The function of the Gadd45 protein family during pluripotency and differentiation

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Manuel Leichsenring

geb. am 11.09.1984 in Ravensburg

Mainz, 2016

Dekan:	-
1. Berichterstatter:	-
2. Berichterstatter:	-
Tag der mündlichen Prüfung:	22.03.2016

Table of contents

1 Summary

Contrary to long held beliefs, DNA methylation is a dynamic epigenetic process and DNA methylation patterns can undergo drastic changes during differentiation and development. The family of Gadd45 (growth arrest and DNA damage) proteins – Gadd45a, Gadd45b and Gadd45g – has been implicated in multiple pathways of active DNA demethylation in previous work, but the physiological role of Gadd45-mediated DNA demethylation is not well understood. Here the potentially redundant functions of the Gadd45 genes during pluripotency, early lineage decision and differentiation are investigated, using mESCs (mouse embryonic stem cells) as a model system.

In *Gadd45a*$^{-/-}$ *Gadd45b*$^{-/-}$ *Gadd45g*$^{-/-}$ triple knockout (TKO) mESCs, a subset of methylation regulated genes – including LIF (leukemia inhibitory factor) target genes and imprinted genes - is downregulated and associated regulatory DNA elements are hypermethylated. Gsk3 and Mek inhibition, which induces "ground state pluripotency", and is accompanied by global DNA demethylation, partially rescues this differential gene expression. Transient overexpression of any single Gadd45 gene in TKOs also reverses gene misexpression, supporting redundant functions of the Gadd45 proteins.

Upon mESC differentiation, the number of differentially expressed genes in Gadd45 TKO cells compared to control cells increases greatly, with a tendency of neuronal and cell motility related genes to be most affected. Interestingly, also a downregulation of trophoblast genes is observed. This is corroborated by a deficiency of Gadd45 TKO mESCs to transdifferentiate into the trophoblast lineage upon BMP (bone morphogenetic protein) treatment. Moreover Gadd45 genes are expressed in partially overlapping expression domains during placental development, indicating a role not only during mESC-to-trophoblast transdifferentiation, but also during physiological trophoblast development. In line with this, trophoblast stem cells devoid of Gadd45 genes show a reduced expression of pluripotency genes.

In summary, these results show both a limited role of the Gadd45 genes in the activation of methylation regulated genes during pluripotency, as well as a broader role during cellular differentiation and trophoblast development.

2 Zusammenfassung

Entgegen langläufiger Meinung handelt es sich bei DNA Methylierung um einen dynamischen epigenetischen Prozess. Daher können DNA Methylierungsmuster drastische Änderungen während Entwicklungs- und Differenzierungsprozessen unterlaufen. Die Gadd45 (Growth Arrest and DNA Damage) Proteinfamilie – bestehend aus Gadd45a, Gadd45b und Gadd45g – wurde in vorangegangenen Studien mit mehreren verschiedenen Mechanismen der aktiven DNA Demethylierung in Verbindung gebracht, doch die physiologische Rolle dieser Mechanismen ist bisher wenig verstanden. In der vorliegenden Arbeit wurden die potenziell redundanten Funktionen der Gadd45 Proteine in pluripotenten, sowie in differenzierenden embryonalen Mausstammzellen untersucht.

In Stammzellen, in denen alle Gadd45 Gene deletiert wurden, erfolgt eine Herunterregulierung von Genen, deren Expression über DNA-Methylierung gesteuert wird – darunter LIF (leukemia inhibitory factor) Zielgene und genomisch geprägte Gene. Zudem sind entsprechende regulatorische Elemente hypermethyliert. Eine Behandlung mit Gsk3 und Mek Inhibitoren, welche die Zellen in die sog. „Ground State Pluripotency“ überführt und mit genomweiter DNA Demethylierung einhergeht, bewirkt eine partielle Korrektur der transkriptionellen Fehlregulation von Gadd45 Zielgenen. Des Weiteren kann die Fehlregulation durch transiente Überexpression eines beliebigen Gadd45 Gens behoben werden – ein Umstand der in der Tat auf eine redundante Funktion dieser Gene hinweist.

Während der *in vitro* Differenzierung der Gadd45-defizienten Stammzellen, kommt es zu noch ausgeprägterer Fehlregulation von Genen, insbesondere bei Genen, die neuronale oder zellbewegungsrelevante Funktionen ausüben. Interessanterweise werden zudem Trophoblast-spezifische Gene weniger exprimiert. Dieser Befund wurde durch die Tatsache bestärkt, dass Gadd45-defiziente Stammzellen nicht experimentell mit Hilfe von BMP (bone morphogenetic protein) zu Trophoblastenzellen transdifferenziert werden können. Zudem werden alle Gadd45 Gene auch während der Plazentaentwicklung in teils überlappenden Domänen exprimiert. Dies macht es wahrscheinlich, dass die Gadd45 Gene nicht nur während Transdifferenzierungsprozessen, sondern auch während der normalen Plazentaentwicklung eine Rolle spielen. In Übereinstimmung mit dieser Hypothese werden in Gadd45-defizienten Trophoblast-Stammzellen Pluripotenzgene herunterreguliert.

Zusammenfassend wird in dieser Arbeit dargelegt, dass die Gadd45 Gene eine beschränkte Funktion in der Aktivierung DNA-methylierungsregulierter Gene in Stammzellen, sowie eine breitere Funktion während der Differenzierung und der Trophoblastenentwicklung ausüben.

3 Introduction

3.1 Early mouse development and its *in vitro* models

3.1.1 The zygote and its first fate decisions

3.1.1.1 Segregation of trophectoderm and inner cell mass

As for most animals, life of a developing mouse starts from a single totipotent cell – the zygote. This single cell will subsequently divide, self-organize and eventually transform into a self-sustaining adult animal, using a set of remarkably accurate mechanisms that confer cells with their identity. In order to achieve this transformation, the zygote has to create both the embryo proper as well as a set of nurturing and protecting extra-embryonic cells that will allow the developing embryo to survive and thrive. The segregation between embryonic and extra-embryonic cells is the first to occur during mammalian development and the decision making process involved in this segregation is therefore of central importance, both for the developing embryo and for the field of developmental biology [1].

After fertilization, the mouse zygote will divide three times over the course of 2.5 days to give rise to the 8-cell stage embryo, or morula (see figure 3.1 for a graphical overview). Up to this stage, the cells of the developing embryo – the blastomeres – remain totipotent and can produce all embryonic and extra-embryonic tissues [2,3]. Subsequently, the morula undergoes compaction through development of tight junctions between the cells [4]. The blastomeres thereby acquire a polarized apical surface that is isolated from the basolateral surface [5]. Already 50 years ago it was hypothesized that the exposure to either the outside or the inside of the embryo might be the cue that distinguishes the first two segregating cell lineages [6]. Indeed, following the 8-cell stage, the next two cell divisions happen asymmetrically. Daughter cells become either outer cells that inherit the polarized apical domain or inner cells that inherit the basolateral surface. This event will define their future fate [7]: On the one hand, outer cells will develop into the extra-embryonic trophectoderm (TE), which later will generate the embryonic part of the placenta and is important for the implantation of the embryo into the uterus. On the other hand, inner cells will give rise to the inner cell mass (ICM) that will later produce all cells of the developing embryo as well as some further extra-embryonic tissues.

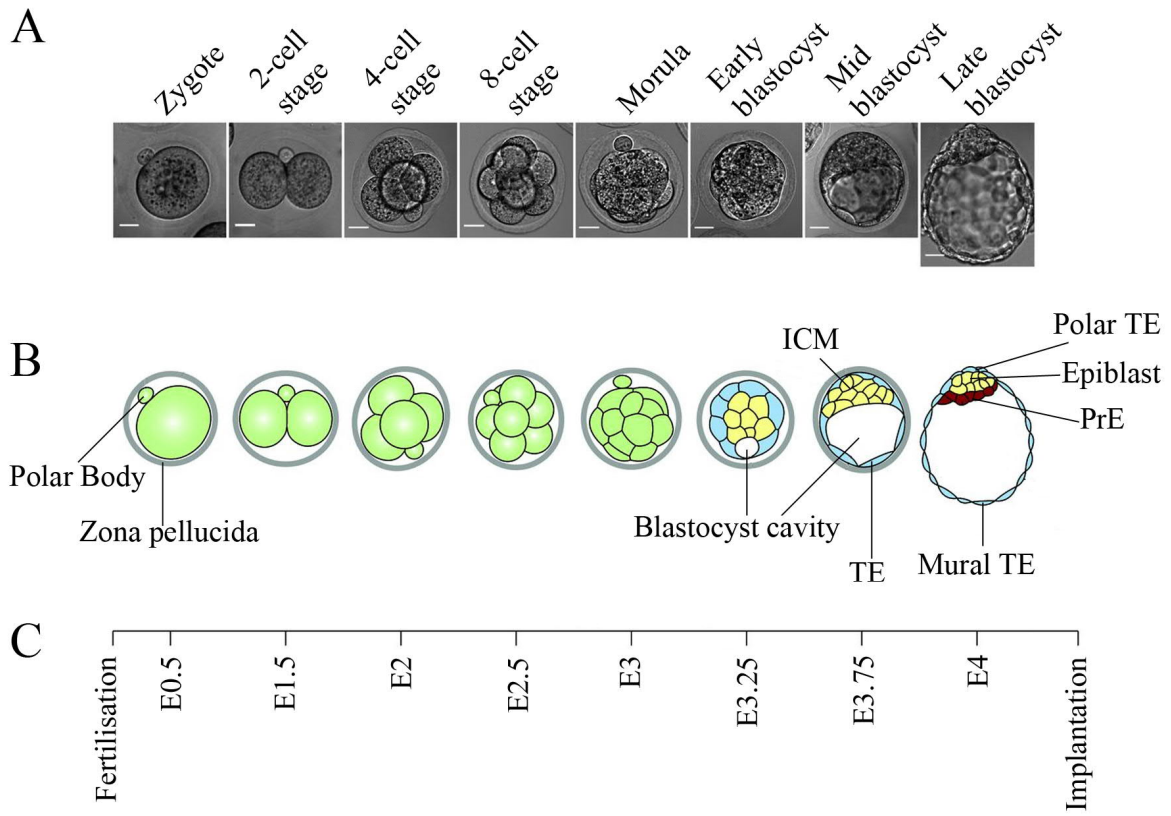


Figure 3.1: Stages of mouse preimplantation development.

A. Microscope images showing pre-implantation mouse embryos at indicated stages. **B.** Schematic illustrations of stages shown in (A). **C.** Time scale indicating days passed since fertilization for stages shown in (A) and (B). Note that the time scale is not linearly scaled. **Abbreviations:** ICM = inner cell mass; TE = trophectoderm; PrE = primitive endoderm; Figure based on Saiz and Plusa (2013)[1]

3.1.1.2 Segregation of primitive endoderm and epiblast

The second morphogenetic event after compaction is cavitation. Starting at embryonic day 3.25 (E3.25) roughly at the 30-cell stage, a small cavity – the blastocoel - is formed inside the developing embryo [8], which is now called the blastocyst. This arising asymmetry is again correlated with cell fate decisions: On the one hand, ICM cells found next to the blastocoel will give rise to the extra-embryonic primitive endoderm (PrE) that will later contribute to the visceral endoderm and the yolk sack. On the other hand, ICM cells further away from the blastocoel will generate the epiblast which will in turn produce the whole embryo proper, as well as the amniotic ectoderm and the extra-embryonic mesoderm (Figure 3.3A). Whether or not the positioning of the cells is the determinant of cell fate, like it is the case at the morula

stage, remains a controversial issue [1]. Original work proposed that cells in contact with the blastocoel give rise to the PrE by virtue of their position [9]. However, later work showed that the ICM is comprised of cells that will become PrE and cells that will become epiblast in a salt-and-pepper fashion and that cell segregation happens only after lineage determination [10]. Which event, if not cell positioning, determines which ICM cells will become PrE and which cells will become epiblast has not been fully resolved yet. Recent studies have indicated that whether cells become inner cells during the 16-cell or the 32-cell stage will determine their fate [11], although another study has found no such connection [12].

3.1.1.3 Transcription factors in charge of early development

The differentiation events in the early embryo are correlated with a concomitant loss of developmental potential. This is required because cells have to commit to increasingly specialized functions. Whereas e.g. the blastomeres at the 8-cell stage are totipotent, because they can give rise to all embryonic and extraembryonic tissues, the epiblast is only pluripotent, because it can no longer give rise to tissues derived from the TE and the PrE. Molecularly, this loss of developmental potential is realized by lineage specific transcription factors that enforce a certain lineage choice and steer the transition towards further differentiation.

Work in the last decades has greatly enhanced our understanding about the nature and function of these transcription factors. On the one hand, a core gene regulatory network surrounding the transcription factors Oct4 (also known as Pou5f1), Sox2 and Nanog is in charge of maintaining pluripotency in the ICM and the epiblast. Zygotic *Oct4* expression starts at the 2-cell stage and is later restricted to the ICM and the epiblast [13]. In its absence, cells fail to maintain ICM identity and differentiate into the TE lineage instead [14]. Sox2 and Nanog are similarly required for epiblast identity [15,16]. TE fate, on the other hand, is initiated by Tead4, which activates the transcription factors *Cdx2* and *Gata3* [17,18], which in turn are responsible for activating further TE-specific downstream genes[18,19].

Importantly, these gene regulatory networks do not exist in isolation, but are tightly interwoven. Oct4 and Cdx2 e.g. reciprocally inhibit each other, thereby limiting the developmental potential of the ICM and the TE respectively [20]. This is reflected by their expression patterns: Whereas both genes are co-expressed at the totipotent 8-cell stage, their expression is subsequently limited to the TE or the ICM respectively, once these more specialized cell populations arise [21]. However, although TE and ICM fate are mutually exclusive, they still depend on each other: Oct4 is needed for the expression of *Fgf4* in the ICM [14], which positively regulates TE proliferation in a paracrine fashion [22,23]. Thus, the described gene regulatory networks do not only allow separation of the different cell lineages, they also ensure their coordinated growth and further development in respect to each other.

3.1.2 Further differentiation of the inner cell mass and development of the embryo proper

As described above, the epiblast is the source for all cells of all embryonic tissues. During differentiation, it will give rise to the three primary germ layers: Ectoderm, mesoderm and endoderm. The ectoderm will later form the epidermis, the nervous system and derivatives of the neural crest. The mesoderm will give rise to the dermis, muscles, cartilage, bone, fat, the cardiovascular system (including the heart) and the kidneys. The endoderm will generate great parts of the digestive system, the lungs, the liver, the pancreas and the thyroid [24].

The initial segregation of epiblast cells into the three primary germ layers happens during gastrulation (Fig. 3.2). During this critical phase, cell movements, transcriptional networks and signaling cascades in concert result in the establishment of cellular identities and the setup of the basic body plan of the embryo. Importantly, all of these processes are tightly controlled not only from within the developing epiblast itself, but also from the surrounding extra-embryonic tissues. Due to their great complexity, these processes can only be described briefly here.

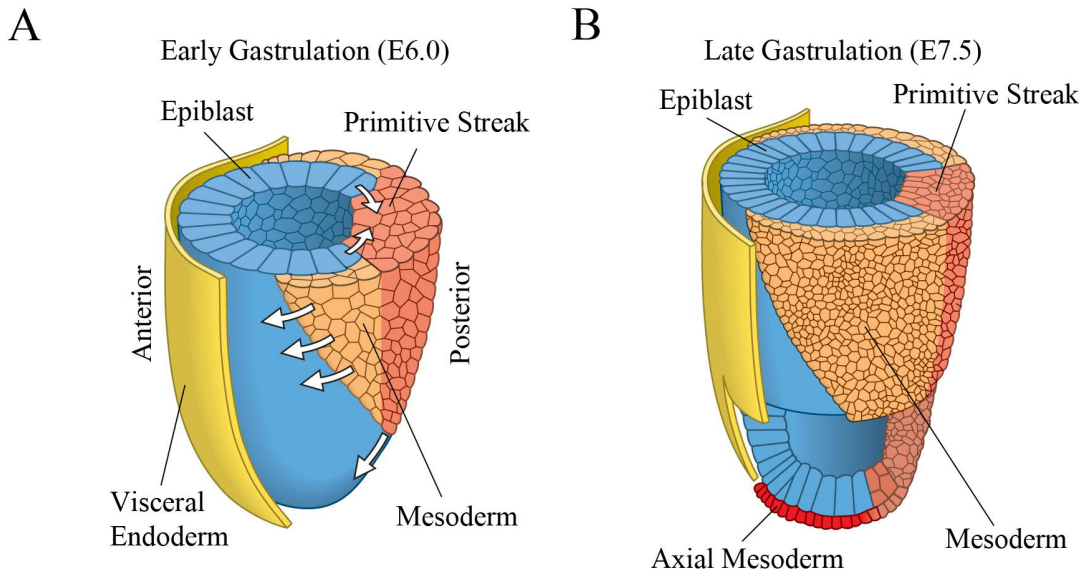


Figure 3.2: Graphic illustration of tissues and cell movements during gastrulation. See main text for detailed description of the shown process. Figure based on Wolpert, Principles of Development.

At the time point of gastrulation, the mouse epiblast is a single, cup shaped cell layer. Gastrulation starts with the formation of the primitive streak at the prospective posterior proximal epiblast, which is the part of the epiblast that is in contact with the TE derived extra-embryonic ectoderm. The primitive streak

functions as a gateway for cell migration. Individual cells of the epiblast undergo epithelial-mesenchymal-transition (EMT) and ingress through the primitive streak to a position below the original epiblast and above the PrE derived visceral endoderm. Following its appearance, the position of the primitive streak will move along the proximal-distal axis and cells will continue to migrate through it. The fate of the migrating cells is tightly linked to the time point and position of ingression: Cells ingressing first will later form the extraembryonic mesoderm, whereas cells ingressing later will in order become cardiac mesoderm, lateral plate mesoderm, paraxial mesoderm and axial mesoderm. Cells ingressing last will become the definitive endoderm cells, whereas cells that do not migrate through the primitive streak at all will become ectoderm [24].

A set of signaling gradients that pattern the developing embryo is interpreted by the cells to gain information about their position within the embryo. This in turn will steer the cells into their precise lineage. Major signaling pathways involved in this decision making process include Bmp4, produced from the extra-embryonic ectoderm at the proximal side, and Nodal, which instructs cells ingressing later at the distal side [25].

3.1.3 Development of the trophectoderm and differentiation of trophoblast lineages

3.1.3.1 General structure of the placenta

Whereas the ICM will ultimately give rise to the embryo itself, the TE will later form great parts of the placenta, which is an important organ supporting the developing embryo. It does not only provide the embryo with the necessary nutrients to grow, by acting as an interface between the fetal and the maternal body, but also acts as a source of hormones and growth factors that affect pregnancy and maternal blood flow [26,27]. In addition, it plays important immunoregulatory roles in order to protect the growing embryo from the maternal immune system [28].

The mature placenta consists of four layers (Fig. 3.3C): Towards the maternal side, the maternal decidua can be found, from where the maternal blood vessels find their way further into the placenta. The decidua is lined by a single layer of trophoblast giant cells (TGCs) that have remodeled the uterus and allowed implantation. Next is the spongiotrophoblast, which forms a structural, supporting layer. Closest to the embryo, the labyrinth layer is located. There, maternal and fetal blood vessels come into close proximity to allow the vital exchange of nutrients, gases and waste products.

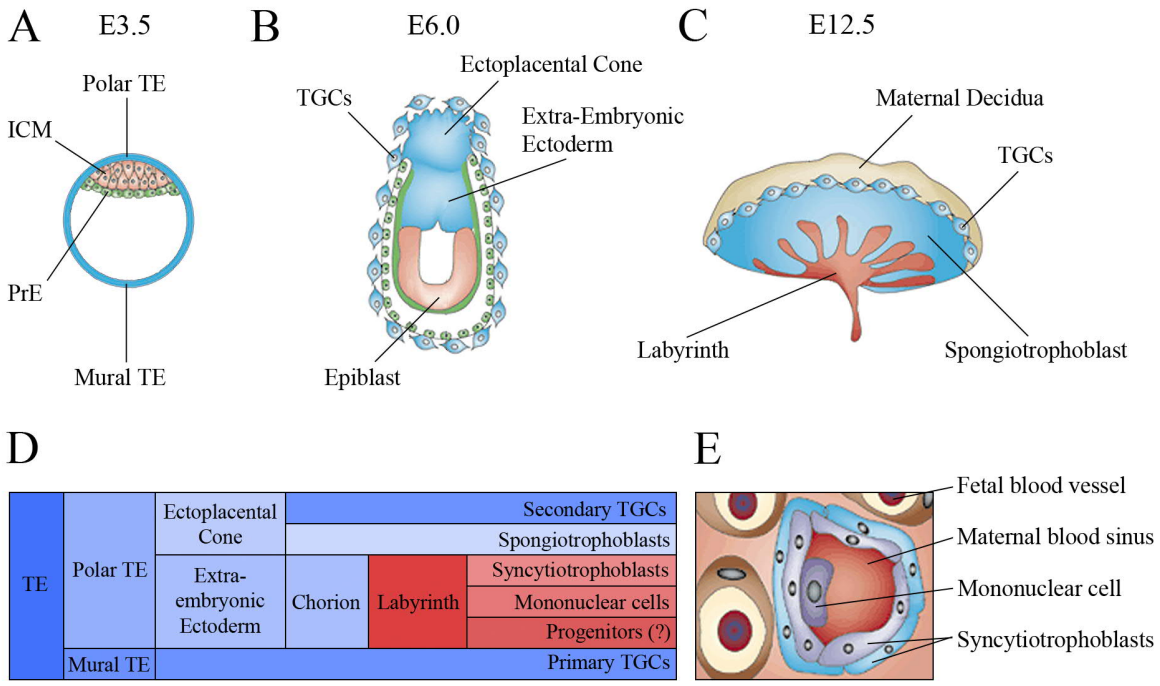


Figure 3.3: Overview over placental development.

A-C. Graphic illustration of placental development at three selected time points. **D.** Lineage scheme indicating the origin of the final cell types in the placenta. Color code based on (C). **E.** Graphic illustration of the labyrinth cell types and their positioning in relation to the maternal and fetal blood vessels. **Abbreviations:** ICM = inner cell mass; TE = trophectoderm; PrE = primitive endoderm; TGCs = trophoblast giant cells; Figure A-C,E based on Rossant and Cross (2001)[29].

Placental cells have three origins that partially relate to which layer they are found in [29]: First, the decidua is of maternal origin and arises from the uterus. Second, embryonic blood vessels within the labyrinth layer are generated from the extra-embryonic mesoderm. Third, the remainder of the placenta, which comprises most of the functional and structural components, originates from the TE and is called the trophoblast.

3.1.3.2 Origin of the different TE derived placental structures

As described in detail in chapter 3.1.1.1, the trophoblast is the first extra-embryonic cell lineage to segregate from the prospective embryo proper and forms a single epithelial structure on the outside of the E3.5 blastocyst, called the TE. One day later, around the time of implantation, the trophoblast becomes clearly asymmetric (Fig. 3.3A). The mural TE that is positioned near the blastocoel and away from the ICM stops dividing and instead forms the first set of polyploid TGCs via endoduplication [30]. TGCs are formed multiple times at different positions during placental development and have important functions

for the implantation of the blastocyst, the remodeling of the uterus and the secretion of hormones [31]. On the other side of the blastocyst, the polar TE that resides on top of the ICM continues to proliferate and forms two new structures: At a position proximal to the ICM the extra-embryonic ectoderm, and more distally the ectoplacental cone are generated [32] (Fig. 3.3B).

The extra-embryonic ectoderm further develops into an epithelial structure called the chorion. The chorion subsequently makes contact with the underlying allantois, which arises from the extra-embryonic mesoderm, in a process called the chorioallantoic fusion [29]. This process marks the beginning of the development of the labyrinth layer. The allantoic part generates the fetal blood vessels, whereas the chorionic part generates the syncytiotrophoblasts. Two layers of these polyploid cells separate the fetal blood vessels from the maternal blood sinuses and function as the main transport surface for nutrients and gases [33]. In addition, the labyrinth contains a type of mononucleated cells of unknown origin or function [29].

The ectoplacental cone gives rise to the spongiotrophoblast that structurally supports the underlying labyrinth layer. Additionally, it generates a secondary set of TGCs that together with the primary TGCs from the mural TE will eventually surround the whole developing embryo [32]. A schematic diagram summarizing the different lineages and origins of trophoblast cells can be seen in figure 3.3D.

3.1.3.3 Regulation of trophoblast stem cell maintenance and differentiation

TGCs, spongiotrophoblasts and syncytiotrophoblasts are terminally differentiated, postmitotic cells [34]. Therefore, in order to create a functional placenta, a balance between differentiation and proliferation of a stem cell pool has to be kept. Initially, the proliferation of stem cells in the polar TE and later in the extra-embryonic ectoderm is maintained by Fgf4 signaling from the ICM and the epiblast [35]. Downstream of the Fgf receptor Fgfr2, which is expressed by the TE [36], Erk2-Mapk signaling ensures expression of the trophoblast stem cell transcription factor *Cdx2* [37]. In addition, Nodal signaling by the epiblast similarly induces stem cell maintenance genes [38]. Cdx2 is important for trophoblast stem cell maintenance and proliferation from the very beginning of the TE lineage on. Other trophoblast stem cell factors, including Eomes, Elf5 and Tcfap2c, fulfill similar roles in concert with Cdx2 during later stages of development, e.g. during the expansion of the extra-embryonic ectoderm [39–41].

Later, when the extra-embryonic ectoderm is no longer in contact with the epiblast, the stem cell inducing factors seem to originate from a transient cavity next to the ectoplacental cone instead [42]. Only around E8.5, when the contact to this ectoplacental cavity is occluded as well, *bona fide* trophoblast stem cells can no longer be identified [42]. The fact that the placenta continues to grow throughout development

despite the absence of trophoblast stem cells suggest the existence of a so far uncharacterized population of progenitor cells [34].

Trophoblast stem cells rapidly differentiate into TGCs in the absence of Fgf signaling, indicating that one important function of Fgf4 is the active suppression of TGC differentiation [23]. Because of this, TGC development is often considered to be the default differentiation pathway of trophoblast stem cells[34]. Differentiating trophoblast cells usually pass through a transient progenitor phase marked by the expression of *Mash2* and *Tpbpa* [43–45], before becoming terminally differentiated TGCs that express effector genes such as hormones or metalloproteases. Whether this is an obligatory step for all TGCs however is not fully resolved, because TGCs are less homogenous than initially thought [46] and are generated from different progenitor cells during development (Fig. 3.3D).

Spongiotrophoblast development is overall only poorly understood [34]. Spongiotrophoblast precursor cells are located in the ectoplacental cone, which is also the source of the secondary TGCs. Several mutations, including one in the gene *Mash2,* have opposing effects on the spongiotrophoblast and TGC populations, suggesting that both might share a common precursor that can differentiate along either lineage [47–49].

The labyrinth, as mentioned in chapter 3.1.3.2, consists of different cell types that occupy fixed positions in relation to the maternal and the fetal blood vessels (Fig. 3.3E). Syncytiotrophoblast precursors most likely originate from the chorion, considering overlapping expression profiles of the chorion and the labyrinth [50–53]. An important regulator of syncytiotrophoblast differentiation is the transcription factor *Gcm1*, which starts to be expressed around E7.5 in the chorion [50]. Gcm1 is necessary to exit cell cycle and to initiate branching morphogenesis, which allows the underlying allantois access into the prospective labyrinth [44,54]. Although its initial inducer is unknown, maintenance of *Gcm1* expression depends on contact to the allantois [55].

Little is known about the origin of the second type of labyrinth cells, the sinusoidal mononuclear cells. They express *placental lactogen II*, a hormone usually expressed by TGCs, leading to the proposal to define these cells as a subpopulation of TGCs [46].

Interestingly, some clusters of cells in the developing labyrinth express the marker *Eomes* [56], and some express the marker *Ehox* [57], although it is currently not known whether these cells are identical. *Eomes* and *Ehox* are usually expressed in the proliferating extra-embryonic ectoderm. This indicates that these cells might represent populations of proliferating precursors also at later stages within the labyrinth.

3.1.4 *In vitro* models of early mouse development

Studying early mouse development is challenging due to only few embryos obtainable per litter and due to the experimental restrictions associated with uterine development. Additionally, the first lineage segregations happen at stages when the embryo contains only a few dozen cells, yielding very little material for experimentation. Fortunately, work in the last decades has shown that *in vitro* cell culture models from all major early lineages can be established and that these models resemble their *in vivo* counterpart remarkably well. Studying these cell lines has greatly enhanced our understanding about early mouse development.

3.1.4.1 Mouse embryonic stem cells

The first cell line established from an early lineage were mouse embryonic stem cells (mESCs). In 1981 Evans and Martin succeeded to derive pluripotent mESCs from the ICM, based on earlier work on embryo carcinoma cells [58,59]. mESCs retain their pluripotency and proliferation potential seemingly indefinitely, even when colonies are grown from single cells. Furthermore they are able to contribute to all embryonic tissues of a chimeric animal after blastocyst injection – even the germ line, a feat that no cell line ever consistently achieved before [60]. Importantly, mESC can be used as a vehicle to introduce gene modifications into animals [61], because cells carrying rare induced genomic recombination events can be selected for *in vitro* and subsequently be used to create transgenic mice strains via injection into blastocysts.

In addition to revolutionizing the field of genome engineering, mESCs also allowed studies of the nature of pluripotency and differentiation. Much of our detailed knowledge about the transcriptional networks in charge of stem cell behavior is derived from experiments in mESCs: Functional screens, overexpression and loss-of-function studies were used to identify key players of pluripotency, whereas approaches such as ChIP-Seq provided unprecedented insight into the mechanisms at work.

A core triad of transcription factors responsible for mESC pluripotency is built from the same factors responsible for the initial fate decisions in the early mouse embryo, highlighting the similarities between the *in vitro* model and the *in vivo* situation: ChIP-Seq experiments revealed that Oct4, Sox2 and Nanog together co-occupy regulatory regions and thereby activate genes that are needed for stem cell maintenance and repress genes needed for differentiation [62–64]. Importantly, they also cross-regulate each other's expression, thereby providing the molecular basis for a sustained stem cell preservation [65]. However, elevating Oct4 or Sox2 levels beyond physiological boundaries induces differentiation, pointing out the importance of keeping an exquisite balance in order to maintain the stem cell state [20,66].

mESC show highest similarity with the early naïve epiblast on a transcriptional level, despite usually being derived from the ICM [67]. Even when compared to the epiblast, however, mESC still differ from the physiological situation in a few respects. First and foremost, the epiblast is only a transient structure, whereas mESCs will continue to proliferate and remain pluripotent throughout their lifetime. This has raised the important question whether mESC are a captured representation of the early epiblast, or whether they are an artificial creation of the culture conditions applied to them [68].

Under classical conditions, mESCs are cultured in the presence of fetal bovine serum and LIF. The cytokine LIF binds to its receptor on mESCs and subsequently leads to an activation of the JAK/Stat signaling pathway, which feeds into the core pluripotency network [69]. The main pluripotency promoting factor found within serum is BMP [70]. BMP alone induces non-neuronal differentiation [71], but promotes pluripotency in combination with LIF by inducing *Id* genes, which suppress differentiation [70]. The main driver of differentiation, which needs to be suppressed, is autocrine Fgf signaling [72], which in the absence of serum induces neuronal development [73]. The pluripotent state of mESCs in regular culture conditions is thereby defined by a balance of pro- and anti-differentiation signals. Importantly however, embryos devoid of LIF or its receptor complex develop normally until the late blastocyst stage, casting doubt on whether these culture conditions faithfully resemble the situation in the early epiblast [74].

Indeed, recent work has shown that different culture conditions result in mESCs that more closely resemble the early epiblast. This is achieved by a combination of Mek/Erk and Gsk3 inhibitors in the presence of LIF, but absence of serum [75]. The Mek/Erk inhibitor suppresses the Fgf-mediated differentiation signals, thereby eliminating the need for BMP. At the same time, the absence of serum eliminates non-autocrine differentiation signals. GSK3 inhibition enhances mESC proliferation and viability, which can be at least partially attributed to its effect on the Wnt signaling pathway [76,77].

Together these two inhibitors in combination with LIF (a treatment dubbed "2i" [75]) are suggested to move the mESCs towards "ground state pluripotency". This state is characterized by profound transcriptomic changes, particularly the downregulation of lineage-affiliated genes and an upregulation of metabolic genes [78]. Importantly, mESCs under regular culture conditions show mosaic expression patterns across the cell population, meaning that some cells show high levels of pluripotency marker gene expressions, whereas others do not [79,80]. This is not the case for mESCs cultured in 2i conditions, which express genes very homogenously and also show a much more homogenous morphology [76]. Last but not least, the epigenome of mESCs in 2i conditions becomes more similar to that of the early epiblast [81–83]. Taken together these results indicate that 2i treated mESCs might represent a naive pluripotent state, i.e. a basal

proliferative state free of epigenetic restrictions and without external stimuli pushing the cells towards differentiation.

3.1.4.2 *In vitro* differentiation of mESCs

In addition to allowing studies about the nature of pluripotency, mESC also allow studies of differentiation processes. Usually, physiological differentiation conditions are mimicked *in vitro* using a combination of different culture conditions (such as cellular aggregates, stromal cell feeder layers or culturing on extracellular matrices) and different signaling molecules. Using such approaches, mESCs can differentiate into virtually any cell type of the body, although efficiency and purity vary greatly depending on the desired cell type [84].

When grown in suspension in the absence of LIF but presence of fetal bovine serum, mESC will form three dimensional floating aggregates, dubbed "embryoid bodies" (EBs) due to their morphological resemblance to blastocysts [85]. Even more, similar to blastocysts, they will self-organize to a certain degree and ultimately give rise to derivatives of all three germ layers [85]. This self-organization is also one of the main advantages of the system, because it allows development of intense cell-cell-interactions. However, EBs on their own will usually not form advanced organ-like or even embryo-like architectures, with few reported exceptions like optic-cup development [86].

Building on the principle of EB differentiation, many studies focused on guiding the differentiating cells within the EBs along desired lineages, either by adding defined growth factors into the culture medium, or by using transgenic mESC lines [87]. Using such strategies, it was possible to generate a plethora of different cell types *in vitro*, including (but not limited to) cardiomyocytes [88], adipocytes [89] and haematopoietic precursors [90]. An exceptionally pure population of neuronal precursors can be created by application of retinoic acid during the EB differentiation [91].

Besides EB based differentiation, mESCs can also be differentiated when attached as a single layer of cells to the cell culture dish, which can be advantageous for imaging based studies. Of particular note is a differentiation protocol that is solely based on endogenous signaling by the mESCs themselves. When grown under LIF and serum free conditions, mESCs will predominantly differentiate along the neuronal lineage, due to autocrine Fgf signaling [73]. This differentiation is less synchronous and produces less pure populations compared to retinoic acid treated EBs, but allows differentiation into mature neurons with greater ease.

When using these models to study embryology, it is of great importance whether or not these *in vitro* differentiation processes represent a physiologically relevant state. Fortunately, it seems as if *in vivo*

developmental programs can be remarkably faithfully preserved in the petri dish, as judged by e.g. the timing of expression of important lineage specific transcription factors [88,92]. Also, the resulting differentiated cells show many hallmarks of functional cells, such as periodic contractions of cardiomyocytes [93] or membrane potentials of neurons [94].

3.1.4.3 Mouse trophoblast stem cells

In a manner highly analogous to mESCs derivation, mouse trophoblast stem cells (mTSCs) can be isolated from the early blastocyst as well [23]. In accordance with their maintenance *in vivo*, indistinguishable mTSCs can be derived during early development from the polar TE and later on from the extra-embryonic ectoderm, up to E8.5 [42]. mTSCs continue to proliferate and maintain their stem cell state apparently indefinitely, if cultured in the presence of Fgf4, serum and mouse embryonic fibroblast (MEF) conditioned medium [95].

The preserved multipotency of mTSCs can most strikingly be shown by their ability to contribute to chimeric placentas upon blastocyst injections [23]. However, similar to mESCs, they also possess the ability to realize their developmental potential *in vitro*: mTSCs will differentiate into spongiotrophoblasts, syncytiotrophoblasts and TGCs, once Fgf4 and MEF conditioned medium are withdrawn [23].

The transcriptional network required for maintenance of mTSCs *in vitro* is highly similar to the one *in vivo*. A core network of transcription factors, including Cdx2, Elf5 and Eomes, is important for mTSC maintenance *in vitro* and mTSCs cannot be derived from blastocysts deficient for these factors [19,39,41]. Moreover, ChIP-on-chip and ChiP-Seq experiments have shown that these factors directly activate each other's expression [96,97]. This provides a mechanistic base for the maintenance of the stem cell state, conceptually similar to the situation in mESCs. Additional targets bound by Cdx2, Elf5 and Eomes include both genes uniquely expressed in mTSCs, as well as genes involved in proliferation and pluripotency also in mESCs. This raises the interesting possibility that an identical network is responsible for the general stem cell state in both mESCs and mTSCs, whereas lineage identity is conveyed by an additional, cell specific network [96].

3.1.4.4 Lineage restrictions of early embryonic cell lines

Analysis of the aforementioned cell lines did not only allow the investigation of physiological pluripotency and differentiation conditions, but also supported studies about lineage commitment and reprogramming. This is possible because the lineage restrictions that are imposed upon the embryo at the time point of cell line derivation are usually faithfully memorized and maintained also *in vitro*.

mESCs, e.g., will readily contribute to all embryonic tissues if injected into a host blastocyst, but they will usually not contribute to the extraembryonic tissues such as the TE or the PrE [98]. This fits with the presumed developmental stage of the mESCs, which is the naïve epiblast stage[67]. Similarly, as a model of TE development, mTSCs will contribute to the placenta if injected into a host blastocyst, but not to the embryonic tissues [23].

Although this concept is generally broadly accepted, it should be noted that it might be oversimplifying to some degree. First, mESCs injected into blastocysts were found to be sometimes contributing also to placental tissues [99]. Although this contribution was only minor, and did not occur if only single mESCs were injected, it nevertheless shows that mESCs keep at least some potential for transdifferentiation into the trophoblast lineage. Second, stem cells isolated from later stages of epiblast development paradoxically show a higher propensity than mESCs to differentiate along the trophoblast lineage *in vitro* [100], even though they should in theory be more restricted in their lineage choice.

Although lineage commitment is mostly permanent under physiological conditions, it can be reversed experimentally, as shown maybe most strikingly by the work in the Yamanaka lab [101]: If *Oct4*, *Sox2* and *Nanog* are overexpressed in terminally differentiated cells together with *c-myc*, cells will revert to a fully pluripotent state called iPSCs (induced pluripotent stem cells). Likewise, if important TE lineage determinants, such as *Cdx2* or *Elf5* are overexpressed in mESCs, cells will transdifferentiate into the trophoblast lineage [102,103]. Transdifferentiation can also be induced by knockdown of *Oct4*, showing that lineage-specific transcription factors are sometimes constantly required to enforce lineage commitment [104]. Interestingly, some recent studies claim that simple stimulation with certain signaling factors, such as Bmp4 or Wnt3a, might be sufficient to similarly induce a transdifferentiation of mESCs into the trophoblast lineage [105–107], although these results might still be considered controversial [108].

One important barrier that needs to be overcome by any transdifferentiation process - be it physiological, transgene or chemically induced – is epigenetic gene regulation, in particular DNA methylation. This concept will be introduced in more detail at the end of the following chapter.

3.2 DNA methylation and demethylation processes and pathways

3.2.1 DNA methylation

DNA methylation is a well described epigenetic mechanism, in eukaryotes associated mostly with gene silencing. In animal DNA, cytosine can be methylated at its carbon 5 primarily in a symmetric fashion in the context of palindromic CpG dinucleotides [109,110]. The initial establishment of this epigenetic mark is

carried out by the *de novo* DNA-methyltransferases Dnmt3a and Dnmt3b [111]. Once set, it is maintained semi-conservatively during DNA replication: Dnmt1 in complex with Uhrf1 recognizes hemi-methylated DNA and faithfully methylates the cytosine in the nascent DNA strand to restore a fully methylated state [112,113].

Once established and maintained, 5-methylcytosine (5mC) can be recognized by methyl-binding proteins, such as MBD2, MBD3 or MeCP2 [114,115]. These proteins in turn recruit various epigenetic modifiers, such as histone deacetylases, which ultimately convey the repressive function of 5mC [116,117]. In addition to this predominant mechanism, 5mC can also directly inhibit binding of methylation-sensitive transcription factors [118] and influence nucleosome positioning [119]. In combination with the abovementioned semi-conservative maintenance, these mechanisms allow stable long-term silencing of genes in a cell and all its progeny.

Methylation patterns in most tissues are bimodal. That is, in any given cell population most CpGs are either strongly methylated or barely methylated, whereas only very few CpGs show intermediate methylation levels on a population scale [120].

In somatic cells, the majority (60-80%) of CpGs are fully methylated [121]. This genome wide transcriptional repression mechanism most likely serves two purposes. First, it represses transcription in transposable and repetitive elements and is therefore of central importance for genomic stability [122,123]. Second, a large fraction of the genome is expressed in a tissue specific manner and therefore needs to be repressed in the majority of situations [124]. Paradoxically however, high methylation levels can also be observed in gene bodies of actively transcribed genes [125]. In contrast to its usually repressive function at regulatory elements, the function of 5mC at these regions remains to be fully understood [126].

Unmethylated CpGs are often located in so called CpG islands [127]. These CpG-rich genomic regions can be found in promoters of transcriptionally active genes, particularly housekeeping genes that are active in a tissue-independent fashion. These regions are protected from methylation via the Dnmt-repulsive action of certain histone marks, e.g. H3K4me3, and histone variants such as H2A.Z [128,129].

Lowly methylated regions (LMRs) which show intermediate levels of methylation are often positioned within CpG-poor distal regulatory elements [130]. They overlap with histone marks typically found at enhancers and tissue specific transcription factor binding sites [130]. LMR methylation levels change dynamically during differentiation, unlike CpG islands which are usually unmethylated irrespective of the cell type analyzed, making it likely that methylation at these sites serves gene regulatory purposes.

3.2.2 DNA demethylation

In the past, DNA methylation has been considered to be a very stable epigenetic modification that only is reversed during very particular developmental time windows (see chapter 3.2.3.1). However, work in the recent years has shown that DNA methylation can undergo dynamic changes similar to other epigenetic modifications.

DNA methylation can be reversed passively by preventing Dnmt1-mediated maintenance methylation. This can occur either genome wide, by e.g. downregulating *Dnmt1* expression, or site-specifically by repelling the Dnmt1 protein [131]. With each subsequent cell division, 5mC will continuously be out-diluted due to the incorporation of unmethylated cytosine into the nascent DNA strand. Due to its very nature, passive DNA demethylation is limited to proliferative cells.

The idea of physiologically relevant active, rather than passive, DNA demethylation has long been a controversial issue. This was both because of the thermodynamic difficulties of removing the methyl group directly and because of countless unsuccessful tries to identify DNA demethylases in animals [132]. By now however, a biochemically and biologically well validated pathway of active DNA demethylation has been established (Fig. 3.4).

Three related DNA dioxygenases (Tet1, Tet2 and Tet3) are able to oxidize 5mC to 5-hydroxymethylcytosine (5hmC) [133,134]. 5hmC can then be further oxidized by the Tet enzymes to 5-formylcytosine (5fC) and subsequently 5-carboxylcytosine (5caC) [135,136]. On the one hand, this can result in passive demethylation even in the presence of the DNA methylation machinery, because Dnmt1 does not efficiently methylate hemi-hydroxymethylated CpGs [137]. On the other hand, 5fC and 5caC can be directly removed by the glycosylase TDG [135,138]. The resulting abasic site is repaired via the base excision repair (BER) pathway, culminating in the incorporation of an unmethylated cytosine and therefore a *bona fide* active DNA demethylation. Moreover, 5hmC, 5fC and 5caC might also be epigenetic marks of their own, rather than being only intermediates of active DNA demethylation, as judged by the fact that they appear to have specific “readers” [139].

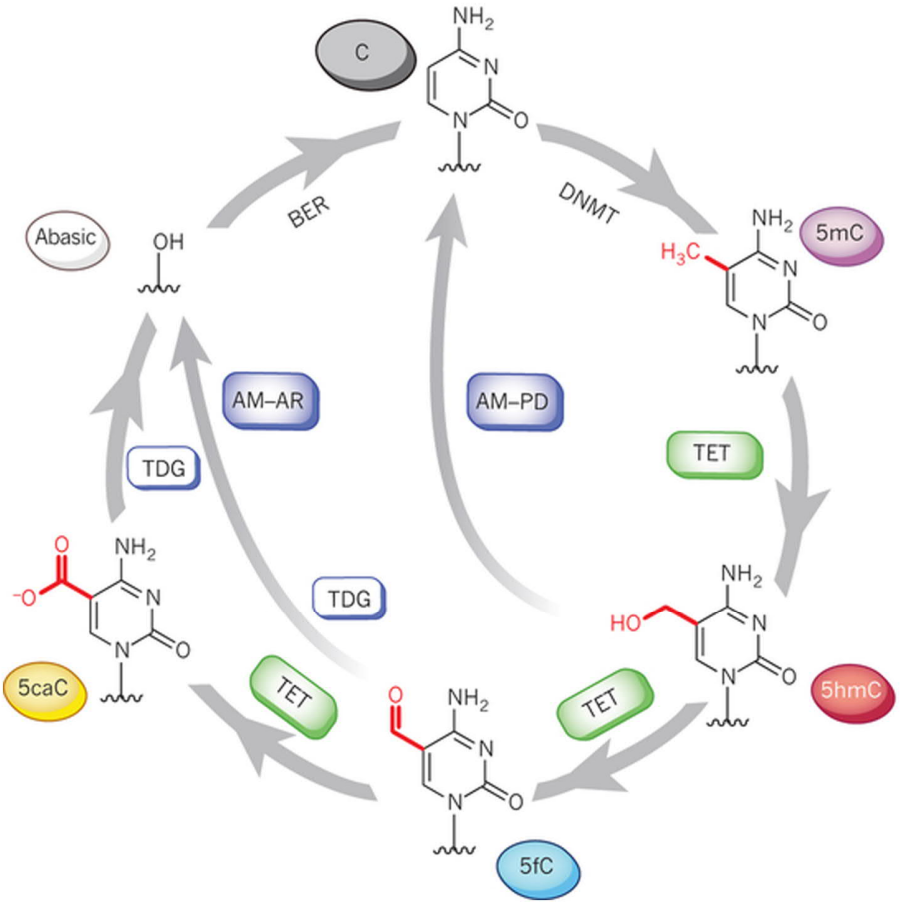


Figure 3.4: The Tet cycle of active DNA demethylation.

