Role of GADD45 during Stress Response and Embryonic Heart Development

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Gaurav Shrikant Joshi

geboren am 11.04.1997 in Ratnagiri, Indien

Mainz, 2024

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 01.10.2024

Table of Contents

1. Summary

Active DNA demethylation plays a crucial role in shaping the epigenetic landscape to achieve spatiotemporal transcription regulation during mouse embryonic development. Recent work from our lab showed that GADD45 (Growth arrest and DNA damage-inducible 45) proteins bind R-loops, non-canonical nucleic acid structures comprising of a DNA: RNA hybrid and a displaced single-stranded DNA (ssDNA). GADD45 proteins recruit TET1 (Ten-eleven translocation) to the R-loop site, facilitating site-specific active DNA demethylation and transcription activation. However, the genome-wide prevalence and underlying mechanism of GADD45 and TET1-mediated DNA demethylation remains elusive. I have addressed this gap in our understanding by studying (I) stress response and (II) embryonic heart development, using mouse embryonic stem cells (mESCs) as an in-vitro model.

(I): Unpublished CUT&Tag sequencing datasets from our lab showed a significant enrichment for heat shock element (HSE) DNA motif at R-loop sites bound by GADD45a and TET1 in mESCs. The HSE motif is commonly observed in promoters of heat shock genes and is recognized by HSF1 to activate gene transcription during heat shock response (HSR). I hypothesized that GADD45 proteins bind the R-loops in heat shock gene promoters and recruit TET1 to aid with transcription activation during HSR. I uncovered thermal stress-specific nuclear translocation of GADD45 proteins during the early phase of HSR. Lack of GADD45 proteins sensitized the *Gadd45a,b,g* triple knockout (*Gadd45* TKO) mESCs to thermal stress. Furthermore, I showed that HSF1 occupancy at heat shock gene promoters was significantly reduced in *Gadd45* TKO mESCs. However, the reduced HSF1 occupancy did not lead to altered heat shock gene expression in *Gadd45* TKO mESCs (with and without R-loop level manipulation). Intrigued by the lack of gene expression changes in Gadd45 TKO mESCs, I revisited our previous GADD45a mapping data and found that the previously identified GADD45a binding sites in mESCs were artefacts of antibody non-specificity. Therefore, I did not pursue this line of research work further.

(II): Unpublished work from our lab showed that GADD45 and TET proteins are required for normal embryonic heart development. However, the underlying mechanism by which GADD45 and TET proteins regulate embryonic heart development remained elusive. I hypothesized that GADD45 proteins balance TET functions (catalytic vs. non-catalytic) to regulate gene transcription during embryonic heart development. In line with my hypothesis, RNA-seq in *Gadd45* TKO mESCs, with deletion of coding sequence (CDS) of all *Gadd45* genes, showed reduced expression of heart development-associated genes. Furthermore, in-vitro cardiomyocyte differentiation of WT and *Gadd45* TKO (CDS) mESCs revealed impaired cardiac differentiation of *Gadd45* TKO (CDS) mESCs. These cardiac differentiation defects in *Gadd45* TKO (CDS) mESCs were supported by

pilot single-cell RNA sequencing (scRNA-seq) analysis. In collaboration with colleagues, we identified differential TET1 binding and differentially methylated sites between WT and *Gadd45* TKO (CDS) mESCs. I showed that the reduced expression of heart development-associated genes was linked to increased TET1 occupancy in *Gadd45* TKO (CDS) mESCs but did not correlate with DNA methylation changes. Instead, the downregulated heart development-associated genes with increased TET1 occupancy were enriched for Polycomb repressive complex (PRC) binding sites. These novel results suggest that GADD45 proteins supress TET1 non-catalytic functions and thereby promote expression of genes crucial for normal heart development.

2. Zusammenfassung

Aktive DNA Demethylierung ist entscheidend für die Ausbildung epigenetischer Muster und damit einhergehend der räumlich-temporalen Regulation der Transkription während der Embryonalentwicklung. Vorausgegangene Studien dieses Labor zeigten Bindung von GADD45-Proteinen (Growth arrest and DNA damage-inducible 45) an R-loops, nicht-kanonische Nukleinsäurestrukturen bestehend aus DNA:RNA-Hybriden und einer dislozierten Einzelstrang-DNA (ssDNA). GADD45-Proteine rekrutieren TET1 (Ten-eleven translocation) an R-loop Stellen, um dadurch lokale aktive DNA Demethylierung und Transkriptionsaktivierung zu ermöglichen. Jedoch sind sowohl die genomweite Bedeutung als auch der zugrundeliegende Mechanismus der GADD45- und TET1-vermittelten DNA Demethylierung weitgehend unklar. Die vorliegender Arbeit befasst sich mit beiden Aspekten mittels Untersuchungen zur (I) Stressreaktion und (II) embryonalen Herzentwicklung in murinen embryonalen Stammzellen (mESCs) als in vitro-Modell.

(I): Unveröffentlichte GADD45a CUT&Tag dieses Labors zeigten eine signifikante Anreicherung des DNA-Sequenzmotivs „Hitzeschock-Element“ (HSE) an R-Loop-Stellen, die von GADD45a und TET1 in mESCs gebunden werden. Das HSE-Motiv wird häufig in Promotoren von Hitzeschockgenen beobachtet und von HSF1 erkannt, um Gentranskription während der Hitzeschockreaktion (HSR) zu aktivieren. Daraus lässt sich die Hypothese ableiten, dass GADD45-Proteine R-Loop Strukturen an den Promotoren von Hitzeschockgenen binden und TET1 rekrutieren, um die Transkriptionsaktivierung während der HSR zu unterstützen. Tatsächlich translozieren GADD45-Proteine Hitzestress-spezifisch und während der frühen Phase der HSR in den Zellkern. Abwesenheit der GADD45-Proteine in Gadd45-TKO Zellen sensibilisierte die mESCs für thermischen Stress. Zudem war die Bindung von HSF1 an Promotoren von Hitzeschockgenen in Gadd45 TKO mESCs reduziert. Die reduzierte HSF1-Bindung führte jedoch nicht zu einer veränderten Expression von Hitzeschockgenen in Gadd45 TKO mESCs, weswegen ich die unpublizierten GADD45a CUT&Tag Daten überprüfte. Dabei fand ich heraus, dass es sich bei den zuvor identifizierten GADD45a-Bindestellen in mESCs vermutlich um Artefakte handelte, weshalb ich dieses Forschungsprojekt nicht weiter verfolgte.

(II): Weitere unveröffentlichte Arbeiten dieses Labors zeigten, dass GADD45- und TET-Proteine für eine normale embryonale Herzentwicklung erforderlich sind. Der zugrundeliegende Mechanismus, durch den GADD45- und TET-Proteine die embryonale Herzentwicklung regulieren, blieb jedoch ungeklärt. Die Arbeitshypothese lautete, dass GADD45-Proteine katalytische und nicht katalytische Funktionen der TET Dioxygenase reguliert, um die Gentranskription während der embryonalen Herzentwicklung zu steuern. Im Einklang mit dieser Hypothese zeigte die Analyse einer Transkriptom-Sequenzierung (RNA-seq) der Gadd45 TKO

mESCs, mit kompletter Deletion der kodierenden Sequenz (CDS) aller Gadd45 Gene, eine reduzierte Expression von mit der Herzentwicklung assoziierten Genen. Darüber hinaus zeigte in vitro-Differenzierung zu Kardiomyozyten eine beeinträchtigte kardiale Differenzierung von Gadd45 TKO (CDS) mESCs. Diese kardialen Differenzierungsdefekte in Gadd45 TKO mESCs wurden mittels Einzelzell-RNA-Sequenzierung (scRNA-seq) verifiziert. In Zusammenarbeit mit Kollegen wurden differentielle TET1-Bindungen und unterschiedlich methylierte Sequenzen zwischen WT- und Gadd45 TKO (CDS) mESCs identifiziert. Die reduzierte Expression von Herzentwicklungsgenen korrelierte mit einer erhöhten TET1-Bindung in Gadd45 TKO (CDS) mESCs, aber nicht mit DNA-Methylierungsänderungen. Stattdessen wiesen die herunterregulierten Herzentwicklungsgene mit erhöhter TET1-Bindung Bindemotive des repressiven Polycomb-Komplexes (PRC) auf. Die Ergebnisse deuten auf eine bislang unbekannte Rolle der GADD45-Proteine hin, die die nicht-katalytischen Funktionen von TET1 unterdrücken und dadurch die Expression von Genen fördern, die für die normale Herzentwicklung entscheidend sind.

3. General Introduction

3.1. Mouse embryonic development

Embryonic development involves a series of precisely orchestrated events, including cell division, differentiation, and morphogenesis (Figure 3-1). It starts with a rapid proliferation of a single-cell embryo (zygote) formed by fertilization. Initially, the resulting cells are totipotent (the capacity to create all of the specialized cell types of the embryo, the extra-embryonic membranes, and the placenta) [1]. As the number of cells increases, they progressively begin to specialize in function during embryonic differentiation. At the late 8-cell stage, cells undergo compaction and polarization, forming the early morula [2]. At the 16-cell morula stage, two asymmetric cell divisions segregate the first two lineages: the trophectoderm (TE), giving rise to the placenta, and the inner cell mass (ICM) [3]. The blastocyst undergoes the process of cavitation, during which the inner cell mass (ICM) differentiates into the epiblast (EPI), which forms the fetal cells, and the primitive endoderm (PE), which becomes part of the extra-embryonic yolk sac. Subsequently, the blastocyst comprises of a fluid-filled cavity, the epiblast and the primitive endoderm surrounded by the trophectoderm. The blastocyst hatches the zona pellucida and is now capable of uterine implantation. The post-implantation phase starts with gastrulation, during which the three primary germ layers: mesoderm, endoderm and ectoderm arise [4]. This lineage segregation is highly dependent on cell movements and surrounding signals. At the onset of gastrulation, the primitive streak forms in the posterior region of the embryo. Cells that migrate through the primitive streak differentiate into the mesoderm and the endoderm, while the cells that do not ingress the primitive streak become the ectoderm. Broadly, the mesoderm gives rise to skeletal and cardiac muscle, endothelial cells and connective tissues. The endoderm contributes to liver, lungs, thyroid and digestive cells, whereas the ectoderm yields the nervous system and the skin.

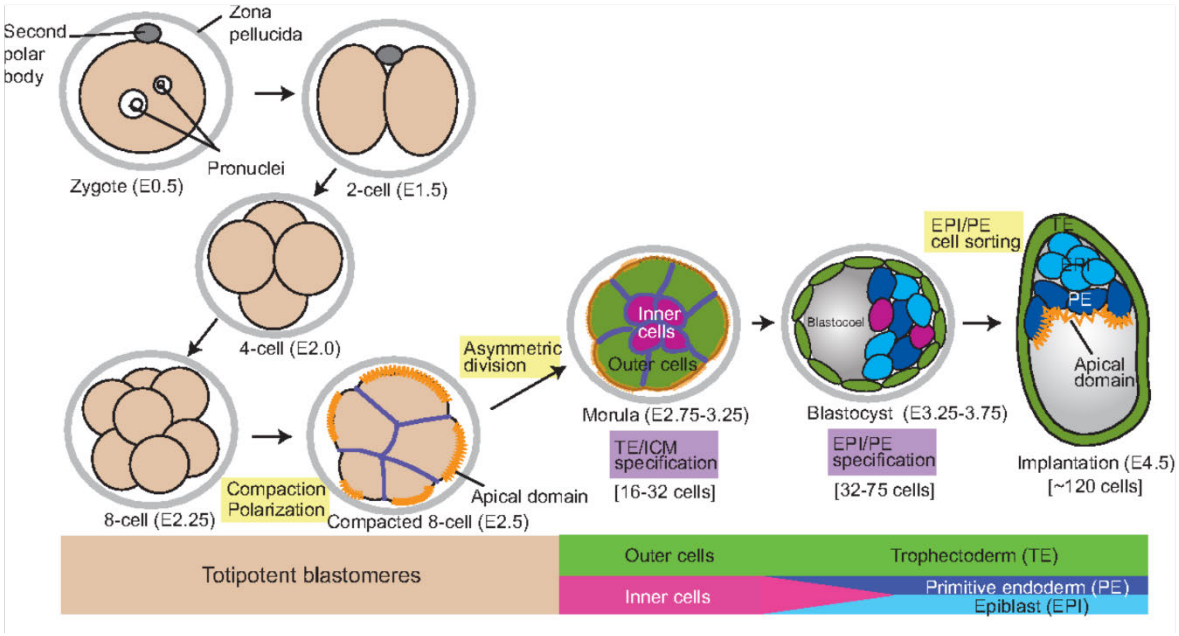

Figure 3-1: Schematic for stages of mouse pre-implantation development. The zygote divides three times in a symmetric manner forming the early 8-cell stage. Subsequent compaction, polarization and asymmetric division result in the morula stage that defines the first segregation between TE and ICM. Further development requires differentiation of the ICM into EPI and PE and a process called cavitation leading to a blastocyst capable of implantation. Green: trophectoderm; light blue: epiblast; dark blue: primitive endoderm; yellow: morphogenetic changes; purple: lineage specification steps; orange: apical domains of cell. Figure from [1].

The process of lineage commitment during gastrulation involves not only the physical differentiation of cells into distinct germ layers but also profound gene expression changes via epigenetic mechanisms [4,5]. Moreover, some epigenetic changes acquired during embryonic development persist into adulthood, influencing gene expression patterns and cellular functions throughout life [6]. Thus, the dynamic interplay between gastrulation-driven lineage commitment and changes in epigenetic landscape provides a roadmap for embryo development.

3.2. Epigenetic mechanisms of transcription regulation

Epigenetic modifications bring about additional levels of DNA organization, ranging from the positioning of chromosomes in the nucleus to modifications of single nucleotides [7]. Such epigenetic mechanisms help in tightly regulating gene expression in cells. Consequently, aberrant DNA organization and modification can lead to a variety of diseases [8].

Nucleosomes consist of a histone core complex that is wrapped by a 147 bp long DNA segment and forms the repeating unit of chromatin. The histone core is an octamer comprising two of each of the core histones H2A, H2B, H3 and H4. The positively charged histones aid in wrapping the negatively charged DNA helix. Various proteins function as epigenetic modifiers by introducing post-translational modifications (PTMs) on histone tails [9]. These histone modifications in-turn are

interpreted by “reader” proteins to regulate cellular processes such as gene transcription, DNA replication, and repair [10]. For instance, active genes typically exhibit H3/H4 acetylation and H3K36/H3K4 trimethylation (me3), while repressed genes bear marks like H3K9me3/H3K27me3. Other known histone modifications include ubiquitination, sumoylation, and poly ADP-ribosylation [9].

The other layer of epigenetic transcription regulation occurs via dynamic chemical modifications of DNA. One of the best characterized DNA modifications is the methylation at C5 position in the pyrimidine ring of cytosine, 5-methylcytosine (5mC) [11]. DNA methylation patterns are established by the de novo DNA methyltransferase 3a and 3b (DNMT3a and DNMT3b) within a CpG dinucleotide context [12]. These DNA methylation patterns are mitotically inherited by the action of DNA methyltransferase 1 (DNMT1) [13].

5mC in promoter regions is associated with transcription repression. Additionally, 5mC is enriched over gene bodies of highly transcribed genes, to prevent spurious initiation of transcription [14]. 5mC can be erased via passive or active DNA demethylation processes [15]. Passive DNA demethylation results from replication without maintenance of the methylation pattern by DNMT1. On the other hand, active DNA demethylation occurs by iterative oxidation of 5mC into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) by ten-eleven-translocation (TET) family of enzymes (TET1, TET2 and TET3) [15–18]. 5fC and 5caC can be excised by base excision repair (BER) initiated by thymine DNA glycosylase (TDG), thereby restoring an unmodified cytosine base [19] (Figure 3-2).

