications **5**: 4000.
Postigo, A. A. and D. C. Dean (2000). "Differential expression and function of members of the zfh-1 family of zinc finger/homeodomain repressors." Proceedings of the National Academy of Sciences of the United States of America **97**(12): 6391-6396.
Qi, J., J. Y. Yu, et al. (2009). "microRNAs regulate human embryonic stem cell division." Cell cycle **8**(22): 3729-3741.
Rada-Iglesias, A., R. Bajpai, et al. (2012). "Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest." Cell stem cell **11**(5): 633-648.
Rahl, P. B., C. Y. Lin, et al. (2010). "c-Myc regulates transcriptional pause release." Cell **141**(3): 432-445.
Robertson, E., A. Bradley, et al. (1986). "Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector." Nature **323**(6087): 445-448.
Rodin, S., L. Antonsson, et al. (2014). "Clonal culturing of human embryonic stem cells on laminin-521/E-cadherin matrix in defined and xeno-free environment." Nature communications **5**: 3195.
Rodriguez-Boulan, E. and I. G. Macara (2014). "Organization and execution of the epithelial polarity programme." Nature reviews. Molecular cell biology **15**(4): 225-242.
Rossant, J. (2008). "Stem cells and early lineage development." Cell **132**(4): 527-531.
Ruby, J. G., C. H. Jan, et al. (2007). "Intronic microRNA precursors that bypass Drosha processing." Nature **448**(7149): 83-86.
Sakai, D., T. Suzuki, et al. (2006). "Cooperative action of Sox9, Snail2 and PKA signaling in early neural crest development." Development **133**(7): 1323-1333.
Salani, S., C. Donadoni, et al. (2012). "Generation of skeletal muscle cells from embryonic and induced pluripotent stem cells as an in vitro model and for therapy of muscular dystrophies." Journal of cellular and molecular medicine **16**(7): 1353-1364.
Sawai, S., A. Shimono, et al. (1991). "Embryonic lethality resulting from disruption of both N-myc alleles in mouse zygotes." The New biologist **3**(9): 861-869.
Schnetz, M. P., L. Handoko, et al. (2010). "CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression." PLoS genetics **6**(7): e1001023.
Schneuwly, S., R. Klemenz, et al. (1987). "Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia." Nature **325**(6107): 816-818.
Schorle, H., P. Meier, et al. (1996). "Transcription factor AP-2 essential for cranial closure and craniofacial development." Nature **381**(6579): 235-238.
Sekido, R., K. Murai, et al. (1994). "The delta-crystallin enhancer-binding protein delta EF1 is a repressor of E2-box-mediated gene activation." Molecular and cellular biology **14**(9): 5692-5700.
Sewell, W. and R. Y. Lin (2014). "Generation of thyroid follicular cells from pluripotent stem cells: potential for regenerative medicine." Frontiers in endocrinology **5**: 96.
Shamir, E. R. and A. J. Ewald (2014). "Three-dimensional organotypic culture: experimental models of mammalian biology and disease." Nature reviews. Molecular cell biology **15**(10): 647-664.
Shapiro, I. M., A. W. Cheng, et al. (2011). "An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype." PLoS genetics **7**(8): e1002218.

Shi, D., F. Xie, et al. (2014). "TFAP2A regulates nasopharyngeal carcinoma growth and survival by targeting HIF-1alpha signaling pathway." Cancer prevention research **7**(2): 266-277.
Sinagoga, K. L. and J. M. Wells (2015). "Generating human intestinal tissues from pluripotent stem cells to study development and disease." The EMBO journal **34**(9): 1149-1163.
Slattum, G. M. and J. Rosenblatt (2014). "Tumour cell invasion: an emerging role for basal epithelial cell extrusion." Nature reviews. Cancer **14**(7): 495-501.
Slorach, E. M., J. Chou, et al. (2011). "Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization." Genes & development **25**(5): 471-484.