Scheme illustrating the cycle of iterative modification of cytosine (C) to 5-methyl-cytosine (5mC), 5-hydroxymethyl-cytosine (5hmC), 5-formyl-cytosine (5fC) and 5-carboxyl-cytosine (5caC) as well as their repair based exchange back to C. **Abbreviations:** BER = base excision repair; AM-PD = active modification – passive dilution; AM-AR = active modification – active removal; DNMT = DNA methyltransferase; TET = ten-eleven-translocation; TDG = thymine DNA glycosylase; Figure from Kholi and Zhang (2013)[140]

Other pathways of active DNA demethylation have been proposed, including a direct removal of the methyl moiety by MBD2 [141], a glycosidic bond cleavage of the 5mC nucleotide itself by TDG or MBD4-coupled BER [142,143], a deamination of 5mC to thymine by AID/APOBEC and subsequent BER of the resulting T:G mismatch [144,145], or a nucleotide excision repair (NER) based removal of 5mC [146,147]. This might indicate that multiple DNA demethylation pathways exist and might be employed in a cell- and context specific manner, although these findings did not receive the same level of corroboration from other studies as did the Tet-mediated DNA demethylation pathway.

3.2.3 DNA methylation and demethylation during differentiation processes

3.2.3.1 DNA methylation during embryonic development

During the development of the mouse embryo, drastic changes in DNA methylation occur. Before fertilization, the sperm and the oocyte DNA are marked by highly specialized methylation patterns that allow expression of germline genes, repress non-germline genes and mark imprinted regions in a parent-specific manner [148]. Also, the overall methylation levels in sperm are considerably higher (around 90% of the CpGs) than in oocytes (around 40% of the CpGs) [149]. Before onset of the first cell division of the fertilized zygote, the 5mC in the male pronuclear DNA undergoes genome wide oxidation mediated by Tet3, potentially to compensate for the initial differences in methylation compared to oocytes [150]. During subsequent cell divisions, both the male and female pronucleus derived chromosomes undergo passive demethylation/demodification due to an exclusion of Dnmt1 from the nucleus [151]. This erasure of 5mC (female) or its oxidative derivatives (male) from the DNA is crucial to reacquire totipotency and occurs genome wide, with the exception of imprinting control regions (ICR) and certain retroviral elements [152,153].

Methylation patterns are then established in their typical bimodal pattern *de novo* by Dnmt3a and Dnmt3b during implantation [111]. Whether a genomic region will be methylated or not depends mostly on the underlying DNA sequence and the CpG density, such that e.g. CpG islands are not methylated [154,155]. Notably, the overall methylation levels are lower in the TE than in the ICM derivatives [156].

During later stages of development, tissue specific *de novo* methylation or demethylation can occur site specifically in order to silence or activate lineage specific genes. This includes e.g. the silencing of pluripotency genes in derivatives of the epiblast [157], the activation of lineage specific genes during hematopoiesis[158] or the activation of stimulus-responsive genes in post mitotic neurons [159,160].

A special case of lineage specific DNA methylation change occurs during the specification and differentiation of primordial germ cells (PGCs). Around E10, after the embryonic bimodal methylation patterns have been established, PGCs undergo genome wide demethylation [161]. Most likely, this includes both Tet1-mediated active demethylation [162] and passive demethylation due to downregulation of *Uhrf1* [163]. The genome wide erasure of methylation ensures that ICRs are reset and can be reestablished in a parent-specific manner. Additionally it allows the reactivation of germ line specific genes, which usually are silenced in somatic cells. During further differentiation, Dnmt3a then remethylates the PGC genome *de novo* to establish sperm or oocyte specific methylation patterns [164].

3.2.3.2 DNA methylation in stem cells and during lineage commitment

Lineage commitment and DNA methylation are tightly interconnected, as shown exemplarily by work in mESCs. *Dnmt1*$^{-/-}$ *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$ triple knockout mESCs are viable and will proliferate without major changes in their transcriptome, despite being completely devoid of DNA methylation [165]. When differentiated to EBs, however, they fail to completely shut down transcription of pluripotency factors such as *Oct4* and *Nanog* [166] and will ultimately undergo apoptosis [167]. Conversely, once methylation marks are established during differentiation, they will impose major barriers to reprogramming [168]. Because of this, reprogramming somatic cells to iPSCs can greatly be enhanced by adding demethylation inducing agents such as 5-azacytidine [169].

Similarly, DNA methylation is a major roadblock for mESC to mTSC transdifferentiation. In particular, *Elf5* has been identified as an important gatekeeper [102]. The *Elf5* promoter is unmethylated and the gene is transcribed in mTSCs, whereas it is repressed and methylated in mESCs. Although mESC-to-mTSCs-transdifferentiation can be induced experimentally by forced transcription factor overexpression, reprogramming is usually incomplete [170]: Promoters of several genes, including *Elf5*, fail to be completely demethylated. Interestingly, mESCs deficient for DNA methylation can differentiate along the trophoblast lineage, potentially due to the lack of methylation at the *Elf5* promoter [102]. *In vivo*, DNA methylation differences at specific important regulatory regions only occur after the segregation of TE and ICM, arguing against a role of DNA methylation in the specification, but rather for a role in the maintenance of TE versus ICM fate [171].

3.3 The family of Gadd45 proteins and their role in active DNA demethylation

3.3.1 General functions of Gadd45

The Gadd45 (Growth arrest and DNA damage-inducible) gene family encodes small, highly acidic proteins with a mass of around 18 kDa and an isoelectrical point of roughly pH 4.1 [172,173]. They belong to the L7Ae/L30e/S12 superfamily which is comprised of RNA binding proteins. In mammals, three Gadd45 members have been described: Gadd45a, Gadd45b (also called MyD118) and Gadd45g (also called CR6). They show a roughly 55% identity in amino acid sequence [173] and are evolutionary conserved, and a Gadd45 homolog exists e.g. in *Drosophila* [174].

Gadd45 proteins do not have any known enzymatic activity, but nevertheless fulfill a plethora of different functions in the cell, mostly mediated via protein-protein interactions:

One well studied function of Gadd45 proteins is the activation of MAPK signaling. Interaction of any family member with MEKK4 induces autophosphorylation of MEKK4, which subsequently activates the p38/JNK signaling branches [175]. Also, a direct activation of p38 by Gadd45a has been reported [176]. Gadd45-mediated activation of this signaling pathway can induce apoptosis or cell cycle arrest upon stress [175], or can regulate differentiation during embryonic development [177].

Another function of Gadd45 is regulation of the cell cycle. All three Gadd45 family members interact with the cdc2/cyclinB1 complex, ultimately leading to inhibition of its function and subsequent activation of the G2/M cell cycle checkpoint [178]. Furthermore, Gadd45 proteins interact with the cyclin-dependent kinase inhibitor p21, which similarly could influence cell cycle progression [179,180].

Importantly, Gadd45 proteins also play a role both in BER- and NER-mediated DNA repair: GADD45A is needed to facilitate removal of abasic sites during BER of methyl methanesulfonate-induced base damage, potentially by promoting APEX1 interaction with PCNA [181], which GADD45A interacts with [182]. Furthermore GADD45A helps to relax the chromatin around sites of UV damage, potentially via direct interaction with histones, and thereby facilitate access of DNA to NER proteins [183]. In line with this, Gadd45a deficiency is correlated with increased UV sensitivity and reduced NER capacity in a number of different cells [184,185].

Whenever specifically tested, the different Gadd45 family members seem to have mostly similar and overlapping functions [186]. However, they show partially different expression patterns during embryonic development. *Gadd45a* is expressed predominantly in the tip of the closing neural tube, in cranial and dorsal root ganglia and in somites [187]. *Gadd45b* is only marginally expressed during the development of the embryo proper [187], limited to the brain and to small areas in the developing limbs, but is expressed in the adult brain [188]. *Gadd45g* is expressed most strongly in neural precursors and post-mitotic neurons, but also in cranial and dorsal root ganglia and the presomitic mesoderm [187]. In agreement with their partially neuronal expression patterns in the mouse embryo, Gadd45a and Gadd45g are required for the exit from pluripotency and the regulation of neural development in *Xenopus* [189].

Consistent with their various functions and their specific expression patterns, mice deficient for the different Gadd45 family members show diverse phenotypes, which are summarized in table 1. They range all the way from genomic instability [190], tumorigenesis [191], immune deficiencies [192,193], neural tube closure defects [190], decreased bone growth [194], altered learning behavior [195,196] to sex reversal [197]. Importantly, triple knockout mice have not been successfully bred in sufficient numbers to analyze their phenotype [198].

Any defect resulting only from a combined loss of all three Gadd45 family members would therefore have been missed so far.

Genotype	Reported Phenotype	Publication
Gadd45a$^{-/-}$	genomic instability, increased radiation carcinogenesis and a low frequency of exencephaly	Hollander et al. (1999)[190]
Gadd45a$^{-/-}$	increased skin UV sensitivity and tumorigenesis	Hildesheim et al. (2002)[191]
Gadd45a$^{-/-}$ *BRCA*$^{\Delta 11/\Delta 11}$	increased exencephaly frequencies	Wang et al. (2004)[199]
Gadd45a$^{-/-}$ MMTV-*Ras*	accelerated Ras-driven mammary tumor formation	Tront et al. (2006)[200]
Gadd45a$^{-/-}$ *Trp53*$^{-/-}$	increased exencephaly frequencies	Patterson et al. (2006)[201]
Gadd45a$^{-/-}$ *XPC*$^{-/-}$	increased XPC-loss induced lung tumor progression	Hollander et al. (2005)[202]
Gadd45a$^{-/-}$, *Gadd45b*$^{-/-}$	increased sensitivity to genotoxic-stress-induced apoptosis in the bone marrow	Gupta et al. (2005)[203]
Gadd45a$^{-/-}$, *Gadd45b*$^{-/-}$	impaired stress responses of myeloid cells to acute stimulation with differentiating cytokines, myelo-ablation and inflammation	Gupta et al. (2006)[204]
Gadd45b$^{-/-}$	defective mineralization and decreased bone growth	Ijiri et al. (2005)[205]
Gadd45b$^{-/-}$	deficit in long-term contextual fear conditioning	Leach et al. (2012)[195]
Gadd45b$^{-/-}$	enhanced persisting memory in tasks of motor performance, aversive conditioning and spatial navigation	Sultan et al. (2012)[196]
Gadd45bg$^{-/-}$	autoimmune lymphoproliferative syndrome and systemic lupus erythematosus	Liu et al. (2005)[192]
Gadd45g$^{-/-}$	comprised T-cell receptor response, reduced interferon gamma production and impaired activation-induced cell death	Lu et al. (2001)[193]
Gadd45g$^{-/-}$	male to female sex reversal	Gier et al. (2012)[197]
Gadd45g$^{-/-}$	development of a systemic lupus erythematosus-like autoimmune disorder	Salvador et al. (2002)[206]

Table 1: Overview of reported Gadd45 knockout mice phenotypes.

3.3.2 Gadd45 in active DNA demethylation

In addition to the aforementioned functions, Gadd45a was identified in a screen as an inducer of active DNA demethylation of transiently transfected *in vitro* methylated plasmids [146]. Plasmid demethylation required the NER protein XPG, which Gadd45a was shown to interact with. This led to the proposed model whereby Gadd45a guides the NER machinery to DNA, where a stretch of nucleotides is excised and the resulting gap refilled, replacing methylated cytosine with unmethylated one in the process.

Similar to its versatile functions outside of DNA demethylation however, work in the following years has shown an involvement of Gadd45 in a number of different DNA demethylation pathways apart from NER: In zebrafish, Gadd45 promotes deamination of 5mC to thymine by Aid/Apobec and its subsequent removal by Mbd4 [207]. In other studies, Gadd45a has been shown to promote Tet activity [208] and to stimulate the Tdg-mediated excision of the resulting 5fC and 5caC [209,210]. Last, it has also been proposed that GADD45A mediates passive DNA demethylation via interaction with the catalytic domain of DNMT1 [211].

Just as varied as the enzymatic cofactors, factors identified and proposed to target Gadd45 itself to specific sites in the genome are diverse. First, GADD45A has been shown to bind hemi-methylated DNA [211], leading to the possibility that Gadd45 could be targeted to sites that have already undergone demethylation at one of their strands. Second, Gadd45a has been characterized as an RNA binding protein [212], making it possible that Gadd45 finds its target sites via base complementarity of bound RNA or via binding nascent RNA. Third, Gadd45 might be targeted to sites of active DNA demethylation via its intense network of protein-protein interactions. Fittingly, site specific Gadd45-dependent active DNA demethylation has been observed in a number of different contexts and the identified applied targeting mechanisms include some of the abovementioned (Table 2).

Gadd45 protein	**Target**	**Targeting mechanism**	**Publication**
GADD45A	*rDNA* promoter	TAF12 (protein)	Schmitz et al. (2009)[147]
GADD45A	*TCF21* promoter	*TARID* (lncRNA)	Arab et al. (2014)[209]
GADD45A	*MAGEB2*, *DHRS2*, *TCEAL7* and others upon UV stimulation	ING1 (protein)	Schäfer et al. (2013)[213]
GADD45A	*RARβ2* upon retinoic acid stimulation	unknown	Le May et al. (2010)[214]
GADD45A	*CD70* in T helper cells *	unknown	Li et al. (2010)[215]
Gadd45a	*Dlx5*, *Runx2*, *Bglap*, and *Osterix* during osteogenic differentiation *	unknown	Zhang et al. (2011)[216]
Gadd45a, Gadd45b	*S100P* during epidermal differentiation *	unknown	Sen et al. (2010)[217]
Gadd45b	*Fgf-1B* and *Bdnf IX* during neuronal stimulation	unknown	Ma et al. (2009)[159]

Table 2: Gadd45-dependent site specific DNA demethylation. Demethylation of targets marked by * might be indirectly depending on Gadd45, as no direct association of Gadd45 with target sites has been demonstrated.

3.4 Aim of the thesis

Previous work both by our group and other laboratories has established an involvement of the Gadd45 proteins in active DNA demethylation via multiple pathways. The importance of Gadd45-mediated demethylation during physiological differentiation processes, especially during early lineage specification, has however received less attention. Importantly, phenotypic characterization of developmental defects of Gadd45 mice mutants has never been carried out with animals devoid of all three Gadd45 genes. A potential redundancy of Gadd45a, Gadd45b and Gadd45g might mask their importance during development as judged by classical loss-of-function experiments.

The aim of this thesis was the characterization of Gadd45-dependent gene regulation during pluripotency and differentiation, using mESCs as a model system. In order to address this goal, two sets of genetically engineered mESC models had to be established and subsequently analyzed: First, using CRISPR/Cas9, mESCs devoid of all three Gadd45 genes were to be generated and studied on a transcriptomic level under various differentiation regimes using RNA-Seq. Second, using homologous recombination, transgenic mESC lines were to be established that featured tagged endogenous Gadd45a or Gadd45g proteins. Using these transgenic mESCs, Gadd45a and Gadd45g DNA binding sites should be identified genome wide via ChIP-Seq, in order to understand which genomic elements undergo Gadd45-mediated active DNA demethylation.

4 Results

4.1 The function of the Gadd45 genes during pluripotency and differentiation

4.1.1 Triple knockout targeting design and screening procedure

In order to study the role of the Gadd45 genes during differentiation processes, two challenges had to be overcome: First, *in vitro* experiments had shown that Gadd45a, Gadd45b and Gadd45g all induce reactivation of methylated reporter plasmids to similar extents (personal communication). It was therefore likely that such a redundancy might also exist *in vivo*. Knocking out only a single Gadd45 family member might therefore not lead to any visible phenotype. Second, the duration of embryonic differentiation processes exceeds the duration of siRNA-mediated knockdowns. Because of this, using transient knockdowns was not a suitable strategy to study the role of the Gadd45 family during such processes either.

The possibility to carry out loss-of-function studies under such conditions came with the application of the CRISPR/Cas9 system for genomic engineering [218]. Using this tool it is now feasible to knockout both alleles of multiple genes each with a single plasmid transfection. I decided to use this method to create *Gadd45a*$^{-/-}$ *Gadd45b*$^{-/-}$ *Gadd45g*$^{-/-}$ triple knockout (TKO) mESCs in order to study their phenotype under pluripotent and differentiating conditions.

Using CRISPR/Cas9, genetic alterations are induced site specifically by targeting the transfected Cas9 nuclease with a co-transfected guide RNA that is partially complementary to the target locus. The Cas9 nuclease subsequently induces a DNA double strand break. The damage can be afterwards repaired by the error-prone non-homologous end joining pathway. This usually results in small deletions or insertions (indels) that might inactivate e.g. an important domain of the protein, or lead to a frameshift that abrogates protein function altogether. However previous work in our lab has shown that deletions and insertions introduced in this manner were often three base pairs long, or multiples thereof (personal communication). Because such mutations will not lead to a frameshift, I therefore decided to use two guide RNAs per Gadd45 gene. In a subset of cases, this would not only generate small indels at the two sites, but a loss of the whole stretch of DNA between the two targets sites.

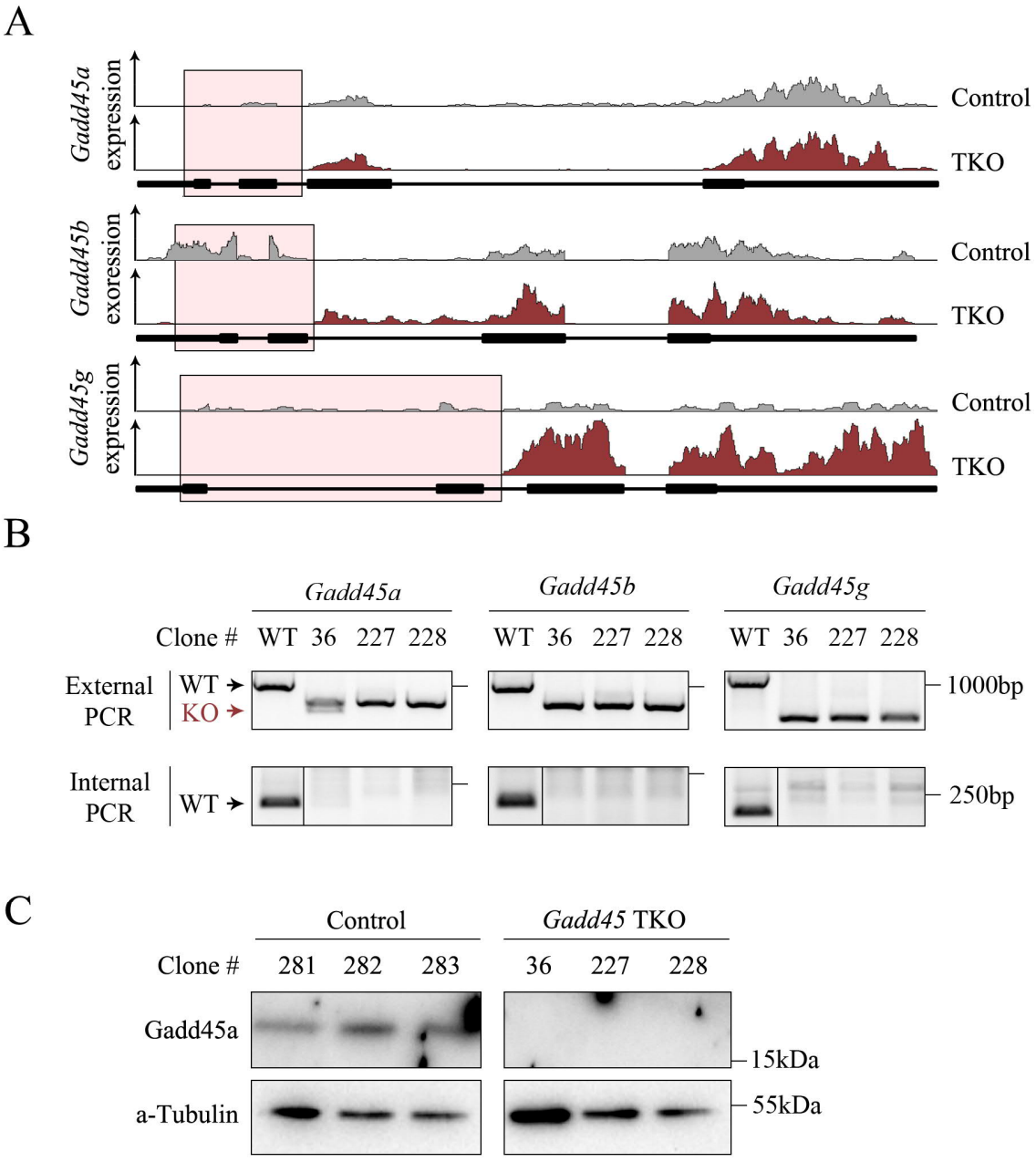


Figure 4.1: CRISPR/Cas9 system allows efficient generation of homozygous triple knockout mESC lines.

A. Scheme indicating desired deletions (in light red) induced by CRISPR/Cas9 at the *Gadd45a*, *Gadd45b* and *Gadd45g* loci. Thin lines indicate introns, medium lines untranslated regions and thick lines coding regions. Overlain are tracks showing expression of the Gadd45 genes in one representative control and one TKO mESC line as measured by RNA-Seq. **B.** External and internal genotyping PCRs to identify mESC clones with the desired *Gadd45a*, *Gadd45b* and *Gadd45g* deletions. WT = Wildtype; KO = Knockout **C.** Western blot showing Gadd45a protein levels in control and Gadd45 TKO mESC clones. Note that endogenous Gadd45b and Gadd45g were not detectable via Western blot in mESCs.

In total, six guide RNAs were designed, all of which were selected for a low probability of off-target binding. For each Gadd45 gene, one guide RNA targeted a region upstream of the start codon, whereas another guide RNA targeted a site within the second intron. The intended deletions would therefore result in a complete loss of the coding part of the first two exons (Fig. 4.1A). The usage of the next start codon following the deletion results in a frameshift and therefore abrogates protein function completely.

E14tg2a mESCs were transfected with six plasmids, each containing a guide RNA and the Cas9 endonuclease sequence. Co-transfected pPGKpuro plasmid allowed for antibiotic selection of successfully transfected cells. For control mESCs, an empty vector coding for Cas9 but no guide RNA was transfected together with pPGKpuro. After selecting with puromycin, individual mESC clones were picked and genotyped.

Genotyping of individual mESC clones was performed using a two-PCR strategy: An external set of PCR primers created a longer PCR product when amplifying a wildtype allele and a shorter product when amplifying a knockout allele. To confirm the genotyping, an additional internal PCR was used. In this case both PCR primers were located within the expected site of deletion and therefore would fail to produce a PCR product in case of a homozygous knockout.

Using this strategy, 280 mESC clones were screened. Individual *Gadd45a*, *Gadd45b* and *Gadd45g* knockouts were observed at comparable rates (Table 3). Importantly, three TKO clones devoid of all six Gadd45 alleles were identified (Fig. 4.1B). It is worth noting that TKOs were obtained at a higher frequency than what would be expected from the rate of single knockouts. Knockouts were therefore not distributed randomly, but "clustered" in particular clones.

Genotype	*Gadd45a*	*Gadd45b*	*Gadd45g*
+/+	172	185	118
+/-	39	28	36
-/-	33	40	56
Failed PCR	34	25	68

Table 3: Frequencies of obtained Gadd45a, Gadd45b and Gadd45g deletions.

As the deletion should lead to a complete frameshift in the coding sequences, it was expected that no functional Gadd45 proteins were translated anymore. Indeed, Western blot analysis showed a complete loss of Gadd45a protein in TKO mESCs (Fig. 4.1C). Unfortunately, Western blots using antibodies for Gadd45b and Gadd45g did not allow detection of these proteins even in control mESCs (data not shown). Ultimately, the occurrence of exactly the intended deletion was later confirmed by RNA-Seq (Fig. 4.1A).

4.1.2 Characterization of Gadd45 TKO mESCs

4.1.2.1 Gadd45 TKO mESCs show regular morphology and pluripotency marker gene expression

All three obtained Gadd45 TKO clones as well as three control clones were expanded for at least 6 passages before further analysis. Morphologically, no striking differences could be observed between TKO and control mESCs. All clones consisted mostly of shiny, round colonies with only minor contribution of flat and separate differentiating cells (Fig. 4.2A). This is indicative of pluripotent, proliferating mESCs. Unaffected stem cell maintenance in Gadd45 TKO mESCs was confirmed on a transcriptional level: Classical markers of pluripotency, such as *Oct4* (*Pou5f1*), *Sox2* and *Nanog* were not misexpressed in the Gadd45 TKO clones (Fig. 4.2B).

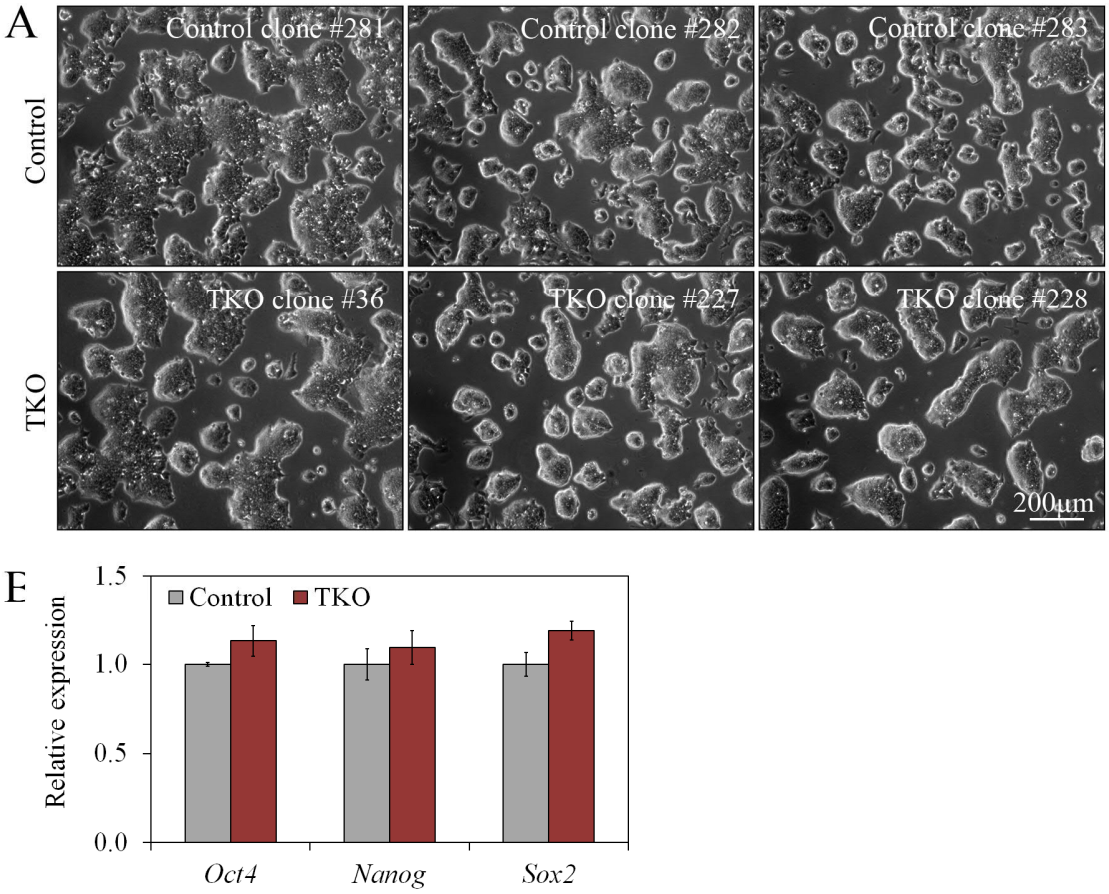


Figure 4.2: Gadd45 TKO mESCs maintain features of pluripotency.

A. Microscope images showing morphology of control (top row) and Gadd45 TKO (bottom row) mESCs. **B.** RNA-Seq expression data showing expression levels of three main pluripotency related transcription factors in control and Gadd45 TKO mESCs (Expression is normalized to the average expression level in control mESCs. Data shown as mean ± SD, n=3).

4.1.2.2 Genome wide transcriptome analysis identifies Gadd45-dependent genes in mESCs

In order to study the impact of the loss of the Gadd45 gene family on the transcriptome in detail, RNA-Seq was performed under six different conditions:

First, RNA-Seq expression analysis was carried out under regular mESC culture conditions, i.e. in presence of serum and LIF, which results in a balance of pro- and anti-differentiation signals (see chapter 3.1.4.1).

Second, the transcriptome was analyzed under two conditions that induce a shift towards "ground state pluripotency". On the one hand, mESCs were cultured for three days in 2i-medium (see chapter 3.1.4.1). On the other hand, cells were treated for three days with vitamin C. Vitamin C works as a cofactor for Fe(II) 2-oxogluterate dioxygenases, including the Tet enzymes. Vitamin C treatment of mESCs therefore increases Tet activity, which leads to activation of germline genes and an overall transcriptome that resembles more closely the blastocyst stage [219]. Notably, both 2i and Vitamin C treatment induce a genome wide DNA demethylation process [78,219], further increasing my interest in the transcriptome of the Gadd45 TKO mESCs under these conditions.

Third, the transcriptome of the Gadd45 TKO mESCs was studied under three different differentiation conditions. This analysis will be introduced in greater detail in chapter 4.1.3. In all cases, transcriptome analysis was performed in triplicates, using three independent TKO and three independent control mESC clones.

Under regular mESC culture conditions, 137 genes were differentially expressed in Gadd45 TKO mESCs compared to control mESCs, with a false discovery rate of 10% (0.1FDR). 40 of these genes were upregulated, whereas the majority (97 genes) was downregulated in the Gadd45 TKOs compared to the controls. Neither down- nor upregulated genes clustered strongly in terms of gene ontology (GO) annotations (data not shown), with the majority of annotation groups containing only very few genes, or having a very general definition (e.g. "cellular process"). Notably, genes downregulated in the Gadd45 TKO mESCs included the genes *Pramel6* and *Pramel7* that have been previously identified as LIF target genes, *Asz1*, which is a known germline marker, and several imprinted genes, namely *H19*, *Xist*, *Mcts2* and *Mest*.

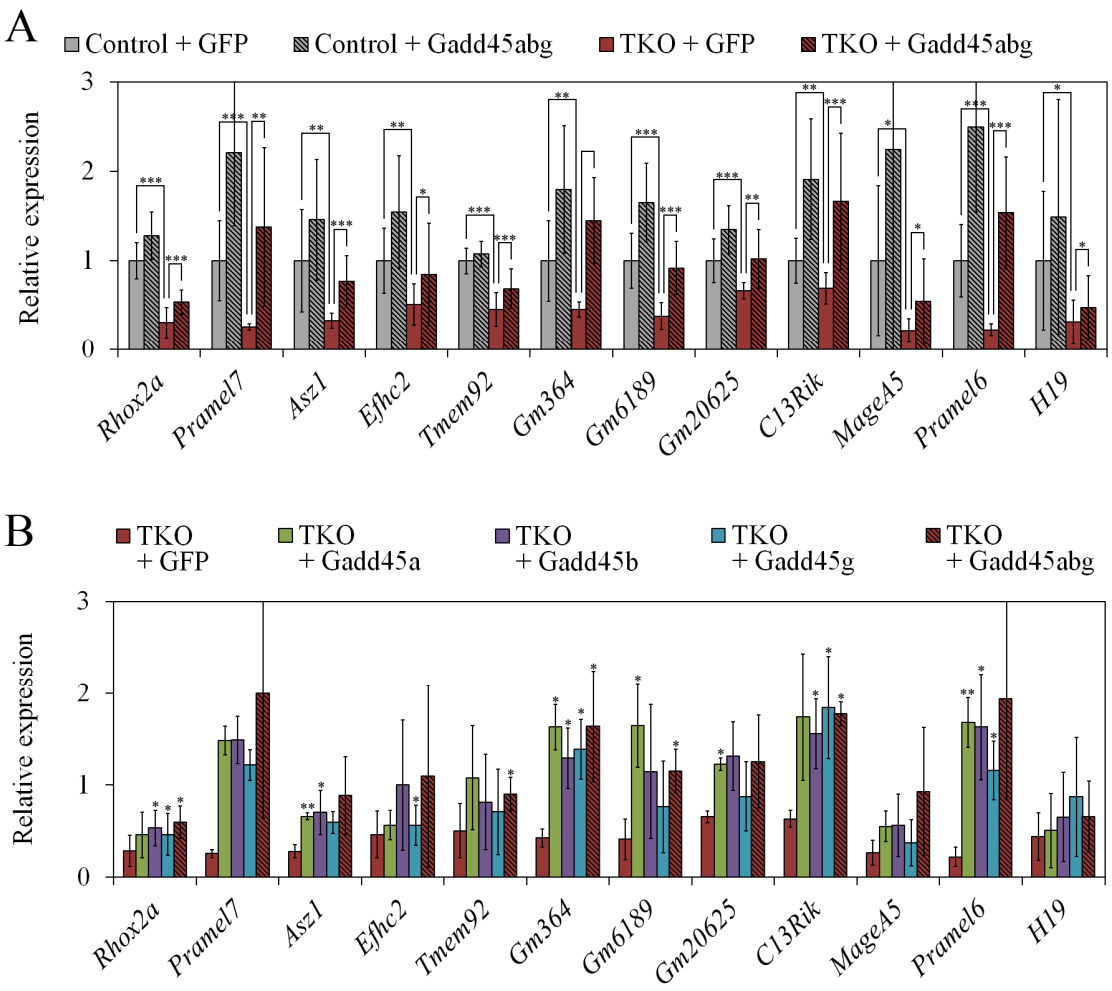


Figure 4.3: Gadd45 overexpression rescues gene deregulation.

A. qPCR expression data showing expression levels of 12 Gadd45 dependent genes in control and Gadd45 TKO mESCs, both after GFP and combined Gadd45a + Gadd45b + Gadd45g transient overexpression. (Expression is relative to *Gapdh* expression and further normalized to the average expression level in control mESCs. Data shown as mean ± SD, n=9. *** = p-value < 0.0005, ** = p-value < 0.005, * = p-value < 0.05 according to student's t-test) **B.** qPCR expression data showing expression levels of 12 Gadd45 dependent genes in Gadd45 TKO mESCs, after overexpression of either GFP, Gadd45a, Gadd45b, Gadd45g or combined overexpression of Gadd45a + Gadd45b + Gadd45g. (Expression is relative to *Gapdh* expression and further normalized to the average expression level in control mESCs. Data shown as mean ± SD, n=3. ** = p-value < 0.005, * = p-value < 0.05 according to student's t-test)

Although guide RNAs were chosen based on a low probability of off-target effects, transfection of six different guide RNAs might increase the likelihood of unwanted mutations [220]. Furthermore, although experiments were performed in triplicates, clone-to-clone variability might lead to a small amount of

identified differentially expressed genes simply due to stochastic reasons. In order to confirm that the identified genes were indeed downregulated due to loss of the Gadd45 family of genes, rescue experiments were carried out. Both control and Gadd45 TKO mESCs were transiently transfected either with a GFP expression construct alone or a combination of *Gadd45a*, *Gadd45b*, *Gadd45g* and GFP expression constructs. Expression levels of twelve Gadd45-dependent genes were then analyzed by qPCR. Strikingly, downregulation of all twelve genes was rescued by transient overexpression of Gadd45a+b+g in TKO mESCs – sometimes partially, sometimes completely (Fig. 4.3A, red bars). Furthermore, the expression of all twelve genes in control mESC was further stimulated by Gadd45 overexpression (Fig. 4.3A, grey bars). Although expression levels were not analyzed globally after rescue experiments, these qPCR results indicate that a vast majority of the observed gene expression changes in TKO mESCs was indeed due to the loss of Gadd45, rather than being caused by off-target effects or clone-to-clone variability.

Previous *in vitro* experiments had indicated a potential redundancy of the different Gadd45 family members. However, this was never shown for regulation of endogenous genes. In order to study which of the three Gadd45 genes was able to induce gene expression of the identified Gadd45-dependent genes, *Gadd45a*, *Gadd45b* and *Gadd45g* were transfected individually in comparison to a combined transfection. Although some slight differences could be observed, overall all three Gadd45 genes were able to rescue target gene expression in Gadd45 TKO mESCs to a highly similar extent (Fig. 4.3B). This shows that in the absence of two Gadd45 genes, the third Gadd45 gene was sufficient to maintain target gene expression levels, at least when overexpressed. In line with these findings, double transfection of any combination of Gadd45 expression constructs showed no synergistic effect (data not shown). Single and double knockout mESC would be required to truly show that all three Gadd45 genes are redundant for mESC gene expression regulation also under physiological conditions. These mESCs would then not be expected to misexpress genes.

4.1.2.3 Gadd45 target genes show hallmarks of methylation regulated genes

Whereas 137 genes were differentially expressed in Gadd45 TKO compared to control mESCs under standard cell culture conditions, this was not the case for 2i and Vitamin C treated mESCs (Fig. 4.4A). Only 15 and 11 genes were differentially expressed between TKO and control mESCs under these conditions respectively, using a 0.1 FDR cut-off. This is in stark contrast to the number of genes which were up- or downregulated upon the treatments itself: 1497 genes were differentially expressed after 3 days of Vitamin C treatment and more than 10.000 genes were differentially expressed after 3 days of 2i treatment compared to standard culture conditions, using the same 0.1 FDR cut-off (data not shown). It

can therefore be concluded that the Gadd45 genes are not necessary for the genome wide transcription changes that occur during the transition towards ground state pluripotency.

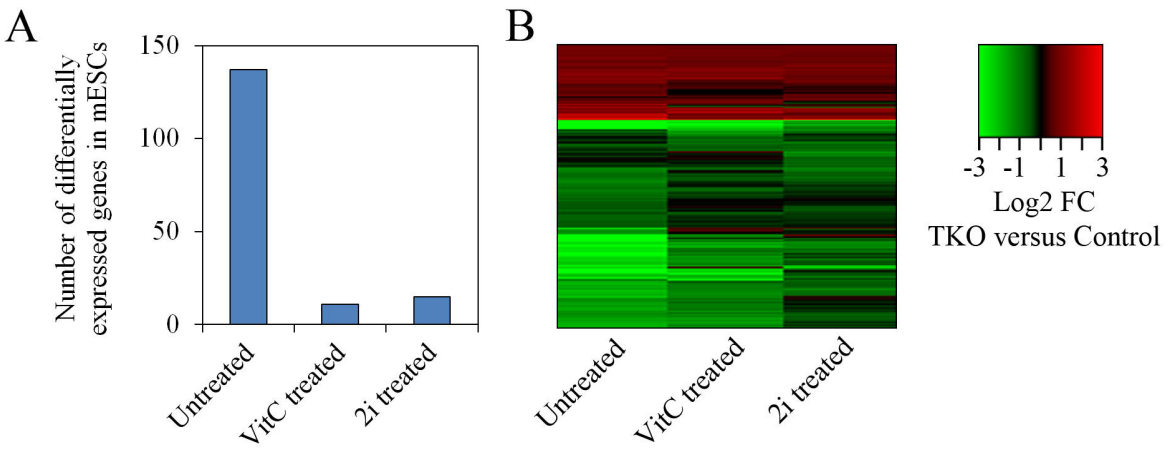


Figure 4.4: Vitamin C and 2i treatment partially rescue gene deregulation in Gadd45 TKO mESCs.

A. Bar chart showing the number of statistically significantly (false discovery rate < 10%) deregulated genes in Gadd45 TKO mESCs compared to control mESCs under three different conditions, as identified by RNA-Seq. **B.** Heatmap indicating expression levels of all genes deregulated in Gadd45 TKO mESCs compared to control mESCs under at least one of the three conditions analyzed. Each line represents a single gene, whereas the logarithmic fold change (FC) in mean expression in Gadd45 TKO mESCs is indicated by the color code shown in the legend. Heatmap generated by M. Mallick (this lab). Primary RNA-Seq data analysis performed by E. Karaulanov (IMB Mainz).

Furthermore, even the genes that were differentially expressed in TKO compared to control mESCs under regular cell culture conditions seemingly were no longer differentially expressed after 2i or Vitamin C treatment. A closer look at these genes revealed that although they were still mildly up- or downregulated in TKO mESCs compared to control mESCs, the differences between TKO and control mESCs were greatly attenuated upon 2i or Vitamin C treatment (Fig. 4.4B) and the observed differences were no longer statistically significant. Therefore, although the Gadd45 genes do not play a role in Vitamin C or 2i induced expression changes, Gadd45-dependent gene expression itself can be rescued by these demethylation inducing treatments.

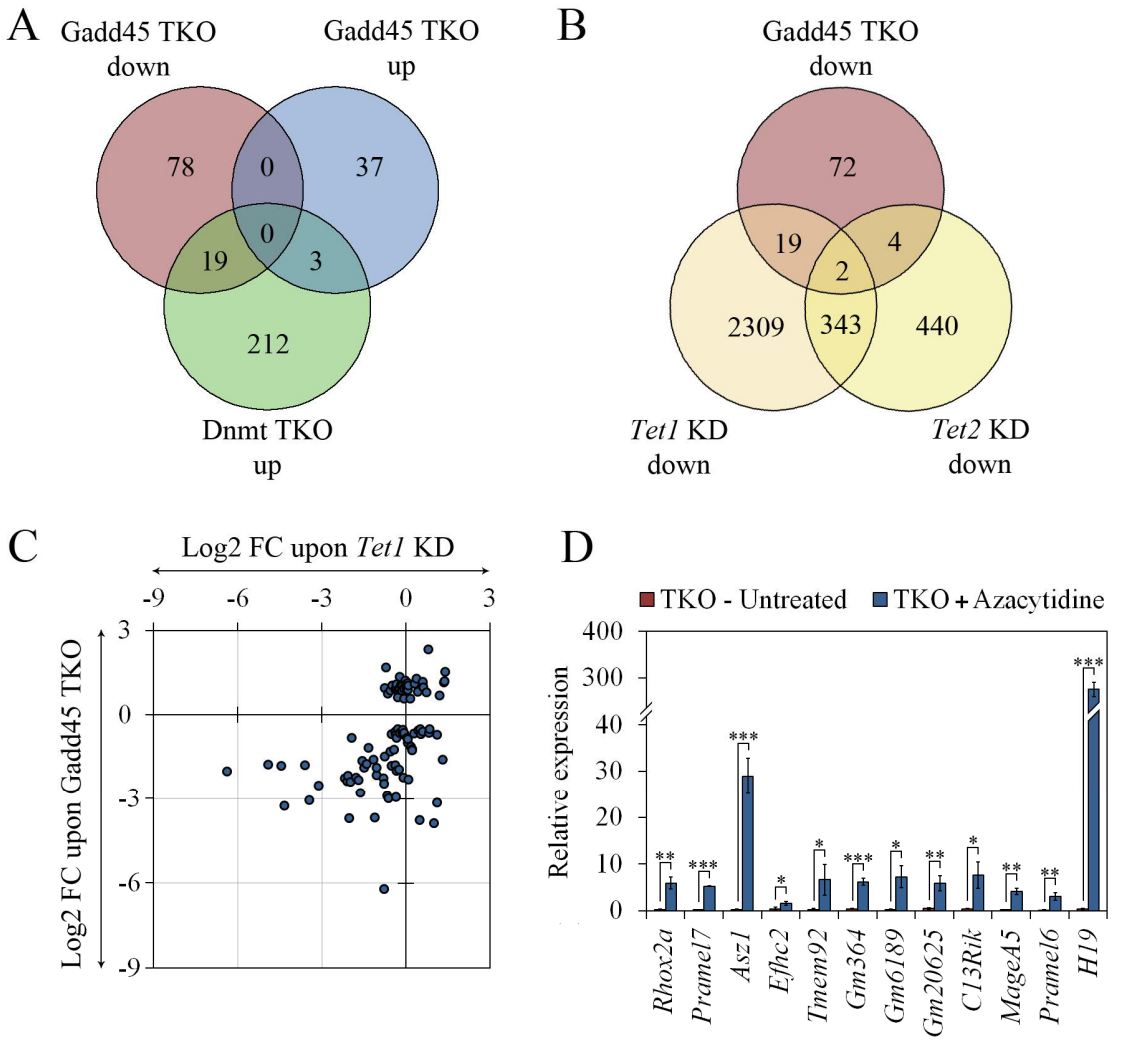


Figure 4.5: Deregulated genes show hallmarks of methylation regulation.

A. Venn diagram showing the overlap of genes up- or downregulated in Gadd45 TKO mESCs compared to control mESCs and genes upregulated in Dnmt TKO mESCs [221] compared to control mESCs. **B.** Venn diagram showing the overlap of genes downregulated in Gadd45 TKO mESCs and genes downregulated upon either *Tet1* or *Tet2* knockdown in mESCs [222]. **C.** Scatterplot comparing logarithmic fold change (FC) in expression levels between *Tet1* KD and control mESCs versus logarithmic fold change (FC) in expression levels between Gadd45 TKO mESCs and control mESCs. Each dot represents a single gene up- or downregulated in the Gadd45 TKO mESCs. **D.** qPCR expression data showing expression levels of 12 Gadd45 dependent genes in Gadd45 TKO mESCs left either untreated or treated with 5'-azacytidine. (Expression is relative to *Gapdh* expression and further normalized to the average expression level in control mESCs without 5'-azacytidine treatment. Data shown as mean ± SD, n=3. Notice that the y-axis is discontinuous. *** = p-value < 0.0005, ** = p-value < 0.005, * = p-value < 0.05 according to student's t-test)

To further corroborate that Gadd45-dependent genes are regulated via methylation, the list of Gadd45 regulated genes was compared to publicly available datasets that aimed at identifying Tet and Dnmt regulated genes in mESCs. In a study by Karimi et al., *Dnmt1*, *Dnmt3a* and *Dnmt3b* triple knockout mESCs were generated and their transcriptome was analyzed genome wide [221]. Indeed, 20% of the genes that are downregulated upon loss of the Gadd45 genes are upregulated upon loss of the Dnmts ($p < 0.0001$, Chi-squared test with Yates' correction, Fig. 4.5A), indicating that these genes are regulated via DNA methylation. As expected, no significant overlap of genes upregulated upon loss of the Dnmts was seen with genes that are upregulated upon loss of the Gadd45 genes (Fig. 4.5A).

Similarly, a study by Huang et al. identified *Tet1*- or *Tet2*-dependent genes by silencing either of the genes using inducible shRNA constructs in mESCs [222]. Again, a minor, but significant overlap between Tet1- and Gadd45-dependent genes was observed ($p < 0.01$, Chi-squared test with Yates' correction, Fig. 4.5B). Notably, genes that were very strongly downregulated upon loss of the Gadd45 genes also tended to be strongly downregulated upon knockdown of *Tet1* (Fig. 4.5C). No such correlation could be observed with the expression changes upon knockdown of *Tet2* (data not shown).

In order to examine directly whether Gadd45-dependent genes in mESCs are regulated via DNA methylation, Gadd45 TKO mESCs were treated with 5-azacytidine to inhibit Dnmt function and Gadd45-dependent gene expression levels were analyzed by qPCR. Indeed, all twelve Gadd45-dependent genes whose expression had been previously rescued by transient Gadd45 overexpression could also be rescued by 5-azacytidine treatment (Fig. 4.5D). Similar to overexpression of Gadd45, 5-azacytidine treatment also further boosted expression levels in control mESCs (data not shown).