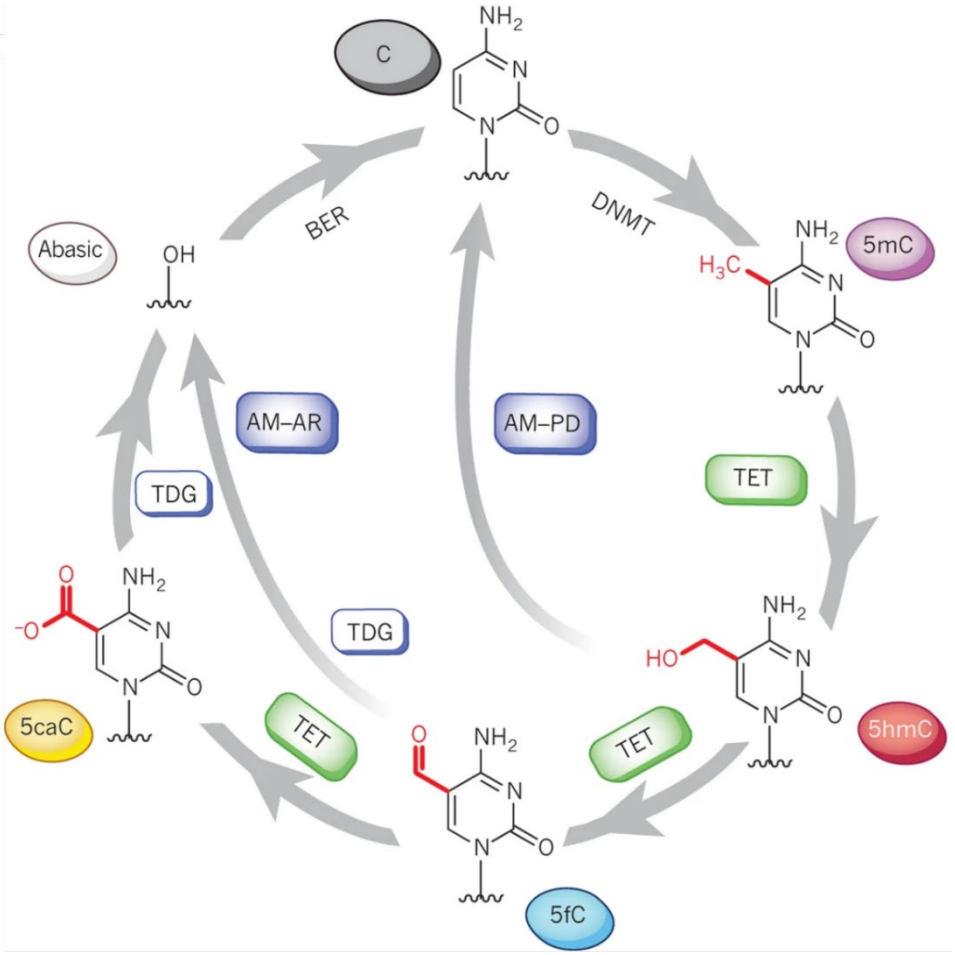

Figure 3-2: TET-mediated active DNA demethylation pathway. Schematic represents the cytosine methylation by DNMTs (5mC) and its iterative oxidation by TET enzymes to 5hmC, 5fC and 5caC. AM-PD represents active modification - passive dilution and AM-AR represents active modification – active removal pathway. 5fC and 5caC can be excised by TDG creating an abasic site, which is restored to unmodified cytosine by BER pathway. Figure from [15].

5mC is involved in genomic imprinting, X-chromosome inactivation and suppression of retrotransposons [4,20]. 5hmC, 5fC and 5caC have been suggested to be regulatory marks and not mere by-products of the DNA demethylation pathway [21–23]. For instance, 5hmC was shown to be enriched over active genes marked by H3K4me2 in neurons and is specifically bound by methyl-CpG binding protein 2 (MeCP2) [24]. In mouse embryonic stem cells (mESC), 5fC preferentially occurs at poised enhancers and coordinates with histone acetyltransferase, E1A-associated protein p300 (EP300), in remodelling epigenetic states of enhancers [25]. Moreover, dynamic interplay between DNA methylation and histone modifications helps fine tune gene expression during fundamental processes of embryonic development, such as cell division, differentiation, and morphogenesis [26–28]. Thus, investigating novel epigenetic mechanisms regulating embryonic development may enable us to discover the factors that govern health and alleviate diseases.

3.3. Role of DNA methylation during embryonic development

DNA methylation patterns influence the accessibility of DNA to transcription factors (TFs) and other regulatory proteins, thereby modulating gene expression programs that drive lineage-specific differentiation [28]. De novo DNA methylation is associated with lineage commitment, while the erasure of DNA methylation is linked to increased developmental potential. Upon fertilization, there is a wave of global DNA demethylation in the zygote, which is necessary for the initiation of embryonic development and the establishment of totipotency [20]. This global DNA demethylation occurs in both the male and female pronuclei, albeit through distinct mechanisms. Maternal DNA demethylation primarily occurs passively via nuclear exclusion of DNMT1, while active demethylation of paternal DNA occurs before replication and is facilitated by TET enzymes [29–31]. Subsequently, as the embryo undergoes lineage specification and gastrulation, there are further changes in DNA methylation landscape that accompany the differentiation of cells into the three germ layers [14,32,33]. Interestingly, the primordial germ cells (PGCs), which are precursors to sperm and eggs, are associated with the second round of global DNA demethylation, essential to erase parental-specific methylation patterns [34,35]. Disruption of DNA methylation patterns results in developmental abnormalities and increases the risk for congenital diseases, as demonstrated by studies involving knockout (KO) mice for proteins in the DNA methylation and demethylation pathways [36–39].

Mutations or genetic ablation of DNMT genes result in post-implantation defects including impaired placental development, immunodeficiency and facial anomalies [40–42]. Given the functional redundancy, *Tet* single knockout (SKO) mice are viable and fertile, while *Tet* double knockout (DKO) mice present more severe phenotypes [43–45]. *Tet1*/2 DKO mice show reduced fertility and smaller ovaries in females [44]. *Tet1*/3 DKO is embryonic lethal, and these embryos exhibit early developmental abnormalities, including delayed or aborted development at the eight-cell stage [45]. Notably, *Tet1*/2/3 triple knockout (TKO) embryos can develop past the implantation stage but show severe defects by E7.5, such as impaired mesodermal migration. By E8.5, *Tet* TKO have no recognizable headfolds, heart and gut tube structures [46].

TET proteins primarily regulate DNA methylation at regulatory regions such as enhancers and promoters, interacting with various proteins and non-coding RNAs to exert their functions [47]. For instance, the transcription factor NANOG interacts with TET1 and TET2, facilitating NANOG-mediated somatic cell reprogramming using TET catalytic activity [48]. Beyond their catalytic roles, TET proteins also regulate gene expression through non-catalytic mechanisms. For example, TET1 interacts with the Polycomb Repressive Complex (PRC) and recruits it to lineage-specific genes, such as *Gata6* and *Sox17*, to achieve transcriptional silencing [49,50]. Additionally, TET1

recruits the Histone Deacetylase Complex sub-unit SIN3a (SIN3A/HDAC complex), to pluripotency and mesendoderm genes, activating their expression during early embryo development [51]. These examples further illustrate the crucial role of active DNA demethylation pathway proteins in regulating gene transcription during mouse embryonic development.

Interestingly, our lab identified a novel inducer of active DNA demethylation, GADD45a (Growth arrest and DNA damage-inducible 45 alpha), which directly binds TET1 and positively regulates TET1-mediated DNA demethylation activity [52,53].

3.4. GADD45 protein family

The GADD45 family comprises of three small (18 kDa), evolutionarily conserved, highly acidic proteins, sharing ~55% similarity in amino acid sequence [54]. GADD45a, GADD45b, and GADD45g belong to the L7Ae/L30e/S12 superfamily of RNA-binding proteins [55–57].

GADD45 proteins lack enzymatic activity and fulfil various functions redundantly through interactions with effector proteins. Additionally, GADD45 proteins can interact with each other and form homo- and hetero-dimers, which are crucial for some effector protein interactions [58,59]. Some of the well-studied functions of GADD45 proteins is their involvement in cellular signalling pathways such as the stress-activated protein kinases, cyclin-dependent kinases (CDKs) and DNA repair pathways. All three GADD45 family members directly interact with mitogen-activated protein kinase kinase kinase 4 (MEKK4) in response to environmental stress and promote apoptosis through the p38/JNK pathways [60]. Despite its role in promoting apoptosis, GADD45b also engages in pro-survival mechanisms by binding to and inhibiting the MAP kinase MKK7 [61]. GADD45a and GADD45b are involved in DNA repair, specifically in nucleotide excision repair (NER), by interacting with Proliferating cell nuclear antigen (PCNA) [62]. This interaction helps recruit repair proteins to DNA damage sites and modulates DNA accessibility to repair proteins. The CDK1/Cyclin B1 complex, crucial for the G2/M checkpoint, is a primary target of GADD45-mediated cell cycle arrest. Induction of cell cycle arrest by GADD45 results in decreased cell proliferation in both normal and cancer cells [63,64].

However, over the past decade, GADD45 proteins have been shown to have functions beyond their roles in cellular signalling pathways (Figure 3-3). GADD45a is involved in TET-TDG-mediated DNA demethylation by promoting TET activity and enhancing the removal of 5fC and 5caC by TDG [65]. Additionally, they interact with various epigenetic reader proteins, and then recruit the DNA demethylation machinery at specific sites in the genome [66,67]. For example, GADD45a binds to distinct genomic loci via the H3K4me3 reader, inhibitor of growth protein 1 (ING1) and ensures gene-specific DNA demethylation [68]. GADD45a binds to the RNA polymerase cofactor TAF12 to cause DNA demethylation in the ribosomal DNA (rDNA) promoter [69]. In mouse embryonic

fibroblasts (MEFs), GADD45a occupies C/EBP (CCAAT/enhancer-binding protein) dependent super-enhancers, and, along with ING1, promotes local DNA demethylation to permit C/EBPβ recruitment [70].

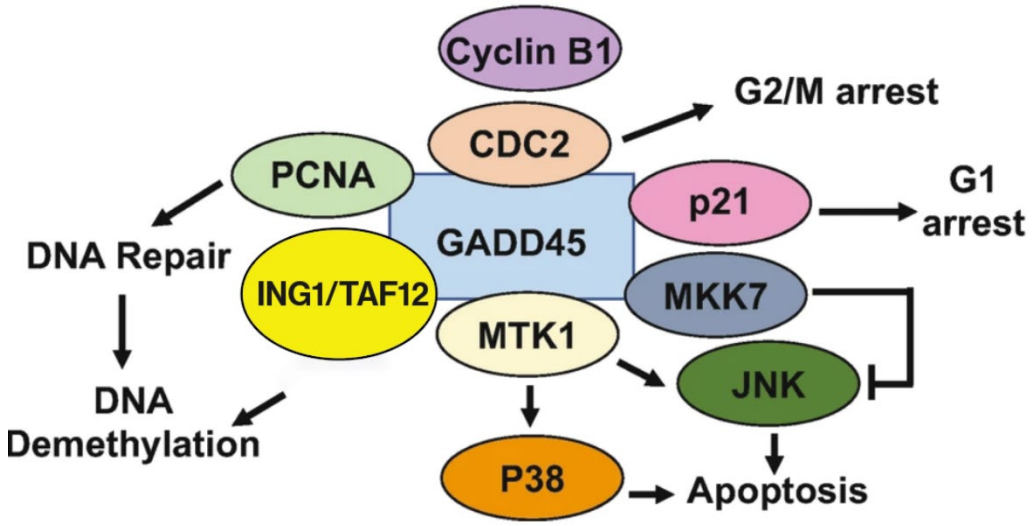

Figure 3-3: GADD45 protein functions. Schematic summarizes the diverse effector proteins GADD45 proteins interacts with to regulate cellular signalling and DNA methylation landscape. Figure adapted from [71]

Given the functional redundancy of GADD45 proteins, *Gadd45* SKO mice are viable, fertile but exhibit mild phenotypes. They range from genomic instability, tumorigenesis, immune deficiencies, decreased bone growth, altered learning behaviour, and sex reversal to accelerated ageing [57,72–76]. On the other hand, a combined deletion of both *Gadd45a,b* genes (*Gadd45a,b* DKO) leads to embryonic sub-lethality, which occurs at later embryonic stages (stages beyond E15.5), with several developmental abnormalities typically categorized as neural tube closure defects [77]. However, the functional redundancy among GADD45 proteins and the inability to produce a sufficient number of *Gadd45* TKO mice for characterization obscure the identification of any additional phenotypes that may result from the combined loss of all three GADD45 proteins.

Studying early mouse development presents challenges due to the limited availability of material from early embryos. While pooling embryos and emerging single-cell analysis techniques offer insights, in-vitro models remain highly valuable. In this thesis, I intend to overcome this limitation by using *Gadd45a,b,g* TKO mESCs and elucidate the genome-wide roles of GADD45 proteins in shaping the DNA demethylation landscape during stress response and embryonic development.

3.5. mESC: A model system of embryonic development

In 1981, Evans and Martins pioneered the isolation of mESC from blastocyst stage embryos, revolutionizing our understanding of early mouse development [78]. Derived from the ICM of pre-implantation embryos, mESCs exhibit pluripotency and immortal self-renewal potential in-vitro.

Transcriptionally resembling the early naïve epiblast, mESCs can differentiate into all three germ layers (mesoderm, endoderm, and ectoderm). Key regulators of pluripotency include OCT4, SOX2, and NANOG, orchestrating the expression of core pluripotency markers and repression of lineage commitment factors. Assessing whether in-vitro differentiation processes represent physiologically relevant states is crucial for embryology studies [79,80]. Fortunately, in vivo developmental programs appear faithfully preserved in-vitro, evidenced by the timing of lineage-specific transcription factor expression and functional characteristics of resulting differentiated cells, such as cardiomyocyte contractions and neuronal membrane potentials. Therefore, mESCs aid in investigating lineage commitment by differentiating into various cell types under specific culture conditions. Methods such as embryoid body (EB) differentiation mimic post-implantation blastocyst development, yielding derivatives of all three germ layers. By modulating growth factors, precursor cell types can differentiate along defined lineages, producing diverse cell types, including cardiomyocytes, hematopoietic precursors, adipocytes, and smooth muscle cells [80].

Additionally, the advent of CRISPR-Cas9 technology has enabled precise and efficient generation of gene knockouts in mESCs. This approach was used to generate *Gadd45* TKO mESC lines and study embryonic developmental processes in-vitro [77]. For instance, *Gadd45* TKO (with exon1, 2 deletion) mESCs exhibit locus-specific DNA hypermethylation at ~7000 sites, enriched for enhancers and loci undergoing TET-TDG-mediated demethylation [77]. They show down-regulation of specific genes found in a rare population of cells with totipotent-like potential, known as 2 cell (2C)-like cells. This example underscores the relevance of the mESC model system and highlights the role of GADD45 proteins as regulators of the genome-wide DNA methylation landscape during embryonic development.

3.6. Structure of the thesis

In this thesis, I present two related projects:

Part - I: Role of GADD45 proteins during stress response in mESCs.

Part - II: Role of GADD45 proteins during embryonic heart development.