Slukvin, II (2013). "Hematopoietic specification from human pluripotent stem cells: current advances and challenges toward de novo generation of hematopoietic stem cells." Blood **122**(25): 4035-4046.
Smith, A. G. (2001). "Embryo-derived stem cells: of mice and men." Annual review of cell and developmental biology **17**: 435-462.
Smith, A. G., J. K. Heath, et al. (1988). "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." Nature **336**(6200): 688-690.
Solter, D., N. Skreb, et al. (1970). "Extrauterine growth of mouse egg-cylinders results in malignant teratoma." Nature **227**(5257): 503-504.
Song, S. J., K. Ito, et al. (2013). "The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation." Cell stem cell **13**(1): 87-101.
Spano, D., C. Heck, et al. (2012). "Molecular networks that regulate cancer metastasis." Seminars in cancer biology **22**(3): 234-249.
Stanton, B. R., A. S. Perkins, et al. (1992). "Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop." Genes & development **6**(12A): 2235-2247.
Stefani, G. and F. J. Slack (2008). "Small non-coding RNAs in animal development." Nature reviews. Molecular cell biology **9**(3): 219-230.
Steiner, D., H. Khaner, et al. (2010). "Derivation, propagation and controlled differentiation of human embryonic stem cells in suspension." Nature biotechnology **28**(4): 361-364.
Stevens, L. C. and C. C. Little (1954). "Spontaneous Testicular Teratomas in an Inbred Strain of Mice." Proceedings of the National Academy of Sciences of the United States of America **40**(11): 1080-1087.
Suh, M. R., Y. Lee, et al. (2004). "Human embryonic stem cells express a unique set of microRNAs." Developmental biology **270**(2): 488-498.
Sun, C., Y. Nakatake, et al. (2009). "Dax1 binds to Oct3/4 and inhibits its transcriptional activity in embryonic stem cells." Molecular and cellular biology **29**(16): 4574-4583.
Takagi, T., H. Moribe, et al. (1998). "DeltaEF1, a zinc finger and homeodomain transcription factor, is required for skeleton patterning in multiple lineages." Development **125**(1): 21-31.
Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.
Takayama, N., S. Nishimura, et al. (2010). "Transient activation of c-MYC expression is critical for efficient platelet generation from human induced pluripotent stem cells." The Journal of experimental medicine **207**(13): 2817-2830.
Taube, J. H., J. I. Herschkowitz, et al. (2010). "Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and

metaplastic breast cancer subtypes." Proceedings of the National Academy of Sciences of the United States of America **107**(35): 15449-15454.
Tavazoie, S. F., C. Alarcon, et al. (2008). "Endogenous human microRNAs that suppress breast cancer metastasis." Nature **451**(7175): 147-152.
Tee, W. W. and D. Reinberg (2014). "Chromatin features and the epigenetic regulation of pluripotency states in ESCs." Development **141**(12): 2376-2390.
Tesar, P. J., J. G. Chenoweth, et al. (2007). "New cell lines from mouse epiblast share defining features with human embryonic stem cells." Nature **448**(7150): 196-199.
Theunissen, T. W. and R. Jaenisch (2014). "Molecular control of induced pluripotency." Cell stem cell **14**(6): 720-734.
Theveneau, E. and R. Mayor (2012). "Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration." Developmental biology **366**(1): 34-54.
Thiery, J. P., H. Acloque, et al. (2009). "Epithelial-mesenchymal transitions in development and disease." Cell **139**(5): 871-890.
Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-1147.
Thuault, S., E. J. Tan, et al. (2008). "HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition." The Journal of biological chemistry **283**(48): 33437-33446.
Tiwari, N., A. Gheldof, et al. (2012). "EMT as the ultimate survival mechanism of cancer cells." Seminars in cancer biology **22**(3): 194-207.
Tiwari, N., V. K. Tiwari, et al. (2013). "Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming." Cancer cell **23**(6): 768-783.