4.1.2.4 Gadd45 TKO mESCs do not show global, but site specific hypermethylation near target genes

Given that methylation regulated genes were deregulated in Gadd45 TKO mESCs, DNA methylation levels were measured in control and Gadd45 TKO mESCs. First, 5mC levels as well as the oxidized derivatives thereof - 5hmC, 5fC and 5caC – were measured globally by quantitative mass spectrometry. Neither of the modifications was significantly altered in TKO mESCs (Fig. 4.6, 0h timepoints), in line with the hypothesis that Gadd45 induces site-specific rather than global demethylation [213]. DNA methylation levels were also measured after 2i and vitamin C treatment. Both treatments resulted in a global reduction of mC, while levels of hmC, caC and fC were rising. These kinetics were unaltered in Gadd45 TKO mESCs (Fig.4.6A-H). This demonstrates that neither the global DNA demethylation nor the global transcription changes (see 4.1.2.3) upon 2i and vitamin C treatment depend on the Gadd45 gene family.

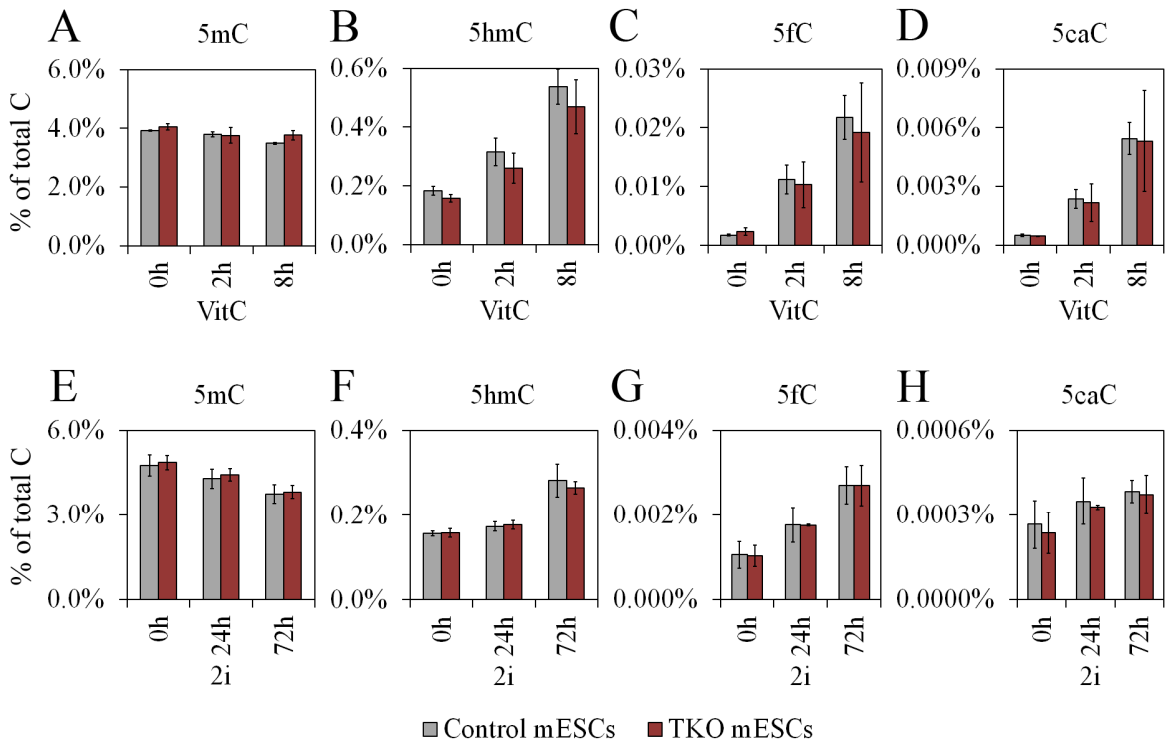


Figure 4.6: Global 5mC and derivative levels are unchanged in Gadd45 TKO mESCs.

A-D. Quantitative mass spectrometry measurements showing the levels of 5mC (A), 5hmC (B), 5fC (C) and 5caC (D) in relation to total amount of C in Gadd45 TKO and control mESCs upon increasing durations of vitamin C treatment (Data shown as mean ± SD, n=3). **E-H.** Quantitative mass spectrometry measurements showing the levels of 5mC (E), 5hmC (F), 5fC (G) and 5caC (H) in relation to total amount of C in Gadd45 TKO and control mESCs upon increasing durations of 2i treatment (Data shown as mean ± SD, n=3). Mass spectrometry data by Michael Musheev (this lab).

To study whether site specific DNA hypermethylation occurs in Gadd45 TKO mESCs, selected regulatory elements of Gadd45-dependent genes were analyzed by bisulfite sequencing (Table 4). In total, 26 promoters, enhancers and imprinting control regions (ICR) of Gadd45-dependent genes, as well as 2 control promoters, were amplified by PCR on bisulfite converted DNA and analyzed by small scale next generation sequencing. All CpGs within these regions were highly covered, allowing a very accurate methylation level calling, with the exception of 3 regions that were subsequently excluded from analysis.

Gene	Regulatory Region	Coverage (raw read count)	Nr. Of CpGs	Average methylation difference in TKO mESCs vs. control mESC
Rhox2a	Promoter	31558	8	15.5% ± 2.19%
Pramel6	Promoter	6686	3	16.5% ± 3.16%
Pramel7	Promoter	1935	3	8.7% ± 4.05%
Pramel7	LMR	16971	4	-1.1% ± 1.16%
Asz1	Promoter	4635	6	6.7% ± 7.89%
Efhc2	Promoter/LMR	188	18	excluded from analysis
Tmem92	Promoter	5682	3	11.5% ± 2.02%
Tmem92	Alt. Promoter	12029	3	7.1% ± 3.28%
Gm364	Promoter	8537	11	11.8% ± 1.87%
Gm364	LMR	9239	15	4.6% ± 4.56%
Gm6189	Promoter	801	3	10.1% ± 1.44%
Gm6189	LMR	7895	3	0.9% ± 1.71%
MageA5	Promoter	6499	4	3.7% ± 1.96%
H19	Promoter	28287	14	9.3% ± 3.45%
H19	Alt. Promoter	8701	8	2.4% ± 1.94%
H19	ICR	5465	20	39.4% ± 3.38%
Calcoco2	Promoter	16076	5	1.6% ± 2.22%
Calcoco2	LMR	15953	9	1.4% ± 1.44%
Igfbp2	Promoter	10236	15	0% ± 0.32%
Igfbp2	LMR	2614	6	-0.1% ± 1.21%
Mest	Promoter	19237	10	15.3% ± 1.58%
Mest	LMR	2633	7	16.5% ± 7.05%
Mest	LMR	22207	14	31.6% ± 6.09%
Mest	LMR	11651	4	19.1% ± 0.99%
Mest	ICR	622	41	excluded from analysis
Xlr3b	Promoter	434	6	excluded from analysis
Sry	Promoter	8074	4	3.6% ± 1.7%
Rerg	Promoter	13291	28	0.1% ± 0.35%

Table 4: Overview of regions analyzed by bisulfite sequencing. LMR = Lowly methylated region [130].

Bisulfite PCR can be biased, because after bisulfite conversion methylated and unmethylated DNA will differ in base composition. To examine to which extent such biases might have occured, spiked-in plasmid DNA was amplified as well. pOctTKGFP and pGL3 plasmids were *in vitro* methylated using *HpaII* or *MsssI* methylases respectively and afterwards mixed at defined ratios with unmethylated plasmid. Then, plasmid DNA was added in the same molarity as the genomic DNA isolated from the mESCs before bisulfite conversion. Bisulfite sequencing ultimately revealed how closely the measured DNA methylation levels correlated with the real methylation levels.

As can be seen in figure 4.7A, the methylation levels of the *HpaII* methylated pOctTKGFP plasmid were very accurately determined, only deviating from the real methylation levels by a margin of a few percent. In case of the *MsssI* methylated pGL3 plasmid however, bisulfite PCR was biased towards amplifying the non-methylated plasmid (Fig. 4.7B). All samples that should contain 20, 40, 60 or 80% methylated DNA were underestimated to contain less methylated DNA. Samples that contained only 100% methylated spiked-in DNA were accurately measured, showing that this bias was indeed PCR based and not a result of incomplete bisulfite conversion, or incomplete *in vitro* methylation. Notably, despite this bias, differences in methylation levels across the standard curve (e.g. 20% versus 40% methylation) could still readily and reliably be identified, although the measured difference was smaller than expected. It can therefore be concluded that PCR biases during this analysis may lead to underestimation of methylation differences between Gadd45 TKO and control mESCs.

Overall, the majority of regulatory elements near Gadd45-dependent genes showed a slight 5-10% hypermethylation across all analyzed CpGs (Table 4) in the Gadd45 TKO mESCs. Even though these differences were mostly too small to be statistically significant based on individual t-tests, the fact that this hypermethylation can be observed consistently across many regions and across all CpGs within a region is noteworthy. Loci which show a stronger, statistically significant hypermethylation were examined more closely: Within the promoters of the genes *Rhox2a*, *Pramel6* and *Gm364*, all analyzed CpGs exhibit 10-20% increased methylation (Fig. 4.7C-E). Interestingly, also the *H19* ICR was strongly hypermethylated, albeit to a more varying degree between individual TKO mESC lines (Fig. 4.8F). No such hypermethylation was observed in the negative control promoters of the genes *Rerg* and *Sry* (Fig. 4.8G-H), which had been chosen as lowly and strongly methylated controls respectively. In summary, the obtained DNA methylation data complements previously obtained gene expression data and is pointing further towards the possibility that these genes are regulated by Gadd45-dependent demethylation processes. However, unlike gene expression changes, hypermethylation could not be rescued by transient, combined *Gadd45a*, *Gadd45b* and *Gadd45g* overexpression (data not shown). Potential reasons for this will be presented in the Discussion.

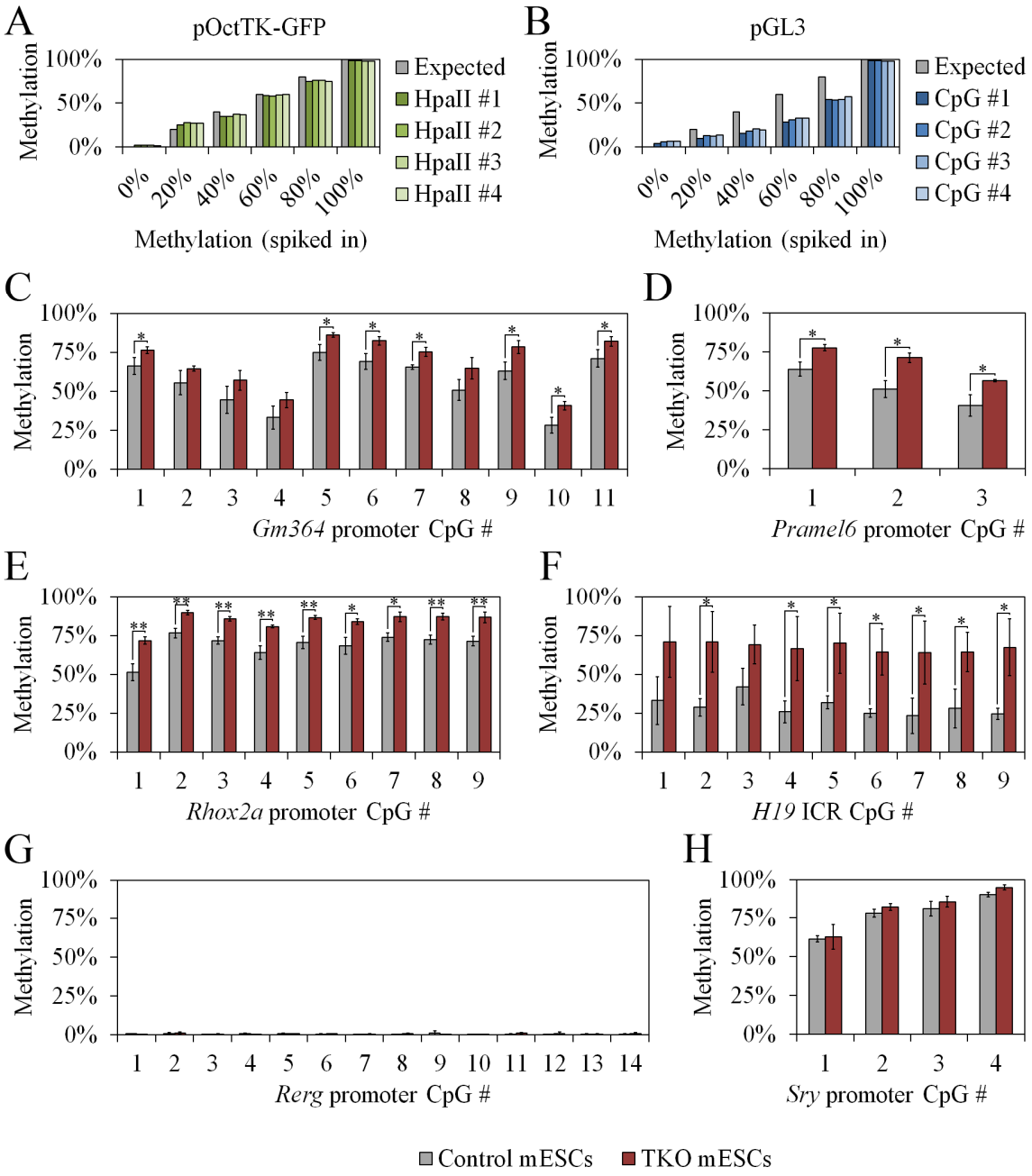


Figure 4.7: Hypermethylation occurs at selected promoters and ICR in Gadd45 TKO mESCs.

A-B. Expected methylation levels of spiked-in *in vitro* methylated pOctTKGFP (A) and pGL3 (B) versus levels experimentally detected via bisulfite sequencing. **C-H**. Methylation levels as measured via bisulfite sequencing of the indicated regulatory elements in Gadd45 TKO and control mESCs (Data shown as mean ± SD, n=3. ** = p-value < 0.005, * = p-value < 0.05 according to student's t-test). Panels G and H show control regions for Gadd45 independent hypo- (G) and hypermethylation (H). Primary analysis of methylation data was performed by M. Mallick (this lab).

4.1.3 Characterization of Gadd45 TKO mESCs under various differentiation regimes

A prime advantage of mESCs as a model cell culture system is their differentiation potential. Rather than constituting a static cell population, mESC can give rise to virtually all cells of a fully developed animal – given the right cues. Therefore the developmental potential of the Gadd45 TKO mESCs was analyzed, focusing on protocols that would allow neuronal differentiation, given the neural expression pattern of the Gadd45 family members, as well as their previously reported functions in *Xenopus* (See chapter 3.3.1).

4.1.3.1 Gadd45 TKO mESCs can differentiate into all three germ layers in teratomas

First, a classical differentiation protocol was applied: In cooperation with EPO Berlin, Gadd45 TKO mESCs and control mESCs were injected into the flanks of immunodeficient mice, where cells were allowed to develop into teratomas. These teratomas were subsequently removed, sectioned and histologically analysed. As judged from exemplary sections, both Gadd45 TKO mESCs and control mESCs were able to contribute to all three germ layers (Fig. 4.8A), as shown by the presence of neural tube like structures, ciliated endodermal cells and chondrocytes. In line with pluripotency marker expression (see chapter 4.1.2.1), this further proved that Gadd45 TKO mESCs remain pluripotent. Furthermore, typical ectodermal (*Pax6*, *Nestin*, *Tubb3*, *Neurog1*), mesodermal (*Desmin*, *Brachyury*), endodermal (*HNFa*, *AFP*) and neural crest (*Snai1*, *Twist1*, *Pax3*) marker genes were expressed similarly in Gadd45 TKO and control teratomas, further validating that overall tissue composition was unchanged in teratomas upon loss of all Gadd45 proteins (Fig. 4.8B).

Closer inspection of the teratoma sections revealed slight differences, however. Notably, all three analyzed Gadd45 TKO teratomas often contained high amounts of nucleated cells within blood islands that typically only consisted of cells without nucleus in control teratomas (Fig. 4.8C). This could be indicative of a leukemia-like overproliferation of hematopoietic stem cells within the teratoma, a finding that is in line with previous studies reporting altered Gadd45 expression in hematopoietic malignancies [223,224].

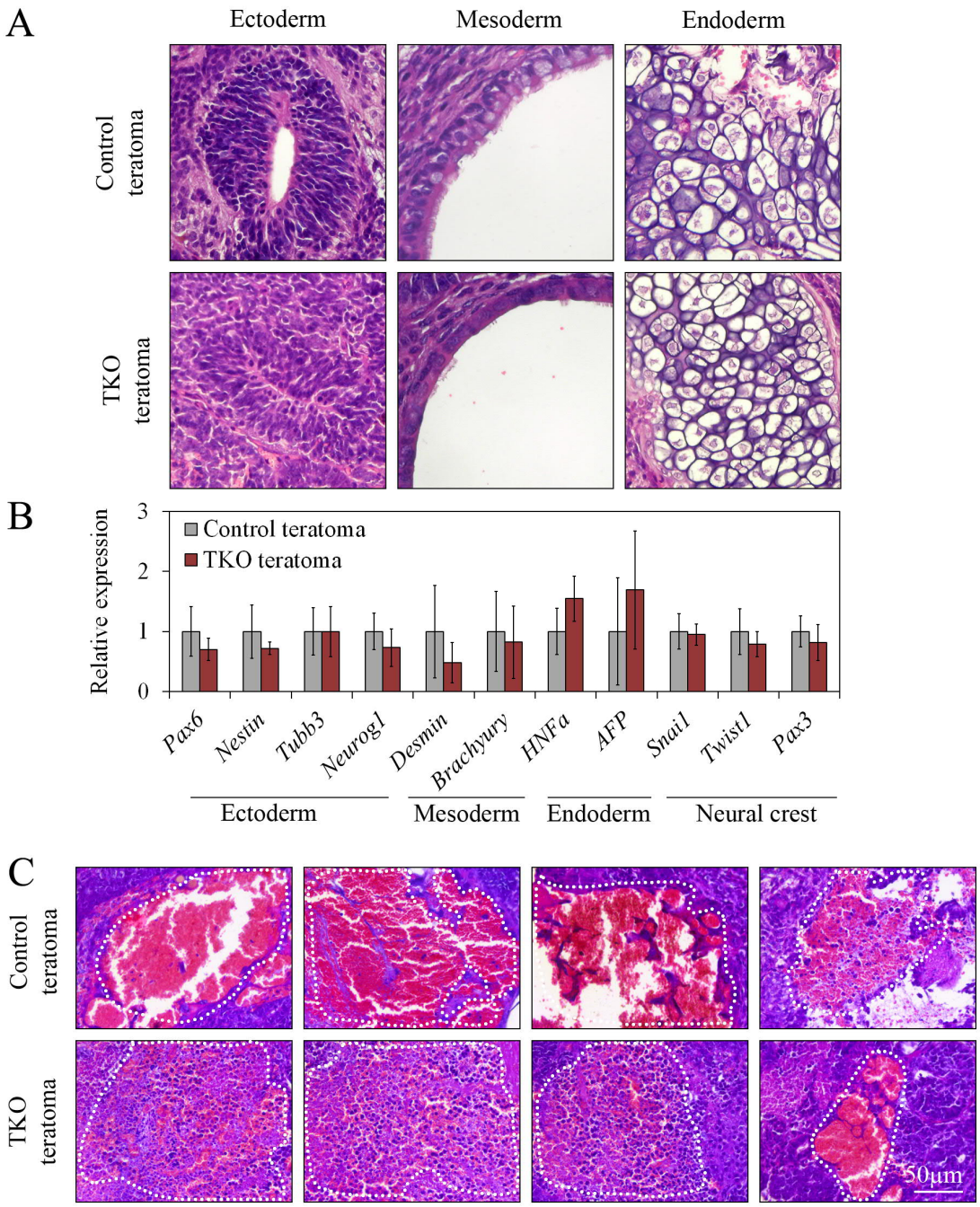


Figure 4.8: Gadd45 TKO mESCs contribute to all three germ layers during teratoma formation.

A. H&E staining of teratomas derived from Gadd45 TKO or control mESCs. Examples depict derivatives of all three germ layers. Pictures taken by W. Haider (EPO Berlin). **B.** qPCR data showing expression levels of lineage marker genes in Gadd45 TKO and control teratoma (Expression is relative to *Gapdh* expression and normalized to the expression in control teratoma. Data shown as mean ± SD, n=9). qPCR carried out by L. Kirchgeorg (this lab). **C.** H&E staining of teratomas derived from Gadd45 TKO or control mESCs. Representative images of blood islands (marked in white lines) are shown.

4.1.3.2 Loss of Gadd45 leads to deregulation of neuronal and cell motility related genes during differentiation

Although teratoma generation is a reliable tool to assess pluripotency, *in vitro* differentiation allows greater control over the direction of differentiation and is more readily scalable and easier to manipulate. Therefore, as mentioned in chapter 4.1.2.2, RNA-Seq was performed to analyze the transcriptome of Gadd45 TKO and control cells after the application of three distinct *in vitro* differentiation protocols: First, cells were cultured for 8 days in suspension as EBs. Under these conditions, mESC will give rise to all three germ layers, allowing interaction between developing tissues within the EBs (see chapter 3.1.4.2). This condition resembles teratoma formation most closely. Secondly, EBs were triggered to induce neurogenesis after 4 days of uninduced development by addition of retinoic acid for further 4 days. These conditions produce a highly homogenous and synchronous population of neural precursors [225]. Thirdly, cells were grown as an adherent monolayer of cells in the absence of serum for 6 days. Without further stimuli that are usually present in serum, cells mostly differentiate in response to autocrine Fgf signaling. Cells differentiated according to this protocol will again mostly differentiate along the neuronal lineage, although less homogenously and less synchronized as compared to retinoic acid treatment. As a result, not only neuronal precursors will arise, but also more mature neurons and glia cells, as well as spurious differentiation along other lineages [73].

First, the expression of the Gadd45 genes themselves was compared across the different protocols (Fig. 4.9A). As reported before [210], *Gadd45a* and *Gadd45b* were the dominantly expressed *Gadd45* genes in mESCs, whereas *Gadd45g* was only lowly expressed. *Gadd45a* expression was very stable and hardly changed more than 2-fold upon the various differentiation procedures. It was expressed strongest in mESCs under regular culture conditions. *Gadd45b* expression was induced by the majority of conditions, with the exception of vitamin C treatment. It was the strongest expressed Gadd45 family member after vitamin C and 2i treatment, and also upon monolayer and EB differentiation. *Gadd45g* was activated mostly by protocols that induce neurogenesis, in line with its expression *in vivo*. However, even after upregulation, its absolute expression levels barely reached those of *Gadd45a* and *Gadd45b*.

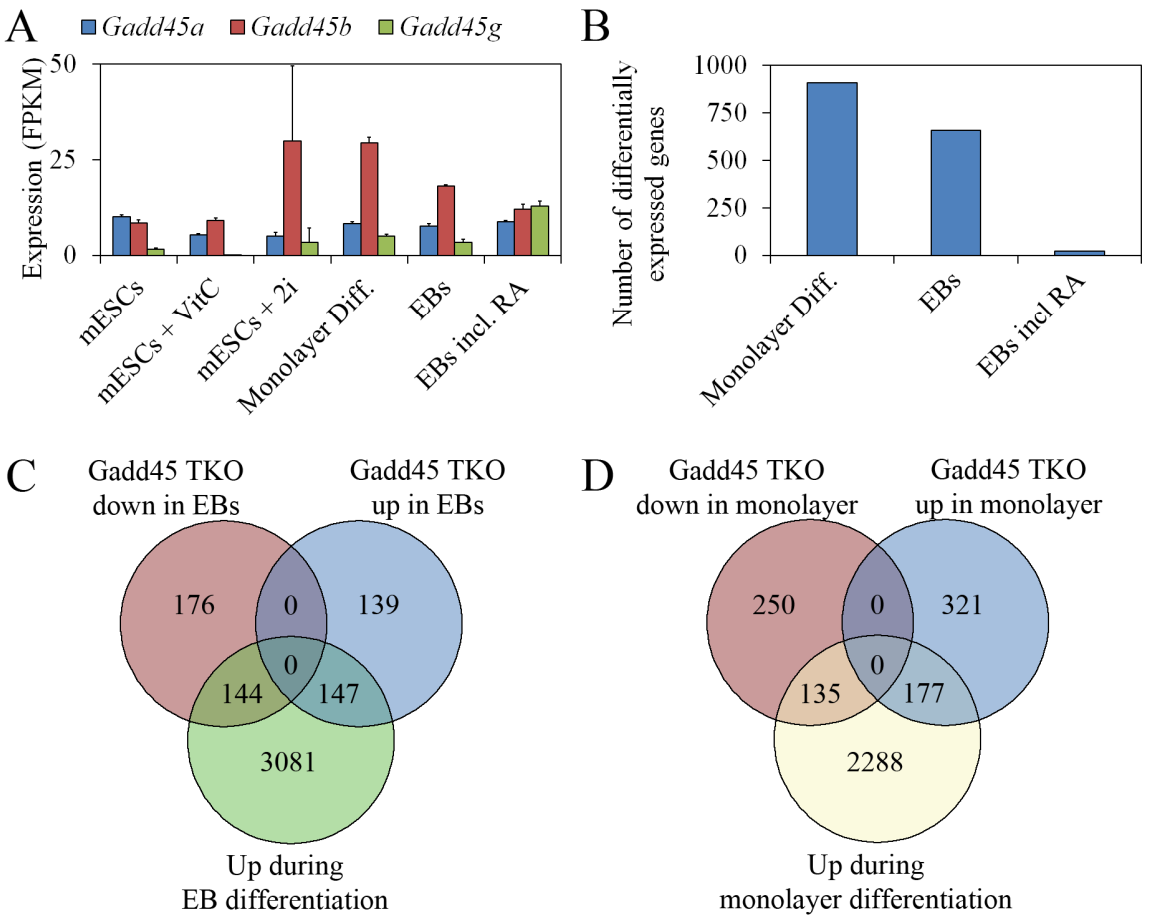


Figure 4.9: The number of deregulated genes in Gadd45 TKO cells increases upon differentiation.

A. Absolute expression levels of the three Gadd45 family members across the different conditions analyzed, as measured via RNA-Seq. (FPKM = fragments per kilobase of exon per million fragments mapped; RA = retinoic acid. Data shown as mean ± SD, n=3). **B.** Number of statistically significantly (false discovery rate < 10%) up- or downregulated genes in Gadd45 TKO cells compared to control cells across the different conditions indicated. **C.** Venn diagram showing the overlap of genes up- or downregulated in Gadd45 TKO EBs compared to control EBs and genes generally induced during EB differentiation. **D.** Venn diagram showing the overlap of genes up- or downregulated in Gadd45 TKO monolayer differentiated cells compared to control monolayer differentiated cells and genes generally induced during monolayer differentiation.

Whereas only 137 genes were differentially expressed in pluripotent Gadd45 TKO mESCs compared to control mESCs, the number of differentially expressed genes was much greater after differentiation (Fig. 4.9B). Notably, similar findings have been made for other cells in which DNA methylation patterns have been disturbed, such as Tet and Dnmt knockout mESCs, which also show stronger phenotypes upon differentiation compared to their pluripotent state [165,226].

In terms of absolute numbers, 909 genes were found to be differentially expressed in monolayer differentiated Gadd45 TKO cells compared to control cells, with 510 genes being upregulated and 399 genes being downregulated. Similarly, 660 genes were differentially expressed in Gadd45 TKO EBs compared to control EBs, of which 313 were upregulated and 347 were downregulated. Notably, only 49 genes were commonly deregulated in Gadd45 TKO cells upon EB and monolayer differentiation, indicating that the function of the Gadd45 genes is highly context dependent.

In stark contrast, only 23 genes were differentially expressed in Gadd45 TKO retinoic acid treated EBs, using the same 0.1FDR cut-off, a finding that will be discussed in chapter 5. Therefore further analysis focused on the untreated EB and monolayer differentiated cells.

Both monolayer differentiation and EB differentiation lead to massive changes of the transcriptome: 2600 genes and 3372 genes were upregulated in comparison to control mESCs respectively, using very stringent cut-offs of a 0.01FDR and at least two-fold upregulation. As expected, these genes were enriched for many developmental regulators (data not shown). Both genes up- and downregulated in Gadd45 TKO cells compared to control cells overlapped significantly with genes that were induced by the two differentiation protocols (Fig. 4.9C-D).

Furthermore, differentially expressed genes in Gadd45 TKO cells were enriched for various developmental and cell biology GO terms: Genes that were downregulated in Gadd45 TKO EBs were enriched for developmental regulators, cell migration related genes and genes involved in nervous system development (Fig. 4.10A). Such genes included e.g. the Ephrin receptors *Epha5* and *Epha7*, the neuronal transcription factors *Zic1* and *Neurod4* and the neural precursor marker *Nes* (Fig.4.10C). Genes that are upregulated in Gadd45 TKO EBs were mildly enriched for GO terms regarding ion transport and localization processes (Fig. 4.10B).

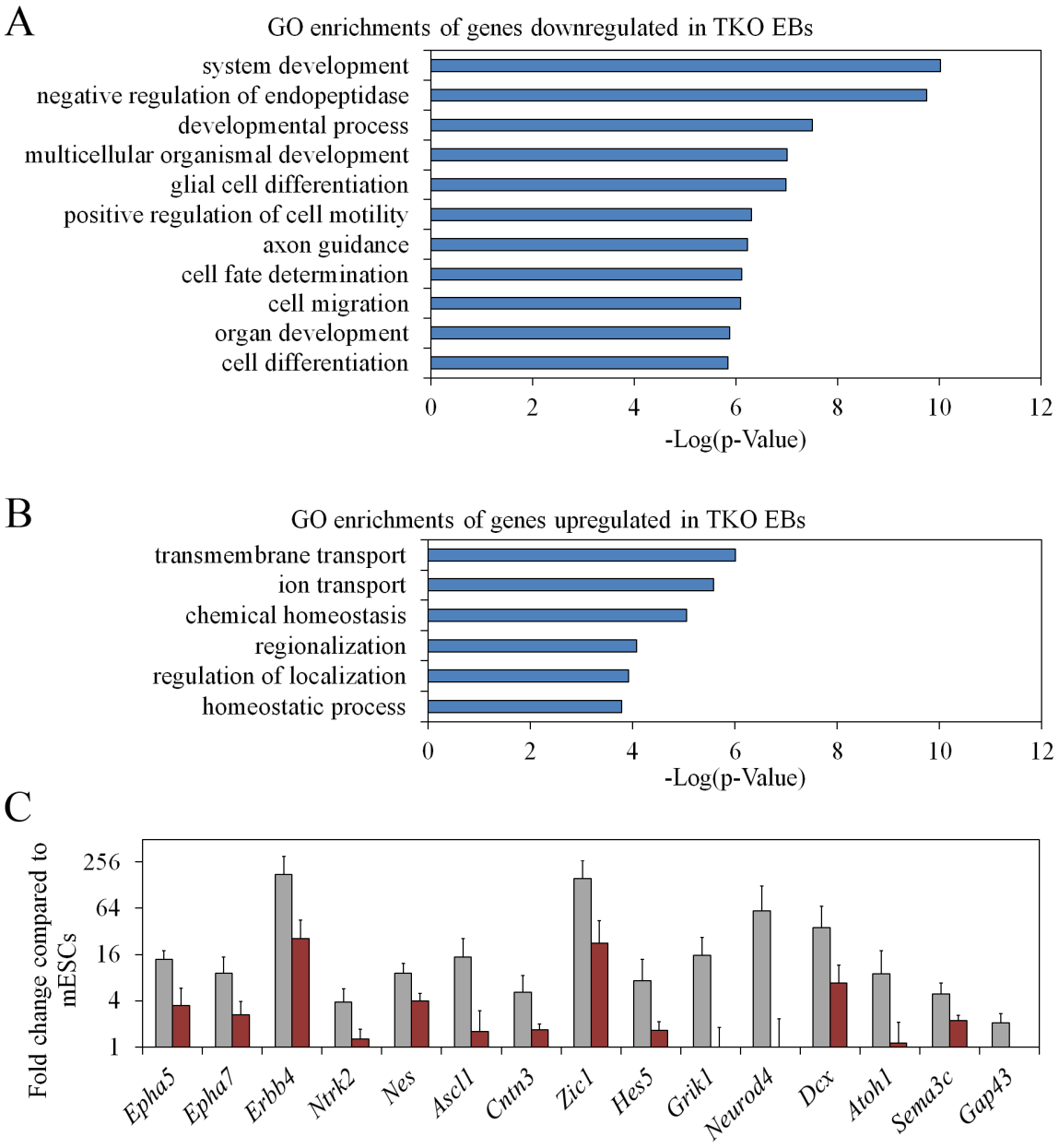


Figure 4.10: Deregulation of neuronal and cell motility related genes in Gadd45 TKO EBs**.**

A. Selected enriched gene ontology (GO) terms associated with genes downregulated in Gadd45 TKO EBs compared to control EBs. **B.** Selected enriched GO terms associated with genes upregulated in Gadd45 TKO EBs compared to control EBs. **C.** Expression levels of genes associated with neuronal GO terms in Gadd45 TKO EBs and control EBs as measured by RNA-Seq. (Expression is normalized to the average expression level in control mESCs. Data shown as mean ± SD, n=3. All genes shown are significantly downregulated in Gadd45 TKO EBs compared to control EBs (False discovery rate < 10%)).

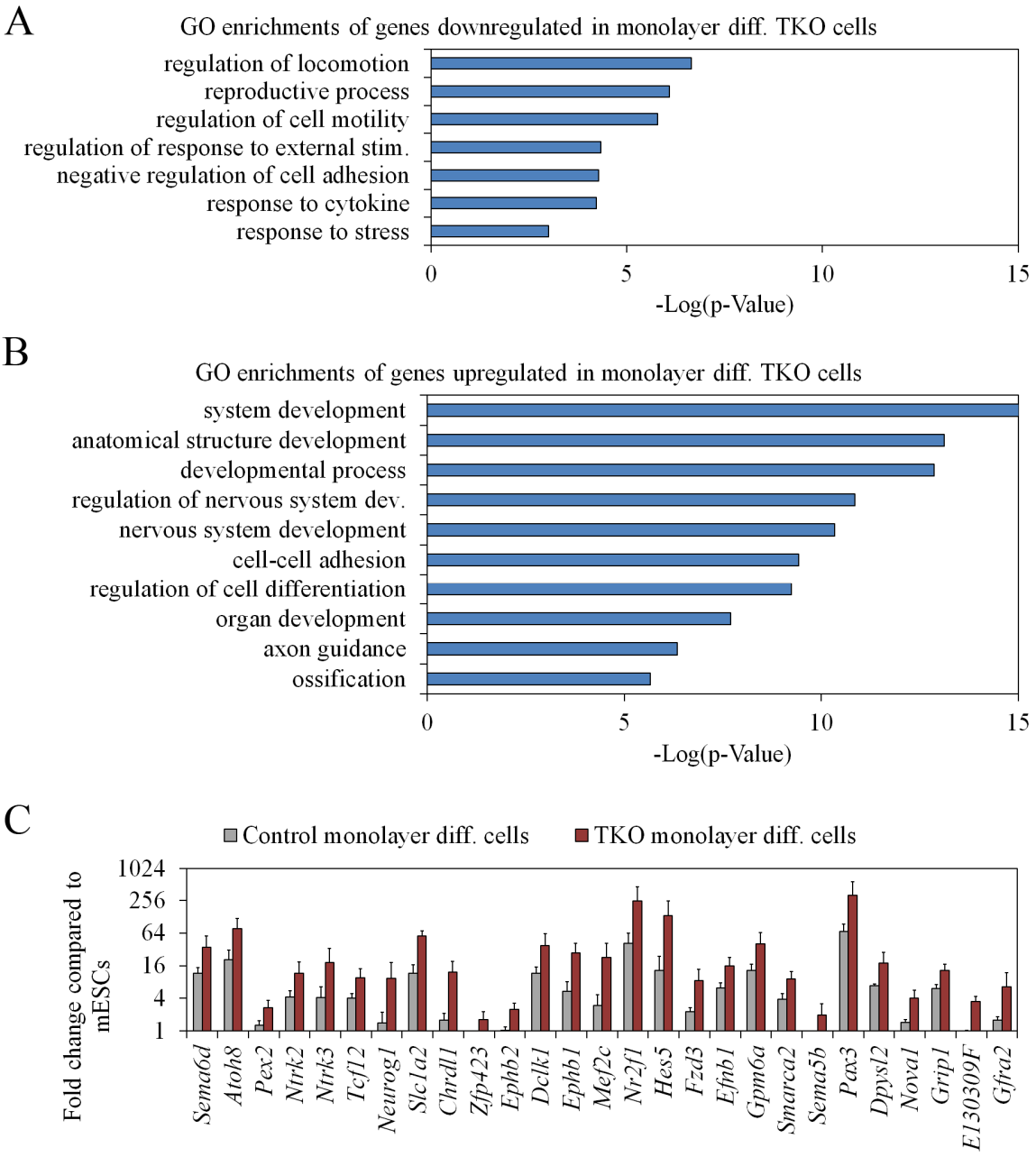


Figure 4.11: Deregulation of neuronal and cell motility related genes in Gadd45 TKO monolayer differentiated cells**.**

A. Selected enriched gene ontology (GO) terms associated with genes downregulated in Gadd45 TKO monolayer differentiated cells compared to control monolayer differentiated cells. **B.** Selected enriched GO terms associated with genes upregulated in Gadd45 TKO monolayer differentiated cells compared to control monolayer differentiated cells. **C.** Expression levels of genes associated with neuronal GO terms in Gadd45 TKO cells and control cells as measured by RNA-Seq (Expression is normalized to the average expression level in control mESCs. Data shown as mean ± SD, n=3. All genes shown are significantly upregulated in Gadd45 TKO cells compared to control cells (FDR < 10%)).

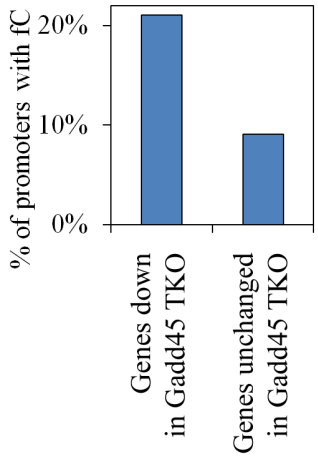
Similar processes were disrupted in Gadd45 TKO cells that have undergone monolayer differentiation. Here, however, genes that were upregulated in the Gadd45 TKO cells compared to control cells were enriched for developmental regulators, particularly those involved in nervous system development (Fig. 4.11B). Examples for these genes are neuronal transcription factors such as *Atoh8*, *Neurog1* and *Hes5* as well as the Ephrin receptors *Ephb1* and *Ephb2* (Fig. 4.11C). Corroborating the findings in the EBs, genes downregulated in Gadd45 TKO monolayer differentiated cells again were enriched for cell motility related genes (Fig. 4.11A), potentially pointing towards a role of the Gadd45 genes in regulating processes such as epithelial-mesenchymal-transition or migration behavior. Considering that neurogenesis is tightly linked to these mechanisms [227], it is possible that the deregulation of neuronal genes and cell migration genes is part of a common phenotype. Notably, roles for the Gadd45 genes in neuronal developmental have been documented before [159,189,196,228–230].

4.1.3.3 Gadd45 target genes overlap with Tet1/2/3 target genes during differentiation

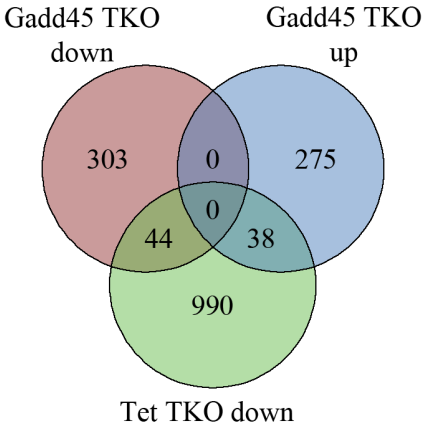
Next, publicly available datasets regarding the function of the Tet or Tdg enzymes were compared to the Gadd45 TKO RNA-Seq datasets. This should help to clarify whether or not the observed deregulated gene expression in Gadd45 TKO cells upon differentiation is due to defects in the DNA demethylation pathway or due to other affected pathways. Unfortunately, only EB differentiation datasets, but not monolayer differentiation datasets, were available.

First, a study by Song et al. identified 5fC at single base resolution and identified promoters enriched for this DNA modification [231]. Arguing that 5fC is an intermediate of DNA demethylation, they defined this set of promoters as undergoing active DNA demethylation upon differentiation. Interestingly, more than 20% of the genes that were downregulated in Gadd45 TKO EBs compared to control EBs are marked by 5fC presence in their promoters (Fig. 4.12A). In contrast, significantly fewer (9%, p-value < 0.0001 Chi-Squared Test with Yates' correction) genes that are unaffected in Gadd45 TKO EBs show this mark in their promoters as well. This shows that Gadd45-dependent genes are frequently regulated by active DNA demethylation.

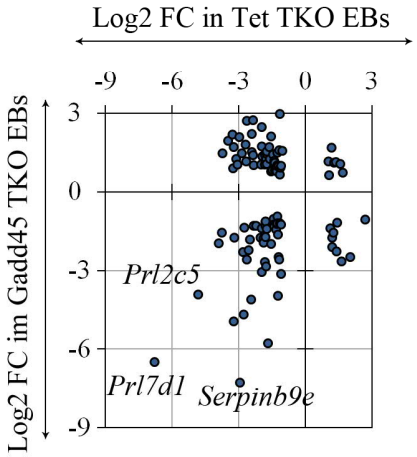
A



B



C



D

Gene	FC ↓	Tissue of main expression
Serpinb9e	156x	Placenta
Mest	153x	Kidney, Brain, Placenta
Psg29	115x	Placenta
Prl7d1	91x	Placenta
Serpinb9f	69x	Placenta
Sst	55x	Pancreas, Brain
Fyb	52x	Testis, Thymus
S100a4	50x	Skin, Placenta
Neurod4	44x	Eye
Anxa13	41x	Pancreas
Serpinb9g	31x	Placenta
Serpinb9c	25x	Ovary, Placenta
Cyp4f14	17x	Liver
Serpinb9d	17x	Placenta
Nap1l5	17x	Brain
Ppbp	16x	Taste Bud, Spleen
Neto1	15x	Brain
Prl2c5	15x	Placenta
Prl6a1	14x	Uterus, Placenta

Figure 4.12: Common Gadd45 and Tet dependent genes are expressed in the placenta.

A. Classification of Gadd45-dependent and Gadd45-independent genes according to the absence or presence of fC in their promoter [231]. **B.** Venn diagram showing the overlap of genes up- or downregulated in Gadd45 TKO EBs compared to control EBs and genes downregulated in Tet TKO EBs compared to control EBs [226]. **C.** Scatterplot comparing logarithmic fold change (FC) in expression levels between Tet TKO EBs and control EBs versus logarithmic fold change (FC) in expression levels in Gadd45 TKO EBs versus control EBs. Each dot represents a single gene up- or downregulated at least 2-fold in the Gadd45 TKO EBs. The lower left square shows genes with correlated downregulation in Gadd45 TKOs and Tet TKOs **D.** List of top downregulated genes in Gadd45 TKO EBs compared to control EBs and their usual tissue of main expression according to Aceview.

Second, a study by Dawlaty et al. analyzed the differentiation potential of *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ *Tet3*$^{-/-}$ TKO mESCs into EBs by RNA-Seq [226]. In total, 950 genes were found to be downregulated in Tet TKO EBs compared to wild type EBs. A significant portion (around 15%, $p<0.0001$ Chi-square with Yates' correction) of both genes up- and downregulated in Gadd45 TKO EBs overlap with these genes (Fig. 4.12B). Importantly, genes that are downregulated strongly in Gadd45 TKO EBs also tend to be more strongly downregulated in the Tet TKO EBs (Fig. 4.12C). Similar to the situation in mESCs (see Fig. 4.5C), this is not the case for genes upregulated in Gadd45 TKO EBs. Notably, genes that are strongly downregulated both in Gadd45 and in Tet TKO EBs include members of the *Prl* and *Serpin* gene families, a finding that will be revisited later.

4.1.3.4 Genes most affected by loss of Gadd45 are usually expressed in the placenta

When comparing the most strongly downregulated Gadd45-dependent genes in EBs with their murine patterns of expression, it becomes obvious that they are predominantly expressed in the placenta (Fig. 4.12D). Genes found in this list include members of the *Serpin*, *Psg* and *Prl* gene families. All of these genes have in common that they are expressed in the late differentiating trophoblast: Serpins are serine protease inhibitors that regulate tissue remodeling during implantation [232], pregnancy specific glycoproteins (Psgs) are cell adhesion molecules that regulate angiogenesis and immune regulation during implanation [233] and *Prl* genes encode for prolactins, a set of pregnancy related hormones, some of which are released from the developing trophoblast [234].

Having this finding in mind, other genes and gene families that play a role during placentation were found to be downregulated in Gadd45 TKO EBs as well, including members of the reproductive homeobox genes *Rhox6* and *Rhox9*, the trophoblast specific protein alpha (*Tpbpa)* as well as the placentally expressed imprinted genes *H19* and *Peg10*. As mentioned above, some of these gene families are similarly downregulated in Tet TKO EBs (Fig. 4.12C), hinting towards a role of DNA demethylation in activating these genes.

Most of these genes were very lowly expressed in EBs (data not shown). Indeed, many were not expressed at all in Gadd45 TKO EBs and were only barely expressed in control EBs, which therefore lead to the observed strong fold-downregulation. This is in accordance with the fact that mESC very rarely transdifferentiate along the trophoblast lineage (see chapter 3.2.3.2). This rare phenomenon might be fully abolished in Gadd45 TKO cells.

Notably, markers of early trophoblast development were not downregulated in Gadd45 TKO EBs (data not shown). These markers however are often also expressed in other lineages, e.g. the developing endo- and mesoderm, unlike the very cell specific markers of late trophoblast differentiation mentioned above. It

is therefore possible that any deficiency of Gadd45 TKO cells to transdifferentiate into early trophoblast cells is masked by the unaltered expression of these markers in the more abundant embryonic germ layers, such as the endo- and mesoderm. Therefore the analysis conducted in EBs does not yet allow the conclusion whether Gadd45 TKO mESCs transdifferentiate less into mTSCs, or whether the arising mTSCs fail to differentiate further and therefore do not express late differentiation markers.

4.1.4 Gadd45 in trophoblast development and mESC transdifferentiation

4.1.4.1 BMP4 induced upregulation of trophoblast genes is impaired in Gadd45 TKO mESCs

As mentioned before, transdifferentiation of mESCs along the trophoblast lineage is a rare event. However, several studies observed mESC to mTSC transdifferentiation without genetic alteration upon addition of growth and signaling factors. Such conditions include WNT3A or BMP4 treatment or growing the cells on defined substrates such as Laminin or Collagen IV [105–107]. Testing these protocols with our E14tg2a mESC strain, a robust upregulation of trophoblast marker genes upon BMP4 treatment could indeed be observed, but less so after treatment of the cells with WNT3A or growing the cells on Collagen IV (data not shown).

In order to corroborate the finding of reduced trophoblast differentiation marker expression in Gadd45 TKO EBs, the ability of Gadd45 TKO mESC to upregulate trophoblast markers upon BMP4 treatment was tested. Gadd45 TKO mESCs and control mESCs were treated with BMP4 for up to 8 days in the presence of Knockout-Serum-Replacement and early and late trophoblast marker gene expression was analyzed via qPCR.