I will present these projects in two separate parts, each containing an Introduction, Aims, Results and Discussion section. Materials and Methods used during both projects are presented together at the end of the thesis.

Part - I

4. Role of GADD45 proteins during stress response in mESCs

4.1. Introduction

4.1.1. R-loops: Biochemical nature and genomic distribution

R-loops are non-canonical nucleic acid structures comprising of a DNA: RNA hybrid and a displaced single-stranded DNA (ssDNA) [81–83]. R-loops form in-cis when the nascent transcribed RNA invades the upstream negatively supercoiled double-stranded DNA (dsDNA). Transcription through GC skew regions also promotes R-loop accumulation because G-rich RNAs bound to C-rich DNA templates have a high thermodynamic stability [84]. R-loops can also form in-trans, as found in the bacterial CRISPR-Cas defence mechanism against invading viral DNA or by TERRA long non-coding RNA (lncRNA) at telomeric DNA repeats [85,86]. Secondary structures of the displaced non-coding DNA strand (e.g. G-quadruplexes) further promote R-loop formation [87].

Initial electron microscopy studies of individual R-loop structures predicted them to be ~200 nucleotides (nt) long [88,89]. However, more recent sequencing-based genome-wide R-loop mapping methods rather show heterogeneity in R-loop sizes, ranging from a few hundred bases to several kilobases (kb). Such genome-wide R-loop mapping techniques primarily use the R-loop recognizing S9.6 antibody, including DNA: RNA immunoprecipitation sequencing (DRIP-seq), sequencing the RNA strand of an R-loop after cDNA synthesis (DRIPc-seq) and recently Cleavage Under Targets and Tagmentation (CUT&Tag) [90]. bisDRIP-seq is one of the strand-specific R-loop mapping method combining the S9.6 antibody DRIP-seq with sodium bisulfite treatment [91]. Lastly, S1-DRIP-seq uses Nuclease S1 to degrade the displaced ssDNA of an R-loop and maps R-loops in a strand-specific manner [92]. It is important to note that individual R-loop mapping approaches have inherent technical biases, some of which stem from distinct biochemical activities involved in R-loop formation [90]. Nonetheless, all these studies revealed that R-loops occupy ~5% of the mammalian genome, with CpG islands in gene promoters highly prone to R-loop formation [81,84,93].

Cells regulate R-loop levels in the genome using diverse enzymes with helicase activity to unwind R-loops (e.g. Senataxin, Aquarius, and DEAD-Box RNA helicases) or R-loop degrading activity (e.g. RNA-specific nucleases RNASEH1 and DNA endonucleases like XPF and XPG) [94,95]. DNA repair factors like BRCA1/2 and Fanconi Anemia (FA) proteins also help prevent R-loop accumulation by promoting "R-loop erasers" recruitment at DNA damage sites [94]. Mutations in

RNASEH2, responsible for R-loop degradation, are linked to Aicardi-Goutieres syndrome (AGS), characterized by genome-wide R-loop accumulation [96]. In diseases like Friedreich's ataxia and fragile X syndrome, R-loops form at repeat expansions, potentially silencing critical disease genes [97]. Deficiencies in Senataxin-associated exonuclease (*San1*) lead to excessive R-loop accumulation and DNA damage, causing early cell cycle exit in cardiomyocytes and resulting in hypoplastic cardiac disorders [98]. Beyond the nucleus, cytoplasmic R-loops can trigger innate immune responses and cell death. Therefore, maintaining a balance between the formation and degradation of R-loops is crucial for normal development [82].

4.1.2. Regulatory R-loops: a new class of epigenetic regulators

Until recently, R-loops were believed to be mere by-products of transcription; however, recent work has shed light on new class of R-loops termed regulatory R-loops [83]. These subset of R-loops are harnessed by cells, primarily to regulate gene transcription and DNA repair. Regulatory R-loops are also initiated by lncRNAs in-cis and in-trans and serve as a platform for recruiting epigenetic modifiers [99].

R-loop levels at specific promoter-proximal sites in mESC directly regulate the binding of activating histone acetyltransferase TIP60–p400 complex and the repressive PRC2 complex [100]. This dynamic interplay between R-loops and chromatin modifiers directly impacts mESC pluripotency and differentiation gene networks. R-loops are also known to induce antisense transcription over RNA polymerase II (RNA Pol II) pause elements at transcription termination sites (TTS), which in turn leads to the recruitment of the G9a histone lysine methyltransferase for efficient transcription termination [101]. R-loops accumulate during stress conditions like hypoxia and contribute to the transcriptional stress response via deposition of heterochromatic H3K9me2 in rRNA promoters [102]. In female mammals, one X chromosome undergoes inactivation to achieve equal gene dosage between sexes. R-loops facilitate this process by recruiting silencing factors to the inactive X chromosome, ensuring its proper function [103]. Similarly, R-loops are crucial for establishing and maintaining genetic imprinting patterns [84,95]. Disruption of R-loop homeostasis can inhibit cellular reprogramming; however, certain TFs such as SOX2 can overcome this inhibition [104].

Interestingly, R-loop accumulation is a characteristic feature of unmethylated gene promoters. R-loop formation in promoters of active genes inhibits DNA methylation by obstructing DNA methyltransferase binding or recruiting TET DNA demethylases, thus regulating gene expression (Figure 4-1 A) [105–107]. For instance, in mESCs thousands of TET1 binding sites (~4% of 90,000 total binding sites) seem to be R-loop dependent [107]. Transcription through 5hmC-rich DNA favours R-loop formation, with TET activity increasing cellular R-loop levels [108]. The oxidised DNA demethylation intermediates 5fC and 5caC are enriched at R-loops, and their levels increase upon

RNASEH1 depletion [107]. However, under stress 5hmC modification negatively regulates R-loop accumulation, highlighting the context-specific interplay between DNA demethylation intermediates and R-loops [109–111].

4.1.3. GADD45a: R-loop reader recruiting DNA demethylation machinery

It still remains unclear how the cellular machinery distinguishes regulatory R-loops from unscheduled R-loops. Do proteins identify some novel consensus DNA sequence in regulatory R-loops? Does it depend on the orientation and position of regulatory R-loops with respect to their nearby gene? Are there specific histone marks or DNA methylation patterns which help with this molecular distinction?

Recently, our lab identified a novel mechanism for GADD45A-mediated recruitment of DNA demethylation machinery in human cells [106,107]. GADD45A was shown to bind R-loops in-vitro and act as an R-loop reader. Additionally, GADD45A recognized an R-loop formed in the promoter region of the *tumour suppressor gene transcription factor 21* (*TCF21*) gene via the antisense lncRNA, *TCF21 antisense RNA inducing demethylation* (*TARID*) (Figure 4-1 A). This leads to GADD45A-mediated TET1 recruitment and local DNA demethylation. However, the mechanism and genome-wide prevalence of GADD45a-mediated regulatory R-loop recognition to potentially promote local DNA demethylation remains elusive.

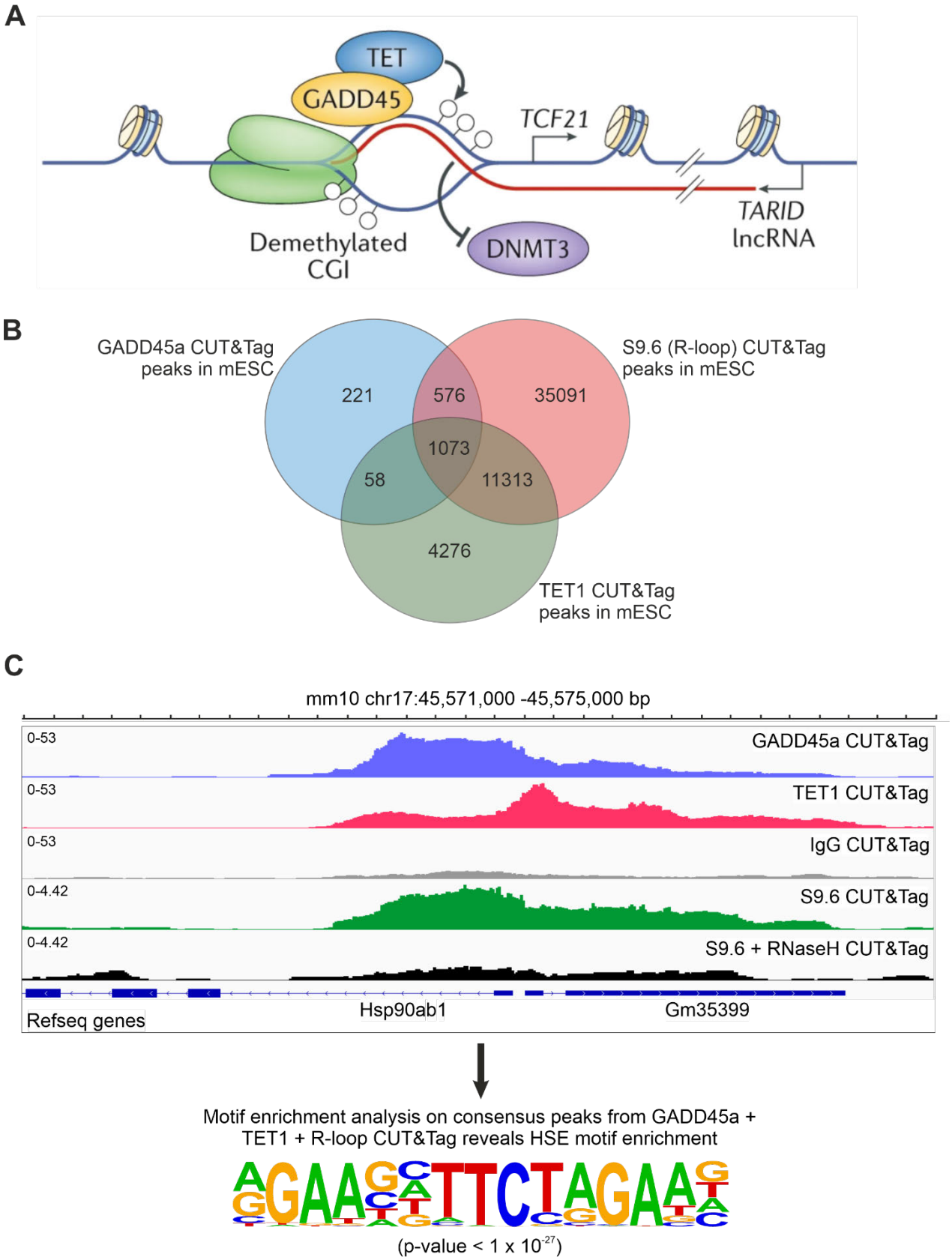

Figure 4-1: Analysis of GADD45a, TET1 and S9.6 (R-loop) CUT&Tag in mESCs. (A) Schematic overview of GADD45 protein binding R-loop at TCF21-TARID locus, which is associated with TET recruitment for active DNA demethylation. Figure from [83]. **(B)** Venn diagram shows the overlap between GADD45a, TET1 and S9.6 (R-loop) CUT&Tag peaks in mESCs. **(C)** Representative genome browser track centred at promoter of *Hsp90ab1*, a heat shock gene with anti-sense transcript *Gm35399*. This representative

CUT&Tag sequencing example illustrates the presence of GADD45a (blue), TET1 (magenta) and R-loop (green) peaks in the promoter of *Hsp90ab1* gene. IgG served as control for peak calling in GADD45a and TET1 CUT&Tag, while S9.6+RNaseH served as control for S9.6 (R-loop) peak calling. The enriched heat shock element (HSE) DNA motif logo is shown at the bottom, along with the FDR corrected p-value.

Additionally, unpublished CUT&Tag sequencing data in mESC revealed ~55% of the total GADD45a peaks (~1100 peaks) overlap with both R-loops (S9.6 CUT&Tag) and TET1 binding sites (Figure 4-1 B). Notably, these common overlapping sites showed significant enrichment for the heat shock element (HSE) DNA motif, comprising inverted repeats of "NGAAN" pentamers (N representing any nucleotide) (Figure 4-1 C). The heat shock element (HSE) motif serves as the recognition site for heat shock transcription factor 1 (HSF1), a crucial regulator of the heat shock gene transcription[112].

4.1.4. Role of HSR during stress response and embryonic development

Organisms have evolved to thrive within a wide range of temperatures [113]. However, even moderately elevated temperatures above the optimal growth range poses challenges to cellular homeostasis. This elevation can lead to protein unfolding, tangling, and non-specific aggregation. Additionally, cells also need to respond to the presence of unfolded proteins resulting from various other stressors like oxidative stress, heavy metals, ethanol, or other toxins [114]. These are some of the context in which cells initiate the highly conserved HSR pathway. The HSR pathway activation is associated with increased transcription of heat shock genes, producing heat shock proteins (HSPs), which act as molecular chaperones to prevent protein mis-folding and maintain proteostasis. This activation of the HSR involves trimerization of HSF1 and specific PTMs such as phosphorylation at Ser326 residue, enabling its binding to the HSE motif in heat shock gene promoters [115]. Following the alleviation of stress conditions, cells downregulate the HSR and consequently reduce the production of HSPs to normal levels. However, HSF1 and the HSR are not solely stress response mechanisms; but also play vital roles in embryonic development.

HSF1 is critical for female fertility, as *Hsf1*−/− females fail to produce viable embryos due to early developmental arrest [116,117]. These mice also show defects in chorioallantoic placenta formation, highlighting importance of HSF1 in extra-embryonic tissue development. HSF1 knockdown in zebrafish disrupts the expression of *Charon* and *Spaw* genes, essential for left-right axis polarity [118]. Maternal HSF1 regulates HSPs, particularly HSP90α, essential for meiotic maturation [119]. HSPs are crucial in early mammalian development, especially for brain and heart formation, basement membrane formation, extracellular matrix production, and neural crest migration [120,121]. Mutation of the HSP90β gene inhibits hepatocyte differentiation by regulating HNF4A protein's half-life [122]. HSP90 maintains pluripotency by stabilizing key factors like OCT4, NANOG, and STAT3, protecting them from degradation [123]. HSP90 inhibition increases mesodermal marker

expression, indicating its role in suppressing mesodermal differentiation from ESCs [123]. HSP90 also regulates DNMT transcription and proteins levels via its chaperone activity [124].

DNA methylation changes induced during embryonic heat conditioning (EHC) of early chick embryos determine their stress resilience and vulnerability in later life. The RE1-silencing transcription factor (REST) forms a complex with TET3 at this regulatory region, bringing about DNA demethylation and histone modification (H3K27ac) during thermal conditioning of embryos [125]. TET-mediated DNA methylation changes during EHC also induce cross-tolerance with the immune system, reducing hypothalamic inflammation in chicks [126]. This highlights the role of DNA methylation in regulation of HSR during embryonic development and differentiation. Lastly, GADD45 proteins are known to play a crucial role in the regulation of HSR, primarily via cellular signalling pathways [127]. However, it remains unclear whether GADD45 proteins can directly regulate heat shock gene transcription via recruitment of TET proteins to heat shock gene promoter R-loops.

4.2. Aim

The ability of regulatory R-loops to act as landmarks for recruitment of epigenetic modifier is a novel epigenetic mechanism regulating gene transcription [83]. Previous studies in our laboratory identified GADD45 proteins as regulators of active DNA demethylation and GADD45a as an R-loop reader [106,107]. Interestingly, the heat shock gene promoters bound by GADD45a are associated with poorly characterized anti-sense lncRNA transcripts which contribute to R-loop formation similar to the TCF21-TARID pair (Figure 4-1 A). However, the genome-wide prevalence and physiological importance of GADD45-mediated DNA demethylation via regulatory R-loops remains elusive. In part - I of my thesis, I aim to address this open question using mESCs as an in-vitro system.