Tiwari, V. K., M. B. Stadler, et al. (2012). "A chromatin-modifying function of JNK during stem cell differentiation." Nature genetics **44**(1): 94-100.
Trelstad, R. L., E. D. Hay, et al. (1967). "Cell contact during early morphogenesis in the chick embryo." Developmental biology **16**(1): 78-106.
van Berlo, J. H. and J. D. Molkentin (2014). "An emerging consensus on cardiac regeneration." Nature medicine **20**(12): 1386-1393.
Van de Putte, T., M. Maruhashi, et al. (2003). "Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome." American journal of human genetics **72**(2): 465-470.
Varlakhanova, N. V., R. F. Cotterman, et al. (2010). "myc maintains embryonic stem cell pluripotency and self-renewal." Differentiation; research in biological diversity **80**(1): 9-19.
Venables, J. P., L. Lapasset, et al. (2013). "MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation." Nature communications **4**: 2480.
Venkov, C. D., A. J. Link, et al. (2007). "A proximal activator of transcription in epithelial-mesenchymal transition." The Journal of clinical investigation **117**(2): 482-491.
Villa-Diaz, L. G., A. M. Ross, et al. (2013). "Concise review: The evolution of human pluripotent stem cell culture: from feeder cells to synthetic coatings." Stem cells **31**(1): 1-7.
Wainger, B. J., E. Kiskinis, et al. (2014). "Intrinsic membrane hyperexcitability of amyotrophic lateral sclerosis patient-derived motor neurons." Cell reports **7**(1): 1-11.

Wang, D., T. H. Shin, et al. (1997). "Transcription factor AP-2 controls transcription of the human transforming growth factor-alpha gene." The Journal of biological chemistry **272**(22): 14244-14250.
Wang, G., X. Guo, et al. (2013). "Critical regulation of miR-200/ZEB2 pathway in Oct4/Sox2-induced mesenchymal-to-epithelial transition and induced pluripotent stem cell generation." Proceedings of the National Academy of Sciences of the United States of America **110**(8): 2858-2863.
Wang, W., L. Lv, et al. (2011). "Reduced expression of transcription factor AP-2alpha is associated with gastric adenocarcinoma prognosis." PloS one **6**(9): e24897.
Wang, X., M. Zheng, et al. (2007). "Kruppel-like factor 8 induces epithelial to mesenchymal transition and epithelial cell invasion." Cancer research **67**(15): 7184-7193.
Wang, Y., S. Baskerville, et al. (2008). "Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation." Nature genetics **40**(12): 1478-1483.
Wang, Y., Y. Li, et al. (2014). "N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells." Nature cell biology **16**(2): 191-198.
Wang, Y., R. Medvid, et al. (2007). "DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal." Nature genetics **39**(3): 380-385.
Warzecha, C. C., P. Jiang, et al. (2010). "An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition." The EMBO journal **29**(19): 3286-3300.
Warzecha, C. C., T. K. Sato, et al. (2009). "ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing." Molecular cell **33**(5): 591-601.
Wellner, U., J. Schubert, et al. (2009). "The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs." Nature cell biology **11**(12): 1487-1495.
Wheelock, M. J., Y. Shintani, et al. (2008). "Cadherin switching." Journal of cell science **121**(Pt 6): 727-735.
Williams, R. L., D. J. Hilton, et al. (1988). "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells." Nature **336**(6200): 684-687.
Williams, T. and R. Tjian (1991). "Characterization of a dimerization motif in AP-2 and its function in heterologous DNA-binding proteins." Science **251**(4997): 1067-1071.
Wolf-Yadlin, A., S. Hautaniemi, et al. (2007). "Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks." Proceedings of the National Academy of Sciences of the United States of America **104**(14): 5860-5865.
Woodcock, C. L. and R. P. Ghosh (2010). "Chromatin higher-order structure and dynamics." Cold Spring Harbor perspectives in biology **2**(5): a000596.