The trophoblast stem cell factors *Cdx2*, *Eomes*, *Elf5* and *Tead4* were all induced by BMP4 treatment. *Eomes* reached maximum expression already after 4 days of treatment, whereas *Cdx2*, *Elf5* and *Tead4* were increased further after 8 days (Fig. 4.13A, comparing black and grey bars). Importantly, upregulation of *Cdx2*, *Elf5* and *Tead4* was impaired in Gadd45 TKO mESCs (Fig. 4.13A, comparing grey and red bars).

Next, expression of late, differentiated trophoblast marker genes was analyzed. As expected, all of these genes were expressed strongest after 8 days of BMP4 treatment (Fig. 4.13B). Notably, induction of all of these genes was severely impaired in Gadd45 TKO cells, corroborating the findings initially made in differentiating EBs (Fig. 4.12). Similarly, induction of two imprinted genes by BMP4 was impaired in Gadd45 TKO mESCs, although to a lesser extent than other late differentiation markers (Fig. 4.13C).

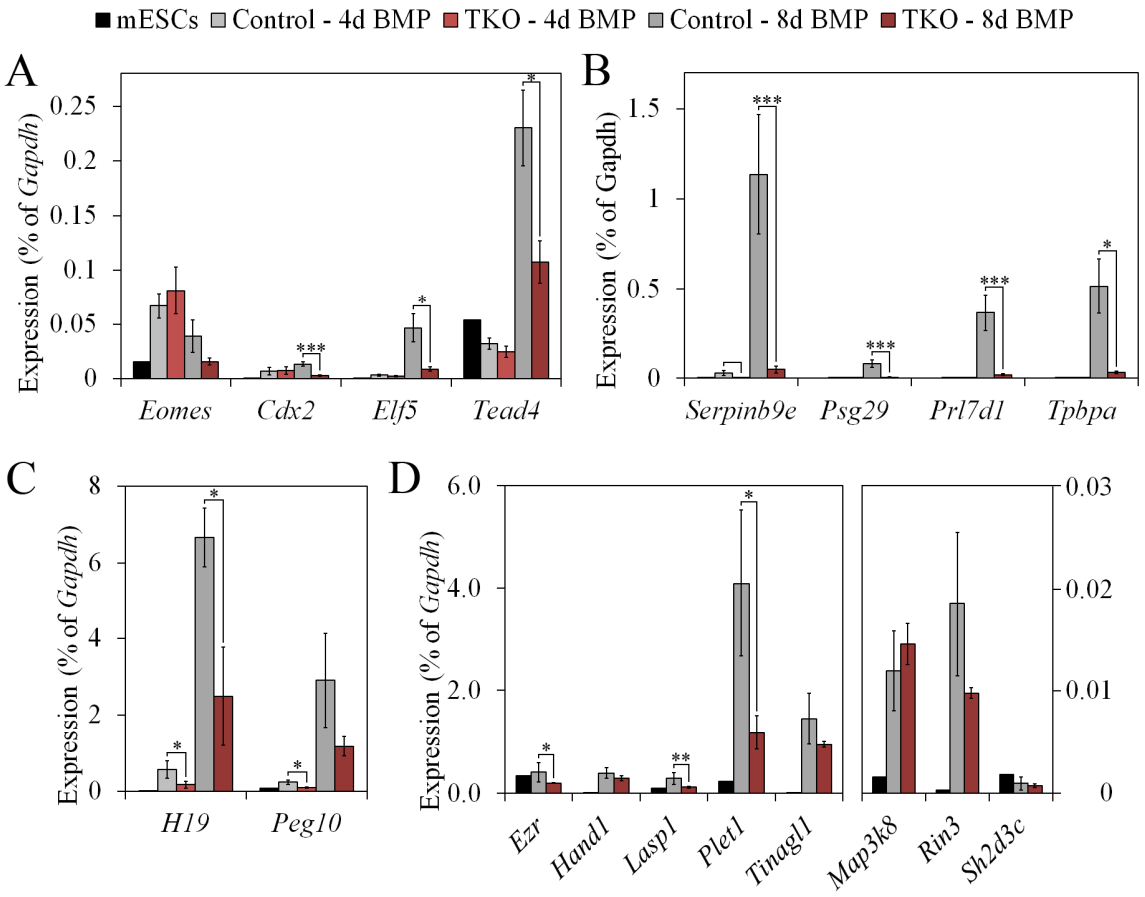


Figure 4.13: BMP4-induced upregulated of trophoblast genes is impaired in Gadd45 TKO mESCs.

A-D. Expression levels of trophoblast specific genes measured via qPCR (Expression is relative to *Gapdh* expression. Data shown as mean ± SD, n=3 for BMP treated mESCs and n=1 for initial expression levels in mESCs. *** = p-value < 0.0005, ** = p-value < 0.005, * = p-value < 0.05 according to student's t-test) **A.** Expression levels of mTSC pluripotency related genes. **B.** Expression levels of TGC related genes. **C.** Expression levels of trophoblast related imprinted genes. **D.** Expression levels of trophoblast related genes known to be regulated via methylation. Note the different y-axis for genes on the left and the right side of the panel.

Whether or not addition of simple signaling factors is enough to trigger mESC to mTSC transdifferentiation has been a matter of debate. Specifically, it has been claimed that due to a partially shared transcriptome between mesoderm and trophoblast, mESC-to-mesoderm differentiation might be mistaken for mESC-to-trophoblast transdifferentiation [108]. Another study aimed at analyzing the efficiency of transgene based transdifferentiation found this process to be incomplete as well, due to insufficient reprogramming of several methylation regulated genes [170]. To analyze whether or not BMP4 treatment is sufficient to induce these highly reprogramming resistant genes, their expression was measured by qPCR

as well. Indeed, 6 out of 8 tested methylation-regulated genes (*Hand1*, *Lasp1*, *Plet1*, *Tinagl1*, *Map3k8* and *Rin3*) were induced by BMP4 treatment, further strengthening the usefulness of this protocol in studying mESC-to-mTSC transdifferentiation (Fig. 4.13D). Moreover, 2 out these 6 genes (*Plet1* and *Lasp1*) showed a significantly reduced induction in Gadd45 TKO mESCs, pointing towards a role of the Gadd45 genes in demethylating the regulatory regions of these genes.

4.1.4.2 Gadd45 genes show distinct but partially overlapping expression domains in the placenta and are upregulated during trophoblast differentiation

The observation that the Gadd45 genes are important for mESC-to-trophoblast transdifferentiation opened up the possibility that they are also important for physiological trophoblast development. In order to gain insights into the potential function of the Gadd45 genes during placenta development, spatial Gadd45 expression patterns were studied in the placenta next. Previous whole mount *in situ* hybridization data from our lab already showed that *Gadd45b* and *Gadd45g* are expressed in tissues that later give rise to the placenta: The chorion and the ectoplacental cone, respectively [187]. *Gadd45a* seems not to be expressed in these tissues. In order to study their further expression during trophoblast differentiation, E10.5 placental sections were analyzed by *in situ* hybridization. Interestingly, all three Gadd45 genes were expressed in partially overlapping domains: *Gadd45a* and *Gadd45g* were expressed both in the trophoblast giant cell layer as well as in the labyrinth layer (Fig. 4.14A, E), whereas *Gadd45b* was expressed exclusively in the labyrinth layer (Fig. 4.14C). *Gadd45g* was also expressed in the maternal decidua (Fig. 4.14E).

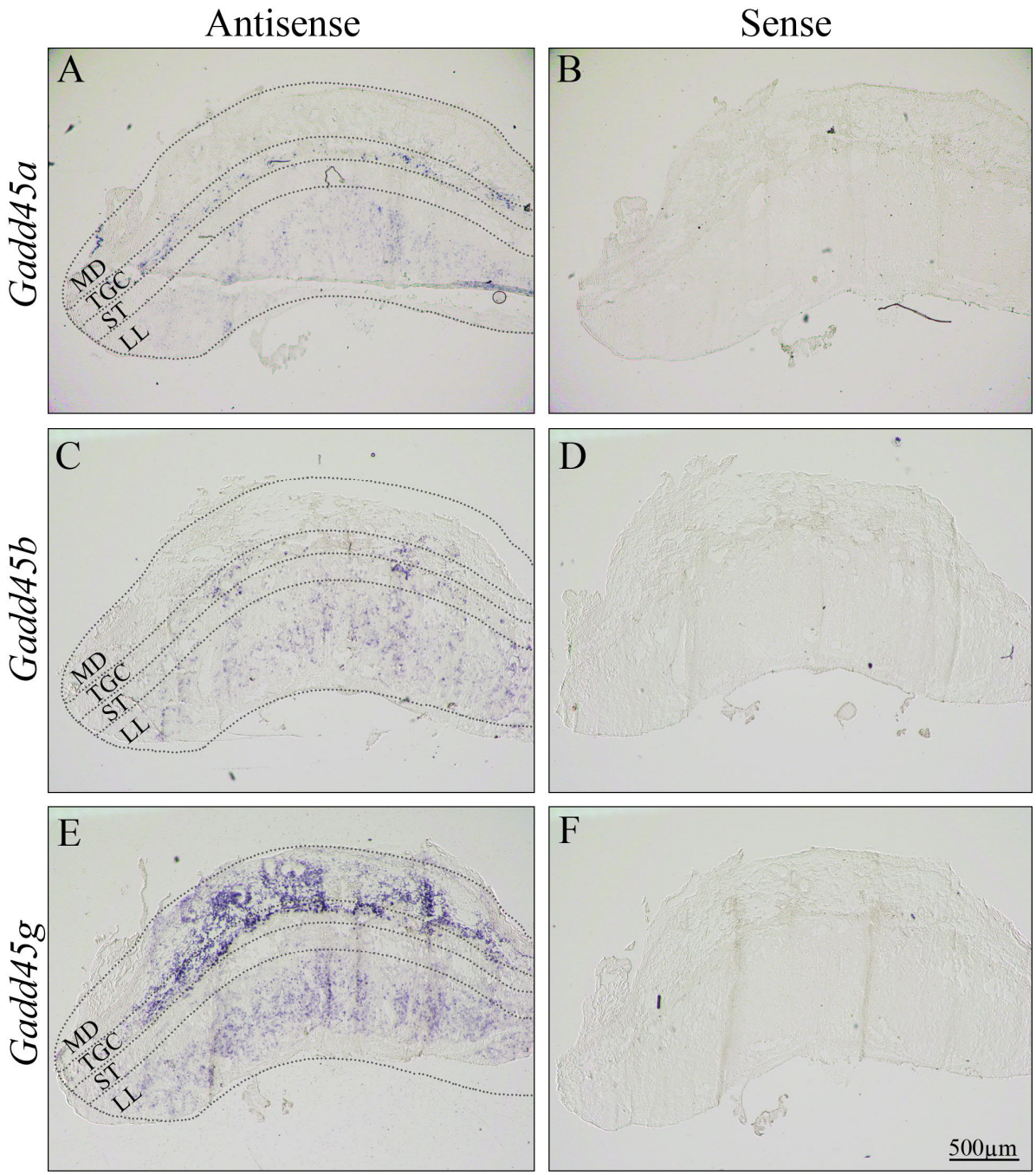


Figure 4.14: *Gadd45a*, *Gadd45b* and *Gadd45g* are expressed in partially overlapping domains in the developing placenta.

A-F. Left column: Microscope images of in-situ hybridizations showing expression domains of *Gadd45a* (A), *Gadd45b* (C) and *Gadd45g* (E) in E10.5 placentas. Right column: Corresponding negative control *in situ* hybridizations performed with sense probes. *In situ* hybridizations carried out by A. von Seggern (this lab). Dashed lines highlight estimated layer structure based on *Gadd45a* expression. (MD = maternal decidua; TGC = trophoblast giant cells; ST = spongiotrophoblast; LL = labyrinth layer)

In order to study the kinetics of Gadd45 gene expression during trophoblast differentiation, mTSCs were differentiated *in vitro* by Fgf4 withdrawal and expression of *Gadd45a*, *Gadd45b* and *Gadd45g* was analyzed by qPCR. All three genes were induced 3-6 fold during mTSC-to-trophoblast differentiation, with *Gadd45b* and *Gadd45g* peaking around day 4 and *Gadd45a* peaking at day 8 of differentiation (Fig. 4.15, blue bars). Notably, all three Gadd45 genes were expressed more strongly in differentiating mTSCs than in mESCs and could be induced by BMP4 treatment of mESCs (Fig 4.15, comparing black and grey bars).

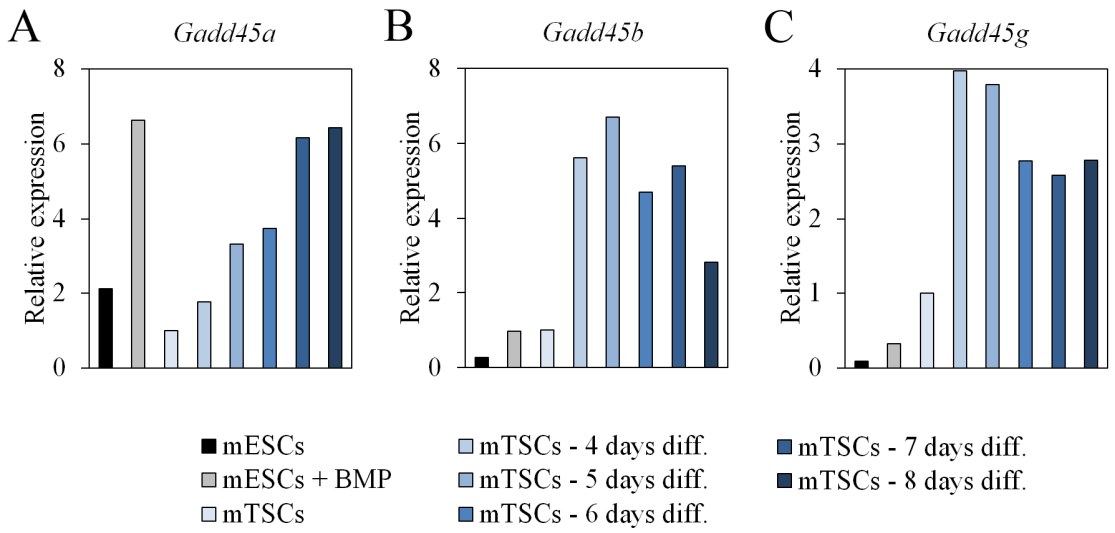


Figure 4.15: *Gadd45a*, *Gadd45b* and *Gadd45g* are induced during trophoblast differentiation.

A-C. Expression levels of *Gadd45a* (A), *Gadd45b* (B) and *Gadd45g* (C) as measured by qPCR under the conditions indicated (Expression is relative to *Gapdh* expression and further normalized to the expression level in undifferentiated mTSCs. n=1).

In summary, all three Gadd45 genes were expressed in the developing placenta in tissues that contain mostly differentiating cells and are induced during differentiation of both mESCs and mTSCs along the trophoblast lineage. Under physiological conditions the Gadd45 genes therefore might play a role during mTSC differentiation, in addition to their role in mESC-to-mTSC transdifferentiation.

4.1.4.3 Loss of Gadd45 genes does not impair trophoblast differentiation

In order to further study the function of the Gadd45 proteins during trophoblast development, loss-of-function experiments were carried out next. An mTSC line derived from E6.5 embryos previously established in a collaborating laboratory [235] was used to create *Gadd45a*$^{-/-}$ *Gadd45b*$^{-/-}$ *Gadd45g*$^{-/-}$ TKO mTSCs, using the same CRISPR/Cas9 knockout design outlined in chapter 4.1.1. However, mTSCs are

less clonogenic than mESCs and more difficult to transfect [236], potentially lowering the probability to receive TKO clones upon a single plasmid transfection. Indeed, only one out of 816 picked clones was identified as being a TKO judged by genotyping PCR (Fig. 4.16A). Western blot analysis confirmed the expected knockout for Gadd45a and Gadd45b (Fig. 4.16B), whereas endogenous Gadd45g remained undetectable also in wild type mTSCs with the available antibodies (data not shown).

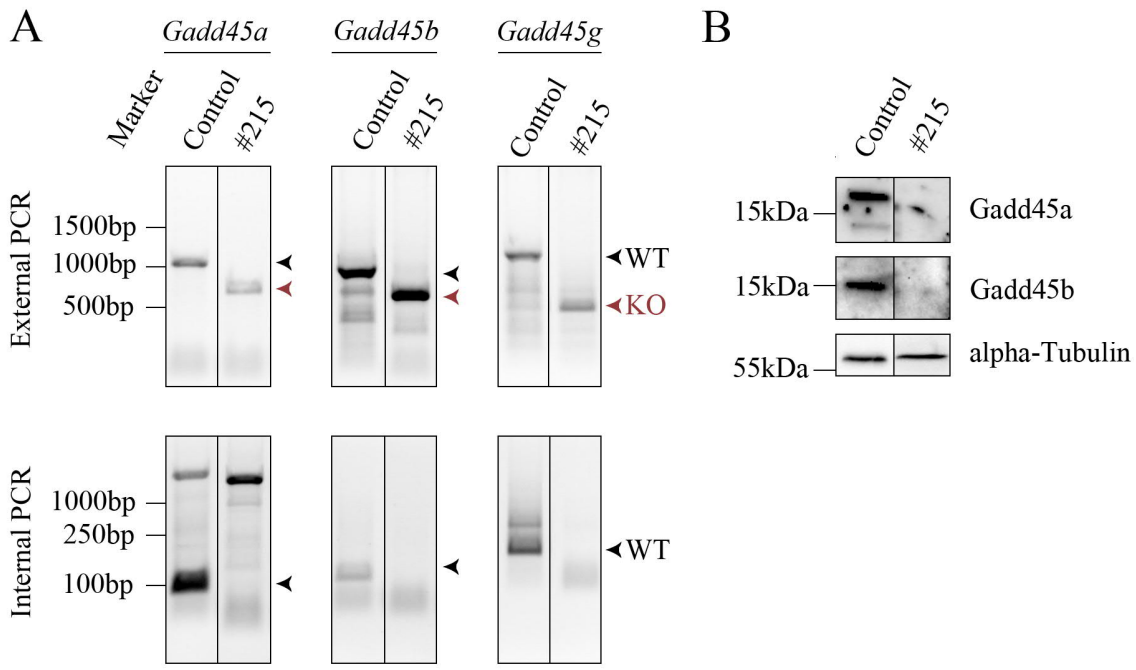


Figure 4.16: Establishment of a single Gadd45 TKO mTSC line.

A. Genotyping PCRs confirming homozygous knockout of *Gadd45a*, *Gadd45b* and *Gadd45g* in clone #215. Arrows indicate expected size of wildtype (WT) and knockout (KO) bands. Top row: PCR using primers outside of the desired deletion. Bottom row: PCR with primers inside of the desired deletion. **B.** Western blot of Gadd45a and Gadd45b in clone #215 and control mTSCs upon 4 days of differentiation and inhibition of the proteasome.

Considering that only a single TKO mTSC clone was obtained at the time point of writing this thesis, the following results have to be taken with caution and should only be considered exploratory in nature and not be judged as final.

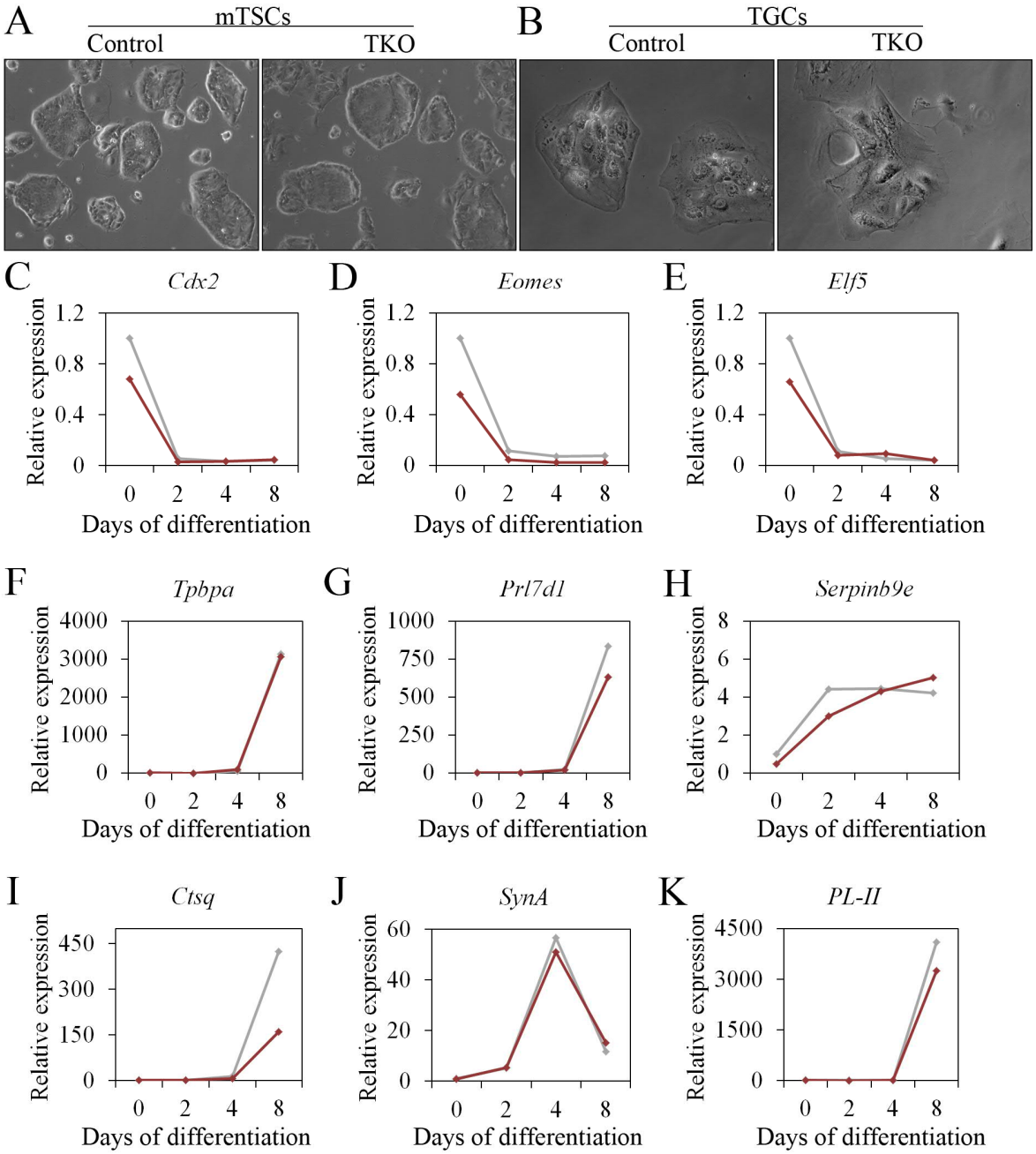


Figure 4.17: Differentiation of a single Gadd45 TKO mTSC line is mostly unimpaired.

A. Microscope images showing morphology of pluripotent Gadd45 TKO and control mTSCs. **B.** Microscope images showing morphology of differentiated Gadd45 TKO and control cells. **C-K.** Expression levels of pluripotency associated (C-E), TGC associated (F-H) and labyrinth associated (I-K) genes as measured via qPCR (Expression is relative to *Gapdh* expression and further normalized to the expression level in undifferentiated mTSCs. n=1).

To study the differentiation potential of the TKO mTSC line compared to the parental mTSC line, mTSCs were differentiated by withdrawal of Fgf4 and conditioned medium. Under these culture conditions mTSC will predominantly differentiate into the trophoblast giant cell lineage, with minor populations of syncytiotrophoblasts arising. The differentiation can readily be observed morphologically, as the small, compact mTSCs will flatten, enlarge tremendously and become polyploid. Neither stem cell morphology (Fig. 4.17A) nor differentiated cell morphology (Fig. 4.17B) was changed in the Gadd45 TKO line compared to the parental line.

RNA was isolated after 0, 2, 4 and 8 days of mTSC differentiation and expression of selected marker genes was analyzed via qPCR. As expected, mTSC stem cell markers (*Cdx2*, *Eomes* and *Elf5*, Fig. 4.17C-E) readily declined during differentiation, whereas both markers for giant cells (*Serpinb9e*, *Tpbpa* and *Prl7d1*, Fig. 4.17F-H) and syncytiotrophoblasts (*Ctsq*, *PL-II* and *SynA*, Fig. 4.17I-K) were induced to various extents. Expression of the mTSC stem cell markers *Cdx2*, *Elf5* and *Eomes* might be reduced in Gadd45 TKO mTSCs, although the observed difference is minor and definitely would need further confirmation from independent clones. More importantly, and unexpectedly, none of the late trophoblast markers that were deregulated after BMP4 treatment or after the EB differentiation (*Serpinb9e*, *Tpbpa*, *Prl7d1* and *SynA*) were deregulated under these more physiological mTSC differentiation conditions in the Gadd45 TKO mTSC line. Indeed, if anything, some of these markers were induced more strongly during Gadd45 TKO mTSC differentiation than in the parental cell line. The only late differentiation marker with reduced expression levels in the Gadd45 TKO mTSCs was *Ctsq*.

As stated before, only limited conclusions can be drawn from these analyses due to the lack of independent clones. Nevertheless, at least this single Gadd45 TKO clone showed no impairment of mTSC differentiation, indicating the Gadd45 genes might be dispensable for the differentiation process itself. If the lowered mTSC pluripotency marker expression in Gadd45 TKO mTSCs can be confirmed in independent clones, this could indicate that the Gadd45 genes are required mainly for maintenance of the mTSC stem cell state. This would fit well to the observed mESC-to-trophoblast transdifferentiation defect.

4.2 Genome wide identification of Gadd45 binding sites

Analysis of CRISPR-mediated triple knockout cells has provided insight into the role of the Gadd45 genes during cellular differentiation. Nevertheless, the experiments described so far have been exclusively descriptive in nature. For example, the observed methylation differences near Gadd45-dependent genes in mESCs (see 4.1.2.4) could be both cause and consequence of the downregulation of these genes. Especially because the Gadd45 genes have multiple, distinct functions, it would be advantageous to gain mechanistic knowledge about the role of the Gadd45 proteins during differentiation. Genome wide identification of Gadd45 binding sites at genomic DNA via ChIP-Seq would provide multiple new insights: First, it would allow distinguishing direct from indirect Gadd45 target genes. Second, it would provide strong evidence for a role of Gadd45-mediated DNA demethylation in target gene regulation, rather than e.g. Gadd45's cytoplasmic role in p38 signaling. Third, many publicly available datasets could be used to provide new clues in regards to potential interaction partners of Gadd45, e.g. by comparing Gadd45 ChIP-Seq profiles with ChIP-Seq profiles for other agents involved in DNA demethylation like Tet or Tdg.

For clarity reasons it should be mentioned that the following experiments and their design were conducted sometimes in parallel and most often even before the results described in chapter 4.1 were obtained. Therefore not all information and tools presented in this thesis so far were available at the time. Importantly, Gadd45 expression levels during mESC and mTSC development were not at hand. Therefore, based on the absence of a strong *Gadd45b* expression during mouse embryonic development according to *in situ* data [187], I focused my studies solely on Gadd45a and Gadd45g.

Endogenous ChIPs have been technically challenging for Gadd45a (personal communication) and due to antibody quality most likely even impossible for Gadd45g. I therefore decided to rather use a tagged Gadd45 protein in combination with an especially robust method of immunoprecipitation. As already outlined in chapter 4.1.1 however, transient overexpression of a tagged gene is unsuitable for studying differentiation processes. Also, because Gadd45 gene expression is dynamically regulated during differentiation and development, a cell line simply overexpressing the tagged Gadd45 version in an ubiquitous fashion was also considered unsuitable. Rather, a tag sequence was introduced to one of the two endogenous alleles of *Gadd45a* or *Gadd45g*, thereby leaving endogenous gene regulation untouched.

A suitable tag that had already been used to great success in a collaborating lab was the V5-3xFlag-BAP tag (M. Treier, MDC Berlin, personal communication). In addition to its regular, antibody-targeted V5 and Flag epitopes, this tag additionally contains a biotinylation acceptor peptide (BAP). This protein domain is targeted and biotinylated by the BirA enzyme. A mESC line ubiquitously overexpressing the BirA gene

was kindly provided by M. Treier. Once the endogenous Gadd45 locus is tagged in this cell line and the resulting fusion protein expressed, BirA will specifically biotinylate the Gadd45 protein. Because this system is derived from *E.coli*, no other targets for BirA are present in the mammalian proteome [237].

The biotinylated protein can subsequently be precipitated by streptavidin coupled beads. The biotin-streptavidin binding is much stronger than regular antigen-antibody binding affinities and therefore allows more stringent washing, therefore - in theory - resulting in more pronounced ChIP enrichments [237].

4.2.1 Gadd45-mediated reporter reactivation is not impaired by epitope tagging

Although the full V5-Flag-BAP tag had already been successfully used in another lab (M. Treier, personal communication), the system might be less suitable for application with the Gadd45 proteins. Gadd45 genes encode small proteins, roughly 160 amino acids in size. The full V5-3xFlag-BAP tag including a small linker would add another 120 amino acids, thereby nearly doubling the size of the protein. Considering that the Gadd45 proteins mostly act via protein-protein interactions, there was a certain likelihood that their function might be impaired in large fusion constructs.

In order to evaluate whether the introduced tag impairs protein function, reactivation of *in vitro* methylated plasmids by transient *Gadd45a* or *Gadd45g* overexpression was analyzed in HEK293T cells. In addition to the original C-terminal V5-3xFlag-BAP tag, shorter C- and N-terminal 1xFlag-BAP and BAP-only tags were tested and compared to myc-tagged Gadd45a and Gadd45g that had been used in the lab previously.

First, *in vitro* methylated luciferase-encoding pGL3 plasmid was co-transfected with the different Gadd45 constructs and their ability to reactivate expression from the pGL3 plasmid was assessed via luminescence measurements. As visualized in figure 4.18A, all constructs induced luciferase expression, however to different extents: On the one hand, the different Gadd45a constructs reactivated the pGL3 plasmid to a similar degree. On the other hand, Gadd45g-mediated reactivation of the luciferase plasmid was impaired by all Flag-containing tags.

To confirm these results, *in vitro* methylated pOctTKGFP plasmid was co-transfected with the different Gadd45 constructs and methylation levels of the -299 site relative to the TSS were measured by methylation-specific PCR, as it has been shown that this site is demethylated in a Gadd45-mediated fashion [213,238]. Indeed, Flag-tag containing Gadd45g fusion proteins induced less demethylation than non-Flag-tag containing fusion proteins (Fig. 4.18B). The full V5-3xFlag-BAP Gadd45g fusion protein did not induce demethylation of this particular CpG at all.

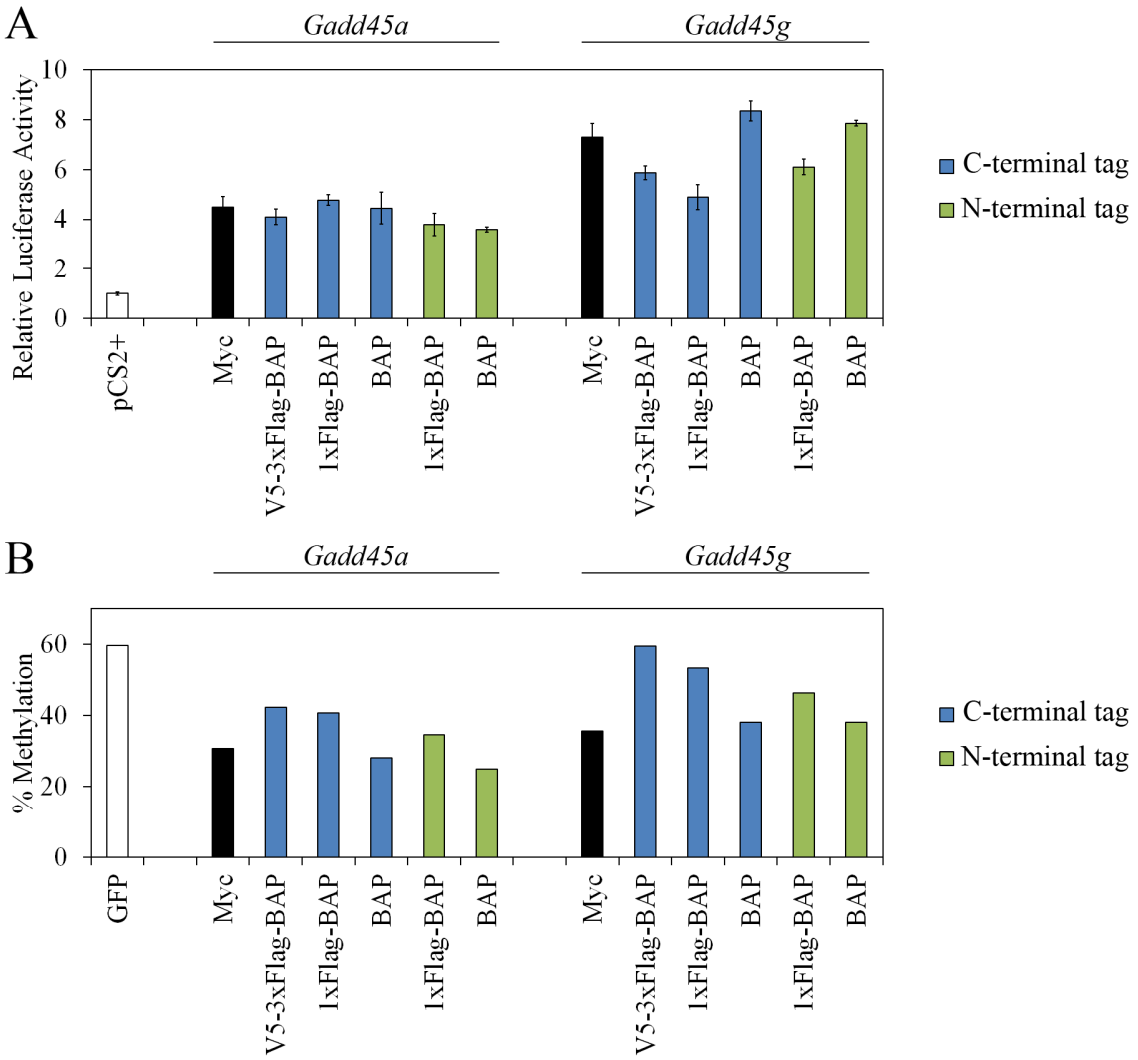


Figure 4.18: Longer tags can negatively affect Gadd45 function in reporter assays.

A. Luciferase activity induced by transfected, *M.sssI in vitro* methylated pGL3 plasmid in HEK293T cells. X-axis indicates co-transfected tagged Gadd45a and Gadd45g constructs. (Luciferase activity is relative to co-transfected pRL renilla luciferase activity and further normalized to the activity level in control transfected HEK293Ts. n=3 technical replicates. Data shown as mean ± SD) **B.** Methylation level of *HpaII in vitro* methylated pOCTTKGFP plasmid 48h after transfection of HEK293T cells as judged by methylation specific PCR. X-axis indicates co-transfected tagged Gadd45a and Gadd45g constructs.

These experiments indicated that changing the tag to a smaller size was beneficial, at least for Gadd45g. However, changing the tag structure might potentially reduce BirA affinity towards the tag, because amino acid composition outside the BAP motif might influence accessibility. To rule this out, the different constructs were transiently co-transfected with BirA in HEK293T cells. Using Western blot, total

Gadd45a and Gadd45g levels were measured using their respective antibodies, whereas biotinylation levels were measured using HRP-coupled streptavidin. Indeed C-terminally BAP-only tagged Gadd45a fusion proteins showed strongly reduced levels of biotinylation and was therefore not suited for further experiments (Fig. 4.19).

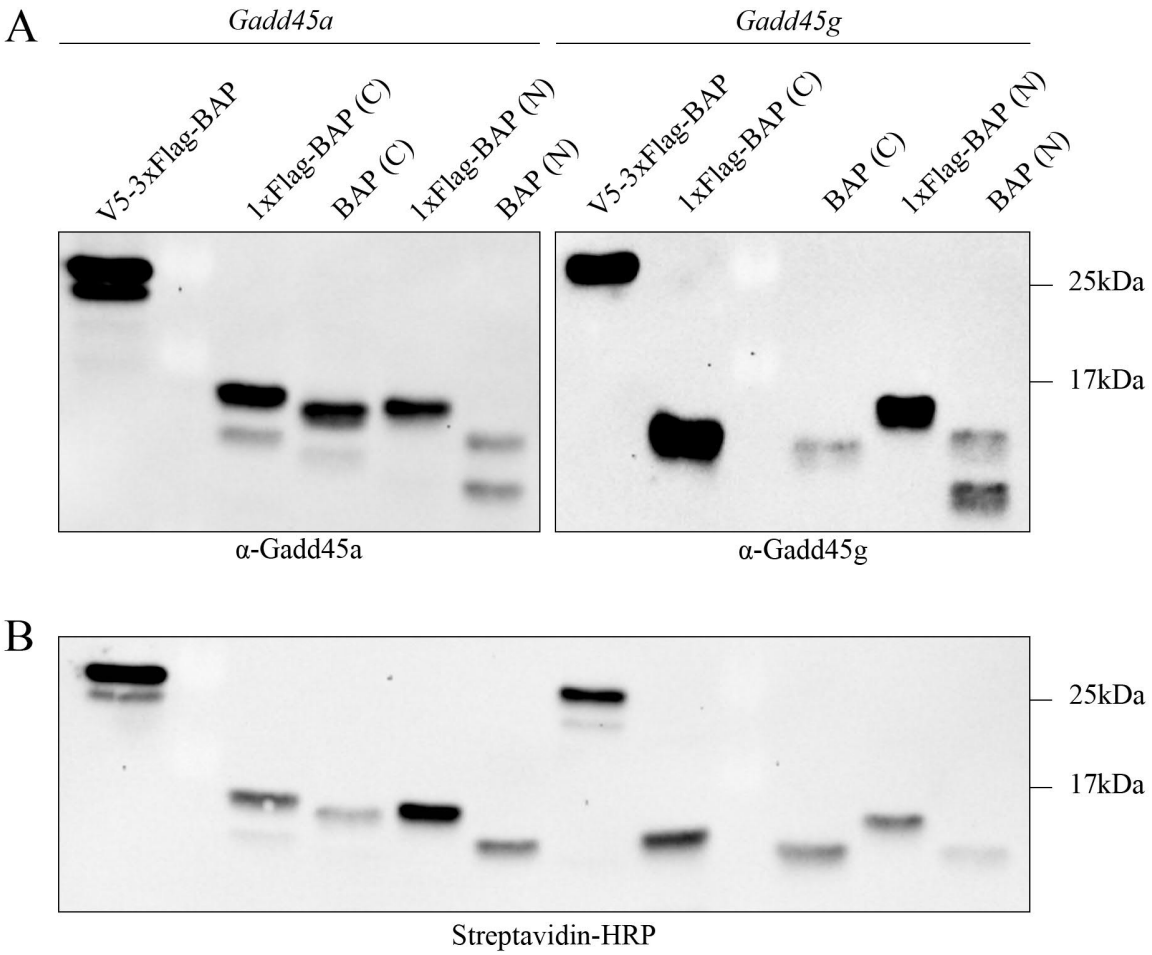


Figure 4.19: Differently tagged Gadd45 fusion proteins are biotinylated with varying efficiency.

A. Representative western blots showing Gadd45a and Gadd45g protein levels after transfection of the indicated constructs in HEK293T cells (C-terminal = (C), N-terminal = (N)). Note that the different constructs are not translated to equal amounts of protein in the cells. **B.** Representative Western blot of the same cell lysates as in (A), but stained with streptavidin-coupled HRP, showing the amount of biotinylated Gadd45a and Gadd45g.

Although Gadd45a tagged with the full V5-3xFlag-BAP tag was still functional in the reporter assays carried out, it was deemed too risky that other, so far untested functions of Gadd45a might be impaired nevertheless.

Taking all this into account, the following tagging strategy was chosen: A 1xFlag-BAP tag was chosen to be used for Gadd45a and the shortest possible tag, the BAP-only version, for Gadd45g. All other constructs lead to diminished reactivation of reporter plasmids and might show even stronger impairment under physiological conditions.

4.2.2 Targeting strategy and establishment of transgenic lines

Because CRISPR-Cas9-mediated gene editing was not developed yet when the following transgenic mESC lines were established, a classical homologous recombination based targeting strategy was chosen (see Fig. 4.20A for a schematic overview). Two targeting vectors were designed and cloned based on a vector backbone kindly supplied by M. Treier. Centrally they feature a transgenic cassette intended to remove the stop codon of *Gadd45a* or *Gadd45g* and replace it by the 1x-Flag-BAP or BAP-only tag respectively, followed by a puromycin resistance cassette for selection.

After electroporation of the targeting vector into BirA overexpressing mESCs, transgenic clones were identified using Southern blot: The transgenic cassette introduced new restriction enzyme sites at the recombined locus (see Fig. 4.20A for an example). The genomic DNA fragment identified by a radioactively labeled probe therefore shifted to a lower size after the respective restriction digest. One out of 48 picked clones and 22 out of 96 picked clones were positive for tagged *Gadd45g* and *Gadd45a* respectively (Fig. 4.20B-C, Fig. 4.21A-B). The strength of the upper and lower bands was roughly equal, indicating a heterozygous addition of the tag.

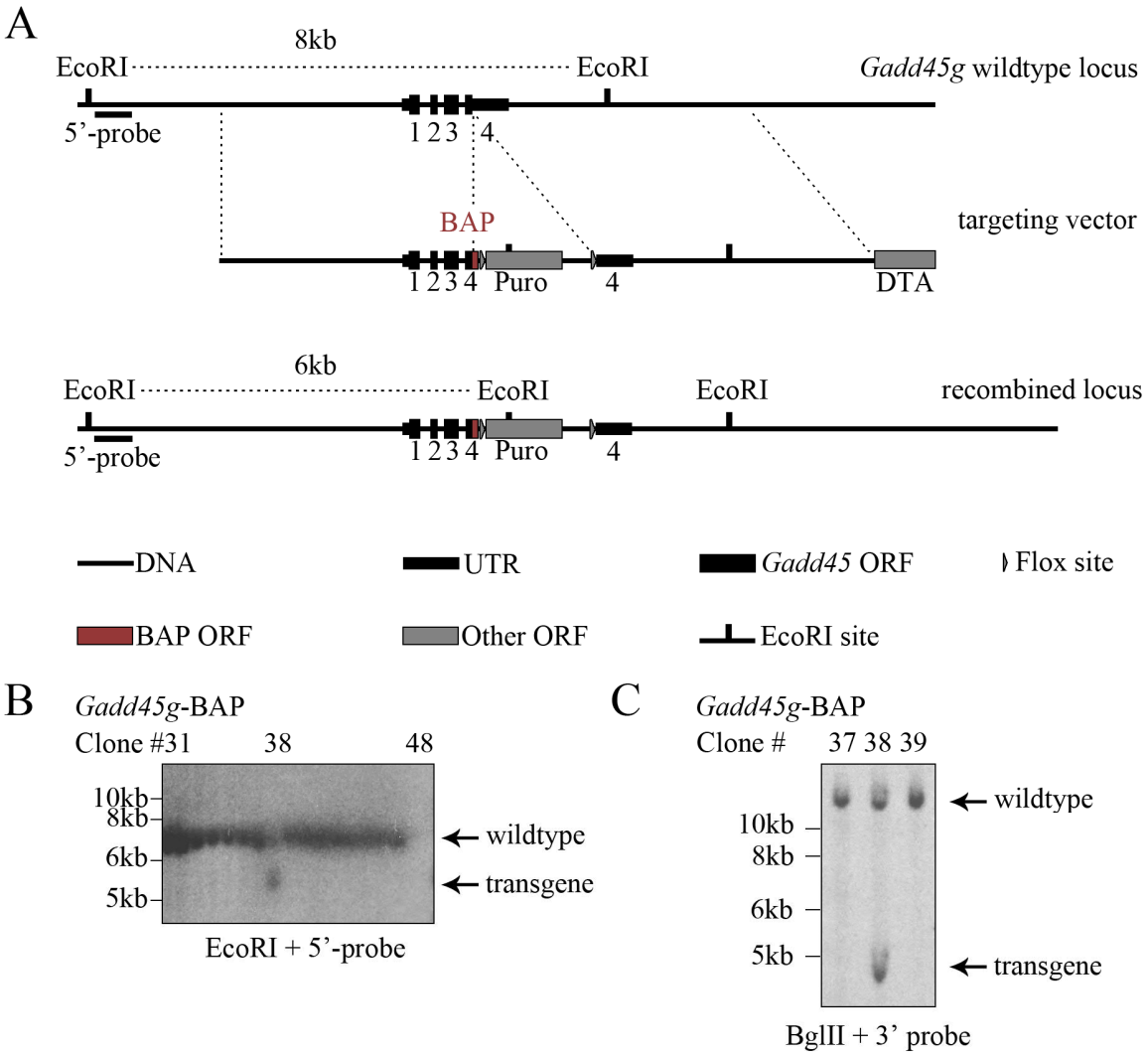


Figure 4.20: *Gadd45g* targeting strategy allows detection of transgenes via southern blot.

A. Scheme depicting the targeting strategy aimed at replacing the endogenous *Gadd45g* gene with a transgenic, BAP-tagged *Gadd45g*. Also shown are the *EcoRI* sites and the position of the radioactive 5'-probe that allow detection of a size shift in southern blots in case of successful recombination. *BglII* sites and the position of the 3'-probe are not shown for clarity reasons. Note that the *Gadd45a* targeting strategy is highly analogous and therefore not shown separately. **B.** Autoradiograph showing the size of the *EcoRI* fragment depicted in (A) in clones 31-48, including the positive clone #38. **C.** Autoradiograph confirming the successful recombination at the *Gadd45g* locus in clone #38, using an independent combination of restriction enzyme and probe.