I planned on using the thermal stress response context to understand the potential mechanism of GADD45 and R-loop mediated transcription regulation. I planned to analyse GADD45 protein dynamics and cellular localization to understand their role during HSR. Furthermore, HSF1 chromatin immunoprecipitation sequencing (ChIP-seq) analysis should reveal whether GADD45 binding R-loops promotes conducive chromatin environment for HSF1 binding during HSR. Manipulating R-loop levels by RNASEH1 fusion protein overexpression (OE) followed by gene expression analysis in WT and *Gadd45* TKO mESCs should give us insight in the transcription regulatory roles of GADD45 proteins at heat shock gene promoter R-loops. I planned on performing TET1 mapping in WT and *Gadd45* TKO mESCs to link potential transcription defects to impaired TET1 recruitment and impaired DNA demethylation at these genes.

4.3. Results and Discussion

4.3.1. GADD45 proteins show thermal stress-specific nuclear translocation

To elucidate the function of GADD45 proteins during HSR, I characterized their protein levels and sub-cellular distribution upon HS exposure (incubating cells at 43°C). Due to lack of specific GADD45a and GADD45b antibodies, I used two separate mESC lines encoding for an in frame C-terminal 3xHA-mAID tag at the endogenous *Gadd45a* or *Gadd45b* loci (a kind gift from [redacted] this lab). The GADD45a-3xHA-mAID and GADD45b-3xHA-mAID proteins were probed using an anti-HA antibody in my downstream assays. GADD45g protein dynamics were not characterized due to low protein levels in mESCs [77].

Western blot analysis of GADD45a and GADD45b in cells exposed to increasing duration of HS revealed dynamic changes in their protein levels (Figure 4-2 A). Both GADD45a and GADD45b protein levels increased until 10 min of HS exposure. Subsequently, both GADD45a and GADD45b protein levels gradually declined upon 15 min and longer HS exposure, with the lowest levels at 120 min HS.

GADD45 proteins are located in both nucleus and cytoplasm, and their sub-cellular localization changes based on cell type and environmental stimuli [128–130]. As I was interested in analysing transcription regulatory roles of GADD45 proteins, it was important to test whether their sub-cellular localization changes during HS to regulate the heat shock gene transcription in the nucleus. I performed nuclear (N) and cytoplasmic (C) protein fractionation of mESCs exposed to increasing durations of HS (2 min to 60 min) (Figure 4-2 B).

HSF1 translocated to the nucleus upon HS and I used it as a positive control for the fractionation experiment [131]. The purity of protein fractions was tested using TUBULIN, a known cytoplasmic protein and TBP, a known nuclear protein. HSF1 levels showed a significant increase in the nuclear (N) fraction with increasing duration of HS. The nuclear translocation of HSF1 occurred as early as 2 min of HS exposure (Figure 4-2 B). GADD45a and GADD45b protein levels increased substantially in the cytoplasmic (C) fraction upon short HS exposure (2-10 min). Interestingly, both GADD45a and GADD45b protein levels increased substantially in the nuclear (N) fraction between 5-10 min HS exposure. At longer duration of HS, there was a significant reduction in their cytoplasmic levels while nuclear levels remained stable (Figure 4-2 B).

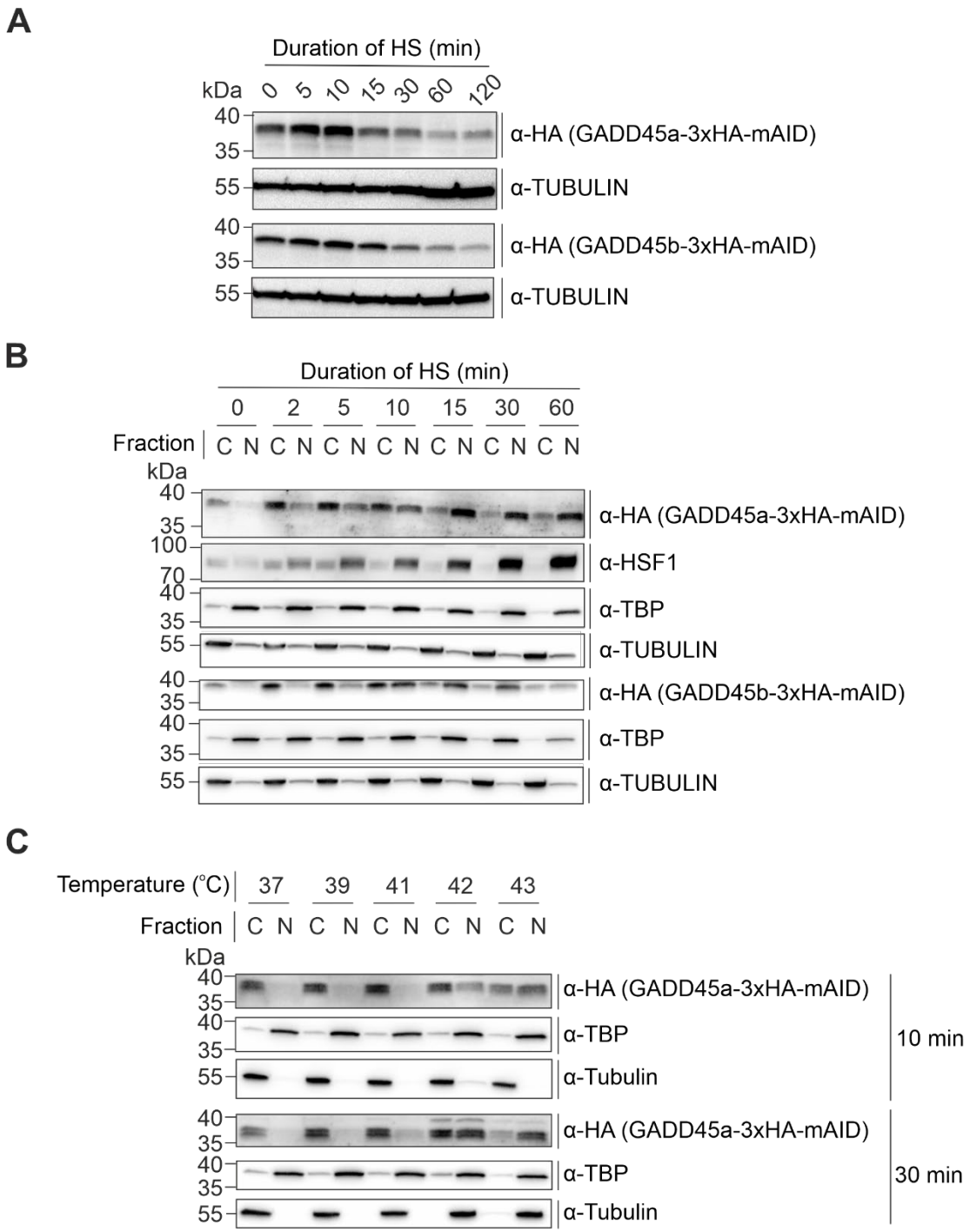

Figure 4-2: GADD45 proteins show dynamic changes in protein levels and nuclear translocation upon HS exposure in mESCs. (A) Western blot analysis of GADD45a and GADD45b proteins with increasing duration of HS (43°C). TUBULIN served as loading control. **(B)** Western blot analysis of GADD45a, GADD45b and HSF1 proteins in cytoplasmic (C) and nuclear (N) protein fractions from mESCs exposed to increasing duration of HS (43°C). **(C)** Western blot analysis of GADD45a protein in cytoplasmic and nuclear protein fractions from mESCs exposed to increasing HS temperature (39°C - 43°C) for 10 min (top) and 30 min (bottom). TUBULIN and TBP served as loading controls for cytoplasmic and nuclear fractions respectively. Expected molecular weight for GADD45a-3xHA-mAID: 37kDa; GADD45b-3xHA-mAID: 37kDa; TUBULIN: 55kDa; TBP: 37kDa; HSF1: 57kDa, but migrates at 80kDa.

I evaluated the correlation between GADD45a nuclear translocation and degree of thermal stress. I performed HS using a temperature range from 39°C to 43°C, for two different durations, 10 and 30 min. The nuclear translocation of GADD45a correlated positively with the degree of thermal stress i.e. nuclear translocation at lower temperatures required longer duration to reach nuclear levels as seen for 10 min HS at 43°C (Figure 4-2 C).

I performed SDS-insoluble protein fractionation of whole cell lysates to exclude the possibility of GADD45a protein aggregation upon HS exposure. Western blot analysis of the soluble protein fraction showed the expected pattern of GADD45a levels, increase at 10 min followed by reduction at longer HS exposure (Figure 4-3 A). On the other hand, the insoluble protein fraction showed a significant GADD45a signal at 60 min of HS, confirming that the nuclear translocation observed at early HS exposure (10 min) was not a protein aggregation artefact (Figure 4-3 B). I analysed localization of some known cytoplasmic proteins (GAPDH, p38, MEKK4) upon HS exposure to ensure that the nuclear translocation of GADD45 proteins was not due to compromised nuclear membrane integrity from HS (Figure 4-3 C). GAPDH and p38 (~35-40 kDa) did not show increase in nuclear levels upon HS exposure. On the contrary, MEKK4 (~180 kDa), a known GADD45a interacting protein, showed significant increase in nuclear levels starting at 15 min of HS exposure [132]. Lastly, the effects of GADD45a nuclear translocation were confirmed via immunofluorescence assay in HS exposed mESCs (Figure 4-3 D, E).

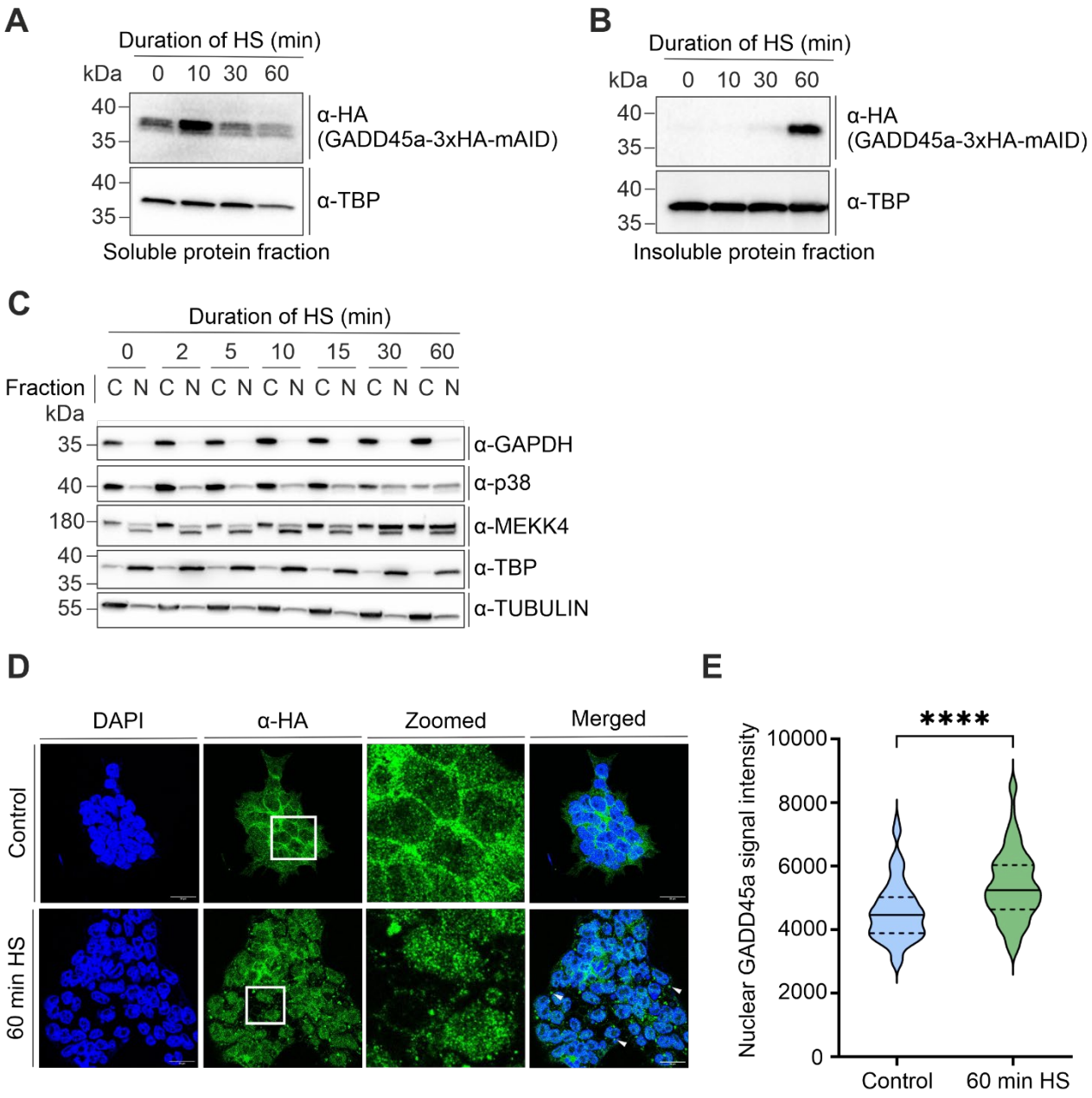

Figure 4-3: GADD45a nuclear translocation is not an artefact of protein aggregation or compromised nuclear integrity. (A) Western blot analysis of GADD45a protein in soluble protein fraction from mESCs exposed to increasing duration of HS (43°C). TBP served as loading control. **(B)** Western blot analysis of GADD45a protein in SDS-insoluble protein fraction from mESCs exposed to increasing duration of HS (43°C). TBP served as loading control. **(C)** Western blot analysis of GAPDH, p38 and MEKK4 proteins in cytoplasmic (C) and nuclear (N) protein fractions from mESCs exposed to increasing duration of HS. TUBULIN and TBP served as loading controls for cytoplasmic and nuclear fractions respectively. **(D)** Representative confocal microscopy images of immunofluorescence staining in control and 60 min HS exposed mESCs. Nuclei were counterstained with DAPI (blue) and GADD45a was probed using anti-HA antibody (green). The zoomed panel shows the distribution of GADD45a under control and HS conditions. The white arrows in merged view show the GADD45a protein aggregates formed upon 60 min HS exposure. Scale bar ~20μm. **(E)** Violin plot shows quantification of nuclear GADD45a signal intensity using confocal microscopy images (~100 nuclei analysed per condition) of control and 60 min HS exposed mESCs. Statistical significance was tested with two-tailed, unpaired Student's t-test (****: p-value <0.0001). Expected molecular weight for GADD45a-3xHA-mAID: 37kDa; TBP: 37kDa; GAPDH: 35kDa, p38: 40kDa and MEKK4: 180kDa.

I investigated the nuclear translocation of GADD45a in mESCs under various stress conditions, to ascertain whether this translocation is unique to thermal stress. Western blot analysis of protein fractions upon exposure to other stress conditions (X-ray irradiation, cold shock at 4°C, and oxidative stress via hydrogen peroxide treatment) showed no GADD45a nuclear translocation (Figure 4-4 A-C). X-ray irradiation led to cytoplasmic increase in GADD45a levels but still had no impact on its nuclear levels (Figure 4-4 A). Hence, the nuclear translocation of GADD45a was specific to thermal stress conditions. However, it remains unclear whether the nuclear translocation of GADD45 proteins is to regulate cell-cycle progression during stress response or as I hypothesized for heat shock gene transcription regulation [130].