Woodfield, G. W., Y. Chen, et al. (2010). "Identification of primary gene targets of TFAP2C in hormone responsive breast carcinoma cells." Genes, chromosomes & cancer **49**(10): 948-962.
Worringer, K. A., T. A. Rand, et al. (2014). "The let-7/LIN-41 pathway regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes." Cell stem cell **14**(1): 40-52.
Xu, C., M. S. Inokuma, et al. (2001). "Feeder-free growth of undifferentiated human embryonic stem cells." Nature biotechnology **19**(10): 971-974.
Xu, R. H., X. Chen, et al. (2002). "BMP4 initiates human embryonic stem cell differentiation to trophoblast." Nature biotechnology **20**(12): 1261-1264.

Yamanaka, S. (2009). "A fresh look at iPS cells." Cell **137**(1): 13-17.
Yamanaka, S. (2012). "Induced pluripotent stem cells: past, present, and future." Cell stem cell **10**(6): 678-684.
Yang, Z., S. Rayala, et al. (2005). "Pak1 phosphorylation of snail, a master regulator of epithelial-to-mesenchyme transition, modulates snail's subcellular localization and functions." Cancer research **65**(8): 3179-3184.
Ye, J. and R. Blelloch (2014). "Regulation of pluripotency by RNA binding proteins." Cell stem cell **15**(3): 271-280.
Yeo, J. C. and H. H. Ng (2013). "The transcriptional regulation of pluripotency." Cell research **23**(1): 20-32.
Yi, R., Y. Qin, et al. (2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." Genes & development **17**(24): 3011-3016.
Yilmaz, M. and G. Christofori (2009). "EMT, the cytoskeleton, and cancer cell invasion." Cancer metastasis reviews **28**(1-2): 15-33.
Yilmaz, M. and G. Christofori (2010). "Mechanisms of motility in metastasizing cells." Molecular cancer research : MCR **8**(5): 629-642.
Ying, Q. L., J. Nichols, et al. (2003). "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3." Cell **115**(3): 281-292.
Yook, J. I., X. Y. Li, et al. (2006). "A Wnt-Axin2-GSK3beta cascade regulates Snail1 activity in breast cancer cells." Nature cell biology **8**(12): 1398-1406.
Young, R. A. (2011). "Control of the embryonic stem cell state." Cell **144**(6): 940-954.
Yu, L., M. J. Hitchler, et al. (2009). "AP-2alpha Inhibits c-MYC Induced Oxidative Stress and Apoptosis in HaCaT Human Keratinocytes." Journal of oncology **2009**: 780874.
Zeng, Y. X., K. Somasundaram, et al. (1997). "AP2 inhibits cancer cell growth and activates p21WAF1/CIP1 expression." Nature genetics **15**(1): 78-82.
Zhang, H., F. Meng, et al. (2011). "Forkhead transcription factor foxq1 promotes epithelial-mesenchymal transition and breast cancer metastasis." Cancer research **71**(4): 1292-1301.
Zhang, J., S. Brewer, et al. (2003). "Overexpression of transcription factor AP-2alpha suppresses mammary gland growth and morphogenesis." Developmental biology **256**(1): 127-145.
Zhang, J., S. Hagopian-Donaldson, et al. (1996). "Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2." Nature **381**(6579): 238-241.
Zhang, J., W. L. Tam, et al. (2006). "Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1." Nature cell biology **8**(10): 1114-1123.
Zhang, K., E. Rodriguez-Aznar, et al. (2012). "Lats2 kinase potentiates Snail1 activity by promoting nuclear retention upon phosphorylation." The EMBO journal **31**(1): 29-43.
Zhong, L., Y. Wang, et al. (2003). "Functional characterization of the interacting domains of the positive coactivator PC4 with the transcription factor AP-2alpha." Gene **320**: 155-164.
Zhou, B. P., J. Deng, et al. (2004). "Dual regulation of Snail by GSK-3beta-mediated phosphorylation in control of epithelial-mesenchymal transition." Nature cell biology **6**(10): 931-940.