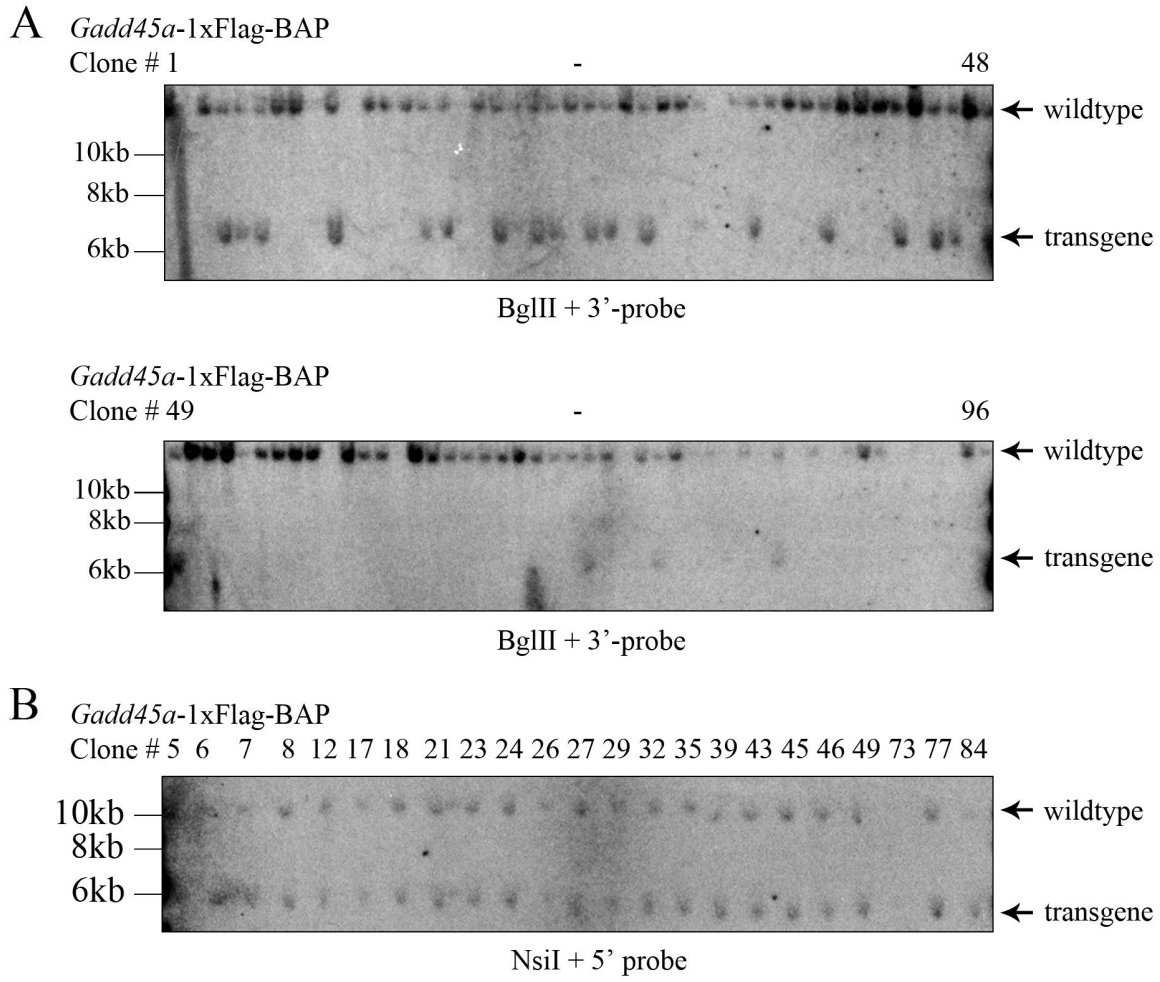


Figure 4.21: Identification of 21 transgenic Gadd45-1xFlag-BAP clones via southern blot.

A. Autoradiograph used to screen for clones with positive recombination events, making use of an incorporated additional *BglII* site. **B.** Autoradiograph confirming the presence of a recombined *Gadd45a* locus in those clones identified as positive in (A), using an independent combination of restriction enzyme and radioactive probe.

Using Western blot, biotinylated protein bands of the expected size were identified in these clones, although endogenous Gadd45 protein levels were too low to detect and bands became only apparent after MG132 based blocking of the proteasome (Fig. 4.22A-B). This confirmed that the targeting had been successful and that the fusion protein was successfully biotinylated.

4.2.3 Gadd45a and Gadd45g ChIP in transgenic mESCs

4.2.3.1 Streptavidin-based precipitation of Gadd45 is only possible under denaturing conditions

Due to their dominant expression in neuronal tissues and their documented role in neuronal development [187,189], it was planned to identify Gadd45a and Gadd45g binding sites in mESC differentiated along the neuronal lineage. Also, because of its relatively strong expression in mESC (Fig. 4.9), identifying Gadd45a binding sites in pluripotent mESCs was deemed interesting. Unfortunately, no known binding sites for Gadd45a or Gadd45g existed in any of these tissues that could be used as a positive control. Therefore, successful ChIP-enrichment could not be tested via qPCR before committing samples to sequencing.

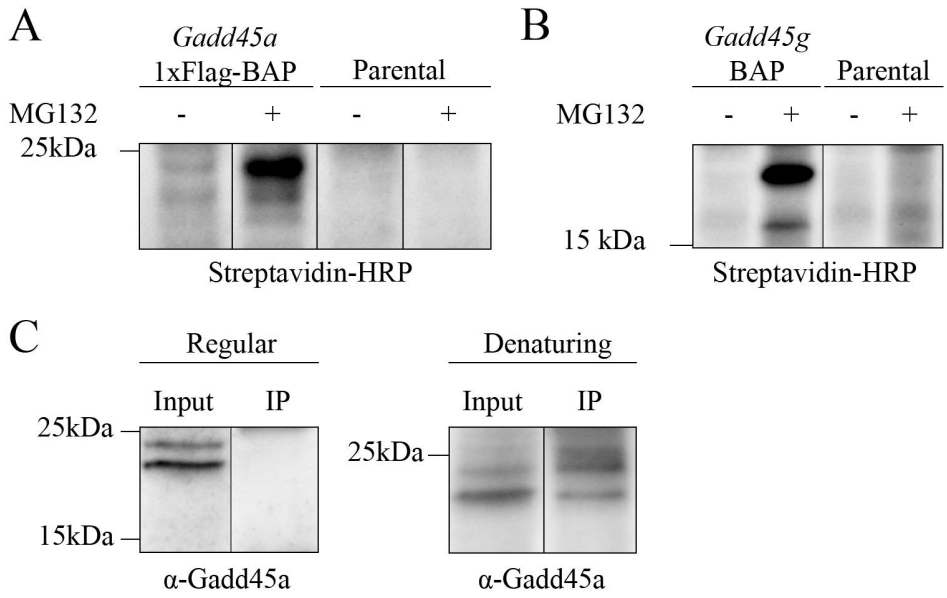


Figure 4.22: Biotinylated Gadd45 can only be successfully immunoprecipitated under harsh denaturing conditions.

A. Western blot showing the presence of an additional biotinylated protein in the lysate of the transgenic Gadd45a-1xFlag-BAP mESC line with a molecular weight corresponding to Gadd45a. **B.** Western blot showing the presence of an additional biotinylated protein in the lysate of the transgenic Gadd45g-BAP mESC line with a molecular weight corresponding to Gadd45g. **C.** Western blots showing immunoprecipitation efficiency under normal and denaturing conditions. Immunoprecipitation was carried out using streptavidin coupled magnetic beads and Western blot detection was done using an antibody against Gadd45a.

In order to evaluate at least immunoprecipitation (IP) efficiency, IP Western blots were carried out, using streptavidin coupled magnetic beads and cell extracts from the transgenic mESC lines containing

biotinylated Gadd45a or Gadd45g. Surprisingly, despite the theoretically strong biotin-streptavidin binding, IP efficiency was low under regular, non-denaturing IP conditions (Fig. 4.22C, left panel). Reasoning that this could indicate that the epitope was hidden and not accessible to the streptavidin-coupled beads, a number of different denaturing IP conditions were evaluated. Indeed, strongly denaturing conditions using the alkylating agent iodoacetamid and heating the samples briefly to 65°C increased IP efficiency significantly (Fig. 4.22C, right panel).

4.2.3.2 Most potential Gadd45 binding sites are identified in monolayer differentiated cells

Using these optimized conditions, five ChIP-Seq experiments were carried out using the transgenic mESC lines and the previously described neuronal differentiation protocols (see chapter 4.1.3.2): Gadd45a ChIPs were conducted in pluripotent mESCs, monolayer differentiated mESCs and EBs treated with retinoic acid. Gadd45g ChIPs were performed in monolayer differentiated cells and retinoic acid treated EBs only, considering the very low *Gadd45g* expression in pluripotent mESCs (Fig. 4.9). As mock-ChIP controls, streptavidin pulldowns were also carried out under these conditions using extracts from cells without biotinylated Gadd45a and Gadd45g, i.e. from the parental BirA overexpressing mESC line.

Notably, DNA was not sheared by sonication, as most commonly done for ChIP, but via enzymatic micrococcal nuclease treatment. Using this method, genomic DNA is cut between the nucleosomes and therefore DNA of a very defined length is created, which is ideal for library preparation [239].

In total, ten libraries were prepared and sequenced on an Illumina HiSeq 2000, yielding on average 10^8 raw 50 bp reads of very good quality per sample (data not shown). Reads were mapped to the mm9 mouse genome and peaks were called using the MACS peak finder algorithm [240]. Both positive and negative peaks were identified, i.e. regions of DNA that are – as expected – enriched in the ChIP versus the mock-ChIP control, as well as regions that are enriched in the mock-ChIP control versus the ChIP. The first represent potential Gadd45 binding sites, whereas the latter provide an estimate of the amount of false positive peaks.

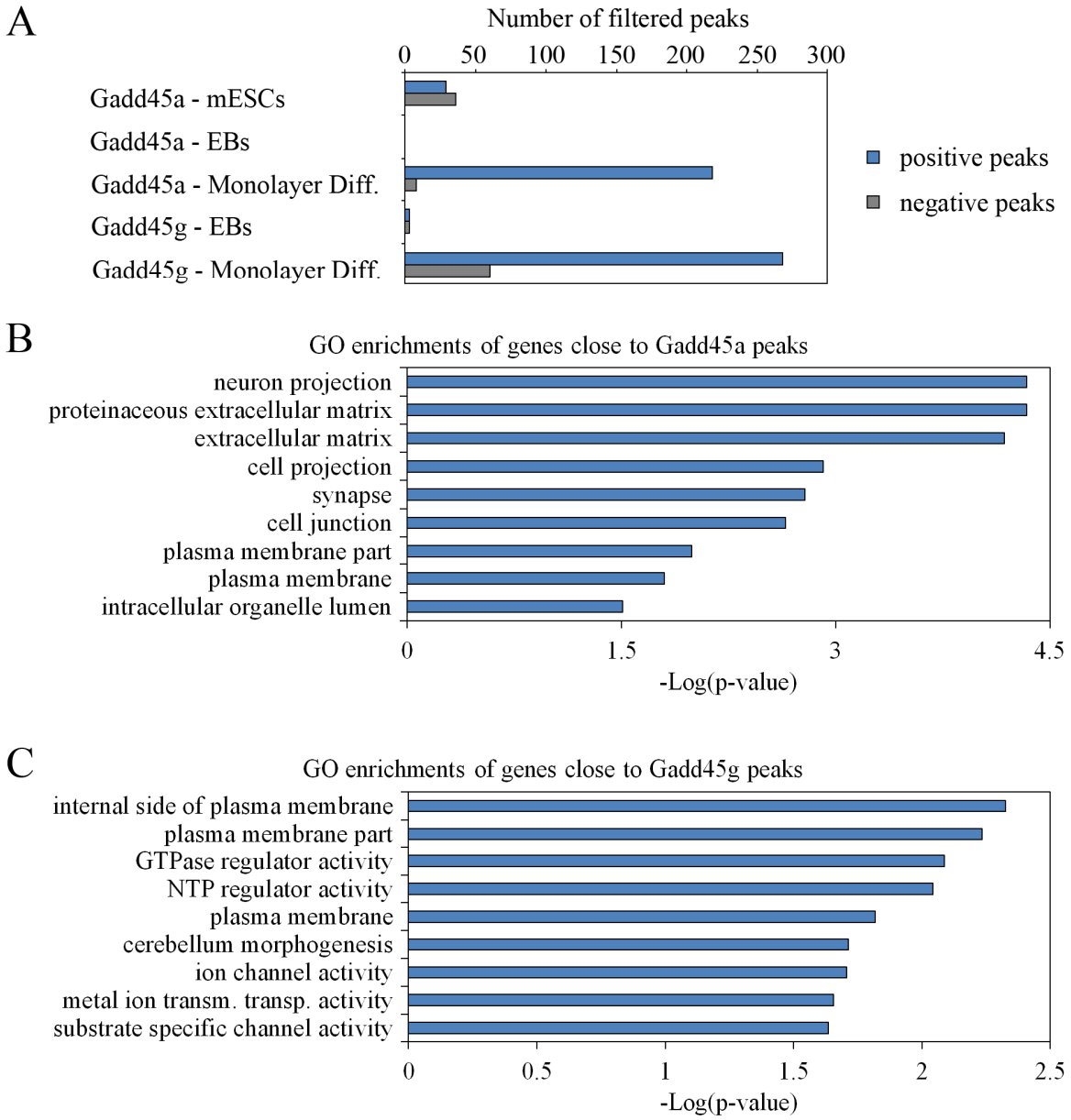


Figure 4.23: Identified potential Gadd45a and Gadd45g binding sites are most numerous upon monolayer differentiation.

A. Number of identified peaks in Gadd45a-1xFlag-BAP and Gadd45g-BAP ChIP-Seq under different cell culture conditions. Positive peaks show at least 10x enrichment of read density in the ChIP in the transgenic line versus mock-ChIP in the parental line. Conversely, negative peaks show at least 10x higher read density in the mock-ChIP in the parental versus the ChIP in the transgenic line. **B.** Enriched gene ontology (GO) terms linked with Gadd45a-peak-associated genes in monolayer differentiated cells. **C.** Enriched GO terms linked with Gadd45g-peak-associated genes in monolayer differentiated cells. ChIP-Seq peak calling and GO analysis performed by E. Karaulanov (IMB Mainz).

The total amount of identified peaks varied greatly across the different conditions (Fig. 4.23A): In pluripotent mESCs, roughly 30 positive and negative peaks were called each, arguing that only few potential Gadd45a binding sites were identified in mESCs, with most of them likely being false positives. In retinoic acid treated EBs, barely any peaks were called either after Gadd45a or Gadd45g ChIP, again arguing that no Gadd45 binding sites could be identified under these differentiation conditions. More promisingly, over 200 positive peaks were called both for the Gadd45a and Gadd45g ChIPs in the monolayer differentiated cells, while the number of negative peaks was considerably lower. Further analysis therefore concentrated on analysis and validation of the potential Gadd45a and Gadd45g binding sites in monolayer differentiated neuronal cells.

Interestingly, genes close to putative Gadd45a binding sites were enriched for genes involved in neuron projection, synapse function and membrane localization (Fig. 4.23B). This observation corresponded well with previous reports hinting towards a function for Gadd45a in neuronal development and particularly neurite outgrowth [241,242]. Genes close to potential Gadd45g binding sites were enriched for functions related to the plasma membrane, particularly ion channels, which are a fundamental element of functioning neurons (Fig. 4.23C).

4.2.3.3 Gadd45 binding sites cannot be confirmed by independent ChIP-qPCR experiments

Closer inspection of these peaks however revealed a significant problem. Due to the use of the micrococcal nuclease, reads were not evenly distributed across the genome but clustered in "mini-peaks" around the position of the nucleosomes (Fig. 4.24A-C). Moreover, the size and distribution of these "mini-peaks" was not always equal across samples. The mESC-mock sample e.g. showed a much smoother distribution compared to other samples (Fig. 4.24A-C, top row). These differences are likely a result of slight differences in nuclease digestion and potentially also differences in nucleosome density or positioning. Importantly, detailed analysis of the positive peaks in the monolayer differentiated samples showed that many of these potential binding sites did not show a significantly higher peak in the ChIP sample, but rather an unusually low peak in the mock-ChIP sample (Fig. 4.24A, marked by arrow).

To limit the problems arising from the uneven nucleosome pattern across samples, potential binding sites were filtered more stringently by selecting only those peaks for further validation that were significantly enriched not only in comparison to its related mock-ChIP sample, but in comparison to all five mock-ChIP samples.

Examples for these potential binding sites are shown in figures 4.24B and 4.24C. In these examples, the peaks were consistently enriched versus all controls; however, the height of the peaks was still only marginally higher than the background. Also, many of these stringently selected peaks were not located

near annotated genes and therefore their function remained unclear. Considering these caveats, it was of vital importance to validate the potential binding sites in independent ChIP experiments.

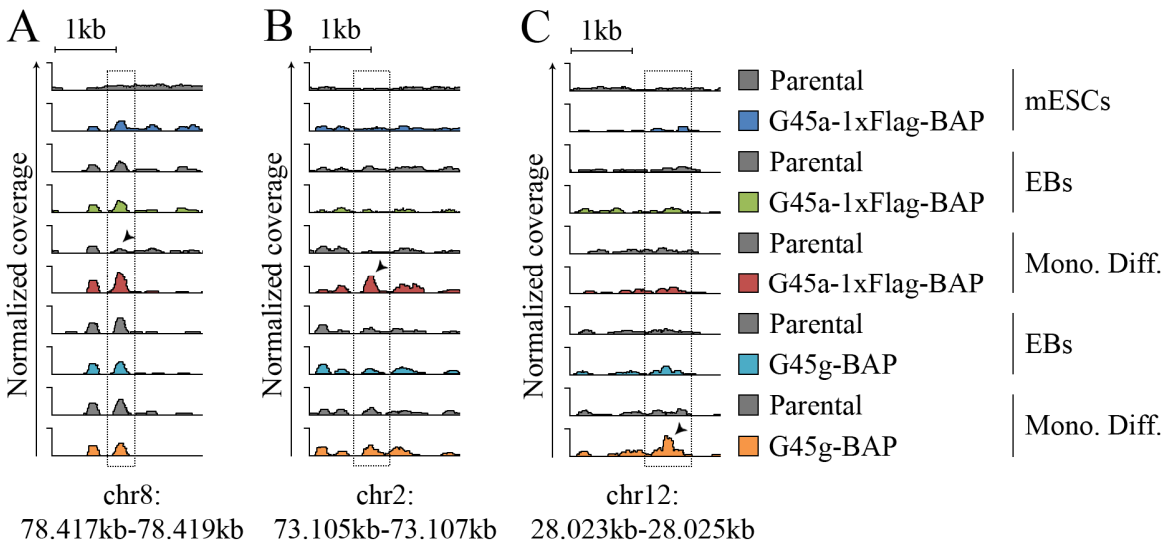


Figure 4.24: Slightly varying nucleosome patterns in ChIP-Seq tracks lead to false-positive ChIP peaks.

A-C. ChIP-Seq tracks from five ChIP-Seq experiments in both transgenic and parental cells at the indicated genomic locations. **A.** Example for a false positive peak. Note that the peak in the Gadd45a-1xFlag-BAP Monolayer differentiation track is enriched versus its own control track (marked by arrow), but not against any of the other tracks. **B.** Example for a potential Gadd45a binding site. Note that the peak marked by the arrow is enriched in comparison to all control tracks. **C.** Example for a potential Gadd45g binding site. Note that the peak marked by the arrow is enriched in comparison to all control tracks. Original ChIP-Seq read density tracks supplied by E. Karaulanov (IMB Mainz).

In total, 49 regions that showed significant enrichment in monolayer differentiated cells compared to all mock control samples were picked for further validation: 11 potential Gadd45a binding sites were selected due to their relatively strong enrichment, irrespective of their position relative to genes. 6 potential Gadd45a sites were selected because they were in close proximity to neuronal genes. Similarly, 17 Gadd45g binding sites were chosen due to their strong enrichment and 7 binding sites were chosen because they were close to neuronal genes. Additionally, 8 potential binding sites were selected for validation because they were close to genes that were identified as being downregulated in *Gadd45g*$^{-/-}$ spinal cords in a microarray study conducted by M. Gierl (this lab, data not shown). As negative controls, 3 regions next to housekeeping genes were included that did not show differences between ChIP and mock-ChIP samples.

First, technical reproducibility of ChIP peaks was analyzed: Using left-over library DNA from next generation sequencing, enrichment of selected regions was analyzed via qPCR. Unfortunately already at this level, the majority of potential binding sites could not be validated. Only 14 out of 49 regions showed an enrichment of at least twofold over mock-ChIP controls in the qPCR (Fig. 4.25 and 4.26), arguing that technical reproducibility was low.

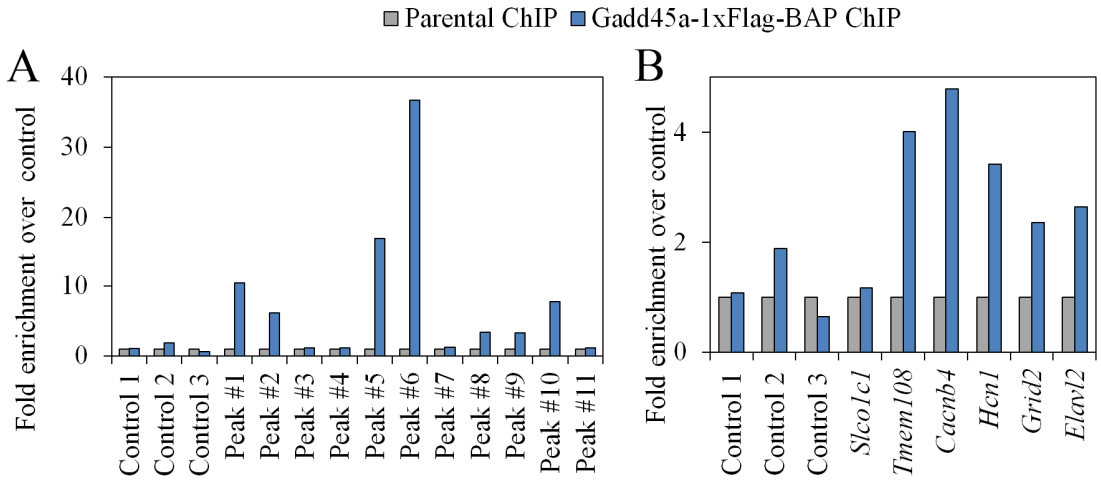


Figure 4.25: Only a subset of identified Gadd45a ChIP-Seq peaks can be technically confirmed by qPCR.

A-B. qPCR on library-DNA measuring enrichment of indicated peaks in the transgenic Gadd45a-1xFlag-BAP cell line versus the parental cell line. (Enrichments were normalized to the mock-ChIP in the parental cell line.) **A.** ChIP-Seq peaks selected based on high enrichment in the ChIP-Seq. **B.** ChIP-Seq peaks selected due to vicinity to neuronal genes.

Nevertheless, two potential Gadd45a binding sites (peak #5 and peak #6) and two potential Gadd45g binding sites (peaks near the *Tcch* and the *Sox21* genes) were then selected for further validation with independent replicates, considering that these regions were enriched also in the qPCR based analysis of the library DNA. Streptavidin based ChIPs of tagged Gadd45a or Gadd45g were carried out using cell extracts from three independent monolayer differentiation cultures and enrichment of potential binding sites was analyzed via qPCR. None of the three Gadd45a-ChIPs was able to confirm the binding to either the peak #5 or peak #6 sites (Fig. 4.27A). In contrast, two out of three Gadd45g ChIPs resulted in a modest twofold enrichment of potential binding sites near the *Sox21* and the *Tcch* genes, whereas the third Gadd45g ChIP did not result in enrichment at these sites (Fig. 4.27B).

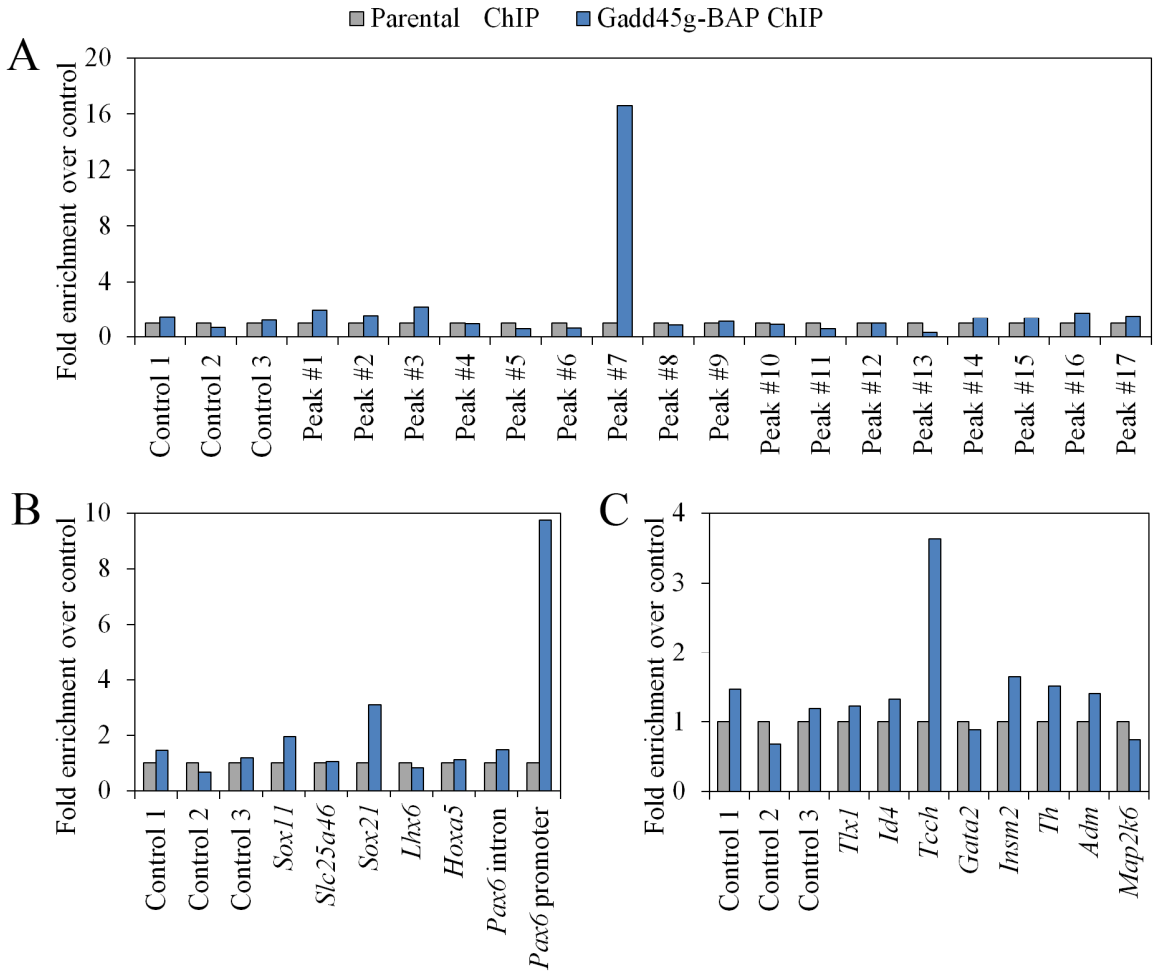


Figure 4.26: Only a minority of identified Gadd45g ChIP-Seq peaks can be technically confirmed by qPCR.

A-C. qPCR on library-DNA measuring enrichment of indicated peaks in the transgenic Gadd45g-BAP cell line versus the parental cell line. (Enrichments were normalized to the mock-ChIP in the parental cell line.) **A.** ChIP-Seq peaks selected based on high enrichment in the ChIP-Seq. **B.** ChIP-Seq peaks selected due to vicinity to neuronal genes. **C.** ChIP-Seq peaks selected due to vicinity to genes deregulated in *Gadd45g*$^{-/-}$ spinal cords.

Most likely due to the very low amounts of DNA being handled during the ChIP procedure, stochastic processes may often lead to minor fluctuations in the measured enrichment of any given DNA locus. A twofold enrichment can therefore hardly be considered convincing evidence of chromatin association. However it is also possible that Gadd45g is only weakly associated with DNA, potentially even only in a subset of cells, which could similarly lead to weak twofold enrichments.

In order to investigate this issue further, more independent ChIP experiments were carried out, also comparing different fixation and immunoprecipitation procedures in hope of increasing enrichments. In summary, only a small fraction of the performed ChIPs confirmed the enrichment at the *Sox21* and *Tcch* sites, and enrichment never exceeded a twofold threshold (data no shown). Ultimately it was therefore concluded that these regions most likely represented falsely identified binding sites. Given the overall poor reproducibility of the ChIP-Seq peaks, this would imply that no binding sites for Gadd45a and Gadd45g were identified using this method. Whether this indicates that Gadd45a and Gadd45g are not associated with chromatin in monolayer differentiated cells or whether the results were a consequence of a suboptimal ChIP strategy cannot be determined conclusively. Arguments for either case will be provided in greater detail in the Discussion.

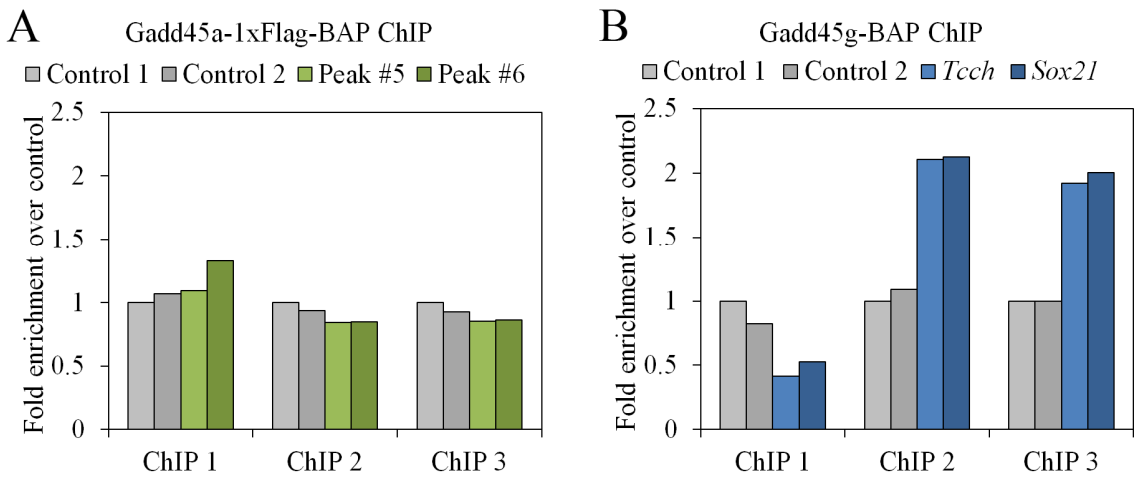


Figure 4.27: Attempts at biological validation of identified potential Gadd45a and Gadd45g binding sites.

A. ChIP-qPCR experiments showing no enrichment for previously strongly enriched peaks #5 and #6 for Gadd45a (compare Fig. 4.25A). (ChIP enrichment relative to mock-ChIP in parental mESCs and normalized to control region 1.) **B.** ChIP-qPCR experiments showing minor enrichments for peaks near the *Sox21* and *Tcch* genes in 2 out of 3 Gadd45g-ChIPs. (ChIP enrichment relative to mock-ChIP in parental mESCs and normalized to control region 1.)

Lastly, potential binding sites of Gadd45a in undifferentiated mESC were analyzed in greater detail. Initially, the similar amount of positive and negative peaks had been discouraging (Fig. 4.23A). Also, the few peaks that had been identified showed signs of being artifacts: The enriched regions had very sharp boundaries and had a considerably higher read coverage than surrounding stretches of DNA (data not shown). Artifacts like these usually arise from wrongly annotated repetitive elements: If an element is

present in a higher copy number than annotated, sequencing reads that come from different physical elements are all mapped to a single annotated region, thereby leading to false enrichment.

However, in this particular case of ChIP enrichment, this artifact could be resolved. All peaks identified after Gadd45a-ChIP in mESCs were part of the rDNA locus, which is partly present in the mm9 genome build at multiple locations throughout the genome. Therefore, ChIP-Seq data was re-mapped to a prototype rDNA locus.

rDNA loci are constructed of multiple similar elements orientated in a tandem fashion. Each element consists of two promoters, the actual rRNA encoding part and a linker [243]. Notably, Gadd45a-ChIP resulted in an enrichment of the promoter and rRNA DNA regions, but not in an enrichment of the linker (Fig. 4.28A). This finding confirmed previous findings of GADD45A binding to the rDNA locus in human cells [147].

Similar to validation efforts of monolayer differentiation ChIP results, validation experiments carried out in mESCs were ambiguous: Initial technical replication of the sequencing result by qPCR on left-over library material confirmed enrichment of the rDNA locus in the Gadd45a-ChIP mESC library DNA (Fig. 4.28B). However, independent ChIP experiments predominantly failed to reproduce the binding of Gadd45a to the rDNA locus (Fig. 4.28C). Notably, ChIPs carried out with cell extracts from mESCs grown on gelatin for increasing times resulted in progressively diminishing enrichments of the rDNA locus (Fig. 4.28D), potentially pointing towards biological variability as a cause for inconclusive ChIPs.

Taken together, the binding of Gadd45a to the rDNA locus in mESCs remains the best candidate for a valid DNA association, especially considering previous results obtained in a different cell line [147].

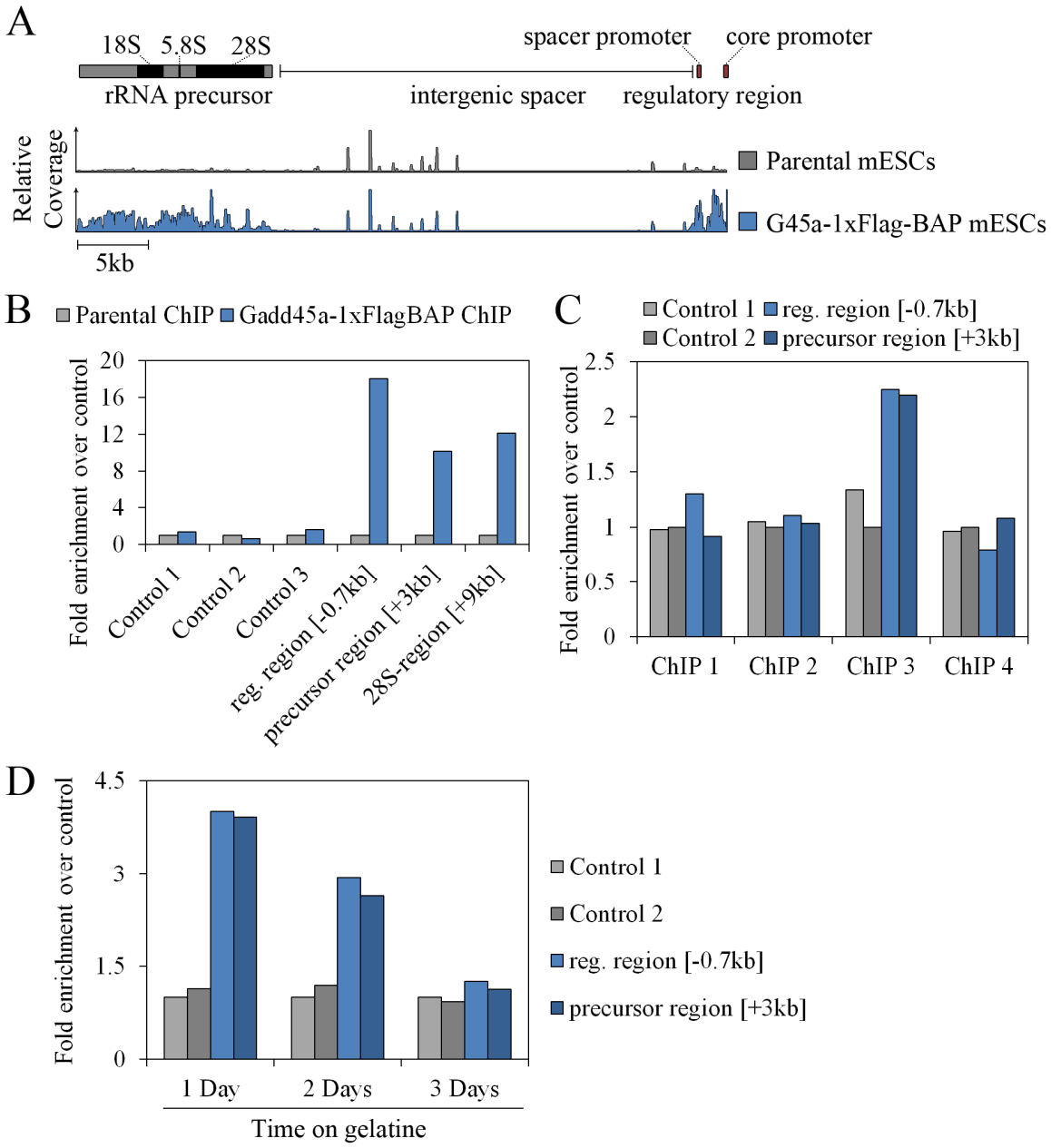


Figure 4.28: Potential Gadd45a binding sites at the rDNA locus.

A. ChIP-Seq tracks in mESCs both from the Gadd45a-1xFlag-BAP transgenic and the parental mESC line. Scheme at the top indicates the position of different functional elements in the rDNA locus. **B.** qPCR on library-DNA technically confirms enrichment at three selected rDNA regions. (Enrichment calculated relative to mock-ChIP in the parental line). **C.** The majority of independent ChIP-qPCR experiments do not show strong enrichment at the rDNA locus. (Enrichment relative to mock-ChIP in the parental cell line and normalized to control region 1). **D.** ChIP-qPCRs using chromatin from cells cultured for increasing times on gelatine. (Enrichment relative to mock-ChIP in the parental cell line and normalized to control region 1).

5 Discussion

5.1 Gadd45-dependent regulation of distinct methylation-sensitive genes in pluripotent mESCs

The functional role of methylated cytosine, its oxidative derivatives and their active removal in regards to gene regulation in mESCs is currently not very well understood. In pluripotent mESCs, *Tet1* and *Tet2* expression levels are relatively high and so are the resulting amounts of 5hmC, 5fC and 5caC in the genome [136]. This sparked great interest, and many studies have subsequently focused on analyzing the distribution of these oxidative 5mC derivatives and the binding sites of the Tet enzymes in the mESC genome. These studies clearly show that 5mC, 5hmC, 5fC and 5caC are not distributed randomly, but cluster at regulatory elements, often correlated to expression levels of nearby genes [222,244,245].

The functional relevance of these epigenetic marks however has not been ultimately elucidated and different studies report sometimes conflicting phenotypes upon loss of the Tet enzymes: Initial work in the Zhang laboratory reported a direct involvement of the Tet enzymes in mESC pluripotency via regulation of *Nanog* [246]. Similarly, another study reported impaired LIF-dependent Stat3 signaling upon loss of Tet1 and an ensuing differentiation phenotype [247]. However multiple reports from the Jaenisch [226,248,249] and the Rao laboratories [222,250] show no diminished expression of pluripotency markers upon loss of the Tet enzymes in mESCs. Moreover, *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ mESC, even though almost completely devoid of 5hmC, can contribute to all germ layers in teratoma assays and in chimeric embryos upon blastocyst injection [248]. Only once all three Tet enzymes are lost is differentiation severely impaired, although mESCs themselves remain viable [226].

Similarly, mESC maintenance is unimpaired upon Tdg depletion [251]. Therefore, active DNA demethylation seems to play only a limited role in mESCs, despite paradoxically high Tet- and consequently high 5hmC levels [124]. Rather, active DNA demethylation might be important for fine-tuning differentiation events upon exit from pluripotency. Indeed, many of the aforementioned studies show skewed differentiation of Tet-depleted mESCs towards the TE lineage [246,248–250].

In light of this, it is fitting that the deregulation of genes in the Gadd45 TKO mESCs was strongest upon differentiation and weakest at ground state pluripotency, whereas the phenotype of Gadd45 TKO mESC cultured in serum is intermediate (see chapter 4.1, particularly Fig. 4.4 and 4.9). Therefore, the Gadd45 genes might be required more for the transition towards differentiation than for mESC maintenance, similar to the Tet genes. This is well in line with the observation that the Gadd45 genes are often

transiently induced upon various stimuli, but usually only mildly expressed under steady-state conditions [252].

Nevertheless, more than a hundred genes were deregulated in Gadd45 TKO mESCs under standard culture conditions (Fig. 4.4), hinting towards the possibility that the Gadd45 genes also have a defined, but limited, role in mESC maintenance. Deregulated genes might e.g. be especially sensitive to changes in DNA methylation levels. Intriguingly, a deregulation of imprinted genes, which are classically regulated via methylation, and a corresponding hypermethylation at the *H19* ICR could be observed in Gadd45 TKO mESCs (Fig. 4.3 and 4.7).

However, the functional impact of this deregulation is currently unclear. Imprinted genes are believed to be a result of the slightly conflicting interests of mother and father: Whereas the mother has an interest in preserving energy for her own survival, the father's interest lies more in the fitness of the offspring. Therefore many imprinted genes encode for growth and metabolism related genes, which typically become only relevant during post-implantation development, when the mother has to nurture the embryo [253]. In line with this, mESCs express many imprinted genes still in a bi-allelic fashion [254]. It is therefore possible that deregulation of imprinted genes in Gadd45 TKO cells would only lead to a visible phenotype *in vivo* after implantation.

The second notable set of genes deregulated in Gadd45 TKO mESCs contained *Pramel6*, *Pramel7* and *Calcoco2*/*Ndp52* (Fig. 4.3 and data not shown). These are three out of the 26 genes identified as being LIF targets in mESCs [255]. Furthermore, *Pramel7* overexpression is sufficient to keep mESCs pluripotent in the absence of LIF [256]. Also, a direct knockdown of *Pramel7* induces upregulation of differentiation genes even in presence of LIF [256]. However, *Pramel7* downregulation in Gadd45 TKO mESCs did not have the same effect, as Gadd45 TKO mESCs remained pluripotent and did not upregulate differentiation related genes (Fig. 4.2). It is therefore likely that the loss of the Gadd45 genes could mostly be compensated for by the highly adaptable pluripotency network.

Additional corroboration for a limited role of the Gadd45 genes in mESC maintenance comes from a recent study by Li et al. that analyzed the methylome of *Gadd45a*$^{-/-}$ *Gadd45b*$^{-/-}$ double knockout (DKO) mESCs via reduced representation bisulfite sequencing (RRBS) [210]. In total, only 68 hypermethylated regions were identified in Gadd45 DKO mESCs, arguing for a restricted and very site-specific role of the Gadd45 genes in DNA demethylation in steady-state mESCs.

However, several points need to be considered regarding this study. First, it is possible that there is compensatory upregulation of *Gadd45g* in the Gadd45 DKO mESCs, potentially attenuating the resulting

DNA hypermethylation. Indeed, an upregulation of the truncated *Gadd45g* transcript could be observed in Gadd45 TKO mESC, indicating that mESCs sense the loss of the Gadd45 genes and try to compensate for it (see Fig. 4.1A).

Second, RRBS enriches for CpG dense regions [257]. DNA hypermethylation at CpG sparse regions, which often carry out regulatory functions, would therefore most likely be missed. Indeed, work in our lab has shown that DNA hypermethylation upon loss of Gadd45a and Ing1 in mouse embryonic fibroblasts primarily occurs at enhancers (personal communication). Thus, to study the impact of the Gadd45 genes on DNA methylation patterns, it would be beneficial to analyze the methylome of Gadd45 TKO cells in an unbiased fashion, e.g. by whole genome bisulfite sequencing.

Notably, the study by Li et al. did not detect DNA hypermethylation at the sites presented in chapter 4.1.2.4, and conversely, genes in vicinity of the 68 hypermethylated sites identified by Li et al. were not deregulated in the Gadd45 TKO mESCs studied here (data not shown). These differences can potentially be explained by the different mESC strains used (V6.5 versus E14tg2a) or the different methods employed (genome wide RRBS might not be covering the sites studied in this work using site specific bisulfite sequencing). Unfortunately, Li et al. do not provide RNA-Seq data, making it impossible to compare the Gadd45 DKO and TKO mESC transcriptomes.

In summary, both my work and the work by Li et al. point towards a measurable, but limited role of the Gadd45 genes in gene regulation via DNA demethylation in pluripotent mESCs. The functional impact of this gene regulation so far remains elusive. Ultimately, both the genes identified during my work as well as the sites identified by Li et al. still need to be shown to be direct Gadd45 targets via ChIP. Furthermore, the exact identity of the DNA demethylation pathway recruited by Gadd45 in mESCs has to be elucidated. Until then it cannot be fully ruled out that Gadd45 regulates these genes via functions independent from DNA demethylation, e.g. by activation of the p38/JNK pathway.

5.2 Diversely impaired differentiation of Gadd45 TKO mESCs

A varied set of differentiation protocols was used in this study to gain an unbiased view on the role of the Gadd45 genes during differentiation. At the same time, however, the diversity of differentiation protocols complicates analysis, because the role of the Gadd45 genes might be context-specific. Indeed, only a minor fraction of genes was commonly deregulated in Gadd45 TKO cells upon undirected EB differentiation and monolayer differentiation (see chapter 4.1.3.2). Additionally, hardly any genes were found to be deregulated upon retinoic acid stimulation in the Gadd45 TKO EBs compared to control EBs (Fig. 4.9). This strongly implies that the function of the Gadd45 genes during cellular differentiation is tissue- and context-dependent.

At this point, it can only be speculated why the number of deregulated genes was lowest in retinoic acid treated EBs. On the one hand it is possible that a strong inducer of differentiation is less susceptible to changes in DNA methylation. Retinoic acid surely is such a strong signal, as it will lead to direct binding of nuclear hormone receptors at their target genes. On the other hand it is possible that the Gadd45 gene family is not needed during this particular differentiation step from mESCs to neuronal precursors. Thus, subsequent analysis focused on the other two employed differentiation protocols.

The undirected EB differentiation has the great advantage that all three germ layers are formed in a semi-physiological context. On the one hand this allows identification of *bona fide* differentiation defects, such as Gadd45 TKO cells failing to differentiate into a certain lineage altogether. On the other hand it allows detection of more subtle tissue contribution biases, when Gadd45 TKO cells might be more prone to differentiate along a certain lineage, if they are given the possibility to "choose". The drawback of the protocol is that results are expected to be noisy, as each arising cell type will potentially have its own set of deregulated genes, which will be superimposed in RNA-Seq. Additionally, any given gene can be induced in different tissue contexts. Therefore the measured fold-change between Gadd45 TKO and controls cells will ultimately result from integration over both affected and non-affected tissues.

In agreement with this, the set of deregulated genes in Gadd45 TKO EBs was complex, with the most significant enrichments clustering around the generic terms "system development" and "developmental process" (Fig 4.10). Conversely, selected gene sets that were subsequently focused on, i.e. neuronal genes, cell motility genes and trophoblast genes, make up only a small portion of all deregulated genes (<5%). Therefore the EB differentiation protocol is most appropriately used in an explorative manner and results obtained should be further corroborated in more controlled environments. On the one hand this was done in experiments aimed at deciphering the role of the Gadd45 genes in trophoblast transdifferentiation, which will be discussed in the next chapter. On the other hand the observation that neuronal and cell motility related genes were deregulated could be confirmed by analysis of the monolayer differentiation RNA-Seq data (Fig. 4.11).

Indeed, neuronal and cell motility related genes were deregulated in Gadd45 TKO cells, both upon EB and monolayer differentiation. This corroborated earlier work in *Xenopus*, which showed that Gadd45a and Gadd45g are necessary for the exit from pluripotency and the entry into neural differentiation programs [189], as well as work in P19 cells that showed that Gadd45g is sufficient to induce neurogenesis if overexpressed [230]. Neuronal differentiation, cell motility and cell cycle progression are very closely linked during development, indicating that the deregulation of these genes might be part of a common phenotype, potentially even indicating that the Gadd45 genes are involved in coordination of these processes.

Intriguingly, *Gadd45a* and *Gadd45g* are both expressed in dorsal root ganglia, which are derivatives of the neural crest and which therefore depend strongly on the coordination of cell migration and neuronal differentiation. Unfortunately, neural crest development is difficult to recapitulate *in vitro* [258,259], making mESCs an unsuitable model to further study a potential role of the Gadd45 genes in this tissue.