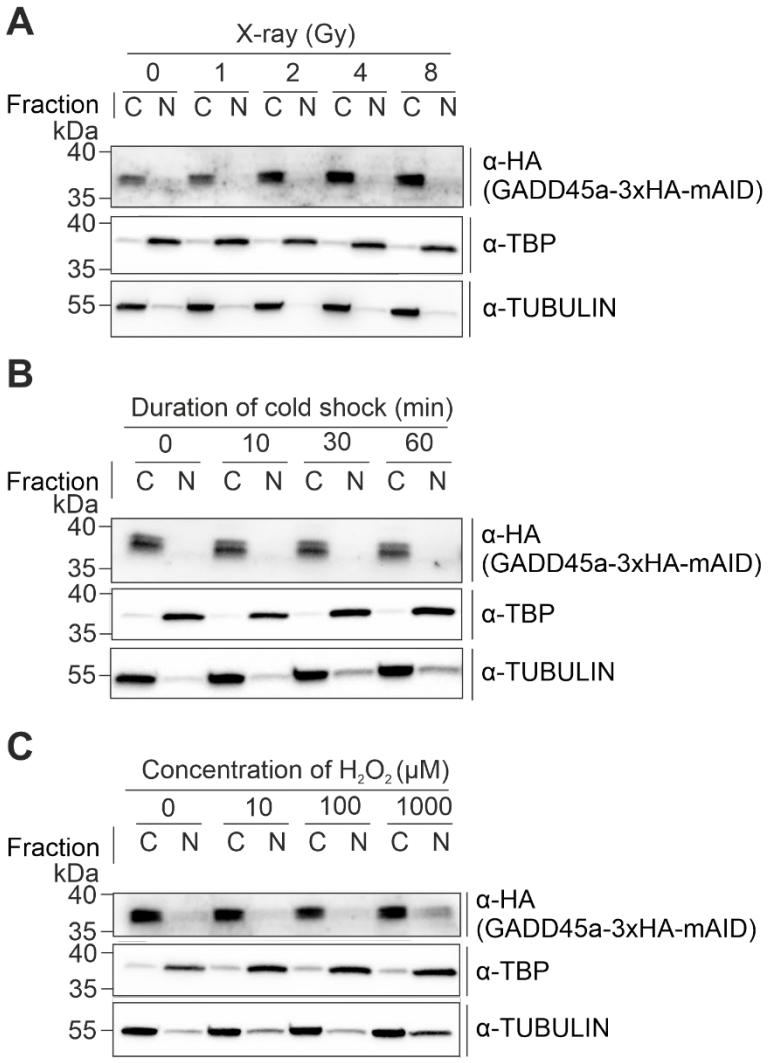

Figure 4-4: GADD45a does not show nuclear translocation upon exposure to other stress conditions in mESCs. (A) Western blot analysis of GADD45a protein in cytoplasmic (C) and nuclear (N) protein fractions from mESCs irradiated with X-rays for 10 min at indicated energy levels. **(B)** Western blot analysis of GADD45a protein in cytoplasmic and nuclear protein fractions from mESCs exposed to cold shock (4°C) for increasing duration. **(C)** Western blot analysis of GADD45a protein in cytoplasmic and nuclear protein fractions from mESCs exposed to increasing concentration of hydrogen peroxide (H_2O_2) to induce oxidative stress. TUBULIN and TBP served as loading controls for cytoplasmic and nuclear fractions respectively. Expected molecular weight for GADD45a-3xHA-mAID: 37kDa; TUBULIN: 55kDa; TBP: 37kDa.

4.3.2. *Gadd45* TKO mESCs show increased sensitivity to thermal stress

I used three independent mESC lines of wild-type (WT281, WT282, WT283, referred as WT in figures) and *Gadd45* TKO mESCs (TKO36, TKO227, TKO228, referred as TKO in figures) to test the functional significance of GADD45 family proteins during HSR [77]. The *Gadd45* TKO lines harbour homozygous deletions of exon 1 & 2 in all *Gadd45* genes. I tested all WT and *Gadd45* TKO mESC clones for their ability to survive thermal stress. The induction of apoptosis upon HS was quantified using the levels of cleaved CASPASE-3, with an antibody recognizing both cleaved (17 kDa) and full-length (35 kDa) forms of CASPASE-3 (Figure 4-5 A). In general, mESCs are highly resilient to thermal stress due to a high basal expression of heat shock genes [133]. In line, I did not observe any apoptosis in WT mESCs for up to 2 hrs of HS exposure (data not shown). Therefore, I performed longer HS exposure (3hrs) followed by a 2 hrs recovery at 37°C. Under these conditions and compared to control (Ctrl) conditions, I detected in WT mESCs cleaved CASPASE-3 levels that were, interestingly, significantly elevated in *Gadd45* TKO mESCs (Figure 4-5 A-B).

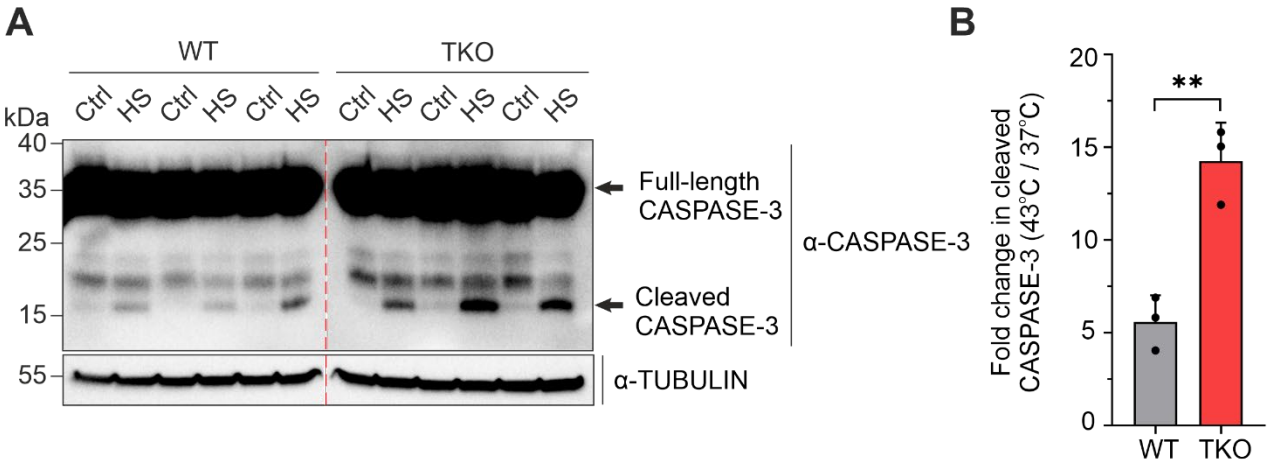

Figure 4-5: *Gadd45* TKO mESCs show increased apoptosis levels under thermal stress. (A) Western blot analysis of CASPASE-3 protein (cleaved and full-length) from whole cell lysates of WT and *Gadd45* TKO mESCs exposed to control (Ctrl) or HS (3hr at 43°C followed by 2hr recovery at 37°C) condition. TUBULIN served as loading control. **(B)** Bar graph represents the fold change in cleaved CASPASE-3 levels upon HS exposure in WT and *Gadd45* TKO mESCs. Statistical significance was tested with two-tailed, unpaired Student's t-test (**: p-value < 0.01). Expected molecular weight for cleaved CASPASE-3: 17kDa; full-length CASPASE-3: 35kDa; TUBULIN: 55kDa.

4.3.3. *Gadd45* TKO mESCs show reduced HSF1 occupancy at heat shock gene promoters

Unpublished GADD45a CUT&Tag analysis from our lab showed that GADD45a binds in the promoter of heat shock genes (Figure 4-1 C). Additionally, my data showed GADD45 nuclear translocation upon HS and increased sensitivity to thermal stress in *Gadd45* TKO mESCs, indicative of impaired HSR (Figure 4-2, 4-5). I hypothesized that GADD45 promotes HSF1

occupancy at its target heat shock gene promoters to initiate an efficient HSR. To test this hypothesis, I performed HSF1 ChIP-seq in Ctrl and 10 min HS conditions to identify differential HSF1 occupancy sites between WT and *Gadd45* TKO mESCs (Figure 4-6 A). I chose 10 min HS exposure since GADD45 proteins translocated to nucleus in the early phase of HSR, and short HS exposure excluded any apoptosis-induced artefacts (Figure 4-2 B, Figure 4-5).

ChIP-seq revealed very few HSF1 peaks under Ctrl conditions in both WT (40 peaks) and *Gadd45* TKO mESCs (39 peaks). These HSF1 peaks under Ctrl condition were situated primarily in distal intergenic regions in both WT and *Gadd45* TKO mESCs (Figure 4-6 B). On the contrary, under 10 min HS condition WT mESCs (738 peaks) showed a significantly higher number of HSF1 peaks compared to *Gadd45* TKO mESCs (164 peaks). These HSF1 peaks were primarily situated in promoters at gene transcription start sites (Figure 4-6 B). Furthermore, *Gadd45* TKO mESCs showed 226 peaks, primarily in heat shock gene promoters, with significantly reduced HSF1 occupancy compared to WT mESCs at 10 min HS (FDR p-value < 0.001). For instance, heat shock genes bound by GADD45a and with promoter R-loop (*Hsp90aa1* and *Hspa1a*) showed a significantly reduced signal for HSF1 occupancy in *Gadd45* TKO mESCs at 10 min HS (Figure 4-6 C, D).

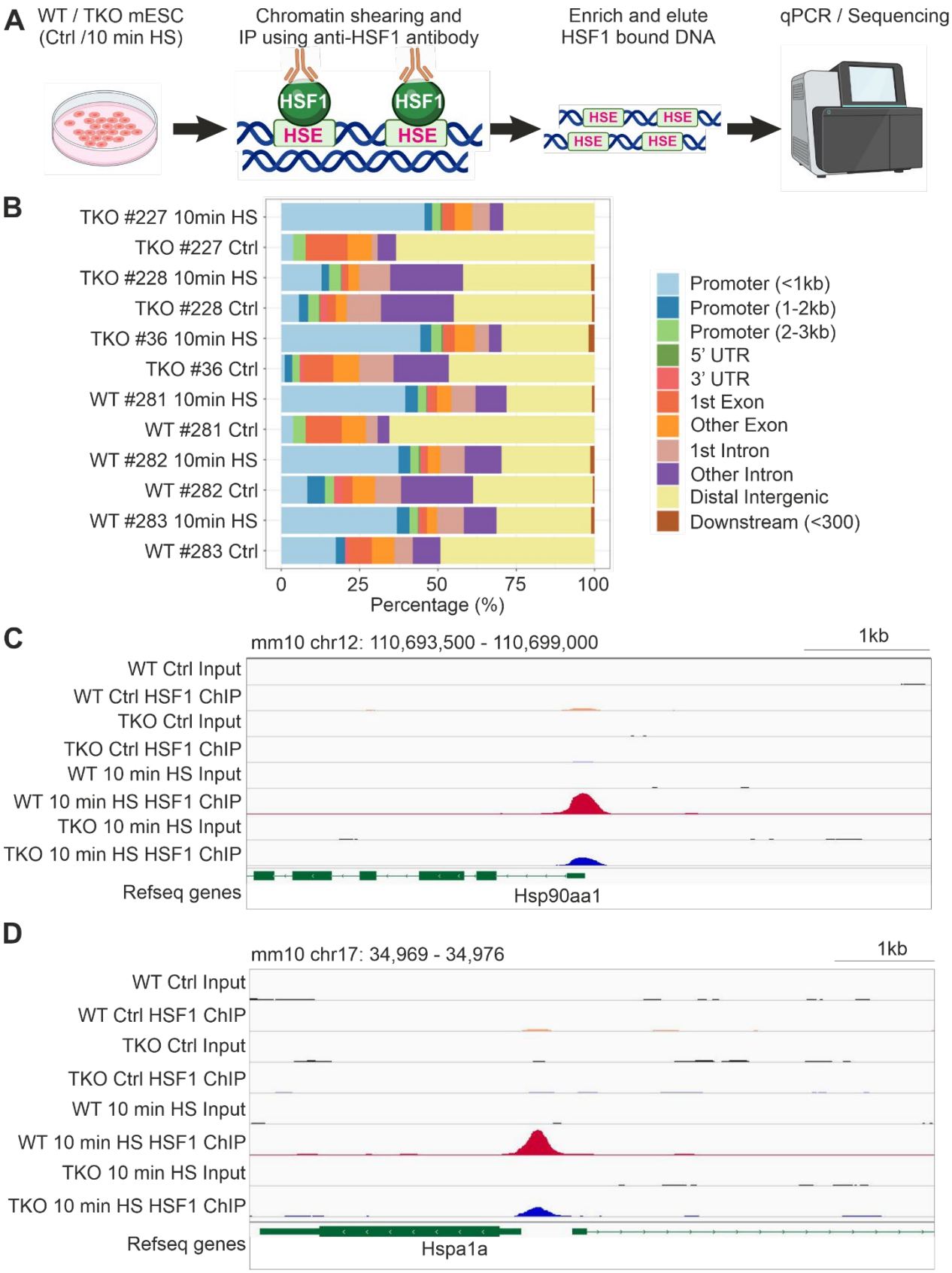

Figure 4-6: HSF1 shows reduced occupancy at target genes in *Gadd45* TKO mESCs at 10 min HS exposure. (A) Schematic represents the workflow for HSF1 ChIP performed in WT and *Gadd45* TKO mESCs under Ctrl or 10 min HS condition. **(B)** Bar graph represents the percentage of the HSF1 ChIP-seq peaks belonging to specific genomic features identified in individual WT and *Gadd45* TKO mESC clones

under Ctrl or 10 min HS condition. **(C)** & **(D)** Representative genome browser snapshot of HSF1 ChIP-seq performed in WT (WT #281) and *Gadd45* TKO (TKO #228) mESCs under control or 10 min HS condition centred at *Hsp90aa1* and *Hspa1a* gene promoters.

I validated the ChIP-seq results at a few representative HSF1 target gene promoters with R-loop bound GADD45a (*Hsp90ab1*, *Dnajb1* and *Hspa1b*). ChIP-qPCR showed no significant enrichment for HSF1 binding at these target loci under the Ctrl conditions, with enrichment comparable to background intergenic region (*Int3*), which served as a negative control (Figure 4-7 A). ChIP-qPCR confirmed the sequencing results showing significantly lower enrichment for HSF1 in *Gadd45* TKO compared to WT mESCs under 10 min HS (Figure 4-7 A).

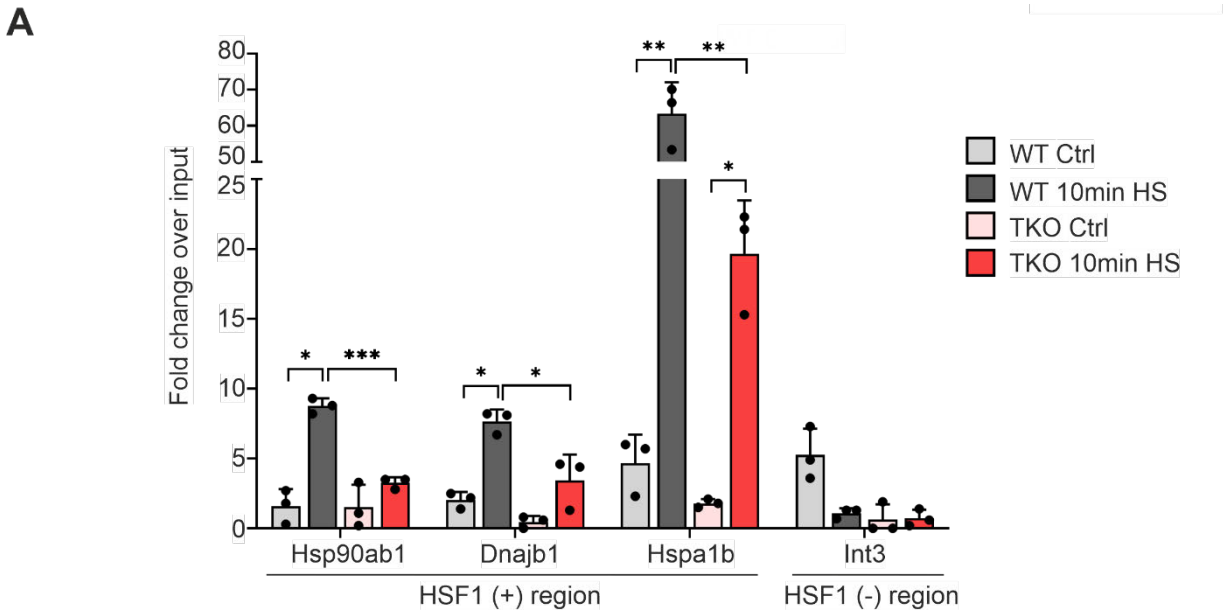

Figure 4-7: HSF1 ChIP-qPCR in WT and *Gadd45* TKO mESCs under Ctrl or 10 min HS condition. (A) Bar graph shows the ChIP-qPCR results from WT and *Gadd45* TKO mESCs under Ctrl and 10 min HS condition. Primers specific to heat shock gene promoters (*Hsp90ab1*, *Dnajb1* and *Hspa1b*) served as positive control. *Int3* served as negative control for the HSF1 ChIP-qPCR. Statistical significance was tested with two-tailed, unpaired Student's t-test (*: p-value <0.05; **: p-value <0.01; ***: p-value <0.001).

Since HSF1 ChIP-seq showed a substantial reduction in number of HSF1 peaks in *Gadd45* TKO mESCs, I analysed the sub-cellular distribution and total protein levels of HSF1 in WT and Gadd45 TKO mESCs. Immunofluorescence assay in 10 min HS exposed WT and *Gadd45* TKO mESCs showed similar nuclear intensity for HSF1 (red channel) (Figure 4-8 A). Hence, the impaired HSF1 occupancy in *Gadd45* TKO mESCs was not due to differential sub-cellular localization of HSF1 protein. Additionally, western blot analysis also confirmed that WT and *Gadd45* TKO mESCs have similar levels of HSF1 under Ctrl and 10 min HS conditions. This also excludes the possibility of differential HSF1 protein levels contributing to differential HSF1 peaks in *Gadd45* TKO mESCs (Figure 4-8 B). Lastly, western blot analysis of the transcriptionally competent HSF1 form (phosphorylated at Ser326) showed similar levels between WT and *Gadd45* TKO mESCs under

Ctrl and 10 min HS (Figure 4-8 C) [115]. Taken together with HSF1 ChIP-seq, my data suggests that GADD45 proteins regulate HSF1 occupancy at heat shock gene promoters during HSR in mESCs.

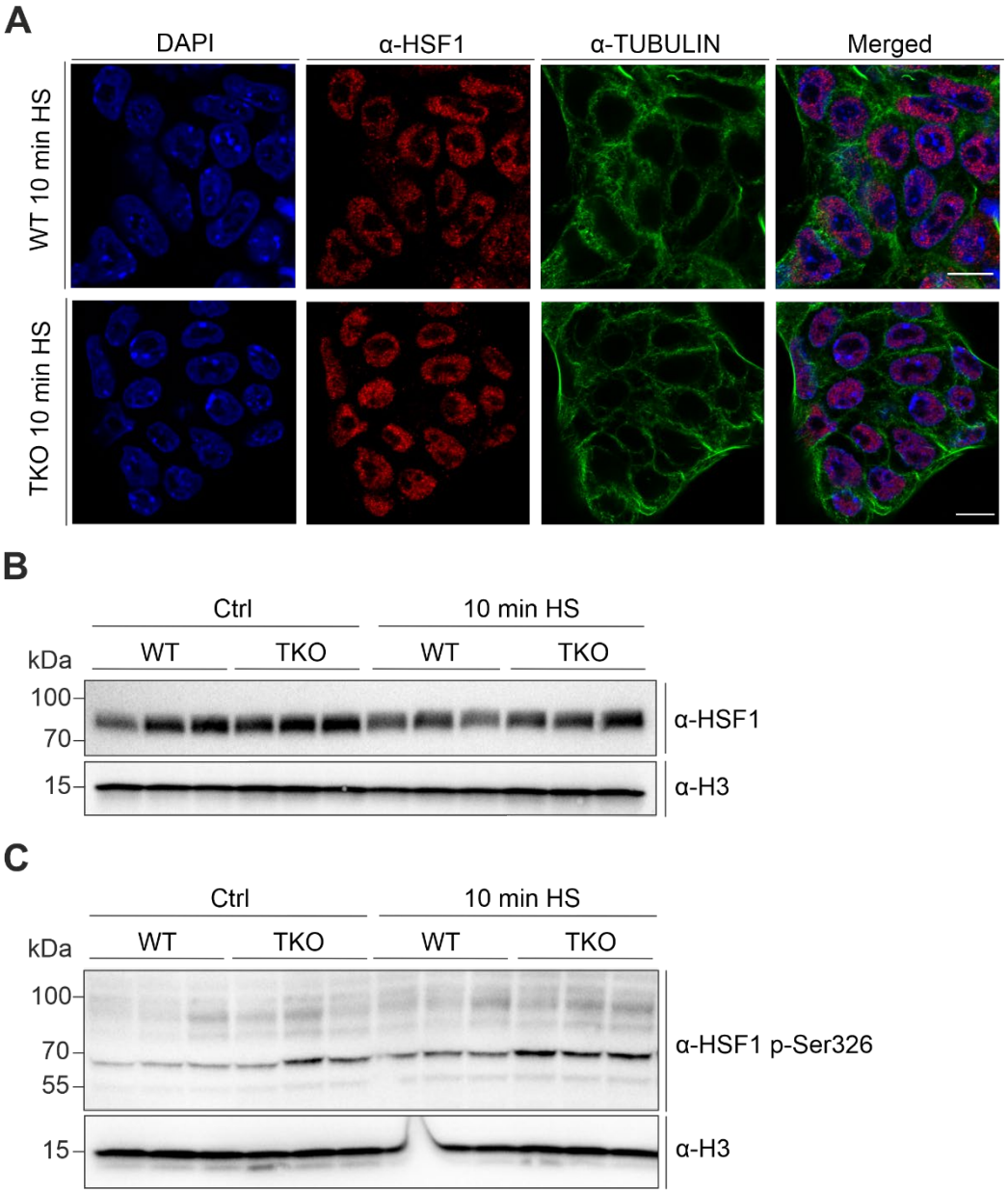

Figure 4-8: HSF1 shows similar cellular distribution and HS induced activation in WT and *Gadd45* TKO mESCs. (A) Representative confocal microscopy images of immunofluorescence staining in 10 min HS WT and *Gadd45* TKO mESCs. Nuclei were counterstained with DAPI (blue) while HSF1 (red) and TUBULIN (green) were probed using their respective antibodies. Scale bar ~20µm. **(B)** Western blot analysis of HSF1 protein from whole cell lysates of WT and *Gadd45* TKO mESCs exposed to Ctrl or 10 min HS. HISTONE H3 served as a loading control. **(C)** Western blot analysis of HSF1 protein phosphorylated at Ser326 from whole cell lysates of WT and *Gadd45* TKO mESCs exposed to Ctrl or 10 min HS conditions. HISTONE H3 served as a loading control. Expected molecular weight for HISTONE H3: 15kDa; HSF1: 57kDa, but migrates at 80kDa; pSer326 HSF1: 57kDa, but migrates at 70kDa.

4.3.4. Manipulating R-loop levels does not affect induction of heat shock genes

Next, I analysed whether the reduced occupancy of HSF1 at GADD45a bound heat shock gene promoters affects the transcription of these genes, leading to impaired HSR in *Gadd45* TKO

mESCs. I performed RNA-seq under Ctrl and 10 min HS conditions in both WT and *Gadd45* TKO mESCs. Both WT and *Gadd45* TKO mESCs showed reduced expression of histone genes, fitting with heat shock induced cell cycle arrest (data not shown) [134]. Interestingly, the downregulation of histone genes was comparable between WT and *Gadd45* TKO mESCs, suggesting that previously observed nuclear translocation of GADD45 proteins was likely not linked to cell cycle regulation upon HS exposure (Figure 4-2 B).

However, there were no differentially expressed heat shock genes between WT and *Gadd45* TKO mESCs under Ctrl or 10 min HS conditions (data not shown). It is important to note that most of the GADD45a bound heat shock genes are among the top expressed genes in mESCs and likely the qPCR / RNA-seq approach was not sensitive enough to measure 10 min HS induced changes. For example, *Hsp90aa1* and *Hsp90ab1* are among the top 10 highly expressed genes in mESCs (data not shown).

I therefore had to expose WT and *Gadd45* TKO mESCs for longer duration of HS without inducing apoptosis (1 hr HS). Additionally, I overexpressed RNASEH1-GFP fusion protein (RNASEH1 OE reduces R-loop levels globally) in WT and *Gadd45* TKO mESCs, to elucidate combined roles of R-loops and GADD45 proteins in regulation of heat shock gene transcription [82,95]. The RNASEH1-GFP positive (reduced R-loop levels) and RNASEH1-GFP negative cells (normal R-loop levels) were separated by flow cytometry from the transfected mESC pool (gating strategy summarized in methods section). This ensured that I analysed pure population of RNASEH1 overexpressing cells and minimized any transfection reagent induced changes to heat shock gene transcription. Sorted cell fractions were exposed to either Ctrl or 1 hr HS conditions in both WT and *Gadd45* TKO mESCs. This was followed with qPCR-based measurement of both mRNA and pre-mRNA (non-spliced transcript) levels of R-loop bound GADD45a target heat shock genes.

Both WT and *Gadd45* TKO mESCs showed similar levels of RNASEH1-GFP transfection efficiency (~45%) (Figure 4-9 A-B). qPCR analysis revealed a significant increase in pre-mRNA levels of heat shock genes at 1 hr HS compared to Ctrl condition in WT and *Gadd45* TKO mESCs. For instance, *Hspa1a* (encoding for HSP70 protein, positive control for heat shock gene induction) pre-mRNA levels increased by ~200 fold upon 1 hr HS compared to Ctrl conditions in both WT and *Gadd45* TKO mESCs (Figure 4-9 C).

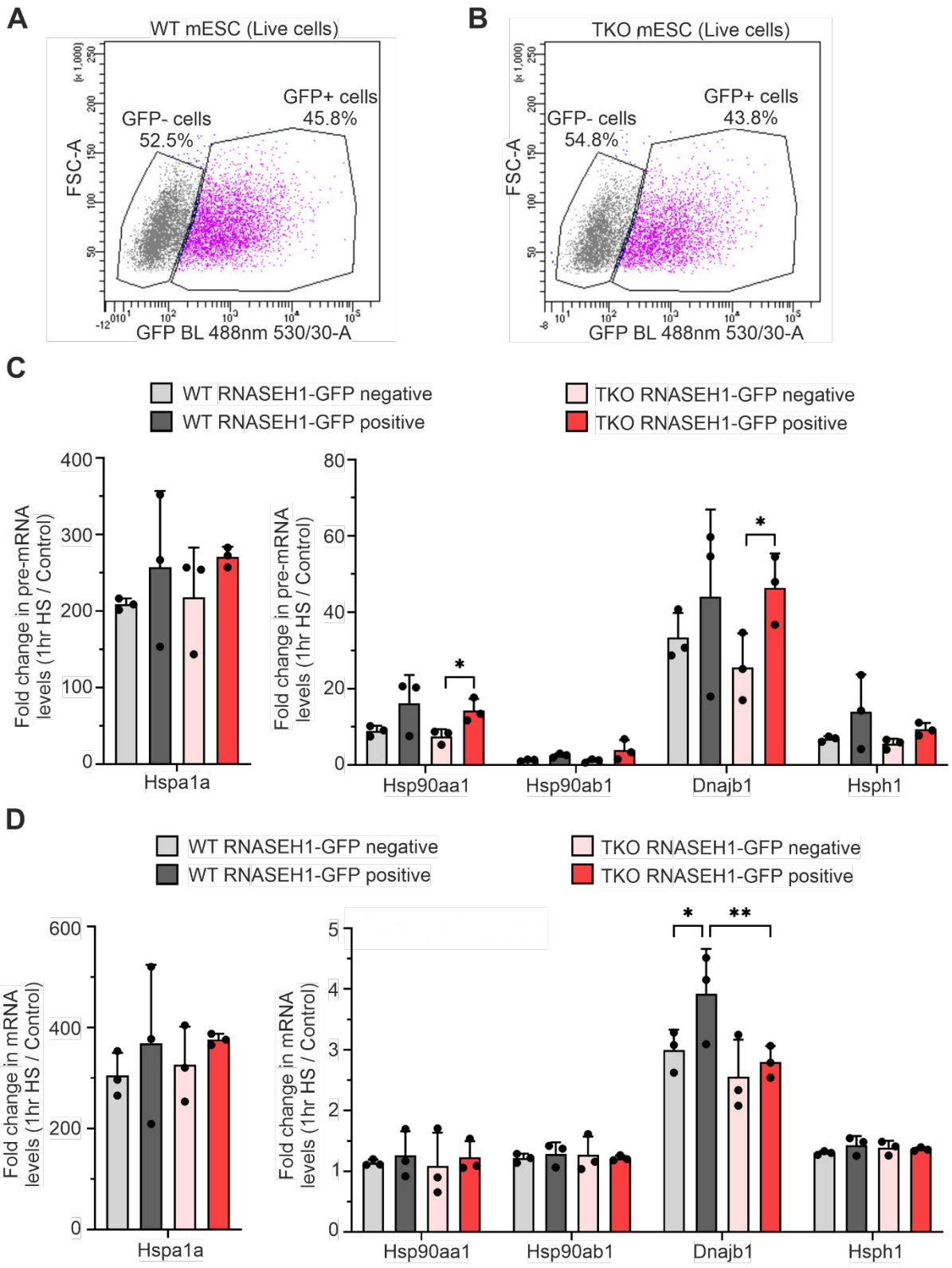

Figure 4-9: Heat shock gene expression is independent of R-loop and GADD45 in mESCs. (A) & (B) FACS gating strategy used to sort RNASEH1-GFP transfected WT and *Gadd45* TKO mESCs. GFP positive cells were detected after excitation with a blue laser (BL 488 nm), passing a 530/30 band pass filter. (FSC-A: Forward Scatter Area) **(C)** Bar graphs shows the fold change in heat shock gene pre-mRNA levels upon exposure to 1 hr of HS in RNASEH1-GFP positive or RNASEH1-GFP negative sorted WT and *Gadd45* TKO mESCs. **(D)** Bar graphs shows the fold change in heat shock gene mRNA levels upon exposure to 1 hr of HS in RNASEH1-GFP positive or RNASEH1-GFP negative sorted WT and *Gadd45* TKO mESCs. Statistical significance was tested with two-tailed, unpaired Student's t-test (*: p-value <0.05; **: p-value <0.01

The HS induced pre-mRNA levels were similar between RNASEH1-GFP positive and RNASEH1-GFP negative WT mESCs (Figure 4-9 C). On the contrary, RNASEH1-GFP positive cells showed significantly higher HS induced pre-mRNA levels compared to RNASEH1-GFP negative cells in *Gadd45* TKO mESCs (*Hsp90aa1* and *Dnajb1*). However, there were no differences in HS induced pre-mRNA levels at R-loop bound GADD45 target heat shock genes (*Hsp90aa1*, *Hsp90ab1*, *Dnajb1*, and *Hsph1*) between WT and *Gadd45* TKO mESCs (Figure 4-9 C).