Although expression data regarding neuronal and cell motility genes in differentiating Gadd45 TKO cells were promising and in accordance with published data, several caveats regarding these experiments have to be considered. First and foremost, whereas neuronal genes tended to be downregulated in Gadd45 TKO EBs compared to control EBs, they tended to be upregulated in Gadd45 TKO cells compared to control cells upon monolayer differentiation (Fig 4.10 and 4.11). However, the genes downregulated in EBs were not identical to the ones upregulated upon monolayer differentiation, making it possible that they do not serve exactly the same function, despite both being involved in neuronal differentiation.

Second, the mechanism by which the Gadd45 genes steer differentiation currently remains unidentified, although efforts were made to find indications whether or not gene deregulation might be coupled to DNA methylation (Fig. 4.12). Considering the pleiotropic functions of the Gadd45 proteins, it seems likely that the observed misexpression of genes is at least partly not attributable to deficiencies in active DNA demethylation.

In summary the RNA-Seq experiments indicated a broad role of the Gadd45 genes during differentiation processes, particularly neuronal differentiation, although the underlying mechanisms are currently unclear. Also, it was not possible to determine whether the Gadd45 genes act in a redundant fashion during differentiation, as it is the case in mESCs. Rather, it is also possible that the observed differentiation phenotype is an accumulation of independent *Gadd45a*, *Gadd45b* and *Gadd45g* knockout phenotypes. It seems likely that the neural differentiation defect, potentially in combination with the cell motility related gene deregulation, is mostly driven by loss of *Gadd45a* and *Gadd45g*, given their expression pattern during embryogenesis and their comparable function in *Xenopus*. Further investigation is required to clarify the exact nature of the role of the Gadd45 genes in neuronal development. Subsequent work, however, focused on the potential involvement of the Gadd45 genes in trophoblast (trans)differentiation.

5.3 The nature of the trophoblast (trans) differentiation defect of Gadd45 TKO cells

Several lines of evidence strongly point towards a function of the Gadd45 genes in trophoblast differentiation and/or transdifferentiation. First, they were found to be expressed during physiological placenta development in partially overlapping expression domains (Fig. 4.14). Second, Gadd45 TKO EBs differentiated less along the trophoblast lineage, as judged from terminal TGC marker gene expression

(Fig. 4.12). Third, BMP-induced transdifferentiation was impaired in Gadd45 TKO mESCs, again judged by TGC marker gene expression (Fig. 4.13).

However, Gadd45 TKO mTSCs were viable and did not show obvious differentiation defects (Fig. 4.17). In addition, the physiological relevance - even existence - of mESC-to-mTSC transdifferentiation has been disputed [108]. How can these findings be reconciled?

First, a recent study helped to shed some light into the controversial issue of whether or not mESCs can still contribute to placental tissues [260]: Macfarlan et al. identified a subpopulation of cells within regularly cultured mESCs that resemble cells at the 2-cell stage and named them “2C-like cells”. Moreover, mESCs constantly fluctuate between this rare 2C-like state and their regular state. Importantly, 2C-like cells are able to contribute to placental tissues upon blastocyst injection. Therefore mESCs retain the potential to differentiate along the trophoblast lineage, although this is true only for a small fraction of the mESC population at any given time.

Interestingly, another study identified differences in gene expression between individual totipotent blastomeres at the 2-cell and 4-cell stages via single cell RNA-Seq [261]. Importantly, these differences in the transcriptome correlate with a predisposition of the individual blastomeres towards the TE or ICM fate. One of the genes identified to be differentially expressed between blastomeres at the 2-cell and 4-cell stage is *Gadd45a*. One can therefore speculate whether Gadd45a is required for the establishment of the 2C-like state in mESC populations.

Another cause of the transdifferentiation deficiency in Gadd45 TKO mESCs might be *H19* expression levels. *H19* is primarily expressed in the TE at implantation *in vivo* [262] and *H19* overexpression is sufficient to induce trophoblast differentiation in mESCs [263]. However so far it is unknown whether endogenous *H19* expression is required for mESC-to-mTSC transdifferentiation. If so, its reduced expression in Gadd45 TKO mESC could be the reason for their restricted transdifferentiation potential.

However, the expression of all three Gadd45 family members in the placenta strongly hints towards a role of these genes also during physiological trophoblast development, and not only during transdifferentiation. Yet, Gadd45 TKO mTSCs differentiate normally along the trophoblast lineage. In this respect it is first important to note that the exact identity of the Gadd45 expressing cells in the placenta has not been ultimately resolved. Gadd45 genes were expressed both in the TGC layer and in the labyrinth layer. These layers consist mostly of differentiated cells, at a first glance making it likely that the Gadd45 genes are expressed in differentiated cells.

However the placenta must contain a population of so far unidentified progenitor cells, which could be the *Ehox* and *Eomes* positive cells in the labyrinth (see chapter 3.1.3.3). Gadd45-positive cells in the labyrinth layer could therefore be either differentiated cells, or progenitors. One indication for a potential role of the Gadd45 genes in progenitor maintenance (rather than differentiation) comes from expression patterns during gastrulation: *Gadd45b* and *Gadd45g* expression already starts in the extra-embryonic ectoderm and the ectoplacental cone, which both consist mostly of proliferating progenitors [187]. Also a minor reduction of stem cell and proliferation markers *Cdx2*, *Eomes* and *Elf5* could be observed in the Gadd45 TKO mTSC clone compared to the parental line (Fig. 4.17).

A model explaining both the seemingly unimpaired Gadd45 TKO mTSC differentiation, the impaired Gadd45 TKO mESC-to-mTSC transdifferentiation and the Gadd45 expression pattern during development is therefore the following: The Gadd45 genes function in mTSC and later progenitor maintenance, rather than terminal differentiation. In regularly cultured mTSCs however, the flexible and adaptable pluripotency gene regulatory network is able to partially compensate the loss of Gadd45 genes. Therefore Gadd45 TKO mTSCs are viable and only show mildly reduced pluripotency gene expression. However, when mESCs transdifferentiate into the trophoblast lineage, this gene regulatory network is not established yet. In the absence of the Gadd45 genes, mESCs upon transdifferentiation might (i) not undergo full reprogramming, failing to establish mTSC identity altogether or (ii) fail to create a temporary pool of progenitors and directly turn into a lower number of terminally differentiated cells. In both cases the final outcome is the same, namely a reduction in terminal differentiation gene expression, which was consistently observed in multiple experiments.

In order to corroborate this model, the exact identity of the Gadd45-positve cells also during later placental developmental stages should be clarified, e.g. by analyzing the co-localization of Gadd45 and different lineage marker genes. Also the exact defect during transdifferentiation of Gadd45 TKO mESCs needs to be pinpointed. This however is very difficult, because of the sporadic nature of the event and the only marginal upregulation of pluripotency markers above background levels.

Considering the overall lower DNA methylation levels in the TE compared to the ICM, one could speculate that Gadd45-mediated active DNA demethylation might be involved in the transdifferentiation event and/or the maintenance of trophoblast progenitors. Similarly, the gatekeeper function of *Elf5* and its regulation via DNA methylation [102] makes it a prime candidate for further investigations. Paradoxically however, Tet-deficient mESCs are more biased towards trophoblast differentiation in teratomas [246,248–250]. The exact relation between active DNA demethylation and TE versus ICM lineage commitment therefore still needs to be elucidated.

Last, it would be interesting to study whether the Gadd45 genes are important for other reprogramming events as well. Even though their role in mESC maintenance itself seems to be limited, it is possible that they are required for the generation of iPSCs - analogous to their limited role in mTSC maintenance but their requirement for mESC to mTSC transdifferentiation. Notably, a similar situation was already observed concerning the Tet enzymes: Tet1 is necessary for reprogramming and can boost reprogramming efficiency if overexpressed [264,265], although it is of limited importance for mESC maintenance itself.

5.4 The mechanism of Gadd45-dependent gene regulation

The majority of data discussed so far is descriptive in nature, leaving the question open which pathway the Gadd45 genes employ to regulate gene expression in the variety of different contexts studied. Because the Gadd45 family members fulfill many different functions in the cell, this question can also not easily be answered. Therefore, it is important to gain further mechanistic insights to understand the exact role of the Gadd45 gene family during development and differentiation. Here the different strategies used in this study to achieve this are discussed.

First, datasets obtained were compared to experiments carried out by other groups that were studying DNA demethylation processes. Intriguingly, gene expression changes in Gadd45 TKO cells were correlated with gene expression changes upon loss of the Tet enzymes under different conditions (Fig. 4.5 and 4.12). Also, genes deregulated in Gadd45 TKO EBs were marked by 5fC in their promoters (Fig. 4.12). This is in line with the described role of Gadd45a in facilitating removal of Tet-mediated DNA demethylation intermediates [210] and enhancing Tet activity directly [208]. Ultimately, however, this is only correlative evidence and primarily indicates that similar gene sets depend on Tet and Gadd45 function. If both protein families act during developmental processes, an overlap in target genes would also be possible if Gadd45 and Tet genes functioned in different mechanistic pathways. The correlational evidence for a role of active DNA demethylation in Gadd45-mediated gene regulation is therefore limited.

Second, changes in DNA modification levels were measured in Gadd45 TKO mESCs. Although global levels of 5mC, 5hmC, 5fC and 5caC were unchanged in TKO mESCs, site specific DNA hypermethylation at regulatory elements of Gadd45-dependent genes could be observed (Fig. 4.7). Taking into account previous studies that have shown a direct involvement of Gadd45 in determining methylation levels (see chapter 3.3.2), this already provides stronger evidence for a function of Gadd45-mediated active DNA demethylation in regulating these genes. However, it cannot be ruled out that DNA hypermethylation occurs only as a consequence of DNA-methylation-independent gene silencing in Gadd45 TKO mESCs.

This is especially true because expression changes but not DNA hypermethylation can be rescued by transient overexpression of the Gadd45 genes (See chapter 4.1.2.4). This could indeed indicate that methylation changes at the analyzed regions occur only as a secondary event after expression changes. However, it is also possible that only a small portion of the transfected cells was actually rescued: In this case, expression changes can be observed by qPCR if only a small population of cells upregulates the analyzed genes strongly enough. In contrast, even strong DNA demethylation in a subpopulation of cells would be masked by the unchanged methylation levels in the larger, not rescued population.

A similar difficulty prevents meaningful DNA methylation analysis in differentiating EBs. Because DNA hypermethylation is likely to occur in a tissue dependent manner, it would be masked by unchanged methylation levels in all unaffected cell lineages. Gene expression measurements on the other hand are frequently unaffected by this issue, because lineages that do not express a given gene cannot mask its up- or downregulation in an affected lineage. Therefore, although it would be interesting to study promoter and enhancer methylation levels near Gadd45-dependent genes also in EBs, this is not possible with the current culture model. Rather it would be required to analyze cell culture models that produce homogenous populations of the tissue of interest, such as the mTSC differentiation model.

Third, ChIP-Seq was applied with the goal to identify direct binding sites of Gadd45a and Gadd45g. This would arguably provide the strongest evidence for an involvement of active DNA demethylation in Gadd45-dependent gene regulation and would allow distinguishing genes directly and indirectly regulated by Gadd45.

Interpreting the obtained Gadd45a and Gadd45g ChIP-Seq results (Fig. 4.23-4.27) is challenging. At first glance, the number of identified positive Gadd45a and Gadd45g peaks versus negative peaks in monolayer differentiated cells was encouraging. If peaks would only have resulted from minor fluctuations in nucleosome occupancy and chromatin accessibility, a similar amount of positive and negative peaks would have been expected. Also, the identified peaks were not randomly distributed across the genome, but were enriched near genes relevant for neural development.

On the one hand, a negative interpretation of the results is that peaks are not an indication of actual binding of Gadd45 to the DNA, but rather a consequence of slightly distinct differentiation efficiencies between transgenic and parental cell lines: If either of the cell lines differentiated more efficiently into the neuronal lineage, this could lead to systemic changes of nucleosomal positioning and chromatin accessibility near neuronal genes. In turn, these changes could lead to false positive peaks, because background read density in ChIPs depends on chromatin structure and DNA accessibility [266,267]. Importantly these false positive peaks would no longer be randomly distributed, but would be clustered in

one cell line (e.g. the transgenic cell line) and near neuronal genes, just as observed in the actual ChIP-Seq data. This could explain the irreproducibility of the Gadd45 binding sites in independent ChIP-qPCR experiments, because reproducibility would depend on recreating the distinct differentiation efficiencies that coincidentally occurred during the initial ChIP-Seq experiment.

On the other hand it is possible that the difficulties in reproducing the ChIP-Seq results are consequences of the challenging Gadd45 ChIP procedure. First and foremost, Gadd45 most likely does not bind DNA or chromatin itself, but rather works as a bridging factor between chromatin associated factors and repair proteins [146,213]. Fixation of proteins to chromatin using formaldehyde, as performed during ChIP, strongly depends on the distance of the fixed proteins. It is therefore expected that ChIP efficiency for Gadd45 will be inferior to transcription factor or histone mark ChIPs. Of note, however, using longer crosslinkers during ChIP did not increase enrichment of potential Gadd45 target sites (data not shown).

Also, being a bridging factor might make Gadd45 less accessible for immunoprecipitation, because it might be hidden within bigger complexes. In agreement with this, immunoprecipitation of biotinylated Gadd45 using streptavidin coupled beads required strongly denaturing conditions, including heat, despite the theoretically very strong affinity of streptavidin towards biotin. However, these strongly denaturing and heated conditions might have been detrimental to ChIP, which is classically best carried out at 4°C throughout the procedure [268]. This potential problem could be monitored by applying the same ChIP protocol to a biotinylated positive control protein, e.g. a transcription factor.

Last, ChIP efficiency might be partially reduced by the used model system. Monolayer differentiation of mESCs results in a relatively heterogenous, albeit predominantly neuronal, cell population [73]. If Gadd45 binding occurs only in a subset of these cells, ChIP efficiency will be significantly reduced. Even more, if the occurrence of this subset of cells was not reproducibly induced during differentiation, this could explain why only a minority of the ChIP-qPCR experiments were able to confirm the findings of the ChIP-Seq experiment. Taking all of this into account, it cannot be conclusively decided whether the Gadd45 ChIPs were technically too inefficient, or whether Gadd45 proteins steer differentiation using mechanisms that do not depend on their association with chromatin.

In summary, this study helped to elucidate the role of the Gadd45 proteins during the pluripotent state as well as during differentiation into various embryonic lineages, particularly the neuronal one. Furthermore it indicated a novel role for the Gadd45 proteins in the establishment and maintenance of the extra-embryonic trophoblast lineage. Thereby, it opens avenues to study Gadd45 functions at the heart of the first lineage segregation. Future work will determine the precise mechanism by which Gadd45 realizes these functions.

6 Material and Methods

6.1 Material

6.1.1 Equipment

-150°C freezer (Sanyo); -80°C freezer (Sanyo); agarose gel chambers (BioRad); bacterial incubators (Thermo Scientific); bacterial shaker (Infors); balances (Sartorius, Kern; BAS IP MS 2025 E multipurpose (GE Healthcare); Bioanalyzer (Agilent); Bioruptor (Diagenode); Bioruptor (Diagenode); blotting apparatus (BioRad); cell counter (BioRad); cell culture incubators (Thermo Scientific); centrifuges (Heraeus); Cryo-Safe Cooler (Belart); E-Gel electrophoresis system (Invitrogen); Genepulser XCell incl CE module (Biorad); heating blocks (Eppendorf); HiSeq 2500 sequencing system (Illumina); hybridization oven (Thermo Scientific); laminar flow hoods (Dometric); LightCycler 480 (Roche); magnetic stirrer (Heidolph); MassSpec; microcentrifuges (Heraeus); microplate reader (Tecan); microscope (Leica); microwave oven (Sharp); MiSeq; multidispenser pipette (Eppendorf); Nanodrop 2000 spectrophotometer (Thermo Scientific); orbital shaker (Neolab); PAGE minigel chambers (BioRad); PCR thermocyclers (Biometra); pH meter (Mettler Toledo); pipet boy (Integra); pipettes (Eppendorf); power supplies (BioRad); Qubit (Thermo Fisher); rotator (Neolab); RPN11647 Hypercassette 20x40cm (GE Healthcare); SpeedVac concentrator (Eppendorf); staining jars (Neolab); stereomicroscope (Leica); Typhoon FLA-9500; ultrapure water purification system (Millipore); UV crosslinker; UV photodocumentation system (BioRad); vortexer (Scientific industries); waterbaths (Neolab); Concentrator Plus Speedvac (Eppendorf)

6.1.2 Chemicals and pre-made buffers

0.1% gelatine in ultrapure water (Millipore); 100x L-Glutamin (Lonza); 100x Non-essential amino acids (Gibco); 100x Penicillin/Streptomycin (Lonza); 100x sodium pyruvate (Gibco); 5-azacytidine (Sigma-Aldirch); Advanced DMEM/F12 (Invitrogen); agarose (Biozym); ammonium acetate (Sigma-Aldrich); ampicillin (Sigma-Aldrich); B-27 supplement (Invitrogen); BCIP (Roche); bicinchoninic acid (Sigma-Aldrich); Böhringer blocking reagent (Roche); bovine serum albumin (Sigma-Aldrich); bromphenol blue (Sigma-Aldrich); CHIR99021 (Biocat); chloroform (Roth); citric acid (Sigma-Aldrich); copper(II) sulfate solution (Sigma-Aldrich); DAPI (Sigma-Aldrich); dCTP, [?-32P]- 3000Ci/mmol (Perkin Elmar); Denhard'ts solution (); dimethylsulfoxide (Sigma-Aldrich); dithiothreitol (Sigma-Aldrich); DMEM with glucose (Lonza); dNTPs (Thermo Scientific); EDTA (Sigma-Aldrich); ES-grade FBS (PAN Biotech); ethanol (Sigma-Aldrich); ethidium bromide (Roth); fetal bovine serum (Lonza); formaldehyde (37%, Roth); formamid (deionized, AppliChem); glycerol (Sigma-Aldrich); glycine (Sigma-Aldrich); Glycogen (Fermentas); goat serum (Invitrogen); hydrochloric acid (Sigma-Aldrich); Iodoacetamid (Sigma-Aldrich);

isopropanol (Sigma-Aldrich); L-ascorbic acid 2-phosphate sesquimagnesium salt (Sigma-Aldrich); lithium chloride (Sigma-Aldrich); magnesium chloride (Sigma-Aldrich); methanol (Sigma-Aldrich); Mitomycin C (Sigma-Aldrich); N-2 supplement (Invitrogen); NBT (Roche); Neurobasal (Invitrogen); nuclease-free water (Qiagen); O.C.T. TissueTek (Weckert Labortechnik); Optimem (Gibco); paraformaldehyde (Merck); PD0325901 (Sigma-Aldrich); phenol (Roth); potassium hydroxide (Sigma-Aldrich); protease inhibitor cocktail tablets (Roche); puromycin (Sigma-Aldrich); Qiazol (Qiagen); Random primers (Invitrogen); retinoic acid (Sigma-Aldrich); Roticlear (Roth); RPMI 1640 (Gibco); Sigmacote (Sigma-Aldrich); skim milk powder (Sigma-Aldrich); sodium chloride (Sigma-Aldrich); sodium citrate tribasic dihydrate (Sigma-Aldrich); sodium dodecyl sulfate (Sigma-Aldrich); sodium hydroxide (Sigma-Aldrich); sodium phosphate (Sigma-Aldrich); sodium phosphate dibasic dihydrate (Sigma-Aldrich); sodium phosphate monobasic monohydrate (Sigma-Aldrich); ß-mercaptoethanol (Sigma-Aldrich); sucrose (Sigma-Aldrich); TEMED (Sigma-Aldrich); triethanolamine (Sigma-Aldrich); Tris base (Sigma-Aldrich); Tris HCl (Sigma-Aldrich); Triton X-100 (Sigma-Aldrich); Trypsin, 0.05% (Lonza); Trypsin, 0.25% (Lonza); Tween-20 (Sigma-Aldrich); Ultrapure water with gelatin 0.1% (Millipore); yeast RNA (Roche)

6.1.3 Kits and enzymes

AccuPrime (Thermo Fisher); AmPure XP beads (Beckman Coulter); BamHI (NEB); BbsI (NEB); C1 streptavidin coupled dynabeads (Invitrogen); ClaI (NEB); DFS Taq (Bioron) ; DIG RNA Labeling Mix 10x conc (Roche); DNAse I (Roche); DNeasy Blood & Tissue kit (Qiagen); Dual Luciferase Assay Kit (Promega); Epitect Bisulfite kit (Qiagen); G-50 columns (GE Healthcare, 27-9240-01); HpaII (Promega); HpaII methylase (Promega); Labelling beads (GE Healthcare, 27-5330); LIF (Homemade cell supernatant, CF IMB); Lipofectamin-2000 (Thermo Fisher); MNase (NEB); MspI (Promega); MsssI methylase (Promega); NEB Next ChIP-Seq Library Prep kit (NEB); Neuromag (OZ Biosciences); NotI (NEB); PacI (NEB); PCR purification kit (Qiagen); pGEM_T-easy (Promega); Proteinase K (Roche); PvuII (Promega); Qiaprep Midiprep kit (Qiagen); Qiaprep Qiaprep Miniprep kit (Qiagen); Qiaquick Gel extraction kit (Qiagen); rhBMP4 (R&D Systems); rhFGF4 (Reliatech); RNAse A (Ambion); RNeasy 96 kit (Qiagen); RNeasy Mini Kit (Qiagen); SbfI (NEB); SP6 RNA polymerase (Ambion); Superscript II Reverse Transcriptase (Invitrogen); SuperSignal West Pico/Femto (Thermo Scientific); SwaI (NEB); T4 Ligase (NEB); T7 RNA polymerase (Ambion); XbaI (NEB); XhoI (NEB); XmaI (NEB); x-treme Gene9 (Roche)

6.1.4 Buffers and cell culture media

6.1.4.1 General molecular biology buffers

10xTBE	1M Tris base, 1M boric acid, 20 mM EDTA, autoclaved
20x TBS	3M NaCl, 5.36 mM KCl, 1M Tris-HCl pH 7.4, autoclaved
4xLämmli	60 mM Tris-Cl pH 6.8, 2% sodium dodecyl sulfate, 10% glycerol, 5% β-mercaptoethanol, 0.01% bromophenol blue
6x DNA loading buffer	60% (w/v) sucrose, 0.25% xylene cyanole, 0.25% bromphenol blue
dNTP mix	5 mM dATP, 5mM dCTP, 5mM dGTP, 5mM dTTP
Luria broth (LB)	10g bactotryptone, 5g yeast extract, 10g NaCl in 1l water, pH7.5, autoclaved
PBS	9.0 g/l NaCl, 144 mg/l KH_2PO_4, 795 mg/l Na_2HPO_4
Tail lysis buffer	50mM Tris pH8.5, 2mM EDTA, 0.1% SDS, 200µg/ml Proteinase K
TE buffer	10 mM Tris pH8.0, 1mM EDTA

6.1.4.2 Southern blot buffers

20xSSC	3M NaCl, 300mM trisodium citrate
Church buffer	250mM sodium phosphate, 1mM EDTA, 1% BSA, 7% SDS, pH7.2
Lysis buffer	100mM Tris pH8.5, 5mM EDTA, 0.2% SDS, 200mM NaCl, 100µg/ml Proteinase-K

6.1.4.3 *In situ* hybridisation buffers

Acetylation buffer	1.33% triethanolamine, 0.06% HCl, 0.25% acetic anhydride
Hybridization buffer	50% deionized formamide, 500mM NaCl, 50mM EDTA, 100mM Tris pH7.5, 50mM $NaH_2PO4{\cdot}H_2O$, 50mM Na_2HPO_4, 1x Denhardt's solution, 10% dextran sulphate, 100µg/ml yeast RNA
In situ blocking buffer	89% MABT, 10% goat serum, 1% Boehringer blocking reagent
In situ wash buffer	50% formamide, 1x SSC, 0.1% Tween-20
MABT	0.1M maleic acid, 150mM NaCl, 0.1% Tween-20, pH7.5
NTMT	100mM Tris pH9.5, 100mM NaCl, 50mM $MgCl_2$, 0.1% Tween-20

6.1.4.4 ChIP buffers

ChIP dilution buffer	16.7mM Tris pH8.2, 167mM NaCl, 1.2mM EDTA, 1.1% Triton X-100, 0.01% SDS
ChIP elution buffer	1% SDS, 50mM Tris pH8.2, 10mM EDTA
ChIP lysis buffer	1% Triton X-100, 2mM beta-mercaptoethanol, 50mM Tris pH7.5
ChIP wash buffer I	2% SDS
ChIP wash buffer II	0.1% sodium deoxychelate, 1% Triton X-100, 1mM EDTA, 50mM HEPES, 50mM NaCl
ChIP wash buffer III	0.5% sodium deoxychelate, 0.5% NP-40, 1mM EDTA, 10mM Tris pH8.2, 250mM LiCl
HNB buffer	15mM Tris pH7.6, 500mM sucrose, 60mM KCl, 250nM EDTA, 125nM EGTA, 0.5% Triton X-100, 1mM DTT
NE buffer	20mM Tris pH7.6, 70mM NaCl, 20mM KCl, 5mM $MgCl_2$, 3mM $CaCl_2$

6.1.4.5 Cell culture media

DMEM+++	DMEM, 10% FBS, 2mM L-Glutamine, 50U/ml Penicillin/Streptomycin
mESC medium	DMEM, 15% ESC-grade FBS, 2mM L-Glutamine, 50U/ml Penicillin/Streptomycin, 1xNEAA, 1mM sodium pyruvate, 100µM β-mercaptethanol, 4% LIF-supernatant
mTSC incomplete	RPMI 1640, 20% ESC-grade FBS, 2mM L-Glutamine, 50U/ml Penicillin/Streptomycin, 1mM sodium pyruvate, 100µM β-mercaptoethanol
mTSC complete	70% conditioned mTSC incomplete medium (see chapter 6.2.2.5), 30% mTSC incomplete medium, 25ng/ml Fgf4, 1µg/ml heparin sodium salt
CA medium	DMEM, 10% ES-grade FBS, 2mM L-Glutamine, 1xNEAA, 50U/ml Penicillin/Streptomycin, 100µM β-mercaptoethanol
N2B27 medium	50% Advanced DMEM/F12, 50% Neurobasal, 2mM L-Glutamine, 50U/ml Penicillin/Streptomycin, 50µg/ml BSA, 0.5x N2-supplement, 0.5x B27-supplement
2i medium	N2B27 medium, 1µM PD0325901, 3µM CHIR99021, 4% LIF-supernatant
KOSR medium	DMEM, 15% KOSR, 2mM L-Glutamine, 50U/ml Penicillin/Streptomycin, 1xNEAA, 1mM Na-pyruvate, 100µM β-mercaptethanol

6.1.5 Antibodies

Antibody	Host species	Company	Catalogue #	Dilution
alpha-tubulin	mouse	Sigma	T5168	1:5000
Gadd45a (H-165)	rabbit	Santa Cruz	sc-797	1:2000
Gadd45b (C-18)	goat	Santa Cruz	sc-8776	1:1000
Gadd45g (gp-II)	guinea pig	homemade	-	1:1000
Streptavidin HRP	-	Thermo Scientific	21126	1:40000
anti-mouse HRP	goat	Dianova	115-035-146	1:10000
anti-rabbit HRP	goat	Dianova	111-035-144	1:10000
anti-guinea-pig HRP	goat	Dianova	106-035-003	1:10000
anti-goat HRP	donkey	Dianova	705-035-003	1:10000
Anti-Digoxigenin-AP	sheep	Roche	11093274910	1:1500

6.1.6 Primer, probe and oligonucleotide sequences

6.1.6.1 qPCR expression primers

Gene	Forward primer	Reverse primer	Probe
0610005C13Rik	ctatgacaccgcttgacacc	caaaatcgctgtggattgg	13
AFP	catgctgcaaagctgacaa	ctttgcaatggatgctctctt	63
Asz1	agttgcctgtaaattctcacaagat	ccagttcttcacaaactgaagtaaaat	7
Brachyury	cagcccacctactggctcta	gagcctggggtgatggta	100
Cdx2	caccatcaggaggaaaagtga	ctgcggttctgaaaccaaat	34
Ctsq	ggctcaggcaacctatcct	gggaggaccacaaatcctg	84
Desmin	gcgtgacaacctgatagacg	tggatttcctcctgtagtttgg	110
Efhc2	gtttgagcctatagagaataattcagg	caggctttttcacacgacttc	31
Elf5	gactgtcacagccgaacaag	ccaggatgccacagttctct	56
Eomes	accggcaccaaactgaga	aagctcaagaaaggaaacatgc	9
Ezr	gagctggagtttgccatcc	gccaatcgtctttaccacct	38
Gadd45a	gctgccaagctgctcaac	tcgtcgtcttcgtcagca	98
Gadd45b	cggccaaactgatgaatgt	atctgcagagcgatatcatcc	79
Gadd45g	gtccgccaaagtcctgaat	gctatgtcgccctcatcttc	71
Gapdh	agcttgtcatcaacgggaag	tttgatgttagtggggtctcg	9
Gm20625	cacagctgcgactgaacaat	gatagggcctcagcacctg	66
Gm364	tgctatgctattttcggctgt	gtcaaaaaggctctccacca	73
Gm6189	cagctgggaaaccctgaa	tcctggttgtagtgcagacg	67
H19	cggtgtgatggagaggaca	agacggcttctacgacaagg	42
Hand1	caagcggaaaagggagttg	gtgcgccctttaatcctctt	51
HNF4a	ccaagaggtccatggtgttt	ccgagggacgatgtagtcat	68
Lasp1	tcgtcctatggtgggtacaag	ggcagcgctgtagtcatacac	66
Map3k8	ggactgctgaactctgtttgc	ttccagtgctcatgtactcca	3
Nestin	tcccttagtctggaagtggcta	ggtgtctgcaagcgagagtt	67
Neurog1	gacctgtccagcttcctcac	tggaggctaggggctgtag	101
Pax3	gcccacgtctattccacaa	gaatagtgctttggtgtacagtgc	69
Pax6	gttccctgtcctgtggactc	accgcccttggttaaagtct	78
Peg10	ggacccctcatccttcgt	ttcttaaaacccgcctgttc	6
Plet1	catccgtgaaaatggaacaa	tcacagttggagtcgtgtttatg	20

PL-II	gggcttctggaaggactga	tgaccatgcagaccagaaag	17
Pramel6	ccaactttggcagtctccag	ccgtaattgacaaggtcttcaaa	27
Pramel7	gatcatttctctgttctgcaaactc	aatggggcctccagacac	94
Prl7d1	tgatccaaccgtgctcct	ctggcatttatgggtgcag	7
Psg29	gcagggggttctactcacag	caatggtgactttggaagtgg	62
Rhox2a	gggagtgagtgaagccacag	gcctttacagcctcttataactcct	91
Rin3	actcaaagccttggtggaac	tgtacaaggctgactccactatg	101
Serpinb9e	caccaaggaagtgcccttta	ttcacataggcatggaaaaatg	4
Sh2d3c	gcacagcttgcaaagatgac	aaggcccttgaacttgaagaa	96
Snai1	gtctgcacgacctgtggaa	caggagaatggcttctcacc	71
SynA	gatgcctgacacccctg	ccacgactggagtgtgtagg	46
Tbp	ccaggaaataattctggctca	ggggagctgtgatgtgaagt	97
Tead4	tctctcaaaacacctaccctgtc	gccctgcaggagactcaa	13
Tinagl1	gggactcactcagtcaagatca	acgagttggcagcagtcc	7
Tmem92	gagagcatttgggtcacctc	tgtgggaggtgtcatctgaa	77
Tpbpa	tgaagagctgaaccactgga	caggcataggatgactaggaagat	107
Tubb3	gcgcatcagcgtatactacaa	ttccaagtccaccagaatgg	104
Twist1	agctacgccttctccgtct	tccttctctggaaacaatgaca	58

6.1.6.2 Bisulfite sequencing primers

Associated gene	Genomic element	Forward primer	Reverse primer
Asz1	Promoter	taaaaagttatttatgagaggagaatat	ctaccaaaaatcacaacccaataat
Calcoco2	Promoter	tagttgggatatttagtgtttttt	ccaaaatattttaatttttaattttt
Calcoco2	LMR	ggtttttgtagatagttttgataggtg	aaaaaacaataaattaaaaaaacaac
Efhc2	Promoter/LMR	ggttaaagttgttgttaggtagaaa	aaaaccaataaaatcaaaaacaaaca
Gm364	Promoter-1	tttaaaagaaagattttatgaggttgaa	caaataaaccaccaaaaaataaaaca
Gm364	LMR	ttaaaggggaaaggggtaagttat	aatccatccacatttacaaaaaaaa
Gm6189	Promoter	ttttttttgaaggaggattaagata	aaaattattcttttatttatttatttattt
Gm6189	LMR	tggattattatattaatgaagggagag	taataaaaccttccaccctttacaa
H19	Promoter-1	ggtttttttaattggagtggttt	cataccctcaaattcccattataat
H19	Promoter-2	gattataatgggaatttgagggtat	aaaaaacaaaacatctctcctatcc
H19	ICR	tagttttgtttttatggttatggg	tcttaaataactccttcaatcttac
Igfbp2	Promoter	gagtttggtttagaaggattgag	tactcccttcaacctaaatttatcc
Igfbp2	LMR	tggtatattatgatgtagggtggtattt	aaatcacttaaaccaaaaatcaaaac
MageA5	Promoter	gggaatttttgatatggttttaaag	ttcaaaataaataatcccccaaata
Mest	Promoter	agaatttgtttatttttagttatttagtg	aaaaaccttcccatatattaatacc
Mest	LMR-1	ttggaatgtttttgtttttggta	aaacaaatacttatataaaacctcc
Mest	LMR-2	tgtagtttttggtttaattttgat	cattaaaaacacaaacctcctttac
Mest	LMR-3	gtttttaaggggtgttggttttat	accccaacttaccatacctaaattt
Mest	Alt. promoter/ICR	ggtgtttgtttgtttttttt	ctttaaataaaaatttttacctccc
pGL3 plasmid	-	tttgttaggtattaggtaaggatatgg	aaaacaaaataaattccaattcaac
pOctTK plasmid	-	gatttgttttgtaggtggagagttttt	aaataaacttcaaaatcaacttacca
Pramel6	Promoter	tttaagaatttggattaaaggtgtg	tttaaacactaacacaaaaaaaacc
Pramel7	Promoter	tttaaatttttttggaagggagaat	aacaacctaatctacaaaatacattcca
Pramel7	LMR	tgaaggtgattttgaattttaggat	aacccaatatcaacaacctctaatc
Rerg	Promoter	agaaaggatttagttgaagtttgtaggt	aactaaaaaactctcccaccctaac
Rhox2a	Promoter	aattaaataaggtaggaggatttag	atcccacatcaaacttataattaac
Sry	Coding Region	tttttgtttgtttgtttgtttttgt	accactcctctataacactttaaccc
Tmem92	Promoter	ttggtaattgaatttaggattttaattt	cacccccacaaaatttctctatata
Tmem92	Alt. promoter	gtttttagagaaattttgttttaaaaa	acatcccaacataccacactattac

6.1.6.3 Genotyping primers

Genotyping PCR	Forward primer	Reverse primer
CRISPR *Gadd45a* external	cgctccgtaccctatcacaa	aactctgccttgctttggtg
CRISPR *Gadd45b* external	catctccagccaatctcagc	caagacccatcgtgcatcag
CRISPR *Gadd45g* external	aaatctgcaggctccagtct	cacgcggggcctttctac
CRISPR *Gadd45a* internal	ggccacttacacgttgagc	ggtaagtgtgcctgccga
CRISPR *Gadd45b* internal	tcatgaccctggaagagctg	acattcatcagtttggccgc
CRISPR *Gadd45g* internal	cgccagatcatgactagggt	cagaagttcgtgcagtgctt

6.1.6.4 ChIP primers

ChIP'ped protein	Target site	Forward primer	Reverse primer	Probe
Gadd45a	Peak #1	cagcaccgaatgctcaag	ggctcacaaggaccagacac	98
Gadd45a	Peak #2	caacagcaattgtgggtcaa	gtcctctgggggtcttttgt	57
Gadd45a	Peak #3	ccgaaacccagaattagctg	ccttcaactagagacagcaaagg	127
Gadd45a	Peak #4	ccacagtatgactccacatgc	acacacagagtctggtatcaaagc	17
Gadd45a	Peak #5	catgccaaaatggagcttct	tgcaggcatatggaaggaa	56
Gadd45a	Peak #6	ccatactcaacaggcgctaaa	gcctgaatgctggttgct	1
Gadd45a	Peak #7	caggaacccaattaaggctct	ttcagagggaaattggcaga	82
Gadd45a	Peak #8	catcccccaaatctgaaaaa	gctttttctcgctacctgga	5
Gadd45a	Peak #9	atgcccaaaccgtatcaaac	gaccctgggacaagatgtgt	111
Gadd45a	Peak #10	gggatagaaagatagcgcaca	gatctcttccagcgctgttc	76
Gadd45a	Peak #11	aaaaagaatggaaacatgcaataat	gccccacacttaacctttctt	143
Gadd45a	rDNA region #1	gtcttccaggccgatgtg	gagggcttctggaggaaaaa	78
Gadd45a	rDNA region #2	gcgcaatgaaggtgaagg	ctaacgcgtacgctcgtg	10
Gadd45a	rDNA region #3	gagttcacggtgggttcg	gagaagaccacgccaacg	51
Gadd45a	*Slco1c1*	ttggcctcgaactcagaga	gccttaaatcccagcacttg	98
Gadd45a	*Tmem108*	ttggtcaccagccctcat	atctggtctcccttgagtcg	22
Gadd45a	*Cacnb4*	atccacaaatcggtgtccat	cagccttggcgctcttatt	22
Gadd45a	*Hcn1*	ggaaggaggagagaaagcaag	cgctgctgcacactaagaac	7
Gadd45a	*Grid2*	attcccctccgagtgacag	cccgagagctagcacagc	52
Gadd45a	*Elavl2*	gttgcctccacagttggtc	tttcccactctgacctttgc	24
Gadd45g	Peak #1	gaaagtcagggctgacaacag	acccgatgctgtgcactc	4
Gadd45g	Peak #2	cgtggagcacattgttcatt	gccccgaacttatctcctgt	98
Gadd45g	Peak #3	ctcacacagctccacagacc	cgactataggtgcccaacaaa	104
Gadd45g	Peak #4	gagccaccgcacagagac	tccacacgggaacctgag	75
Gadd45g	Peak #5	tgaaacaggctctgctgaca	aagcaggggactggtagagg	38
Gadd45g	Peak #6	gaagcgtgtttggtttgtcc	atgaccagctgagcaagca	34
Gadd45g	Peak #7	cgtgacctgcagagtaagagc	gtgtggcctgaggtggag	25
Gadd45g	Peak #8	acatctcccaagccaacct	agtccctcctccgacttga	78
Gadd45g	Peak #9	agggtcacgtcctcttgtgt	cccatccctagacacacctc	7
Gadd45g	Peak #10	ccctccagaagctgttgct	gcctgcagattgccttgt	8
Gadd45g	Peak #11	ggtgctgtggactcagtgac	cctgcaggagtatcgaatgg	12
Gadd45g	Peak #12	ctctgaccagccatgcact	gcccacttccatgatagcc	14
Gadd45g	Peak #13	tcccatggtggaaagagaac	tgcaggctagcacttgacac	127
Gadd45g	Peak #14	ggccgttagcttgaagcac	aaaaggcaaaatggggtctc	68
Gadd45g	Peak #15	taatgtgcgacccaggaag	tacactggcggggaacag	85
Gadd45g	Peak #16	caaagctgagggaggttgg	ggctgtgtcccttagtctgtg	53
Gadd45g	Peak #17	tgcaggctcctagccaaa	attggatgcccaaggacag	37
Gadd45g	*Sox11*	tggacactagctccagactcg	ggtgtcctcccttctcacc	58

Gadd45g	*Slc25a46*	caaggtgtgtcccctgcta	aacagtccgcggtcactc	66
Gadd45g	*Sox21*	cgctttggtctgtctggtc	gggacaccgaggagaagttt	55
Gadd45g	*Lhx6*	gatgctgcctccaggtgt	ctagcacccaagttcgtggt	102
Gadd45g	*Hoxa5*	ctgtccccaacatcctgtg	gaaactgactggggccttg	68
Gadd45g	*Pax6* intron	cctcaggacccaagacagc	cctgtgggaaggaaaggact	37
Gadd45g	*Pax6* promoter	ttaagcgttggcctggag	tgtcaataaagggctcgttagg	22
Gadd45g	*Tlx1*	cggtctctgtgctacccact	cctggcgctcagaaagag	147
Gadd45g	*Id4*	gagtgtcccttgcatcacag	taggcaggcgagcagagt	60
Gadd45g	*Tchh*	cgggacagaaaattccatga	ctcctccaccctcaaacg	69
Gadd45g	*Gata2*	tctggctggcatttcctg	ctagagccgtatgccctgag	62
Gadd45g	*Insm2*	agcttcgcagatgaggtgac	cacagctggcagatgaactc	63
Gadd45g	*Th*	catccctctgcctacttgct	cctgtagaccagaccctctcc	104
Gadd45g	*Adm*	ggagctcggtagagccaaa	ggaagcaaacagaacccaaa	12
Gadd45g	*Map2k6*	agggagcaatatttgtcatcg	aaacagctggtgcaactgg	80
-	Control region #1	tgatactgtggacatttccttgtag	gatacagtcaaagctgtgatccaa	102
-	Control region #2	ttgtcttctgcttctgttctgg	gcaggtgggcttttattaggt	87
-	Control region #3	tctggtctgtcttgtctgcaat	acccccaagaatgtgaaaga	47

6.1.6.5 Cloning primers

Primer name	Primer sequence
Gadd45a_cloning_f	atcgatatgactttggaggaattctcgg
Gadd45a_cloning_r	ctcgagccgttccgggagattaatca
Gadd45g_cloning_f	atcgatatgactctggaagaagtccgt
Gadd45g_cloning_r	ctcgagctcgggaagggtgatgct
V5FlagBioTag_cloning_f	ctcgaggggaagctgggcggcgg
V5FlagBioTag_cloning_r	tctagattagctgccgccggcgtt
Gadd45a_5probe_f	ggtctttgtagagaaaggaccaga
Gadd45a_5probe_r	ccttccccatgatggactaa
Gadd45a_3probe_f	ttgaccaggacaccaggact
Gadd45a_3probe_r	agacactagaagtttgcccaca
Gadd45g_5probe_f	tcaccagttctctcccatcc
Gadd45g_5probe_r	tcctttgttcactcccaagg
Gadd45g_3probe_f	ctagctccagtcccaggttg
Gadd45g_3probe_r	aactcagttgccttggttgg
FLBIO_f	ggatccatggattacaaggatga
BIO_f	ggatccatgcccgggggcggcagc
FLBIO_and_BIO_r	atcgatgctgccgccggcgtt
G45a_STOP_r	ctcgagttaccgttccgggagattaatca
G45g_STOP_r	ctcgagttactcgggaagggtgatgct
BirA_CDS_f	ggatccatgaaggataacaccgtgcca
BirA_CDS_r	ctcgagttatttttctgcactacgcagg
HR_Gadd45a_5'_f	ggccgcggatttaaattcgtcagtcaggaatgaacca
HR_Gadd45a_5'_r	ccgtcgacgcctgcaggccgttccgggagattaatcac
HR_Gadd45a_3'_f	cggtcgacgcggccgctctagatggcatccgaatggaaataac
HR_Gadd45a_3'_r	cgggtaccttaattaatgtctccgctcctgcctccag
HR_Gadd45g_5'_f	cggtcgaggcctgcaggctcgggaagggtgatgctggg
HR_Gadd45g_5'_r	ggccgcggatttaaatcctaccgagctatatccccac
HR_Gadd45g_3'_f	cggtcgacgcggccgctctagacagcctggcagggaccttggt
HR_Gadd45g_3'_r	cgggtaccttaattaactgagccatctctaggctgca

6.1.6.6 Probes

Probe name	Sequence
Southern blot *Gadd45a* 5' probe	ggtctttgtagagaaaggaccagaaatttttagggcttatatttttgaagtgatgttaaaggtgacttttctgagaattcatgcac aggactgtagggtcatgaggtggtgattaattctcagattctttttaattcatgacacccactgtactactgaatgtagggtgttg aggacaccaagctgatgaagtcatagtctctgctgtcacatcatcttctaattctttgtcagctgcatccttccagagaacttcta agaaatgtctaatgggtagagccctggagttctagggaacacagttcattagtctccaaatggtgactcggagtggtttgtgt cattatggtattttaaagatggtattttagtgccagttacctgtattgttcctggaacaaaattaagaaaggaagcgtttattcctg ggtacagttagtccatcatggggaagg
Southern blot *Gadd45a* 3' probe	ttgaccaggacaccaggactgcacatttcttgtcctcaatttaaatctaagcctcatgccaggaattaggaggaggtggtcatt aaacctcttacttgccacccccccccaagaatctcttttgctgcatctcactattgtctccgatttcctatcgaggctacatggc tgagtccagcttgttggaaccagagcagctccaggaacactggggctgtagcttagtggtacagtgcaaaggggaaacaa aagaacacttctgcatatataaattatacatatatttgtgtagtacaaccgaaaaaggagaaatgggatattatcacactctgag ggaatctgggaaagatgatgcaacgaacggcaaaatcctgagaggttagaaagttcagccaattgaaattttgttcacacaa tgtttaacaatttaaaagaaatcagtagtaaaacacttaaatgtgggcaaacttctagtgtct
Southern blot *Gadd45g* 5' probe	tcaccagttctctcccatccctccccagcaagtcagacctgactgtaagttgtgagtggtggatgtggaccgtgtaacaaact catttctcggttgcatgggatccccggagccattttagaggacttccccagaagtggagcttttctgagtgaaggctgaccgt cttcttctacctctgacttccagacttgttgcagccttcagccgcggggcacattctgtctctttgcccctgtctctccttctcacc ctctctcctgcgtcctgcagctgctccagttttttccccctgcagctgagccagccttctttcctgatctctaattgaaaaagaat ggctcgctcttcctcctccccacttccctcctccccacctccctcctccctgcagctgccaggctttccctgcttgccactcaa agttgttttttttttcctcttaaaatctaattaagtaccaaccagagcccccacccactctcctcaagaaaagagaaaaagaaag aaaagtagcagaagaaactcccttgggagtgaacaaagga
Southern blot *Gadd45g* 3' probe	ctagctccagtcccaggttgctttgaatggtcccaagggaactgtccttccacaggccccacgggtgctttatgattgggga cagtgaaaacacggtttccaatggcaagccttaggttgggagagcagctgagctggctgcagggacaatggggactgtg gggtgctggcaaccaaccctatcagggctttggtttccgactatagcaattttacatttggcctttggaacttgccctgcttttaa gttgaattgaagcctttatttcttctaaggcttgccaaattatttcacctttctaggtaaaaataaataaataagcaaataaataaa ataaaaataaaataaaagctctccaggaaaggatgaaggctacctgggggatgggaagtctaagaggaggctgggtag tcttagggcacttggcatccggccaaccaaggcaactgagtt
In situ hybridization *Gadd45a* probe	tttgagggcataaagaccaaaaaaaaaaaagaaaaaaaatacccaaactattgctatacatttaaagtttcaacaaagcaata tttggcttatttgaaagtaacctggccatcctaaattagcagttttctttcttgcagtgctttgtagtttttgtccctttttgccttggatt tctcttttctcgcagcttccttcttcaggctcacctctctctccctctgcaaagtcatctctgagccctcgtgcagttgaactcgg ccccttgacatcagtttctgtaatccttgcatcagcgtggggagtgactgcttgagtaactacaaaggtatttcaaaacttcagt gcaatttggttcagttatttccattcggatgccatcaccgttccgggagattaatcacgggcacccactgatccatgtagcgac tttcccggcaaaaacaaataagttgacttaaggcaggatccttccattgtgatgaatgtgggttcgtcacca