As anticipated, the mature mRNA levels showed lower fold induction upon exposure to 1 hr HS compared to pre-mRNA levels (except *Hspa1a*, positive control for heat shock gene induction) (Figure 4-9 D). The HS induced mRNA levels of R-loop bound GADD45 target heat shock genes (*Hsp90aa1*, *Hsp90ab1*, *Dnajb1*, and *Hsph1*) were similar between WT and *Gadd45* TKO mESCs and independent of RNASEH1-GFP OE (except for marginal increase in *Dnajb1* mRNA levels upon RNASEH1-GFP OE in WT mESCs) (Figure 4-9 D). In collaboration with [redacted] (this lab), we stabilized R-loops using catalytically dead RNASEH1 (dRNASEH1-EGFP-dCas9) fusion protein OE [135]. However, this approach also did not yield any differences in expression of heat shock genes in WT and *Gadd45* TKO mESCs (data not shown). Additionally, I also tested whether GADD45 proteins regulate heat shock gene expression in thermal stress independent context in WT and *Gadd45* TKO mESCs. For instance, HSP90α chaperone activity inhibition using 17-AAG treatment was previously reported to induce HSR independent of thermal stress [136,137]. However, there were also no significant differences in mRNA levels of heat shock genes between WT and *Gadd45* TKO mESCs in these thermal stress independent conditions (data not shown).

In conclusion, R-loops and GADD45 proteins do not regulate heat shock gene expression in mESCs, despite the reduced HSF1 occupancy at the heat shock gene promoters observed in *Gadd45* TKO mESCs (Figure 4-6 C).

4.3.5. GADD45a CUT&Tag peaks are artefacts of anti-HA antibody's non-specificity

To understand the physiological relevance of nuclear translocation of GADD45 proteins, in collaboration with [redacted] we performed GADD45a CUT&Tag under Ctrl and 10 min HS conditions. However, there was no difference in GADD45a occupancy at target sites identified previously under Ctrl conditions (data not shown). Furthermore, no additional sites were bound by GADD45a under 10 min HS conditions, despite the substantial increase in nuclear levels upon 10 min HS (data not shown). In parallel, I performed Gadd45a CUT&Tag to identify R-loop dependent GADD45a binding sites using in-vitro RNaseH treatment during CUT&Tag procedure. I realized that an important control for the non-specific chromatin binding of the anti-HA antibody used to map GADD45a binding sites was not included in earlier CUT&Tag sequencing experiments. Therefore, I used E14 (untagged) mESCs to exclude any false positive peaks contributed by anti-

HA antibody non-specificity. Only genuine GADD45a binding sites should be significantly enriched in *Gadd45a-3xHA-mAID* mESCs over E14 (untagged) mESCs. Surprisingly, CUT&Tag qPCR results from GADD45a-3xHA-mAID mESCs showed no enrichment over E14 (untagged) mESCs at previously identified sites (Figure 4-1 B, C; Figure 4-10 A). Taken together with the lack of GADD45's role in regulation of heat shock gene expression in my previous qPCR experiments (Figure 4-9), these CUT&Tag qPCR results led me to reconsider the validity of the previously mapped GADD45a binding sites in mESCs. I hypothesized that all the peaks called against IgG negative control (including those at heat shock gene promoters) were artefacts of the anti-HA antibody's non-specific binding to certain DNA sequences or open chromatin regions [138].

I tested independent anti-HA antibodies (monoclonal, polyclonal from different host species), different detergent in CUT&Tag buffers (Tween20, NP40, TritonX-100), different DNA polymerases to amplify CUT&Tag libraries (Q5 polymerase, NEBNext Ultra II Q5 polymerase) and different cross-linkers (Dimethyl adipimidate and formaldehyde). However, none of these combinations yielded an enrichment in *Gadd45a-3xHA-mAID* over E14 mESCs at heat shock gene promoters (data not shown).

In collaboration with [redacted] we generated new mESC lines encoding an in frame N-terminal 3xHA tag at the endogenous *Gadd45a* locus (*3xHA-Gadd45a* mESC), to exclude any antibody binding issues due to the bulky C-terminal 3xHA-mAID tag in *Gadd45a-3xHA-mAID* mESCs. These new tagged *3xHA-Gadd45a* mESCs also showed nuclear translocation of GADD45a following 10 min of HS, suggesting that nuclear translocation of GADD45 proteins upon HS exposure is independent of the protein tag (Figure 4-10 B). In collaboration with [redacted] (this lab), we performed GADD45a CUT&Tag sequencing in 3x*HA-Gadd45a* mESCs but this time using E14 (untagged) mESCs as control for peak calling (Lane1 and Lane2; Figure 4-10 C). There were no peaks identified in this CUT&Tag sequencing, except for few peaks in blacklisted repetitive regions. Lastly, I performed ChIP-seq using three different anti-HA antibodies in 3xHA-*Gadd45a* and E14 (untagged) mESCs (Lane 3 and 6: Abcam Cat. no.ab9110; Lane 4 and 7: BioLegend Cat. no. 901501; Lane 5 and 8: Epicypher Cat. no. 13-2010) (Figure 4-10 C). ChIP-seq also did not yield any peaks for GADD45a binding sites over E14 (untagged) mESCs, further supporting the hypothesis that previously identified GADD45a CUT&Tag peaks were anti-HA antibody artefacts (Figure 4-10 C). The lack of GADD45a bound sites in mESCs could be due to multiple factors such as low GADD45 protein levels in mESCs, transient or indirect binding to chromatin.

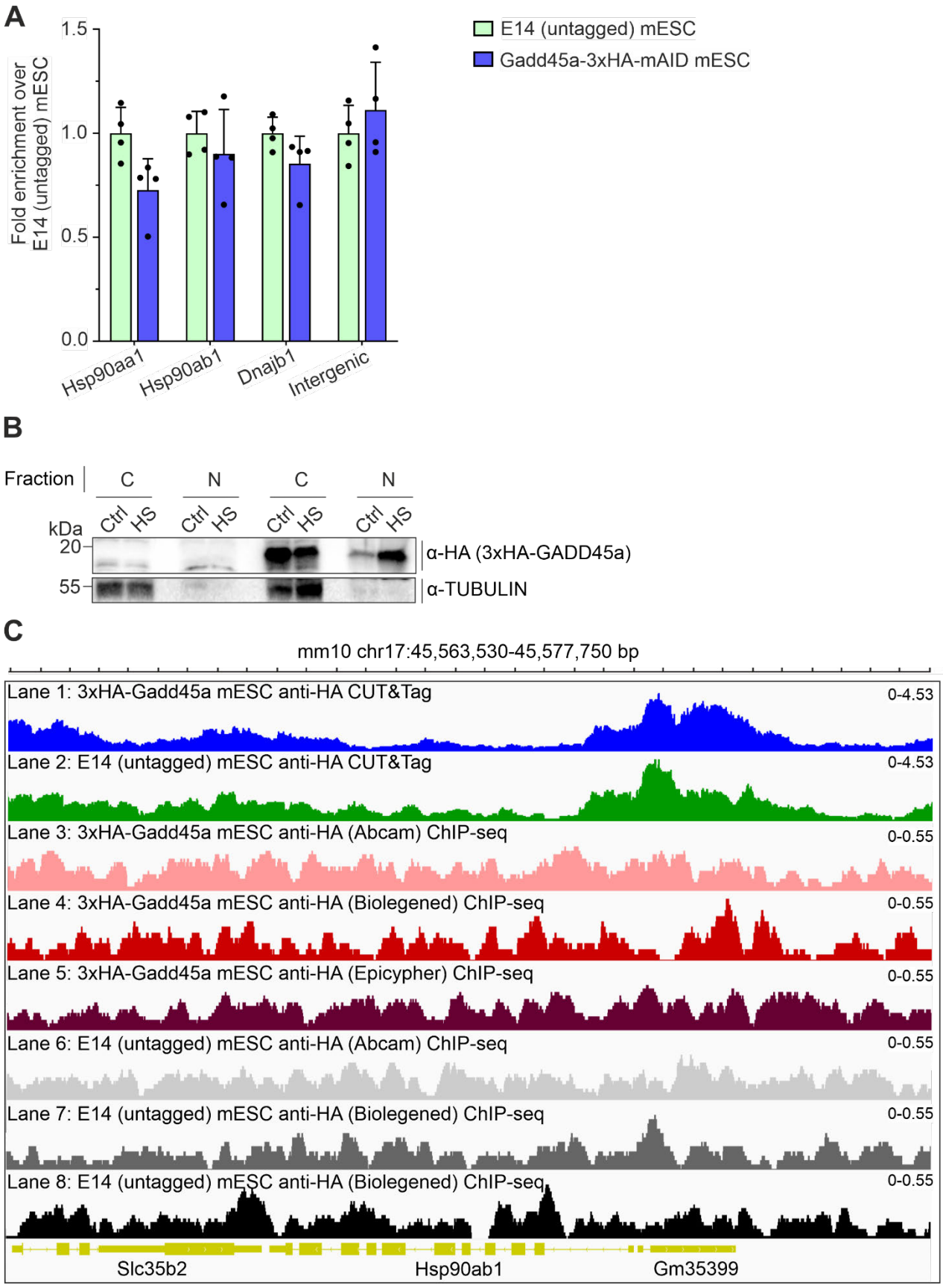

Figure 4-10: Analysis of GADD45a CUT&Tag and ChIP-seq peaks. (A) Bar graph shows the CUT&Tag qPCR results performed in E14 (untagged) and *Gadd45a-3xHA-mAID* mESCs. The fold enrichment over E14 mESCs is presented called heat shock gene promoter peaks. Primers specific to heat shock gene promoters (*Hsp90aa1*, *Hsp90ab1*, *Dnajb1*) were used to validate the previously identified GADD45a peaks.

Intergenic primer served as negative control for the CUT&Tag qPCR, where no GADD45a peak was identified. **(B)** Western blot analysis of GADD45a (~20kDa) protein in cytoplasmic and nuclear protein fractions from 3xHA-GADD45a mESCs exposed to Ctrl or 10 min HS conditions. TUBULIN (~55kDa) served as cytoplasmic (C) loading control. **(C)** Genome browser snapshot of GADD45a CUT&Tag sequencing (Lane 1 and 2) and ChIP-seq (Lane 3 to 8) centred at *Hsp90ab1* gene locus. Expected molecular weight for 3xHA-GADD45a: 20kDa; TUBULIN: 55kDa.

In summary, I showed that GADD45 proteins translocate to the nucleus specifically upon exposure to HS, independent of the protein tag (C-terminal 3xHA-mAID or N-terminal 3xHA) (Figure 4-2 B, 4-10 B). GADD45 nuclear translocation was not required for inducing cell cycle arrest upon HS exposure, based on the RNA-seq data in WT and *Gadd45* TKO mESCs (data not shown). Furthermore, I showed that *Gadd45* TKO mESCs are sensitive to HS exposure (Figure 4-5) and have impaired HSF1 occupancy at heat shock gene promoters upon HS exposure (Figure 4-7 A). I also ruled out the possibility that changes in HSF1 protein levels was leading to impaired HSF1 occupancy in *Gadd45* TKO mESCs (Figure 4-8). However, the heat shock gene transcription upon HS exposure did not correlate with the impaired HSF1 occupancy in *Gadd45* TKO mESCs (Figure 4-9 C, D). Manipulating R-loop levels via global erasure or stabilization did not yield any effects on expression of heat shock genes (Figure 4-9 C, D). Notably, I disproved the former GADD45a CUT&Tag results (Figure 4-1 B, C) by using appropriate controls for antibody non-specificity in CUT&Tag and ChIP-seq experiments (Figure 4-10 A, C). In retrospect, the unaffected expression of heat shock genes in *Gadd45* TKO mESCs, could be attributed to lack of GADD45 protein occupancy at these heat shock gene promoters. As I was specifically interested in studying direct transcription regulatory roles of GADD45 proteins, I did not pursue this line of research further. Therefore, the mechanistic basis of my novel findings such as impaired HSF1 occupancy in *Gadd45* TKO mESCs and the physiological relevance of nuclear translocation of GADD45 proteins upon HS exposure remains to be studied.

Part - II

5. Role of GADD45 proteins in embryonic heart development

5.1. Introduction

5.1.1. Mouse embryonic heart development

The heart is the first organ to become functional during mouse embryonic development and develops through the coordinated action of various cell types, each contributing to its structure and function (Figure 5-1). The embryonic heart originates from mesodermal cardiac progenitor cells (CPCs), mapped to the anterior region of the primitive streak during gastrulation [139–141].

In the mouse embryo at embryonic day 7.0 (E7.0), these cardiac precursors begin migrating in response to endodermal signals, moving anteriorly and laterally to create a crescent shape in the cephalo-medial region [142,143]. This crescent comprises two progenitor populations: the first heart field (FHF) and the second heart field (SHF) [144]. The FHF forms the left ventricle and some atrial tissue, while the SHF contributes to the right ventricle, outflow tract, interventricular septum (IVS), and atria [145]. By E8.0, cells from the FHF coalesce at the embryo's midline, forming the primitive heart, an endothelial tube surrounded by a single layer of myocardial cells [139]. The primitive heart tube subsequently enlarges by incorporating a second group of progenitors that migrate toward both the arterial pole (anteriorly) and the venous pole (posteriorly), contributing to the future outflow tract, right ventricle, and parts of the atria [146,147]. Concurrently, the heart tube undergoes a rightward looping, transforming the initial antero-posterior axis into a left-right orientation. During this looping process, differential growth and specific cell shape changes in designated myocardial regions, regulated by a complex interplay of genetic and mechanical factors, determine the myocardium of the future heart chambers [148]. The process results in the rightward bending of the tube, positioning the future atria and ventricles appropriately [139]. Following the looping, these myocardial regions expand to form the primitive ventricles, atria and trabeculae, a complex sponge-like network formed by ventricular cardiomyocytes. The septation of the heart into four chambers starts with the formation of endocardial cushions in the atrioventricular (AV) canal and outflow tract regions around E9.0 [149]. The E13.5 stage is characterized by the maturation of the heart chambers, the formation of the septa, and the refinement of the conduction system [150].