In situ hybridization *Gadd45b* probe	caagcgatctgtcttgctcagcacaatattaaaaatagatgtgctgtagctgcgaagtcccgggctcgagtcccgggtcccg cggcgagggtcgtcgccctccgctgacttatgcacagttcactgtcccgcagggtgggggacgcctcaccccacccctgc ggcggtgtccagaccgtctgcggctgcgagcagaggctggctgcaggactgccaggccttggctctaaagtctcagtctc ctcttgcctgaggtgccctccttccgacctcggcgaccacctgtggtgctgctcctggctccacagcccagagggcgggc gggcgcacgtgcccctctcctcatctcagtctcggtgactccgcccctcccagaaggtatcacgggtagggtagcctttga gggattgttctgggggatgaggggctgtcaaatatgtattttatttacttttttaggggacagcaactcgacagttgctttaga tgtttggagtgggtctcagcgttcctctagagagatataggggacccattggttattgcctctgctctcttcacagtaactggc cacctccaccaagccttggcttttccaggaatctgtatgacagttcgtgaccaggaggcagtgcaggtctcgggcttcggtt gtgcccaatgtctccgccggctccccccaggagctgcgccagcctctgcatgcctgatacccggacgatgtcaatgtcattg tcgcagcagaacgactggatcagggtgaagtgaatctgcagagcgatatcatcctcctcttcttcgtctatggccaggagg cacaagaccacgctgtcggggtccacattcatcagtttggccgcctcgtacacccccacggtgaggcgatcctgacgctg cgcggccaccagcagctgctccaccgcggcagtcaccgcctgcatcttctgaaccgcgttgtcgctcgccaccagctcttc cagggtcatgatgcagtcacactccacagcggacccccacggaagatccccaagagcaaagtacaagtccagat
In situ hybridization *Gadd45g* probe	tcctgcagtggccgctcctgccgctgtaccaagtccaggcggtcgcctgcctttctttcccctcctttcactggaaggctgctt gctgagtcacacctgcctcgctggtcggttcttcccggccctggggccaacgtcctcgcctaggccttcctgtgctagcag cctctccggacactggcagagtcctgggctcggaacaaggtgtccccgctgccgccagaccccccttcttgggaggccgc gctcctcgctctccaggcaacatcgcctgccagtctccacgcgcactcgagcgccctcggagggccactccagagccag cagcgcgctagagtgtcaccaagtcgatcagaccaaggtccctgccaggctgtcactcgggaagggtgatgctgggcac ccagtcgttgaagctgcggctctcctcgcagaacaaactgagcttctccaaggcagggtccttccatgtgtcctcattaggat tcgaaatgaggatgcaatgcaggtctcccggcgcgcccccctcttcgtcggcgcccacgatcgccgccagcctctgcac gtctcccacgcgcacgatatcaatgtcgttctcacagcagaacgcctgaatcaacgtgaaatggatctgcagcgctatgtcg ccctcatcttcttcatcggcagccagcacgcaaaaggtcacattgtcagggtccacattcaggactttggcggactcgtaga cgccagcggtcagacagccctggccgtgcgccgacagcagaagttcgtgcagtgctttcccggcgccctgcatcctggc tgtgctttccggaactgtatcctggccacggacttcttccagagtcattgtgcgatccacgaacagcaagttatccaaaagag tgcggagg

6.1.6.7 Guide RNA sequences

Target	Reverse off-target frequency[269]	Guide RNA Sequence	PAM Motif
Gadd45a Exon1	99	GCCCCCACAGTGCGCGGCGC	TGG
Gadd45a Intron 2	91	GGTGAGTTGCATACGCTCAG	GGG
Gadd45b Exon 1	90	GAAGTCCCACCGCCTCCGGA	AGG
Gadd45b Intron 2	86	GCACCCCCCTTTCTCGGGCG	TGG
Gadd45g Exon 1	86	GTTATCCAAAAGAGTGCGGA	GGG
Gadd45g Intron 2	92	GTAGAAAGGCCCCGCGTGCA	GGG

6.1.6.8 Homology arm genomic locations

Homology arm	Location
Gadd45a 5' homology arm	chr6:66,985,656-66,989,641 (mm9)
Gadd45a 3' homology arm	chr6:66,980,590-66,985,652 (mm9)
Gadd45g 5' homology arm	chr13:51,939,287-51,943,345 (mm9)
Gadd45g 3' homology arm	chr13:51,943,349-51,947,437 (mm9)

6.1.6.9 Oligonucleotides used for cloning

6.1.6.9.1 Guide RNA insert oligonucleotides

Oligonucleotide name	Oligonucleotide sequence
ML_CRISPR_Gadd45a_Exon1_forward	CACCGCCCCCACAGTGCGCGGCGC
ML_CRISPR_Gadd45a_Exon1_reverse	AAACGCGCCGCGCACTGTGGGGGC
ML_CRISPR_Gadd45a_Intron2_forward	CACCGGTGAGTTGCATACGCTCAG
ML_CRISPR_Gadd45a_Intron2_reverse	AAACCTGAGCGTATGCAACTCACC
ML_CRISPR_Gadd45b_Exon1_forward	CACCGAAGTCCCACCGCCTCCGGA
ML_CRISPR_Gadd45b_Exon1_reverse	AAACTCCGGAGGCGGTGGGACTTC
ML_CRISPR_Gadd45b_Intron2_forward	CACCGCACCCCCCTTTCTCGGGCG
ML_CRISPR_Gadd45b_Intron2_reverse	AAACCGCCCGAGAAAGGGGGGTGC
ML_CRISPR_Gadd45g_Exon1_forward	CACCGTTATCCAAAAGAGTGCGGA
ML_CRISPR_Gadd45g_Exon1_reverse	AAACTCCGCACTCTTTTGGATAAC
ML_CRISPR_Gadd45g_Intron2_forward	CACCGTAGAAAGGCCCCGCGTGCA
ML_CRISPR_Gadd45g_Intron2_reverse	AAACTGCACGCGGGGCCTTTCTAC

6.1.6.9.2 Tag linker oligonucleotides

Oligonucleotide name	Oligonucleotide sequence
BAP-only_linker_oligo1	TCGAGGGCC
BAP-only_linker_oligo2	CCGGGGCCC
1xFlag-BAP_oligo1	TCGAGGATTACAAGGATGACGATGACAAGC
1xFlag-BAP_oligo2	CCGGGCTTGTCATCGTCATCCTTGTAATCC

6.1.6.9.3 Targeting vector tag linker oligonucleotides

Oligonucleotide name	Oligonucleotide sequence
BAP-only_tarvec_oligo1	GGGGGCC
BAP-only_tarvec_oligo2	CCGGGGCCCCCTGCA
1xFlag-BAP_tarvec_oligo1	GGGGATTACAAGGATGACGATGACAAGC
1xFlag-BAP_tarvec_oligo2	CCGGGCTTGTCATCGTCATCCTTGTAATCCCCTGCA

6.1.7 Plasmids

Name	Source	Description
pX330	Addgene	#42230, as published in [270]
px330_AE	Cloned myself	Gadd45a exonic guide RNA, see 6.2.1.1
px330_AI	Cloned myself	Gadd45a intronic guide RNA, see 6.2.1.1
px330_BE	Cloned myself	Gadd45b exonic guide RNA, see 6.2.1.1
px330_BI	Cloned myself	Gadd45b intronic guide RNA, see 6.2.1.1
px330_GE	Cloned myself	Gadd45g exonic guide RNA, see 6.2.1.1
px330_GI	Cloned myself	Gadd45g intronic guide RNA, see 6.2.1.1
pCS2+_myc-Gadd45a	Mathias Gierl	N-terminally myc-tagged mmGadd45a
pCS2+_myc-Gadd45b	Mathias Gierl	N-terminally myc-tagged mmGadd45b
pCS2+_myc-Gadd45g	Mathias Gierl	N-terminally myc-tagged mmGadd45g
pOctTKGFP	Andrea Schäfer	As published in [146]

pGL3	Sabine Kienhöfer	Promega firefly luciferase reporter, E1751
pRL	Sabine Kienhöfer	Promega renilla luciferase reporter, E2261
pGEM_Gadd45a.10.9	Mathias Gierl	Gadd45a antisense probe transcribed with T7
pGEM_Gadd45b.19.20	Mathias Gierl	Gadd45b antisense probe transcribed with SP6
pGEM_Gadd45g.27.28	Mathias Gierl	Gadd45g antisense probe transcribed with SP6
pCS2+_Gadd45a-V53xFlagBAP	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_Gadd45a-1xFlagBAP	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_Gadd45a-BAP	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_1xFlagBAP-Gadd45a	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_BAP-Gadd45a	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_Gadd45g-V53xFlagBAP	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_Gadd45g-1xFlagBAP	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_Gadd45g-BAP	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_1xFlagBAP-Gadd45g	Cloned myself	Overexpression vector, see 6.2.1.1
pCS2+_BAP-Gadd45g	Cloned myself	Overexpression vector, see 6.2.1.1
Targeting vector	Mathias Treier	Targeting vector backbone to introduce BAP
Gadd45a Targeting vector	Cloned myself	Targeting vector, see 6.2.1.1
Gadd45g Targeting vector	Cloned myself	Targeting vector, see 6.2.1.1
pGEM_A3P	Cloned myself	Gadd45a 3' southern probe, see 6.2.1.1
pGEM_A5P	Cloned myself	Gadd45a 5' southern probe, see 6.2.1.1
pGEM_G3P	Cloned myself	Gadd45g 3' southern probe, see 6.2.1.1
pGEM_G5P	Cloned myself	Gadd45g 5' southern probe, see 6.2.1.1
pCS2+_myc-BirA	Cloned myself	BirA overexpression plasmid, see 6.2.1.1
pCS2+_GFP	Andrea Schäfer	GFP overexpression plasmid
pPuro	Lars Schomacher	Puromycin resistance plasmid

6.2 Methods

6.2.1 General molecular biology

General molecular biology methods including preparation of chemically competent XL1-blue *Escherichia coli* bacteria, plasmid amplification in *Escherichia coli*, spectrophotometric quantification of DNA and RNA, restriction digests, DNA ligations, PCR, agarose gel electrophoresis, and SDS-PAGE were carried out as previously described [271]. All used oligonucleotides were synthesized by Sigma Aldrich. Plasmid DNA was sequenced by GATC Biotech.

6.2.1.1 Cloning of plasmids

6.2.1.1.1 Guide RNA plasmids

For guide RNA inserts, guide RNA insert oligonucleotides (see chapter 6.1.6.9.1) were annealed by heating to 99°C for 5 minutes in a heating block and then switching off the heating block and letting the samples cool down to room temperature passively.

px330 vector (Addgene) was cut with *BbsI* restriction enzyme (NEB) for 1h at 37°C. Annealed oligonucleotides were ligated in presence of restriction enzyme into the px330 vector using T4 ligase

(NEB), as the *BbsI* restriction site is disrupted by successful ligation. Positive clones were identified by analytical *BbsI* digestion, and plasmid clones that could no longer be linearized by *BbsI* were sequenced to confirm correct insertion.

6.2.1.1.2 BirA and tagged Gadd45a and Gadd45g overexpression plasmids

V5-3xFlag-BAP tag was amplified from a targeting vector obtained from Mathias Treier by PCR (V5FlagBioTag_cloning forward/reverse primers, see chapter 6.1.6), cloned into pGEM_T-easy and subcloned into pCS2+ using *XbaI* and *XhoI* restriction endonucleases. Gadd45a and Gadd45g were amplified without stop codons from pCS2+_mycGadd45a and pCS2+_mycGadd45g (Gadd45a_cloning forward/reverse primers, Gadd45g_cloning forward/reverse primers, see chapter 6.1.6), cloned into pGEM T-easy and subcloned into pCS2+_V5-3xFlag-BAP using *ClaI* and *XhoI* restriction endonucleases to obtain C-terminally V5-3xFlag-BAP tagged Gadd45a and Gadd45g.

For shortened C-terminal tags, pCS2+_Gadd45g_V5-3xFlag-BAP vector was cut with *XmaI* and *XhoI* endonucleases to remove the 3xFlag and the V5 parts of the tag. Oligonucleotides encoding for the 1x-Flag or a linker (see chapter 6.1.6.9.2) were annealed and cloned into the gap, thus giving rise to the 1xFlag-BAP and BAP-only tagged Gadd45g plasmids, respectively.

pCS2+_Gadd45g-1xFlagBAP and pCS2+_Gadd45g-BAP were cut using *ClaI* and *XhoI* endonucleases, thus removing the Gadd45g open reading frame. Gadd45a open reading frame was cut from pCS2+_Gadd45a-V5-3xFlag-BAP using *ClaI* and *XhoI* and cloned into the resulting gap, thus creating pCS2+_Gadd45a-1xFlagBAP and pCS2+_Gadd45a-BAP.

For N-terminal tags, 1x-Flag-BAP and BAP-only tag sequences were amplified via PCR (FLBIO and BIO forward primer, FLBIO_BIO reverse primer, see chapter 6.1.6) and cloned into pCS2+ using *BamHI* and *ClaI*, thus creating pCS2+_n1xFlag-BAP and pCS2+_nBAP plasmids. Gadd45a and Gadd45g including stop codons were amplified using PCR (Gadd45a_cloning forward and Gadd45g_cloning forward primers, G45a_STOP and G45g_STOP reverse primers, see chapter 6.1.6), cloned into pGEM_T-easy and subcloned into the pCS2+_n1xFlag-BAP and pCS2+_nBAP plasmids using *ClaI* and *XhoI*, thus creating pCS2+_1xFlagBAP-Gadd45a, pCS2+_BAP-Gadd45a, pCS2+_1xFlagBAP-Gadd45g and pCS2+_BAP-Gadd45g.

BirA open reading frame was amplified from *Escherichia coli* genomic DNA (BirA_CDS forward/reverse primers, see chapter 6.1.6), cloned into pGEM_T-easy and cloned into pCS2+_myc using *BamHI* and *XbaI*.

6.2.1.1.3 Gadd45a and Gadd45g targeting vectors

In order to exchange the V5-3xFlag-BAP tag in the targeting vector, the targeting vector was cut open with *XmaI* and *SbfI* endonucleases to remove the 3x-Flag and V5 tags. Oligonucleotides encoding for the 1x-Flag or a linker only (see chapter 6.1.6.9.3) were annealed and cloned into the gap, restoring the *XmaI* and *SbfI* restriction sizes and ensuring that the BAP-tag was still in frame with the targeted gene.

Gadd45a and Gadd45g 3' and 5' homology arms were amplified via PCR (HR set of primers, see chapter 6.1.6), cloned into pGEM_T-easy and sequenced. Homology arms without PCR errors within coding elements were cloned into the modified targeting vectors using *SbfI/SwaI* and *NotI/PacI* restriction endonucleases.

6.2.1.1.4 Southern probe plasmids

Southern probe sequences were amplified from mouse genomic DNA via PCR (Gadd45a_5'probe, Gadd45a_3'probe, Gadd45g_5'probe and Gadd5g_3'probe forward/reverse primers, see chapter 6.1.6) and cloned into pGEM_T-easy.

6.2.1.2 Amplification and purification of plasmids

For plasmid amplification, plasmids were transformed into *Escherichia coli* XL-1 blue chemically competent bacteria by a 40 s heatshock at 42°C, 5 min incubation on ice and subsequent incubation in LB containing an appropriate selection antibiotic overnight at 37°C. Plasmid DNA was purified from bacteria using Qiagen Miniprep or Midiprep kits according to the manufacturer's recommendation. DNA amount and purity were estimated on a Nanodrop 2000 spectrophotometer.

6.2.1.3 *In vitro* methylation

For *in vitro* methylation of plasmids, 40µg of plasmid DNA were incubated with 8µl *HpaII* or *MsssI* methylase (Promega) and 12µl SAM in a 300µl reaction volume with the corresponding *HpaII* or *MsssI* buffer (Promega) for 4h at 37°C. After 4h, 2µl additional methyltransferase and 3µl additional SAM were added and the reaction was incubated at 37°C overnight. On the next day, plasmid DNA was purified via phenol-chloroform extraction and methylation efficiency was judged by *HpaII* and *MspI* endonuclease digestion with subsequent agarose gel electrophoresis.

6.2.1.4 Western blot

Protein concentrations were estimated by BCA assay. For this purpose, 2 µl of the protein sample was incubated with 200 µl of bicinchoninic acid + 2% Cu(II)SO_4 at 37°C for 30 minutes. Absorption at 562 nm was measured with an Infinite 200 Pro microplate reader (Tecan) and compared to a standard curve of

bovine serum albumin samples with known protein concentrations that was run in parallel to samples of interest.

Defined protein amounts (usually 20µg) were adjusted to equal volumes (usually 15µl) using the current sample buffer, then mixed with 4x Lämmli sample buffer (usually 5µl) and incubated for 5 min at 95°C. SDS-PAGE, transfer to polyvinylidene difluoride (PVDF) membranes and Western Blotting was performed according to standard protocols [271]. Signals were developed with SuperSignal West Pico or Femto Chemiluminescent Substrate (Thermo Scientific) and analyzed using a ChemiDoc (BioRad) with ImageLab software.

6.2.1.5 Southern blot

10µg of purified genomic DNA were digested with 50U of appropriate restriction enzyme in a volume of 200µl overnight at 37°C. DNA was precipitated by addition of 600µl ethanol and 20µl 3M sodium acetate, pH 5.2, and incubation at -20°C for at least one hour. DNA was pelleted for 1h at 4°C and 21.000g. Supernatant was removed, the pellet air dried for 3 minutes and resuspended in 15µl TE buffer for at least 1h at 4°C. DNA was then separated on a 0.8% agarose gel overnight.

The agarose gel was briefly rinsed in water, depurinated in 0.25M HCl for 30 minutes and denatured three times in 0.4M NaOH for 20 minutes each, shaking at 40rpm. Afterwards, the agarose gel was stored in 20xSSC buffer until further processing. Positively charged nylon membranes were briefly soaked in 0.4M NaOH. Whatman paper was soaked in 20xSSC. A basin was filled with 20xSSC and the southern blot sandwich was assembled on top of the basin, from bottom to top: A single, large Whatman paper in contact with the 20xSSC in the basin, gel, membrane, three layers of Whatman paper, a stack of dry paper towels, a gel casting case and a single 250ml glass bottle filled with 100ml water. After adding the gel layer, the whole setup was wrapped in saran foil and cut open only directly above the gel to prevent liquid from by-passing the gel. Blotting was performed overnight at room temperature.

The blot was disassembled and the membrane was dried between two sheets of Whatman paper for at least 30 minutes at room temperature. Once completely dry, the membrane was crosslinked with 120.000µJ/cm^2 in an UV crosslinker and stored until further use at room temperature.

Probe DNA was cut from the corresponding plasmid, gel purified and 25-50ng of purified probe were denatured for 3 minutes at 99°C in a volume of 45µl ddH_2O and immediately put on ice. Probe was labelled by resupending one labelling bead (GE Healthcare) in the denatured probe, the addition of 5µl dCTP [α-32P] and incubation at 37°C for 30 minutes. Probe was purified by centrifuging through a G-50 column (GE Healthcare) for 2 minutes at 3.000rpm. Labelling efficiency was measured in a scintillator.

The membrane was wetted in 6xSSC and afterwards pre-hybridized for 30-60 minutes in 20ml Church buffer at 68°C in a rolling hybridization oven. Probe was denatured for 2 minutes at 95°C, cooled on ice for 2 minutes and mixed with 20ml Church buffer to a final concentration of 1.000.000 cpm/ml. Pre-hybridization buffer was decanted and the membrane was incubated in the probe containing Church buffer at 68°C for roughly 17 hours in the rolling hybridization oven.

The membrane was briefly rinsed in 2xSSC/0.1%SDS at 68°C, washed for 30 minutes in 2xSSC/0.1%SDS at 68°C, washed three times 30 minutes in 0.2xSSC/0.1%SDS at 68°C and briefly rinsed in 2xSSC at room temperature. Excess liquid was removed and the membrane was welded into plastic foil. The membrane was incubated with a BAS IP MS 2025 E multipurpose phosphor screen (GE Healthcare) in an intensifier hypercassette (GE Healthcare) overnight. The phosphor screen was imaged using a Typhoon FLA-9500.

6.2.1.6 *In situ* hybridization

For probe transcription, 5µg plasmid DNA was linearized with 50U of appropriate restriction enzyme for 2 hours at 37°C. Linearized plasmid was purified using the PCR purification kit (Qiagen) according to the manufacturer's instructions. RNA was transcribed from 2µg linearized plasmid using 4µl SP6 or T7 RNA polymerase (Ambion) in presence of 5µl DIG-labeled NTPs (Roche), 2µl RNAse Inhibitor and the corresponding buffer for 2 hours at 37°C. After *DNAseI* digestion (2µl, Roche) for 15 minutes at 37°C, probes were purified using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions. Probe quantity and quality was afterwards examined on a 2% agarose gel.

Placental cryo-sections were thawed and air dried for 2 hours at room temperature, fixed for 10 minutes in 4% PFA, washed 3 times 5 minutes each in PBS-DEPC, incubated 10 minutes in acetylation buffer, washed 3 more times 5 minutes each in PBS-DEPC and incubated in hybridization buffer for 2 hours at room temperature.

2µl DIG-labelled probe was added to 150µl hybridization buffer, denatured for 10 minutes at 80°C, put on ice and added to cryo-sections. Sigmacote coated coverslips were added on top and slides were incubated at 65°C overnight in a 5xSSC/50%formamide filled incubation chamber in a humidified incubator.

Slides were washed once 5 minutes in 5xSSC at room temperature, twice 30 minutes in wash buffer at 65°C, once 5 minutes in wash buffer at room temperature and three times in MABT at room temperature. Slides were blocked 2 hours in blocking solution at room temperature. Afterwards, slides were incubated with AP-conjugated anti-DIG antibody diluted 1:1500 in blocking solution overnight at 4°C, rotating.

Slides were washed 3 times 15 minutes each in MABT at room temperature and washed once in NTMT at room temperature. Slides were stained with staining solution (1% NBT (Roche), 1% BCIP (Roche) in NTMT) at 4°C until sufficient staining was visible (up to 3 days). Staining reaction was stopped by washing slides twice in tap water for 10 minutes each. Coverslips were mounted with Mowiol and slides were imaged with a light microscope.

6.2.1.7 Luciferase assay

10.000 HEK293T cells per well were seeded in 96-well plates and transfected the next day with 50ng pGL3, 1ng pRL, 10ng GFP, 51ng pBS and 5ng Gadd45 using 0.35µl Xtreme Gene9 as described in chapter 6.2.2.10. 48 hours after transfection, luciferase activity was measured using the Dual Luciferase Assay Kit (Promega) according to the manufacturer's instructions. Briefly, cells were lysed in 50µl 1x Promega lysis buffer for 15 minutes at room temperature, shaking at 400rpm. Lysate was frozen at -20°C for at least 30 minutes to further facilitate cell lysis. Cells where thawed and 20µl lysate per sample was transferred to white cliniplates in technical duplicates. Renilla luciferase and firefly luciferase substrate buffers were prepared according to the manufacturer's instructions and then diluted further with an equal volume of ddH_2O. Using a Tecan M2000Pro plate reader, 50µl firefly luciferase substrate was automatically added to lysed cells, incubated and luminescence was measured. Afterwards, 50µl renilla luciferase substrate was automatically added to lysed cells, incubated and luminescence was measured. Relative firefly luciferase activity was calculated by normalizing firefly luciferase activity to renilla luciferase activity.

6.2.1.8 Nucleic acid extraction and purification

RNA was isolated using the RNeasy mini and RNeasy mini (96-well) kits from Qiagen according to the manufacturer's instructions. The optional DNAse I on-column digestion was carried out and RNA was eluted in 30µl ddH_2O, measured on the Nanodrop 2000 and diluted to 125ng/µl in ddH2O, if possible.

DNA was isolated using the DNeasy mini kit from Qiagen according to the manufacturer's instruction. If intended for mass spectrometry, cells were resuspended in 100µl PBS including 5µl of RNAse A, vortexed, squashed by 10 minutes of 21.000xg centrifugation at 4°C, further incubated for 15 minutes on ice and afterwards mixed with 100µl of buffer ATL (Qiagen) before proceeding with the kit's regular instructions. Furthermore, DNA was further purified after elution: DNA was precipitated with 3.3x volumes of 100% ethanol (-20°C), 0.1x volume of 7.5M ammonium acetate and centrifugation at 21.000xg, 4°C for at least 1 hour. The DNA pellet was washed once with 70% ethanol (-20°C) and re-centrifuged. Supernatant was removed and the pellet was dried for at least 1 hour at room temperature and resuspended in 20µl ddH_2O.

6.2.1.9 qPCR (expression analysis)

cDNA synthesis was carried out using Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's instructions. Briefly, 500ng RNA, 1µl 5mM dNTPs and 120ng random primers were mixed in a final volume of 12µl. After denaturation for 5 min at 65°C, the mixture was cooled on ice. 4µl 5x FS buffer, 2µl 0.1M DTT, 1 µl Ribolock and 1 µl Superscript II Polymerase were added and samples were incubated for 10 min at 25°C, 90 min at 42°C for cDNA synthesis and 5 min at 72°C for enzyme inactivation. Resulting cDNA was diluted 1:6 in ddH_2O. All incubation steps were carried out in a PCR cycler.

For each 11µl qPCR reaction, 5µl cDNA, 5.5µl 2x Probes master, 0.11µl 100µM forward/reverse primer mixture, 0.11µl UPL probe and 0.28µl ddH_2O were mixed. PCR reactions were carried out in a 384-well format in the Roche Light Cycler 480 using the following PCR program:

	PCR step	Temperature	Time	Ramp rate
1	Denaturation	95°C	10 min	4.8 °C/sec
2	Denaturation	95°C	10s	4.8 °C/sec
3	Annealing	60°C	20s	2.5 °C/sec
4	Elongation + Signal acquisition	72°C	1s	4.8 °C/sec
5	Go to step [2], 49x times	-	-	-
6	Cooling	4°C	1s	2.5 °C/sec

Cp values were determined using the Roche Lightcycler software and relative expression values were calculated according to the ddCp method [272], normalizing to *Gapdh* or *Tbp* as housekeeping genes.

6.2.1.10 qPCR (Methylation-specific PCR)

80% confluent HEK293T cells were split 1:10 onto 6-well dishes and transfected on the next morning with 1.5µg *in vitro HpaII*-methylated pOctTK-GFP and 0.5µg Gadd45 construct or pCS2+ using 6µl Xtreme Gene9 (Roche) as outlined in 6.2.2.10. 48 hours later, cells were washed once with PBS and harvested by scraping and centrifuged for 5 minutes at 300xg at 4°C. Total DNA was extracted from the cell pellet using the DNeasy mini kit (Qiagen) according to the manufacturer's instructions and eluted in 50µl ddH_2O. 1µl DNA was digested with 0.5µl *HpaII* (Promega) or 0.5µl *PvuII* (Promega) in a 10µl reaction for at least 3 hours at 37°C. DNA digest was diluted 1:1000 with ddH_2O and used as a qPCR template as described in 6.2.1.9, normalizing to the *PvuII* control digest instead of housekeeping genes.

6.2.1.11 Genotyping PCRs

For CRISPR/Cas9 clone genotyping, cells in one 96-well-plate well were lysed in 50µl tail lysis buffer for at least 3 hours at 37°C, for up to one week. Afterwards, 25µl lysed sample was transferred to a 96-well PCR plate pre-loaded with 75µl ddH_2O. Proteinase K was inactivated by incubation at 95°C for 10

minutes. Per 20µl genotyping PCR reaction, 12.5µl ddH2O, 1µl DMSO, 1µl Bioron complete buffer, 1µl 5mM dNTPs, 1µl 10µM forward/reverse primer mixture, 0.5µl homemade Taq polymerase and 2µl diluted DNA sample were mixed. DNA was amplified using the following PCR program:

	PCR step	Temperature	Time
1	Denaturation	94°C	5 min
2	Denaturation	95°C	30s
3	Annealing	59°C	30s
4	Elongation	72°C	2 min (external) or 1 min (internal)
5	Go to step [2], 39x times	-	-
6	Final Elongation	72°C	2 min
7	Cooling	4°C	1 min

6.2.1.12 Bisulfite sequencing

Genomic DNA was isolated from control and TKO mESCs using the DNeasy mini kit (Qiagen) according to the manufacturer's instructions. Defined ratios of *in vitro* methylated and non-methylated pOctTK-GFP and pGL3 plasmids were mixed, and 1pg of these mixtures was added as a standard curve to specific genomic DNA samples. Bisulfite conversion reactions using the EpiTect kit (Qiagen) were assembled according to the manufacturer's instructions. Bisulfite conversion was carried out using the following steps in a PCR cycler: 5 min at 95°C, 25 min at 60°C, 5 min at 95°C, 85 min at 60°C, 5 min at 95°C, 175 min at 60°C, 5 min at 95°C, 120 min at 60°C and 20°C overnight. Bisulfite converted DNA was purified according to the manufacturer's instruction.

50µl PCR reactions were assembled, using 40µl ddH_2O, 5µl AccuPrime buffer II, 1.5µl 1mM MgCl2, 1µl AccuPrime Taq, 1.5µl bisulfite converted DNA and 1µl 10µM forward/reverse primer mix. For template amplification, the following PCR program was used:

	PCR step	Temperature	Time
1	Denaturation	95°C	5 min
2	Denaturation	95°C	35s
3	Annealing	52°C-56°C	30s
4	Elongation	68°C	50s
5	Go to step [2], 40x times	-	-
6	Final Elongation	68°C	5 min
7	Cooling	4°C	1 min

5µl of each PCR reaction were run on a 1% agarose gel and band intensity was quantified with the ImageLab 3.0 software (BioRad) using the 1kb DNA ladder (Ambion) for normalization. For each condition (i.e. control mESC line #281), amplicons were mixed in equimolar amounts to yield 2µg total DNA, purified using the PCR purification kit (Qiagen) according to the manufacturer's instructions and eluted in 15µl ddH_2O. From each amplicon pool, libraries were prepared and sequenced on an Illumina

MiSeq by IMB genomics core facility. Methylation levels were called from sequencing data by Medhavi Mallick and data was subsequently analyzed in Microsoft Excel.

6.2.1.13 Streptavidin based chromatin immunoprecipitation (ChIP)

Per 5 ChIPs, one 10cm dish of cells was fixed by addition of formaldehyde to a final concentration of 1% directly to the cell culture medium and incubation at room temperature for 10 minutes, shaking at 30rpm. Formaldehyde was quenched by addition of glycine to a final concentration of 125mM and incubated for further 5 minutes at room temperature, shaking at 30rpm. Cells were washed twice in ice-cold PBS, scraped in ice-cold PBST, centrifuged at 800g at 4°C and snap-frozen in liquid nitrogen until further use. All steps following freezing were carried out in presence of Protease Inhibitor Cocktail (Roche).

In case of later sonication, cells were lysed in 1ml ChIP lysis buffer for 5 minutes on ice. 50µl 20% SDS were added and samples were incubated for further 5 minutes on ice. Cells were sheared for 20 cycles of 30s-on/30s-off in the Bioruptor (Diagenode) or for 840s in the Covaris S2 sonicator. Further 50µl 20% SDS were added and samples were heated to 65°C for 5 minutes and immediately put on ice afterwards. 22µl 500mM iodoacetamid were added and samples were incubated for 30 minutes at 37°C in the dark.

In case of later MNase digestion, cells were lysed in 1ml HNB buffer and incubated on ice for 20 minutes. Nuclei were centrifuged for 5 minutes at 240g and 2 minutes at 2400g at 4°C. Supernatant was discarded and nuclei pellet resuspended in 150µl NE buffer. 1µl of MNase (NEB) was added and samples were incubated for 30 minutes at 37°C. Digestion was stopped by addition of 1.5µl 0.5M EDTA and 3µl 0.25M EGTA. 20µl 20% SDS, 20µl 10% Triton X-100 and 2.8µl 1% β-mercaptoethanol were added. Samples were heated to 65°C for 5 minutes and immediately put on ice afterwards. Samples were centrifuged for 5 minutes at 21.000g at room temperature and the pellet was discarded. 4µl 500mM iodoacetamid were added and samples were incubated for 30 minutes at 37°C in the dark.

In both cases, IPs were set up by combining 200µl (sonication) or 170µl (digestion) fragmented chromatin with 1800µl ChIP dilution buffer and 50µl C1 streptavidin coupled Dynabeads (Invitrogen). 10µl input control was stored at -20°C during the IP procedure. IP was carried out overnight at 4°C on a vertical rotor. On the next day, beads were washed for 8 minutes each with 1ml of the washing buffers I, II, III, and twice with TE buffer, while settling beads for 2 minutes between washing steps on a magnetic rack.

After washing, DNA was eluted from half of the beads by incubation in 300µl elution buffer at 65°C overnight while shaking (800rpm), whereas proteins were eluted from the other half of the beads by incubation in 50µl 1x Lämmli sample buffer at 99°C for 10 minutes while shaking (800rpm). IP efficiency was judged by Western blot.

The Dynabeads from the ChIP sample were extracted and discarded, 300µl TE buffer and 3µl RNAse A (Ambion) were added and samples were incubated at 37°C for 30 minutes. Subsequently, 2µl glycogen (Fermentas) and 6µl Proteinase K (10mg/ml) were added and samples were incubated for further 2 hours at 37°C. DNA was afterwards purified using the PCR purification kit (Qiagen) according to the manufacturer's instructions and eluted in 50µl ddH_2O.

In case of performing ChIP-qPCR, DNA was further diluted depending on the amount of primers subsequently used for analysis, to a maximum volume of 100µl. qPCR was then carried out as described in 6.2.1.9 using ChIP DNA instead of cDNA and normalization either to input control or Mock IP (as indicated in the individual experiment).

6.2.1.14 ChIP-Seq library preparation

ChIP DNA fragmentation efficiency was estimated using a DNA High Sensitivity chip (Bioanalyzer) and DNA quantity was measured using Qubit. 20ng of DNA were diluted in 44µl ddH_2O. DNA libraries were prepared using the NEBNext ChIP-Seq Library Prep Master Mix (NEB) according to the manufacturer's instructions. Briefly, DNA was end-repaired with 1µl proprietary end repair enzyme mix, cleaned up using AmPure XP beads (Beckmann), A-tailed using 1µl of Klenow fragment, cleaned up again, ligated to 1µl 1.5µM Illumina adaptors using 4µl T4 ligase and cleaned up once more.

Library DNA was size-selected for 200-300 bp long fragments using the E-Gel system (Invitrogen) according to the manufacturer's instruction. Collected DNA was reduced to approximately 50 µl using a Speedvac.

23µl adaptor ligated, size selected DNA were combined with 25µl NEB PCR master mix, 1µl Illumina PE1 and 1µl Illumina PE2 primer and amplified via the following PCR program:

	PCR step	Temperature	Time
1	Denaturation	98°C	30s
2	Denaturation	98°C	10s
3	Annealing	65°C	30s
4	Elongation	72°C	30s
5	Go to step [2], 14x times	-	-
6	Final Elongation	72°C	5 min
7	Cooling	4°C	1 min

Libraries were cleaned up using AmPure XP beads (Beckmann) and subsequently sequenced on an Illumina HighSeq2000 sequencer.

6.2.2 Cell culture

6.2.2.1 Gelatinizing of cell culture vessels

To gelatinize cell culture vessels, surface area was completely covered in 0.1% gelatin in ultrapure water (Millipore) and incubated for at least 15 minutes at room temperature. Gelatin was aspirated immediately before plating cells.

6.2.2.2 Regular culture of mESCs

mESCs were regularly cultured in serum-containing mESC medium at 37°C, 5% CO_2 and 21% O_2, with daily medium changes. Cells were passaged every 2 days once reaching approximately 60% confluency, by washing the cells with PBS once, loosening cells with 0.25% trypsin for 2 minutes at 37°C, quenching trypsin with mESC medium, spinning down the cells for 3 minutes at 300g, resuspending cells in an adequate amount of mESC medium and plating $1/8^{th}$ on a new cell culture vessel. For E14tg2a mESCs, gelatinized cell culture vessels were sufficient. For R1/E mESCs, cell culture vessels were covered in a layer of feeder cells (see 6.2.2.5).

6.2.2.3 Regular culture of mTSCs

mTSCs were regularly cultured in mTSC-complete medium at 37°C, 5% CO_2 and 21% O_2, with medium changes every second day. Cells were passaged every 3-4 days, once reaching approximately 70% confluency, by washing cells with PBS twice, loosening cells with 0.25% trypsin for 3 minutes at 37°C, quenching trypsin with mTSC-incomplete, spinning down the cells for 3 minutes at 300g, resuspending cells in a small amount of mTSC-incomplete and plating $1/5^{th}$ on a new, non-gelatinized cell culture vessel.

6.2.2.4 Regular culture of HEK293T

HEK293T were regularly cultured in DMEM+++ medium at 37°C, 5% CO_2 and 21% O_2, without changing medium between passaging. Cells were passaged every other day, once reaching approximately 80% confluency, by washing cells with PBS once, loosening cells with 0.05% trypsin for 1 minute at 37°C, quenching trypsin with DMEM+++ and plating $1/6^{th}$ on a new, non-gelatinized cell culture vessel.

6.2.2.5 Feeder cell and feeder-cell-conditioned medium production

Mouse embryonic fibroblasts were regularly cultured in DMEM+++ medium at 37°C, 5% CO_2 and 5% O_2, without changing medium between passaging. Cells were passaged every 2-3 days, once reaching 100% confluency, by washing cells with PBS once, loosening cells with 0.25% trypsin for 2 minutes at 37°C, quenching trypsin with DMEM+++, spinning down cells for 3 minutes at 300g, resuspending cells in

DMEM+++ and expanding cell culture surface by a factor of 4. Cells were expanded for four passages after thawing, thereby yielding 50x 15cm dishes with confluent fibroblasts.

After reaching 100% confluency, cells were arrested by 10µg/ml mitomycin C in DMEM+++ for 3 hours at 37°C. Afterwards, cells were washed intensively with PBS three times.

For long-term storage, feeder cells were loosened as described above, resuspended in 10% DMSO, 50% DMEM+++, 30% FBS and slowly frozen to -80°C at a cooling rate of -1°C/min in freezing containers filled with isopropanol. Subsequently, cells were transferred to -150°C.

For conditioned medium preparation, 30ml mTSC-incomplete medium was added to each 15cm dish of inactivated feeders and incubated for 2 days. Conditioned medium was collected and new mTSC-incomplete was added. Three batches of conditioned medium were collected in such a way from each feeder dish, subsequently pooled and filtered before usage.

6.2.2.6 Embryoid body differentiation

Regularly cultured E14tg2a mESCs were directly suitable for differentiation. R1/E mESCs were passaged twice on gelatinized cell culture vessels, instead of feeder cell covered cell culture vessels, before proceeding with differentiation.

For EB differentiation, $3.5x10^6$ mESCs were plated on non-adherent 10cm bacterial dishes (Greiner) in 15ml CA medium. Cell culture medium was changed every other day by collecting EBs in 50ml falcon tubes and letting them settle by gravity for 2-3 minutes. Supernatant was aspirated, EBs carefully resuspended in 15ml fresh CA medium and re-plated on a new 10cm non-adherent bacterial dish.

For retinoic acid induced EB differentiation, CA medium was supplemented with retinoic acid to a final concentration of 5µM at days 4 and 6 of differentiation.

6.2.2.7 Monolayer differentiation

Regularly cultured E14tg2a mESCs were directly suitable for differentiation. R1/E mESCs were passaged twice on gelatinized cell culture vessels, instead of feeder cell covered cell culture vessels, before proceeding with differentiation.

For monolayer differentiation, approximately 5.000 mESCs per cm^2 were plated on gelatinized cell culture vessels in regular mESC medium on the evening before differentiation. On the next morning, cells were washed twice with PBS to remove remaining LIF, and N2B27 medium was added. Medium was changed on days 3, 5, 6 and 7 of differentiation.

6.2.2.8 BMP4 induced mESC-to-mTSC transdifferentiation

mESCs were centrifuged for 3 minutes at 300g and supernatant was completely removed. Cells were resuspended in KOSR medium including 10ng/ml recombinant hBMP-4 (R&D Systems) and plated at a density of 10^4 cells per cm² on gelatinized cell culture vessels. Medium was changed every other day.

6.2.2.9 mTSC differentiation

For mTSC differentiation, mTSCs were plated in mTSC-incomplete instead of mTSC-complete after passaging. Medium was changed every other day and cells were no longer passaged.

6.2.2.10 Vitamin C, 2i and 5-azacytidine treatment

For vitamin C treatment, 230.000 mESCs per well were plated in 6-well-plates in regular mESC medium. The next day, 1:500 50mg/ml L-ascorbic acid 2-phosphate sesquimagnesium salt was added to a final concentration of 100µg/ml. Cells were incubated in vitamin C containing medium for up to 72h and passaged once during that time.

For 2i treatment, 230.000 mESCs per well were plated in 6-well-plates in regular mESC medium. On the next day, cells were washed twice with PBS to remove traces of serum. Cells were afterwards incubated up to 72h in 2i medium.

For 5-azacytidine treatment, mESCs were incubated in 10µM 5-azacytidine in mESC medium for 48h.

6.2.2.11 Transfection

mESCs were transfected with Lipofectamine 2000 (Thermo-Fischer). For a 12-well format, 1.25-2.5x10^5 mESCs were plated and let settle for 3h. 250ng of plasmid DNA were prepared in 50µl TE buffer and 4µl of Lipofectamine 2000 were mixed with 46µl OptiMEM medium. Both were mildly vortexed and incubated for 5 minutes at room temperature. Afterwards plasmid containing TE buffer and Lipofectamine 2000- containing OptiMEM were mixed, mildly vortexed and incubated at room temperature for 20 minutes. Subsequently the transfection mix was applied dropwise to the cells and distributed by gentle swirling. Medium was changed on the next day and cells were usually harvested 48h post transfection.

HEK293T cells were transfected with Xtreme Gene9 (Roche) 24 hours after cell passaging. For a 12-well format 480ng of plasmid DNA were prepared in 50µl OptiMEM medium, and 1.584µl Xtreme Gene9 was prepared in 50µl OptiMEM. Both solutions were mildly vortexed and incubated for 5 minutes at room temperature. Afterwards plasmid-containing OptiMEM Xtreme and Gene9-containing OptiMEM were mixed, mildly vortexed and incubated at room temperature for 20 minutes. Subsequently the transfection mix was applied dropwise to the cells and distributed by gentle swirling. Cells were usually harvested 48h post transfection.

6.2.2.12 Establishment of CRISPR/Cas9-mediated knockout mESC and mTSC lines

$1.45x10^6$ mESCs were seeded on 10cm dishes each and transfected on the next day with either i) 13.5µg empty px330 vector and 1.5µg pPuro, ii) 13.5µg px330 and 1.5µg pCS2+_GFP, iii) 2.25µg of each of the six guide RNA containing plasmids (see 6.2.1.1) and 1.5µg pPuro using 45µl Lipofectamine 2000 (Thermo-Fischer) in a volume of 1.5ml OptiMEM, but otherwise as described in 6.2.2.10. mESCs were passaged from 1x10cm to 2x15cm dishes the following day. Cells were selected with 2µg/ml puromycin from 48h post-transfection and kept under selection until the day of freezing.

Forming mESC colonies were picked 6 days after passaging. Using the microscope, single colonies were detached using a 100µl pipette set to 20µl by a combination of gentle pushing and pipetting-induced medium currents. Colonies were subsequently transferred to gelatinized 48-well plates. On the next day, colonies were washed once with 500µl PBS, dissociated with 100µl 0.25% trypsin for 90s at 37°C, quenched with 800µl of mESC medium and plated on a new 48-well plate.

Three days later, cells were dissociated as described above, but quenched with 200µl [70% mESC medium, 30% FBS]. 100µl of the resulting 300µl suspension were transferred to a new 48-well plate containing 900µl mESC medium and incubated at 37°C. The remaining 200µl were mixed with 200µl [20% DMSO, 50% mESC medium, 30% FCS], briefly stored on ice, wrapped in plastic foil, transferred into a 4°C cold Styrofoam box filled with paper towels, and stored at -80°C.

The cells still in culture were grown until confluency and subjected to genotyping PCR as described in chapter 6.2.2.11. Positive clones were thawed by transferring the 48-well plate from -80°C on ice, adding warm mESC medium into the intended well and collecting the thawed cells in a 1.5ml reaction tube. This step was repeated until the complete well was thawed. Cells were centrifuged 5 minutes at 300g, resuspended in mESC medium and plated on a 48-well plate. Cells were subsequently regularly cultured, passaged and expanded as described in chapter 6.2.2.2.

CRISPR/Cas9-mediated Gadd45 knockout in mTSCs was performed similarly. mTSCs were transfected with Neuromag (OZ Biosciences), rather than Lipofectamine 2000, with either i) 15µg empty px330 vector and 2.3µg pPuro, ii) 15µg px330 and 2.3µg pCS2+GPF, or iii) 2.4µg of each of the six guide RNA-containing plasmids and 2.3µg pPuro in a volume of 1ml OptiMEM, using 33.4µl Neuromag and keeping cells on a magnet overnight. Because of the slower cell cycle of mTSCs compared to mESCs, colonies were only picked 12 days after transfection. Lastly, colonies were picked and subsequently cultured in 96-well plates, rather than 48-well plates.

6.2.2.13 Establishment of transgenic mESC lines via homologous recombination

80µg targeting vector was linearized using 300U *SwaI* restriction endonuclease in a 600µl reaction for 3 hours at 25°C, subsequently purified using phenol:chloroform extraction and ethanol precipitation. DNA concentration was measured using the Nanodrop2000. R1/E BirA mESCs corresponding to 50% confluent 40cm^2 were resuspended in 750µl PBS and mixed with 35µg linearized targeting vector in 250µl PBS. Cell/plasmid suspension was transferred to a Biorad 0.4cm Genepulse cuvette and electroporated using the Genepulser XCell incl CE module (Biorad) with the following settings: 240V, 500µF, ∞ resistance. Cells were allowed to recover for 20 minutes at room temperature before plating them on 8x10cm dishes containing puromycin resistant feeder cells.

24 hour after transfection, cells were selected with 1µg/ml puromycin. 10 days after starting selection, individual colonies were picked, transferred to 48-well plates, dissociated, passaged and frozen as described in chapter 6.2.2.12. Rather than growing cells for DNA extraction on 48-well plates, cells were grown on 24-well plates. Once grown to confluency, cells were lysed in southern lysis buffer. Lysed cells were transferred to 1.5ml reaction tubes, and DNA was purified by addition of 250µl 5M NaCl, vigorous shaking and centrifugation at 21.000g for 5 minutes. The pellet was discarded and the DNA in the supernatant precipitated by addition of 615µl isopropanol and centrifugation at 21.000g for 15 minutes at 4°C. The DNA pellet was washed once in 70% ethanol, air dried for 2-3 minutes and dissolved in 50µl TE buffer. Positive clones were identified via Southern blot (see chapter 6.2.1.5) and thawed as described in chapter 6.2.2.12.

6.2.3 Microscopy and histology

6.2.3.1 Cryosectioning

Dissected placentas were fixed overnight in freshly prepared 4% paraformaldehyde in PBS. Following several washes for a few hours in PBS, tissues were infiltrated overnight in 15% sucrose and subsequently in 30% sucrose. Tissues were immersed in O.C.T. TissueTek embedding medium in a mold and frozen using a slurry of dry ice and isopropanol. Cryoconserved blocks were stored at -80°C.

Sections of 10 µm diameter were produced with a Leica CS3050S cryostat chilled tissue-dependently to -35°C to -18°C. Sections were taken up on Superfrost Plus glass slides (Thermo Fisher) and dried for 2 hours at room temperature before long-term storage at -80°C.

6.2.3.2 Microscopy

After *in situ* hybridization, images were taken using a Leica DM2500 upright microscope. Cell culture images were taken using a Leica DMIL LED, directly imaging living cells in cell culture vessels.

6.2.4 Bioinformatics and statistics

6.2.4.1 Genome coverage track visualization

RNA-Seq and ChIP-Seq bigWig files pre-processed by Emil Karaulanov were visualized using the UCSC genome browser (http://genome.ucsc.edu/). Images were further processed using Photoshop CS5 (Adobe) to separate track visualizations from background and align multiple tracks.

6.2.4.2 Venn diagrams

Venn diagram overlaps were calculated using Venny 2.0 (http://bioinfogp.cnb.csic.es/tools/venny/) and images were further processed using Photoshop CS5 (Adobe). Overlaps between different datasets were analyzed using RefSeq or Ensemble identifiers if available, and gene symbols if not.

6.2.4.3 GO enrichment analysis

Gene ontology (GO) enrichment analysis was carried out using GOrilla (http://cbl-gorilla.cs.technion.ac.il/), using the parameter "Two unranked lists of genes" and the whole mouse transcriptome as a background list. For the generation of lists of selected GO enrichments, partly redundant GO terms (i.e. 'regulation of nervous system development' and 'negative regulation of nervous system development') were manually filtered to only contain the more general term (i.e. 'regulation of nervous system development').

6.2.4.4 Student's t-test and Chi-squared test

Student's t-tests were carried out using the TTEST() function in Excel. Chi-squared tests with Yates' correction were carried out using GraphPad (http://graphpad.com/quickcalcs/contingency1.cfm).

6.2.4.5 Expression pattern analysis

Expression values in different tissues according to GenBank/dbEST were queried using AceView (http://www.ncbi.nlm.nih.gov/ieb/research/acembly/).

6.2.5 Techniques carried out by collaborators

6.2.5.1 Mass spectrometry (Michael Musheev)

DNA was degraded to nucleosides with nuclease P1 (Roche), snake venom phosphodiesterase (Worthington) and alkaline phosphatase (Fermentas). An equal volume of isotopic standard mixture (15N3-C (Silantes), 2H3-5mC (TRC) and self-synthesized 15N3-5hmC, 15N3-5fC and 15N3-5caC) was added to the DNA and about 100 ng of total DNA was injected for LC-MS/MS analysis. Quantitative analysis was performed on an Agilent 1290 Infinity Binary LC system (Agilent technologies) using a ReproSil 100 C18 column (Jasco) coupled to an Agilent 6490 triple quadrupole mass spectrometer

(Agilent technologies). Running buffers were 5 mM ammonium acetate pH 6.9 (A) and Acetonitrile (B). Separations were performed at a flow rate of 0.5 ml/min using the following gradient: 0% of solvent B from 0 min to 8 min, linear increase to 15% solvent B for the next 16 min. Washing and reconditioning of the column was performed with a flow rate of 1.0 ml/min with 15% solvent B for one minute and 100% buffer A for additional 5 min. During the last minute the flow rate was linearly decreased to the initial value of 0.5 ml/min. Quantification of highly abundant C and 5mC was performed using 100x diluted samples. The data were analyzed with the Agilent MassHunter Quantitative Analysis software version 26 B.05.02 (Agilent technologies) using isotopic standards to confirm the peak identity. Areas of the integrated peaks were exported into Microsoft Excel with which the areas were normalized to the area of the corresponding isotopic standard. Absolute values for the nucleosides were calculated using linear interpolation from a standard curve. Linear interpolation was performed using the two closely matching data points from the standard curve. Standards were spiked into the mixture of isotopic standards to normalize for ionization variability. The standard curve for every nucleoside was prepared to cover the amount of the corresponding nucleoside in the DNA sample analyzed. The linearity of standard curves over each region was monitored after every run and confirmed to be between 1-0.996 (R2-values) within a concentration range of at least two orders of magnitude. The technical standard deviation was < 7%. Standard curves were newly prepared with every new dataset (As described in Schomacher et al., Nature Struct. Mol. Biol., 2015, in submission).

6.2.5.2 Teratoma formation, fixation and staining (EPO Berlin GmbH)

Three control and three Gadd45 TKO mESC lines were sent cryo-conserved to EPO Berlin GmbH, where they were thawed and passaged twice before transplantation. Cells were resuspended in PBS, mixed with Matrigel and transplanted into the flanks of three NSG mice per mESC clone. Tumor weight and size were measured twice per week. Animals injected with the same mESC clone were sacrificed once the average tumor size reach approximately 1.0 cm^3. Tumors were excised, weighed and cut apart. $1/3^{rd}$ was shock frozen and $2/3^{rds}$ were fixed in formalin. Formalin fixed tumors were paraffin embedded, sectioned, stained with hematoxylin and eosin and evaluated via microscopy by Dr. med. vet. Wolfram Haider.

6.2.5.3 ChIP-Seq peak calling (Emil Karaulanov)

FastQ files were demultiplexed and quality was controlled using FastQC. Reads were mapped to the mm9 genome build using Bowtie, using the following parameters: '-n 3 -e 120 -l 51 -m 1 --best –strata'. Preliminary, at least 10-fold enriched peaks were called using MACS, comparing each ChIP and MockIP sample side by side (e.g. Streptavidin-ChIP in Gadd45-1Flag-BAP mESCs versus Streptavidin-ChIP in parental mESCs). Peaks were then further filtered by extracting read coverage in SeqMonk and only selecting peaks that show at least 2x higher coverage in the ChIP than in any MockIP (e.g. only peaks that

show a 2x higher coverage in the Streptavin-ChIP in Gadd45-1Flag-BAP mESCs than in the Steptavin-ChIP in parental mESCs, parental monolayer differentiated cells and parental retinoic acid treated EBs).

6.2.5.4 Primary RNA-Seq analysis (Emil Karaulanov)

FastQ files were merged per sample and quality was controlled using FastQC. Reads were mapped to the mm9 genome build using TopHat with allowing up to 5 multi-hits, using the iGenomes (Ensembl) GTF annotation. Strandness and duplicate rates were deemed acceptable. Reads per gene were counted using HTSeq-count with the parameter '-s reverse' and multi-hits were discarded. Pairwise differentially expressed gene analysis was carried out using edgeR. Browser tracks were generated by converting BAM to BedGraph files and subsequently to reads-per-million-bps-normalized BigWig files. Absolute RPKM values were calculated using the edgeR module rpkm().

6.2.5.5 Bisulfite sequencing methylation analysis (Medhavi Mallick)

The FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to perform quality control checks on raw sequenced data. For data analysis, a pipeline combining TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and Bismark (http://www.bioinformatics.babraham.ac.uk/projects/bismark/), which is an integrated tool from mapping to methylation analysis, was used to identify methylation level at single base resolution. In order to utilize reads with the best quality, i.e. phred quality score above 30, sequences were trimmed using the TrimGalore tool. Bisulfite treated DNA reads were aligned over the reference amplicons using the mapping algorithm of Bismark. To improve sensitivity, only uniquely aligned reads were used in downstream analysis. These steps were followed by estimation of the methylation levels for every CpG in the amplicons using the bismark_methylation_extractor and bedGraph2cytosine algorithms from the Bismark tool.

6.2.5.6 Isolation of mouse embryonic fibroblasts (Bernadette Mekker)

Timed mouse matings were set up overnight for embryo dissections and isolation of mouse embryonic fibroblasts. In case of successful detection of vaginal plugs in the females, the previous midnight was defined as the time of conception. Pregnant females were sacrificed 15.5 days post coitum. Uteri were isolated, disinfected in 70% ethanol and kept in ice-cold PBS during the dissection procedure. Embryos were liberated from the uteri, and their placenta, amniotic sac, head, limbs and inner organs were removed. A genotyping sample was taken from the discarded tissues. The remaining embryo carcass was minced with a scalpel and transferred to a Falcon tube containing sterile glass beads and 2 ml 0.25% Trypsin and 100U DNaseI. After 30 minutes of incubation, trypsinization was stopped by adding 5 ml DMEM+++ with 20% FBS and transferring the cell suspension to a 10 cm cell culture dish containing 15

DMEM+++. Cells were incubated from here on at 37℃, 5% CO2, 5% O2. MEFs were passaged to 4x 10 cm plates after 24h, and frozen to 8 cryovials after additional 48h.

7 References

1. Saiz, N. & Plusa, B. Early cell fate decisions in the mouse embryo. *Reprod. Camb. Engl.* **145,** R65–80 (2013).
2. Kelly, S. J. Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. *J. Exp. Zool.* **200,** 365–376 (1977).
3. Suwińska, A., Czołowska, R., Ozdzeński, W. & Tarkowski, A. K. Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. *Dev. Biol.* **322,** 133–144 (2008).
4. Ducibella, T. & Anderson, E. Cell shape and membrane changes in the eight-cell mouse embryo: prerequisites for morphogenesis of the blastocyst. *Dev. Biol.* **47,** 45–58 (1975).
5. Ziomek, C. A. & Johnson, M. H. Cell surface interaction induces polarization of mouse 8-cell blastomeres at compaction. *Cell* **21,** 935–942 (1980).
6. Tarkowski, A. K. & Wróblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *J. Embryol. Exp. Morphol.* **18,** 155–180 (1967).
7. Fleming, T. P. A quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse blastocyst. *Dev. Biol.* **119,** 520–531 (1987).
8. Smith, R. & McLaren, A. Factors affecting the time of formation of the mouse blastocoele. *J. Embryol. Exp. Morphol.* **41,** 79–92 (1977).
9. Rossant, J. Investigation of the determinative state of the mouse inner cell mass. II. The fate of isolated inner cell masses transferred to the oviduct. *J. Embryol. Exp. Morphol.* **33,** 991–1001 (1975).
10. Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev. Cell* **10,** 615–624 (2006).
11. Morris, S. A. *et al.* Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 6364–6369 (2010).
12. Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Dev. Camb. Engl.* **137,** 715–724 (2010).
13. Rosner, M. H. *et al.* A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345,** 686–692 (1990).
14. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95,** 379–391 (1998).
15. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17,** 126–140 (2003).
16. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631–642 (2003).
17. Yagi, R. *et al.* Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Dev. Camb. Engl.* **134,** 3827–3836 (2007).
18. Ralston, A. *et al.* Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Dev. Camb. Engl.* **137,** 395–403 (2010).
19. Strumpf, D. *et al.* Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Dev. Camb. Engl.* **132,** 2093–2102 (2005).
20. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123,** 917–929 (2005).

21. Dietrich, J.-E. & Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. *Dev. Camb. Engl.* **134,** 4219–4231 (2007).
22. Goldin, S. N. & Papaioannou, V. E. Paracrine action of FGF4 during periimplantation development maintains trophectoderm and primitive endoderm. *Genes. N. Y. N 2000* **36,** 40–47 (2003).
23. Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282,** 2072–2075 (1998).
24. Tam, P. P. L. & Loebel, D. A. F. Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8,** 368–381 (2007).
25. Bénazéraf, B. & Pourquié, O. Formation and segmentation of the vertebrate body axis. *Annu. Rev. Cell Dev. Biol.* **29,** 1–26 (2013).
26. Achen, M. G., Gad, J. M., Stacker, S. A. & Wilks, A. F. Placenta growth factor and vascular endothelial growth factor are co-expressed during early embryonic development. *Growth Factors Chur Switz.* **15,** 69–80 (1997).
27. Soares, M. J. *et al.* Differentiation of trophoblast endocrine cells. *Placenta* **17,** 277–289 (1996).
28. Arck, P. C. & Hecher, K. Fetomaternal immune cross-talk and its consequences for maternal and offspring's health. *Nat. Med.* **19,** 548–556 (2013).
29. Rossant, J. & Cross, J. C. Placental development: lessons from mouse mutants. *Nat. Rev. Genet.* **2,** 538–548 (2001).
30. Sherman, M. I., McLaren, A. & Walker, P. M. Mechanism of accumulation of DNA in giant cells of mouse trophoblast. *Nature. New Biol.* **238,** 175–176 (1972).
31. Nadra, K. *et al.* Differentiation of Trophoblast Giant Cells and Their Metabolic Functions Are Dependent on Peroxisome Proliferator-Activated Receptor β/δ. *Mol. Cell. Biol.* **26,** 3266–3281 (2006).
32. Gardner, R. L., Papaioannou, V. E. & Barton, S. C. Origin of the ectoplacental cone and secondary giant cells in mouse blastocysts reconstituted from isolated trophoblast and inner cell mass. *J. Embryol. Exp. Morphol.* **30,** 561–572 (1973).
33. Gong, R. *et al.* Syncytin-A mediates the formation of syncytiotrophoblast involved in mouse placental development. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* **20,** 517–526 (2007).
34. Simmons, D. G. & Cross, J. C. Determinants of trophoblast lineage and cell subtype specification in the mouse placenta. *Dev. Biol.* **284,** 12–24 (2005).
35. Rappolee, D. A., Basilico, C., Patel, Y. & Werb, Z. Expression and function of FGF-4 in peri-implantation development in mouse embryos. *Dev. Camb. Engl.* **120,** 2259–2269 (1994).
36. Haffner-Krausz, R., Gorivodsky, M., Chen, Y. & Lonai, P. Expression of Fgfr2 in the early mouse embryo indicates its involvement in preimplantation development. *Mech. Dev.* **85,** 167–172 (1999).
37. Saba-El-Leil, M. K. *et al.* An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. *EMBO Rep.* **4,** 964–968 (2003).
38. Guzman-Ayala, M., Ben-Haim, N., Beck, S. & Constam, D. B. Nodal protein processing and fibroblast growth factor 4 synergize to maintain a trophoblast stem cell microenvironment. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15656–15660 (2004).
39. Russ, A. P. *et al.* Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404,** 95–99 (2000).
40. Werling, U. & Schorle, H. Transcription factor gene AP-2 gamma essential for early murine development. *Mol. Cell. Biol.* **22,** 3149–3156 (2002).

41. Donnison, M. *et al.* Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. *Dev. Camb. Engl.* **132,** 2299–2308 (2005).
42. Uy, G. D., Downs, K. M. & Gardner, R. L. Inhibition of trophoblast stem cell potential in chorionic ectoderm coincides with occlusion of the ectoplacental cavity in the mouse. *Dev. Camb. Engl.* **129,** 3913–3924 (2002).
43. Hu, D. & Cross, J. C. Ablation of Tpbpa-positive trophoblast precursors leads to defects in maternal spiral artery remodeling in the mouse placenta. *Dev. Biol.* **358,** 231–239 (2011).
44. Hughes, M. *et al.* The Hand1, Stra13 and Gcm1 transcription factors override FGF signaling to promote terminal differentiation of trophoblast stem cells. *Dev. Biol.* **271,** 26–37 (2004).
45. Carney, E. W., Prideaux, V., Lye, S. J. & Rossant, J. Progressive expression of trophoblast-specific genes during formation of mouse trophoblast giant cells in vitro. *Mol. Reprod. Dev.* **34,** 357–368 (1993).
46. Simmons, D. G., Fortier, A. L. & Cross, J. C. Diverse subtypes and developmental origins of trophoblast giant cells in the mouse placenta. *Dev. Biol.* **304,** 567–578 (2007).
47. Tamai, Y. *et al.* Cytokeratins 8 and 19 in the mouse placental development. *J. Cell Biol.* **151,** 563–572 (2000).
48. Shi, W. *et al.* Choroideremia gene product affects trophoblast development and vascularization in mouse extra-embryonic tissues. *Dev. Biol.* **272,** 53–65 (2004).
49. Guillemot, F., Nagy, A., Auerbach, A., Rossant, J. & Joyner, A. L. Essential role of Mash-2 in extraembryonic development. *Nature* **371,** 333–336 (1994).
50. Basyuk, E. *et al.* Murine Gcm1 gene is expressed in a subset of placental trophoblast cells. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **214,** 303–311 (1999).
51. Li, Y. & Behringer, R. R. Esx1 is an X-chromosome-imprinted regulator of placental development and fetal growth. *Nat. Genet.* **20,** 309–311 (1998).
52. Morasso, M. I., Grinberg, A., Robinson, G., Sargent, T. D. & Mahon, K. A. Placental failure in mice lacking the homeobox gene Dlx3. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 162–167 (1999).
53. Luo, J. *et al.* Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature* **388,** 778–782 (1997).
54. Anson-Cartwright, L. *et al.* The glial cells missing-1 protein is essential for branching morphogenesis in the chorioallantoic placenta. *Nat. Genet.* **25,** 311–314 (2000).
55. Stecca, B. *et al.* Gcm1 expression defines three stages of chorio-allantoic interaction during placental development. *Mech. Dev.* **115,** 27–34 (2002).
56. Wu, L. *et al.* Extra-embryonic function of Rb is essential for embryonic development and viability. *Nature* **421,** 942–947 (2003).
57. Jackson, M. *et al.* Expression of a novel homeobox gene Ehox in trophoblast stem cells and pharyngeal pouch endoderm. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **228,** 740–744 (2003).
58. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 7634–7638 (1981).
59. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154–156 (1981).
60. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309,** 255–256 (1984).
61. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* **323,** 445–448 (1986).

62. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
63. Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38,** 431–440 (2006).
64. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134,** 521–533 (2008).
65. Young, R. A. Control of the embryonic stem cell state. *Cell* **144,** 940–954 (2011).
66. Zhao, S., Nichols, J., Smith, A. G. & Li, M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Mol. Cell. Neurosci.* **27,** 332–342 (2004).
67. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* **16,** 516–528 (2014).
68. Martello, G. & Smith, A. The nature of embryonic stem cells. *Annu. Rev. Cell Dev. Biol.* **30,** 647–675 (2014).
69. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460,** 118–122 (2009).
70. Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115,** 281–292 (2003).
71. Wiles, M. V. & Johansson, B. M. Embryonic stem cell development in a chemically defined medium. *Exp. Cell Res.* **247,** 241–248 (1999).
72. Kunath, T. *et al.* FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Dev. Camb. Engl.* **134,** 2895–2902 (2007).
73. Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21,** 183–186 (2003).
74. Nichols, J., Chambers, I., Taga, T. & Smith, A. Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Dev. Camb. Engl.* **128,** 2333–2339 (2001).
75. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
76. Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. *Biochem. Soc. Trans.* **38,** 1027–1032 (2010).
77. Shy, B. R. *et al.* Regulation of Tcf7l1 DNA binding and protein stability as principal mechanisms of Wnt/β-catenin signaling. *Cell Rep.* **4,** 1–9 (2013).
78. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149,** 590–604 (2012).
79. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Dev. Camb. Engl.* **135,** 909–918 (2008).
80. Hayashi, K., Lopes, S. M. C. de S., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3,** 391–401 (2008).
81. Ficz, G. *et al.* FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13,** 351–359 (2013).

82. Leitch, H. G. *et al*. Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20,** 311–316 (2013).
83. Habibi, E. *et al*. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13,** 360–369 (2013).
84. Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev.* **19,** 1129–1155 (2005).
85. Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J. Embryol. Exp. Morphol.* **87,** 27–45 (1985).
86. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472,** 51–56 (2011).
87. Desbaillets, I., Ziegler, U., Groscurth, P. & Gassmann, M. Embryoid bodies: an in vitro model of mouse embryogenesis. *Exp. Physiol.* **85,** 645–651 (2000).
88. Maltsev, V. A., Wobus, A. M., Rohwedel, J., Bader, M. & Hescheler, J. Cardiomyocytes differentiated in vitro from embryonic stem cells developmentally express cardiac-specific genes and ionic currents. *Circ. Res.* **75,** 233–244 (1994).
89. Dani, C. *et al.* Differentiation of embryonic stem cells into adipocytes in vitro. *J. Cell Sci.* **110 (Pt 11),** 1279–1285 (1997).
90. Wiles, M. V. & Keller, G. Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture. *Dev. Camb. Engl.* **111,** 259–267 (1991).
91. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.-A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.* **2,** 1034–1043 (2007).
92. Boheler, K. R. *et al.* Differentiation of pluripotent embryonic stem cells into cardiomyocytes. *Circ. Res.* **91,** 189–201 (2002).
93. Hescheler, J. *et al.* Embryonic stem cells: a model to study structural and functional properties in cardiomyogenesis. *Cardiovasc. Res.* **36,** 149–162 (1997).
94. Miles, G. B. *et al.* Functional properties of motoneurons derived from mouse embryonic stem cells. *J. Neurosci. Off. J. Soc. Neurosci.* **24,** 7848–7858 (2004).
95. Latos, P. A. & Hemberger, M. Review: the transcriptional and signalling networks of mouse trophoblast stem cells. *Placenta* **35 Suppl,** S81–85 (2014).
96. Kidder, B. L. & Palmer, S. Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. *Genome Res.* **20,** 458–472 (2010).
97. Chuong, E. B., Rumi, M. A. K., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45,** 325–329 (2013).
98. Rossant, J. Stem cells and early lineage development. *Cell* **132,** 527–531 (2008).
99. Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Dev. Camb. Engl.* **105,** 733–737 (1989).
100. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448,** 191–195 (2007).
101. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
102. Ng, R. K. *et al*. Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat. Cell Biol.* **10,** 1280–1290 (2008).

103. Tolkunova, E. *et al.* The caudal-related protein cdx2 promotes trophoblast differentiation of mouse embryonic stem cells. *Stem Cells Dayt. Ohio* **24,** 139–144 (2006).
104. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24,** 372–376 (2000).
105. Schenke-Layland, K. *et al.* Collagen IV induces trophoectoderm differentiation of mouse embryonic stem cells. *Stem Cells Dayt. Ohio* **25,** 1529–1538 (2007).
106. He, S., Pant, D., Schiffmacher, A., Meece, A. & Keefer, C. L. Lymphoid enhancer factor 1-mediated Wnt signaling promotes the initiation of trophoblast lineage differentiation in mouse embryonic stem cells. *Stem Cells Dayt. Ohio* **26,** 842–849 (2008).
107. Hayashi, Y. *et al*. BMP4 induction of trophoblast from mouse embryonic stem cells in defined culture conditions on laminin. *In Vitro Cell. Dev. Biol. Anim.* **46,** 416–430 (2010).
108. Bernardo, A. S. *et al*. BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* **9,** 144–155 (2011).
109. Sinsheimer, R. L. The action of pancreatic deoxyribonuclease. II. Isomeric dinucleotides. *J. Biol. Chem.* **215,** 579–583 (1955).
110. Cedar, H., Solage, A., Glaser, G. & Razin, A. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic Acids Res.* **6,** 2125–2132 (1979).
111. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–257 (1999).
112. Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317,** 1760–1764 (2007).
113. Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71,** 865–873 (1992).
114. Boyes, J. & Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64,** 1123–1134 (1991).
115. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18,** 6538–6547 (1998).
116. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,** 386–389 (1998).
117. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19,** 187–191 (1998).
118. Watt, F. & Molloy, P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* **2,** 1136–1143 (1988).
119. Chodavarapu, R. K. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466,** 388–392 (2010).
120. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).
121. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).
122. Gopalakrishnan, S., Sullivan, B. A., Trazzi, S., Della Valle, G. & Robertson, K. D. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. *Hum. Mol. Genet.* **18,** 3178–3193 (2009).

123. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20,** 116–117 (1998).
124. Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156,** 45–68 (2014).
125. Yang, X. *et al*. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26,** 577–590 (2014).
126. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).
127. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–1022 (2011).
128. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448,** 714–717 (2007).
129. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456,** 125–129 (2008).
130. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).
131. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).
132. Ooi, S. K. T. & Bestor, T. H. The Colorful History of Active DNA Demethylation. *Cell* **133,** 1145–1148 (2008).
133. Tahiliani, M. *et al*. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
134. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).
135. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–1307 (2011).
136. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).
137. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40,** 4841–4849 (2012).
138. Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286,** 35334–35338 (2011).
139. Spruijt, C. G. *et al*. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).
140. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502,** 472–479 (2013).
141. Bhattacharya, S. K., Ramchandani, S., Cervoni, N. & Szyf, M. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* **397,** 579–583 (1999).
142. Zhu, B. *et al.* 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res.* **28,** 4157–4165 (2000).
143. Zhu, B. *et al.* 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5135–5139 (2000).

144. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279,** 52353–52360 (2004).
145. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463,** 1042–1047 (2010).
146. Barreto, G. *et al.* Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445,** 671–675 (2007).
147. Schmitz, K.-M. *et al.* TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Mol. Cell* **33,** 344–353 (2009).
148. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* **43,** 811–814 (2011).
149. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **8,** e1002440 (2012).
150. Gu, T.-P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477,** 606–610 (2011).
151. Carlson, L. L., Page, A. W. & Bestor, T. H. Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes Dev.* **6,** 2536–2541 (1992).
152. Birger, Y., Shemer, R., Perk, J. & Razin, A. The imprinting box of the mouse Igf2r gene. *Nature* **397,** 84–88 (1999).
153. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genes. N. Y. N 2000* **35,** 88–93 (2003).
154. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20,** 320–331 (2010).
155. Krebs, A. R., Dessus-Babus, S., Burger, L. & Schübeler, D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife* **3,** e04094 (2014).
156. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* **241,** 172–182 (2002).
157. Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol.* **8,** 188–194 (2006).
158. Ko, M. *et al.* Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 14566–14571 (2011).
159. Ma, D. K. *et al.* Neuronal Activity–Induced Gadd45b Promotes Epigenetic DNA Demethylation and Adult Neurogenesis. *Science* **323,** 1074–1077 (2009).
160. Yu, H. *et al.* Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. *Nat. Neurosci.* **18,** 836–843 (2015).
161. Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* **278,** 440–458 (2005).
162. Hackett, J. A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339,** 448–452 (2013).

163. Kurimoto, K. *et al.* Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* **22,** 1617–1635 (2008).
164. Kato, Y. *et al.* Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* **16,** 2272–2280 (2007).
165. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells Devoted Mol. Cell. Mech.* **11,** 805–814 (2006).
166. Schmidt, C. S. *et al.* Global DNA Hypomethylation Prevents Consolidation of Differentiation Programs and Allows Reversion to the Embryonic Stem Cell State. *PLoS ONE* **7,** e52629 (2012).
167. Panning, B. & Jaenisch, R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev.* **10,** 1991–2002 (1996).
168. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
169. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat. Biotechnol.* **26,** 916–924 (2008).
170. Cambuli, F. *et al.* Epigenetic memory of the first cell fate decision prevents complete ES cell reprogramming into trophoblast. *Nat. Commun.* **5,** 5538 (2014).
171. Nakanishi, M. O. *et al.* Trophoblast-specific DNA methylation occurs after the segregation of the trophectoderm and inner cell mass in the mouse periimplantation embryo. *Epigenetics* **7,** 173–182 (2012).
172. Zhan, Q. *et al.* The gadd and MyD genes define a novel set of mammalian genes encoding acidic proteins that synergistically suppress cell growth. *Mol. Cell. Biol.* **14,** 2361–2371 (1994).
173. Liebermann, D. A. & Hoffman, B. Gadd45 in stress signaling. *J. Mol. Signal.* **3,** 15 (2008).
174. Peretz, G., Bakhrat, A. & Abdu, U. Expression of the Drosophila melanogaster GADD45 homolog (CG11086) affects egg asymmetric development that is mediated by the c-Jun N-terminal kinase pathway. *Genetics* **177,** 1691–1702 (2007).
175. Takekawa, M. & Saito, H. A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. *Cell* **95,** 521–530 (1998).
176. Bulavin, D. V., Kovalsky, O., Hollander, M. C. & Fornace, A. J. Loss of oncogenic H-ras-induced cell cycle arrest and p38 mitogen-activated protein kinase activation by disruption of Gadd45a. *Mol. Cell. Biol.* **23,** 3859–3871 (2003).
177. Thalheimer, F. B. *et al.* Cytokine-regulated GADD45G induces differentiation and lineage selection in hematopoietic stem cells. *Stem Cell Rep.* **3,** 34–43 (2014).
178. Vairapandi, M., Balliet, A. G., Hoffman, B. & Liebermann, D. A. GADD45b and GADD45g are cdc2/cyclinB1 kinase inhibitors with a role in S and G2/M cell cycle checkpoints induced by genotoxic stress. *J. Cell. Physiol.* **192,** 327–338 (2002).
179. Azam, N., Vairapandi, M., Zhang, W., Hoffman, B. & Liebermann, D. A. Interaction of CR6 (GADD45gamma) with proliferating cell nuclear antigen impedes negative growth control. *J. Biol. Chem.* **276,** 2766–2774 (2001).
180. Vairapandi, M., Balliet, A. G., Fornace, A. J., Hoffman, B. & Liebermann, D. A. The differentiation primary response gene MyD118, related to GADD45, encodes for a nuclear protein which interacts with PCNA and p21WAF1/CIP1. *Oncogene* **12,** 2579–2594 (1996).
181. Jung, H. J. *et al.* Base excision DNA repair defect in Gadd45a-deficient cells. *Oncogene* **26,** 7517–7525 (2007).

182. Smith, M. L. *et al.* Interaction of the p53-regulated protein Gadd45 with proliferating cell nuclear antigen. *Science* **266,** 1376–1380 (1994).
183. Carrier, F. *et al.* Gadd45, a p53-responsive stress protein, modifies DNA accessibility on damaged chromatin. *Mol. Cell. Biol.* **19,** 1673–1685 (1999).
184. Gupta, M. *et al.* Hematopoietic cells from Gadd45a- and Gadd45b-deficient mice are sensitized to genotoxic-stress-induced apoptosis. *Oncogene* **24,** 7170–7179 (2005).
185. Smith, M. L. *et al.* Antisense GADD45 expression results in decreased DNA repair and sensitizes cells to u.v.-irradiation or cisplatin. *Oncogene* **13,** 2255–2263 (1996).
186. Ma, D. K., Guo, J. U., Ming, G. & Song, H. DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle Georget. Tex* **8,** 1526–1531 (2009).
187. Kaufmann, L. T., Gierl, M. S. & Niehrs, C. Gadd45a, Gadd45b and Gadd45g expression during mouse embryonic development. *Gene Expr. Patterns GEP* **11,** 465–470 (2011).
188. Matsunaga, E., Nambu, S., Oka, M. & Iriki, A. Comparative analysis of developmentally regulated expressions of Gadd45a, Gadd45b, and Gadd45g in the mouse and marmoset cerebral cortex. *Neuroscience* **284,** 566–580 (2015).
189. Kaufmann, L. T. & Niehrs, C. Gadd45a and Gadd45g regulate neural development and exit from pluripotency in Xenopus. *Mech. Dev.* **128,** 401–411 (2011).
190. Hollander, M. C. *et al.* Genomic instability in Gadd45a-deficient mice. *Nat. Genet.* **23,** 176–184 (1999).
191. Hildesheim, J. *et al.* Gadd45a protects against UV irradiation-induced skin tumors, and promotes apoptosis and stress signaling via MAPK and p53. *Cancer Res.* **62,** 7305–7315 (2002).
192. Liu, L. *et al.* Gadd45 beta and Gadd45 gamma are critical for regulating autoimmunity. *J. Exp. Med.* **202,** 1341–1347 (2005).
193. Lu, B. *et al.* GADD45gamma mediates the activation of the p38 and JNK MAP kinase pathways and cytokine production in effector TH1 cells. *Immunity* **14,** 583–590 (2001).
194. Ijiri, K. *et al.* A novel role for GADD45beta as a mediator of MMP-13 gene expression during chondrocyte terminal differentiation. *J. Biol. Chem.* **280,** 38544–38555 (2005).
195. Leach, P. T. *et al.* Gadd45b knockout mice exhibit selective deficits in hippocampus-dependent long-term memory. *Learn. Mem. Cold Spring Harb. N* **19,** 319–324 (2012).
196. Sultan, F. A., Wang, J., Tront, J., Liebermann, D. A. & Sweatt, J. D. Genetic deletion of Gadd45b, a regulator of active DNA demethylation, enhances long-term memory and synaptic plasticity. *J. Neurosci. Off. J. Soc. Neurosci.* **32,** 17059–17066 (2012).
197. Gierl, M. S., Gruhn, W. H., von Seggern, A., Maltry, N. & Niehrs, C. GADD45G functions in male sex determination by promoting p38 signaling and Sry expression. *Dev. Cell* **23,** 1032–1042 (2012).
198. Cai, Q. *et al.* Effects of expression of p53 and Gadd45 on osmotic tolerance of renal inner medullary cells. *Am. J. Physiol. Renal Physiol.* **291,** F341–349 (2006).
199. Wang, X. *et al.* Genetic interactions between Brca1 and Gadd45a in centrosome duplication, genetic stability, and neural tube closure. *J. Biol. Chem.* **279,** 29606–29614 (2004).
200. Tront, J. S., Hoffman, B. & Liebermann, D. A. Gadd45a suppresses Ras-driven mammary tumorigenesis by activation of c-Jun NH2-terminal kinase and p38 stress signaling resulting in apoptosis and senescence. *Cancer Res.* **66,** 8448–8454 (2006).
201. Patterson, A. D., Hildesheim, J., Fornace, A. J., Jr & Hollander, M. C. Neural tube development requires the cooperation of p53- and Gadd45a-associated pathways. *Birt. Defects Res. A. Clin. Mol. Teratol.* **76,** 129–132 (2006).

202. Hollander, M. C. *et al.* Deletion of XPC leads to lung tumors in mice and is associated with early events in human lung carcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 13200–13205 (2005).
203. Gupta, M. *et al.* Hematopoietic cells from Gadd45a- and Gadd45b-deficient mice are sensitized to genotoxic-stress-induced apoptosis. *Oncogene* **24,** 7170–7179 (2005).
204. Gupta, S. K., Gupta, M., Hoffman, B. & Liebermann, D. A. Hematopoietic cells from gadd45a-deficient and gadd45b-deficient mice exhibit impaired stress responses to acute stimulation with cytokines, myeloablation and inflammation. *Oncogene* **25,** 5537–5546 (2006).
205. Ijiri, K. *et al.* A novel role for GADD45beta as a mediator of MMP-13 gene expression during chondrocyte terminal differentiation. *J. Biol. Chem.* **280,** 38544–38555 (2005).
206. Salvador, J. M. *et al.* Mice lacking the p53-effector gene Gadd45a develop a lupus-like syndrome. *Immunity* **16,** 499–508 (2002).
207. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135,** 1201–1212 (2008).
208. Kienhöfer, S. *et al.* GADD45a physically and functionally interacts with TET1. *Differ. Res. Biol. Divers.* (2015). doi:10.1016/j.diff.2015.10.003
209. Arab, K. *et al.* Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol. Cell* **55,** 604–614 (2014).
210. Li, Z. *et al.* Gadd45a promotes DNA demethylation through TDG. *Nucleic Acids Res.* **43,** 3986–3997 (2015).
211. Lee, B., Morano, A., Porcellini, A. & Muller, M. T. GADD45α inhibition of DNMT1 dependent DNA methylation during homology directed DNA repair. *Nucleic Acids Res.* **40,** 2481–2493 (2012).
212. Sytnikova, Y. A., Kubarenko, A. V., Schäfer, A., Weber, A. N. R. & Niehrs, C. Gadd45a is an RNA binding protein and is localized in nuclear speckles. *PloS One* **6,** e14500 (2011).
213. Schäfer, A., Karaulanov, E., Stapf, U., Döderlein, G. & Niehrs, C. Ing1 functions in DNA demethylation by directing Gadd45a to H3K4me3. *Genes Dev.* **27,** 261–273 (2013).
214. Le May, N. *et al.* NER Factors Are Recruited to Active Promoters and Facilitate Chromatin Modification for Transcription in the Absence of Exogenous Genotoxic Attack. *Mol. Cell* **38,** 54–66 (2010).
215. Li, Y. *et al.* Overexpression of the growth arrest and DNA damage-induced 45alpha gene contributes to autoimmunity by promoting DNA demethylation in lupus T cells. *Arthritis Rheum.* **62,** 1438–1447 (2010).
216. Zhang, R., Shao, J. & Xiang, L. GADD45A protein plays an essential role in active DNA demethylation during terminal osteogenic differentiation of adipose-derived mesenchymal stem cells. *J. Biol. Chem.* **286,** 41083–41094 (2011).
217. Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L. & Khavari, P. A. DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **463,** 563–567 (2010).
218. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
219. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500,** 222–226 (2013).
220. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24,** 132–141 (2014).
221. Karimi, M. M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements and chimaeric transcripts in mouse ES cells. *Cell Stem Cell* **8,** (2011).

222. Huang, Y. *et al*. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 1361–1366 (2014).
223. Liebermann, D. A. *et al*. Gadd45 Stress Sensors in Malignancy and Leukemia. *Crit. Rev. Oncog.* **16,** 129–140 (2011).
224. Chen, Y., Yang, R., Guo, P. & Ju, Z. Gadd45a deletion aggravates hematopoietic stem cell dysfunction in ATM-deficient mice. *Protein Cell* **5,** 80–89 (2014).
225. Bibel, M. *et al.* Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat. Neurosci.* **7,** 1003–1009 (2004).
226. Dawlaty, M. M. *et al.* Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* **29,** 102–111 (2014).
227. Götz, M. & Huttner, W. B. The cell biology of neurogenesis. *Nat. Rev. Mol. Cell Biol.* **6,** 777–788 (2005).
228. Gohlke, J. M. *et al.* Characterization of the proneural gene regulatory network during mouse telencephalon development. *BMC Biol.* **6,** 15 (2008).
229. Kawaue, T. *et al.* Neurogenin2-d4Venus and Gadd45g-d4Venus transgenic mice: visualizing mitotic and migratory behaviors of cells committed to the neuronal lineage in the developing mammalian brain. *Dev. Growth Differ.* **56,** 293–304 (2014).
230. Huang, H. S. *et al.* Direct transcriptional induction of Gadd45gamma by Ascl1 during neuronal differentiation. *Mol. Cell. Neurosci.* **44,** 282–296 (2010).
231. Song, C.-X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153,** 678–691 (2013).
232. Silverman, G. A. *et al.* The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J. Biol. Chem.* **276,** 33293–33296 (2001).
233. Ha, C. T. *et al.* Human Pregnancy Specific Beta-1-Glycoprotein 1 (PSG1) Has a Potential Role in Placental Vascular Morphogenesis. *Biol. Reprod.* **83,** 27–35 (2010).
234. Simmons, D. G., Rawn, S., Davies, A., Hughes, M. & Cross, J. C. Spatial and temporal expression of the 23 murine Prolactin/Placental Lactogen-related genes is not associated with their position in the locus. *BMC Genomics* **9,** 352 (2008).
235. Kubaczka, C. *et al.* Derivation and maintenance of murine trophoblast stem cells under defined conditions. *Stem Cell Rep.* **2,** 232–242 (2014).
236. Himeno, E., Tanaka, S. & Kunath, T. Isolation and manipulation of mouse trophoblast stem cells. *Curr. Protoc. Stem Cell Biol.* **Chapter 1,** Unit 1E.4 (2008).
237. Kim, J., Cantor, A. B., Orkin, S. H. & Wang, J. Use of in vivo biotinylation to study protein–protein and protein–DNA interactions in mouse embryonic stem cells. *Nat. Protoc.* **4,** 506–517 (2009).
238. Schäfer, A., Schomacher, L., Barreto, G., Döderlein, G. & Niehrs, C. Gemcitabine Functions Epigenetically by Inhibiting Repair Mediated DNA Demethylation. *PLoS ONE* **5,** e14060 (2010).
239. Brind'Amour, J. *et al.* An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* **6,** 6033 (2015).
240. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
241. Yamauchi, J. *et al*. Gadd45a, the gene induced by the mood stabilizer valproic acid, regulates neurite outgrowth through JNK and the substrate paxillin in N1E-115 neuroblastoma cells. *Exp. Cell Res.* **313,** 1886–1896 (2007).

242. Sarkisian, M. R. & Siebzehnrubl, D. Abnormal levels of Gadd45alpha in developing neocortex impair neurite outgrowth. *PloS One* **7,** e44207 (2012).
243. Kuhn, A. & Grummt, I. A novel promoter in the mouse rDNA spacer is active in vivo and in vitro. *EMBO J.* **6,** 3487–3492 (1987).
244. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–348 (2011).
245. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473,** 389–393 (2011).
246. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–1133 (2010).
247. Freudenberg, J. M. *et al.* Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res.* **40,** 3364–3377 (2012).
248. Dawlaty, M. M. *et al.* Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24,** 310–323 (2013).
249. Dawlaty, M. M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9,** 166–175 (2011).
250. Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8,** 200–213 (2011).
251. Cortázar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470,** 419–423 (2011).
252. Niehrs, C. & Schäfer, A. Active DNA demethylation by Gadd45 and DNA repair. *Trends Cell Biol.* **22,** 220–227 (2012).
253. Peters, J. The role of genomic imprinting in biology and disease: an expanding view. *Nat. Rev. Genet.* **15,** 517–530 (2014).
254. Latos, P. A. *et al.* An in vitro ES cell imprinting model shows that imprinted expression of the Igf2r gene arises from an allele-specific expression bias. *Development* **136,** 437–448 (2009).
255. Cinelli, P. *et al.* Expression profiling in transgenic FVB/N embryonic stem cells overexpressing STAT3. *BMC Dev. Biol.* **8,** 57 (2008).
256. Casanova, E. A. *et al.* Pramel7 mediates LIF/STAT3-dependent self-renewal in embryonic stem cells. *Stem Cells Dayt. Ohio* **29,** 474–485 (2011).
257. Smith, Z. D., Gu, H., Bock, C., Gnirke, A. & Meissner, A. High-throughput bisulfite sequencing in mammalian genomes. *Methods San Diego Calif* **48,** 226–232 (2009).
258. Aihara, Y. *et al.* Induction of neural crest cells from mouse embryonic stem cells in a serum-free monolayer culture. *Int. J. Dev. Biol.* **54,** 1287–1294 (2010).
259. Kawaguchi, J., Nichols, J., Gierl, M. S., Faial, T. & Smith, A. Isolation and propagation of enteric neural crest progenitor cells from mouse embryonic stem cells and embryos. *Dev. Camb. Engl.* **137,** 693–704 (2010).
260. Macfarlan, T. S. *et al.* ES cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).
261. Biase, F. H., Cao, X. & Zhong, S. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. *Genome Res.* **24,** 1787–1796 (2014).
262. Poirier, F. *et al.* The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo. *Development* **113,** 1105–1114 (1991).

263. Fujimori, H. *et al*. The H19 induction triggers trophoblast lineage commitment in mouse ES cells. *Biochem. Biophys. Res. Commun*. **436,** 313–318 (2013).
264. Hu, X. *et al.* Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14,** 512–522 (2014).
265. Gao, Y. *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12,** 453–469 (2013).
266. Teytelman, L. *et al.* Impact of chromatin structures on DNA processing for genomic analyses. *PloS One* **4,** e6700 (2009).
267. Ramachandran, P., Palidwor, G. A. & Perkins, T. J. BIDCHIPS: bias decomposition and removal from ChIP-seq data clarifies true binding signal and its functional correlates. *Epigenetics Chromatin* **8,** 33 (2015).
268. Dedon, P. C., Soults, J. A., Allis, C. D. & Gorovsky, M. A. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. *Anal. Biochem.* **197,** 83–90 (1991).
269. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).
270. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
271. Sambrook, J., Frisch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor, N.Y., 2001).
272. Livak, K. J. & Schmittgen, T. D. 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 (2001).

8 List of abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AID	Activation-induced deaminase
BAP	Biotinylation acceptor peptide
BER	Base excision repair
BMP	Bone morphogenetic protein
bp	Base pair
BSA	Bovine serum albumin
cDNA	Copy DNA
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation followed by massive parallel sequencing
CpG	Cytosine-Guanine dinucleotide
DKO	Double knockout
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
Dnmt	DNA methyltransferase
dNTP	Nucleoside triphosphate
DTA	Diphtheria toxin A
DTT	Dithiothreitol
EB	Embryoid Body
EMT	Epithelial-mesenchymal-transition
FBS	Fetal bovine serum
FC	Fold change
FDR	False discovery rate
FPKM	Fragments per kilobase of exon per million fragments mapped
Gadd45a	Growth arrest and DNA-damage-inducible protein 45 alpha
Gadd45b	Growth arrest and DNA-damage-inducible protein 45 beta
Gadd45g	Growth arrest and DNA-damage-inducible protein 45 gamma
GFP	Green fluorescent protein
GO	Gene ontology
H&E	Hematoxylin & Eosin
H3K4me1,2,3	Histone 3 mono-/di-/trimethylated at lysine 4
HEK293T	Human Embryonic Kidney 293T cells
HRP	Horseradish peroxidase
ICM	Inner cell mass
ICR	Imprinting control region
IgG	Immunoglobulin G
IMB	Insitute of molecular biology
ING1	Inhibitor of growth protein 1
IP	Immunoprecipitation

JAK/Stat	Januskinase/Signal transducers and activators of transcription
JNK	c-Jun N-terminal kinase
JNK	c-Jun N-terminal kinases
kb	Kilobasepair
KD	Knockdown
KO	Knockout
LIF	Leukemia inhibitory factor
LMR	Low methylated region
MACS	Model-based analysis for ChIP-Seq
MAPK	Mitogen-activated protein kinase
MDC	Max-Delbrück Centrum
MEF	Mouse embryonic fibroblast
mESC	Mouse embryonic stem cell
min	Minute
mRNA	Messenger RNA
MS-PCR	Methylation sensitive PCR
mTSC	Mouse trophoblast stem cell
NER	Nucleotide excision repair
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PGC	Primordial germ cell
PrE	Primitive Endoderm
qPCR	Quantitative PCR
RAR	Retinoic acid receptor
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
rpm	Rounds per minute
RRBS	Reduced representation bisulfite sequencing
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TAF12	Transcription initiation factor TFIID subunit 23
TARID	TCF21 activating RNA inducing demethylation
TBP	TATA binding protein
TDG	Thymine DNA glycosylase
TE	Trophectoderm
TET	Ten-eleven translocation methylcytosine dioxigenase
TGC	Trophoblast giant cell
TKO	Triple knockout
TSS	Transcription start site
UPL	Universal probe library
UTR	Untranslated region
UV	ultra violet
Wnt	Wingless-type MMTV integration site family member
WT	Wildtype