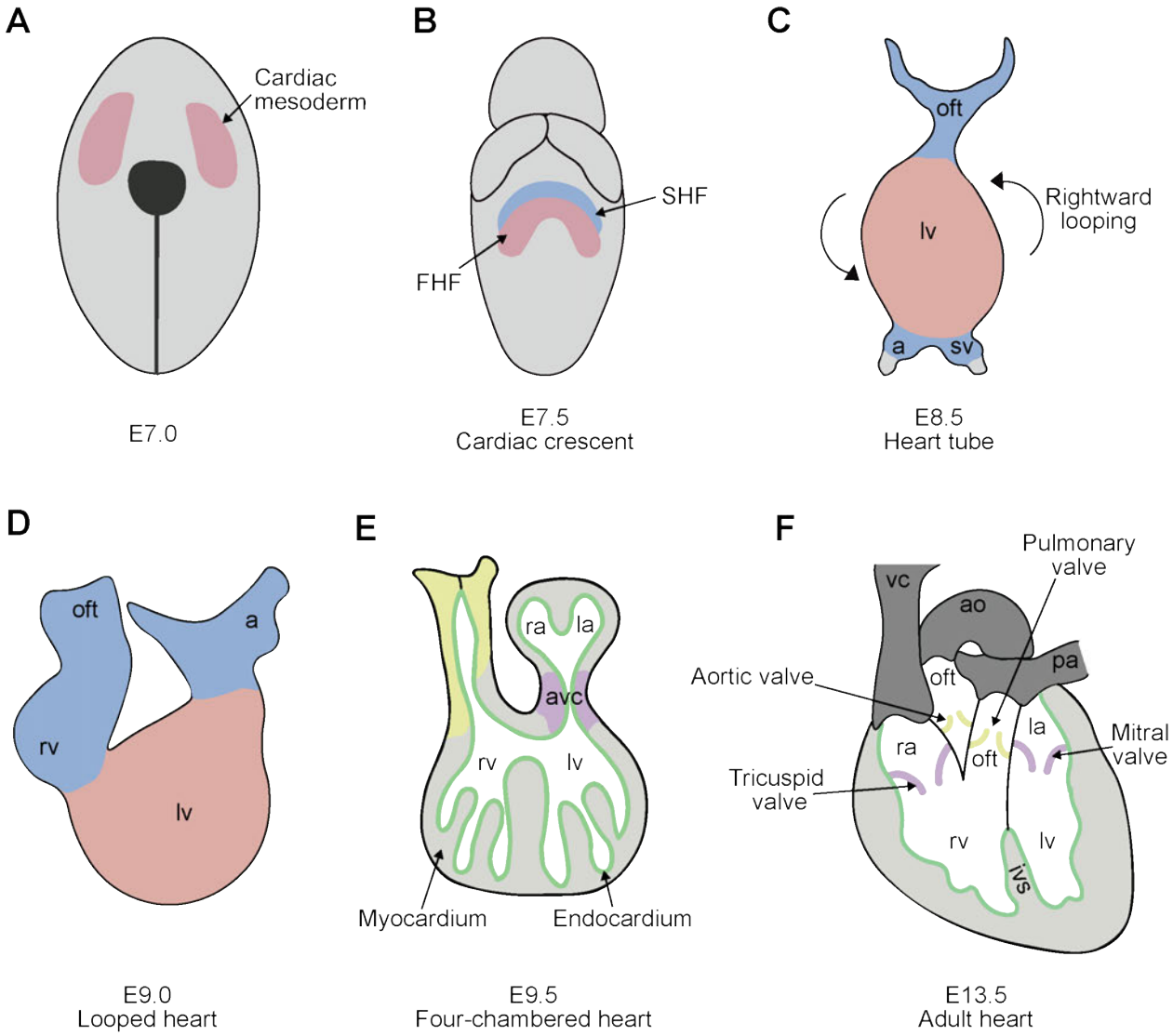

Figure 5-1: Schematic for stages of mouse embryonic heart development. **(A)** At E7.0, mesoderm-derived cardiac progenitors (shown in pink) migrate to the anterior region of the embryo to form the cardiac crescent. **(B)** By E7.5, there are two distinct cardiac progenitor populations: the FHF (pink) and the SHF (blue). **(C)** At E8.5, the heart tube undergoes rightward looping (indicated by arrows), **(D)** resulting in a looped heart structure at E9.0; outflow tract (oft), atrium (a), left ventricle (lv), right ventricle (rv), sinus venosus (sv). **(E)** A transversal section of an E9.5 heart shows four anatomically distinct regions: atrium [right atrium (ra), left atrium (la)], atrioventricular canal (avc, purple), ventricles (lv and rv), and the outflow tract (oft, yellow). Endocardial cells (green) outline the myocardium of the heart chambers (grey). **(F)** A transversal section of an E13.5 heart shows the four chambered heart with mature interventricular septum (ivs), valves and blood vessels [vena cava (vc), pulmonary artery (pa), aorta (ao)]. Figure adapted from [151].

5.1.2. Diverse cell types of an embryonic heart

Cardiomyocytes are the primary contractile cells of the heart. In the embryonic stage, they proliferate rapidly and differentiate to form the myocardium, the muscular layer of the heart responsible for pumping blood. The sarcomere structural proteins are fundamental to the function of cardiomyocytes, ensuring the heart's ability to contract efficiently and maintain its structural integrity [152]. For instance, ACTC1 is the cardiac-specific isoform of actin, which is a major

component of the thin filament in sarcomeres [153]. Actin filaments serve as tracks along which myosin heads move, enabling muscle contraction. MYH6 is the alpha heavy chain isoform of cardiac myosin and helps form the thick filaments of the sarcomere and interact with actin filaments to produce contraction through the ATP-dependent power stroke mechanism [154]. MYL7 is the atrial-specific isoform of myosin light chain, which are essential for modulating the interaction between actin and myosin heavy chains, thereby influencing the contractile properties of cardiomyocytes [155]. Additionally, TNNT2 is the cardiac isoform of Troponin T, a key component of the troponin complex involved in the regulation of muscle contraction [156]. The troponin complex, which also includes Troponin I and Troponin C, is responsible for translating calcium signals into mechanical force. Troponin T binds to tropomyosin, anchoring the troponin complex to the thin filament and facilitating the conformational changes required for muscle contraction and relaxation.

Endothelial cells line the interior of blood vessels, including the coronary vessels of the heart. During embryonic development, they originate from mesodermal progenitors and form the endocardium, the inner lining of the heart chambers [157]. Endothelial cells are essential for angiogenesis, the formation of new blood vessels, which is crucial for supplying oxygen and nutrients to the growing embryo [157]. Smooth muscle cells in the embryonic heart are primarily involved in the formation of the coronary arteries and the smooth muscle layer of the outflow tract. These cells ensure the structural stability and proper functioning of the blood vessels, regulating blood flow and pressure within the developing heart [158].

Cardiac fibroblasts play a significant role in the structural integrity and extracellular matrix (ECM) production within the heart [159]. They arise from the epicardium and contribute to the formation of the cardiac fibrous skeleton, providing mechanical support and facilitating electrical insulation between different parts of the heart. Fibroblasts also secrete growth factors and cytokines that influence cardiomyocyte behaviour [160]. Interestingly, neural crest cells, multipotent migratory cells originating from ectoderm, also contribute to embryonic heart development during the formation of the cardiac outflow tract and the septation of the heart chambers [161].

The development of the mouse embryonic heart therefore is a highly coordinated process involving the specification, differentiation, and morphogenesis of various cell populations. Key TFs and signalling pathways guide these processes, with disruptions often leading to congenital heart defects [162].

5.1.3. TFs and signalling pathways regulating embryonic heart development

Embryonic heart development is associated with the expression of several key TFs, such as NKX2-5 (NK2 homeobox 5), GATA4 (GATA binding protein 4), TBX5 (T-box transcription factor 5), HAND1 (heart and neural crest derivatives expressed 1) and HAND2 (heart and neural crest

derivatives expressed 2) [162–164]. These TFs spatiotemporally regulate expression of target cardiac-specific genes to promote cell type specification, differentiation, and morphogenesis in the developing heart. For instance, NKX2-5 is one of the earliest markers of cardiac progenitor cells and is essential for the proper formation of the heart tube [165,166]. GATA4 is a zinc-finger transcription factor crucial for cardiac morphogenesis, which regulates formation of the cardiac septa and the atrioventricular cushions [167,168]. The basic helix-loop-helix TFs HAND1 and HAND2 are critical for the development of the ventricular chambers. HAND2 is predominantly expressed in the right ventricle, while HAND1 is expressed in the left ventricle [169,170]. TBX5 is a T-box transcription factor that plays a significant role in the formation of the cardiac conduction system, particularly in the development of the atrioventricular node and the His-Purkinje network [171,172].

Additionally, these TFs in co-ordination with signalling pathways also regulate embryonic heart development. For instance, TEAD1 (TEA Domain Transcription Factor 1) is a crucial component of the Hippo signalling pathway and is essential for myocardial differentiation and morphogenesis [173,174]. BMP (Bone Morphogenetic Protein) signalling, through factors like BMP2 and BMP4, promotes cardiac mesoderm differentiation [175]. Conversely, the Wnt signalling pathway has a biphasic role, initially inhibiting and later promoting cardiogenesis [176,177]. The RB1 (Retinoblastoma protein) is a key regulator of cell cycle progression and has been implicated in the proliferation of cardiac progenitor cells. RB1 controls the balance between proliferation and differentiation of these progenitor cells, ensuring proper myocardial growth and chamber formation [178]. ROBO1 (Roundabout Guidance Receptor 1) is primarily known for its role in neuronal guidance, but recent studies have elucidated its significance in the alignment of cardiac neural crest cells during septation of cardiac outflow tract [179,180]. Endocardial cushions serve as primordial structures for the development of the septa and valves. They are formed by the endothelial-to-mesenchymal transition (EMT) of endocardial cells, driven by signals such as TGF-β and Notch [181].

Mutations or altered expression of these TFs and signalling pathway proteins can lead to various forms of cardiomyopathy. For instance, loss of GATA4 function leads to hypoplasia of the ventricular chambers and defects in septation, indicating its essential role in the development of the heart chambers [167,182]. Mutations in NKX2-5 are associated with congenital heart defects such as atrial septal defects and tetralogy of Fallot, ventricular septal defect, pulmonary stenosis, right ventricular hypertrophy, and an overriding aorta [149,183]. Genetic studies have shown that the loss of HAND2 results in right ventricular hypoplasia, whereas HAND1 deficiency affects the development of the left ventricular structures [184]. Mutations in TBX5 result in Holt-Oram syndrome, characterized by atrial septal defects and abnormalities in the conduction system [185]. BMP2 knockout mice exhibit failure in heart development due to defects in mesoderm formation [186]. Disruptions in Wnt signalling components can lead to significant heart defects. For example,

mutations in WNT3a result in a lack of posterior somites and severe truncations in the embryonic axis, affecting heart development [187,188]. Mutations in NOTCH1 are linked to bicuspid aortic valve disease and calcific aortic valve disease [189,190]. Mutations or dysregulation of TEAD1 can result in congenital heart defects, including hypoplastic left heart syndrome and cardiomyopathies. In addition to single-gene mutations, chromosomal abnormalities such as Down syndrome (trisomy 21) are frequently associated with congenital heart defects (CHDs). Approximately 40-50% of individuals with Down syndrome have CHDs, with atrioventricular septal defects being the most common [191].

5.1.4. Changes in epigenetic landscape during embryonic heart development

The dynamic and reversible nature of epigenetic modifications allows for the precise control of gene expression (including expression of cardiac TFs) necessary for the proper formation and function of the heart [192–195]. Among these epigenetic mechanisms, DNA methylation and histone modification changes undergo extensive reprogramming during embryonic heart development [193,196–201]. Furthermore, the dynamic DNA methylation landscape also regulates the binding of TFs to their target gene promoters [198–202].

Transcriptome and DNA methylation analysis in cardiomyocytes shows increased expression and associated decreased CpG methylation at genes expressing the sarcomere proteins [203]. DNA demethylation of enhancers, containing motifs characteristic of cardiac-specific TFs factors such as GATA4 and NKX2-5, is essential for cardiomyocyte differentiation [197,202,204]. Additionally, large demethylated regions, correlating with active histone marks like H3K4me3 and H3K27ac, are present on entire gene bodies of cardiomyocyte marker genes, including *Myh6* and *Tnnt2* [194,196,205]. Studies have shown that histone deacetylase (HDAC) inhibitor treatment to ameliorate cardiovascular diseases by balancing histone acetylation levels essential for proper cardiac morphogenesis [195].

Interestingly, DNA methylation mediated genetic imprinting also plays crucial role in cardiac development [206–208]. The imprinted *Dlk1-Dio3* gene locus (~200 kb) contains a cluster of distinct regulatory ncRNAs known to associate with PRC proteins [209]. Therefore, DNA methylation-dependent expression of these ncRNAs plays crucial roles in establishing genome-wide H3K27me3 marks during cardiac development [210]. In line with role of H3K27me3, EZH2 (histone methyltransferase sub-unit of PRC) knockout mice exhibited hypoplastic ventricles, increased trabeculation, and septal defects [211]. These mice also showed upregulation of non-cardiac genes like *Pax6*, *Six1*, and *Isl1*, indicating role of PRC proteins in repressing inappropriate lineage gene expression. Lastly, cardiac DNA methylation landscape undergoes dynamic changes during

prenatal and postnatal development, primarily to turn off the transcription of developmental signalling pathways [212].

Cardiac-specific deletion of *Tet2* and *Tet3* in mice (*Tet2/3*-DKO) was shown to cause ventricular non-compaction cardiomyopathy (NCC) due to impaired DNA demethylation and altered recruitment of the epigenetic modifier Ying-Yang1 (YY1) [201]. Additionally, TET proteins were shown to control DNA methylation dynamics at sequential steps of human cardiac development, wherein knocking out TET proteins (*Tet* TKO) in human embryonic stem cells (hESCs) led to impaired cardiac mesoderm differentiation [213]. *Tet1*/2/3 TKO at E7.5 show impaired mesodermal migration and by E8.5, *Tet* TKO have no recognizable heart structures [46]. However, our mechanistic understanding on factors modulating the epigenetic landscape during embryonic heart development remains limited.

Notably, GADD45b and GADD45g were reported to play a role in cardiomyocyte apoptosis and heart failure [214,215]. GADD45b is frequently associated with hypoxia-induced apoptosis of cardiomyocytes and is used as a biomarker for hypoxia-induced damage in cardiomyocytes, both in-vitro and in animal models [216]. Furthermore, cardiomyocyte de-differentiation in the mouse model of ischemic injury is characterized by decrease in *Gadd45* mRNA levels, whereas re-differentiation is featured by increase in *Gadd45* mRNA levels [217]. Publicly available GeneChip data shows a significant decrease in *Gadd45b* mRNA levels in patients with hypertrophic cardiomyopathy (HCM) compared to healthy controls [218]. An association study demonstrated that the GADD45B haplotype (rs2024144–rs3783501) influences the thickness of the interventricular septum in patients with HCM [218]. *Gadd45g* mRNA levels are reduced in the hearts of E13.5 and E15.5 embryos from diabetic mice compared to their respective controls [219]. These associations suggest a role for GADD45 proteins in regulating cardiac development and homeostasis, however, the potential mechanism by which GADD45 proteins regulate embryonic heart development remains elusive.

To address this gap in knowledge, [redacted] (this lab) investigated the effects of GADD45a and TET1 depletion on mouse embryonic development (Figure 5-2) (unpublished data). GADD45a and TET1-deficient mice (*Gadd45a/Tet1* DKO) were shown to be embryonic sub-lethal (Figure 5-2 A), while the surviving embryos had midline abnormalities such as facial-cleft closure defect, spina bifida and curled tail during embryogenesis (data not shown). Interestingly, the *Gadd45a/Tet1* DKO embryos showed defects in embryonic heart development with reduced myocardium (mc) thickness in right ventricle (rv), altered right atrium (ra) morphology and reduced IVS thickness (Figure 5-2 B). These defects were not due to developmental delays, since the development of other organs (kidney, lungs and limbs) in the *Gadd45a/Tet1* DKO embryos was unaffected (data not shown). Additionally, *Gadd45a* SKO embryos did not show embryonic heart

development defects, while *Tet1* SKO embryos showed a trend for reduced myocardial thickness (Figure 5-2 B). These novel findings from in-vivo *Gadd45a*/*Tet1* DKO mice hinted at a synergistic functional interplay between GADD45a and TET1 proteins during embryonic heart development.

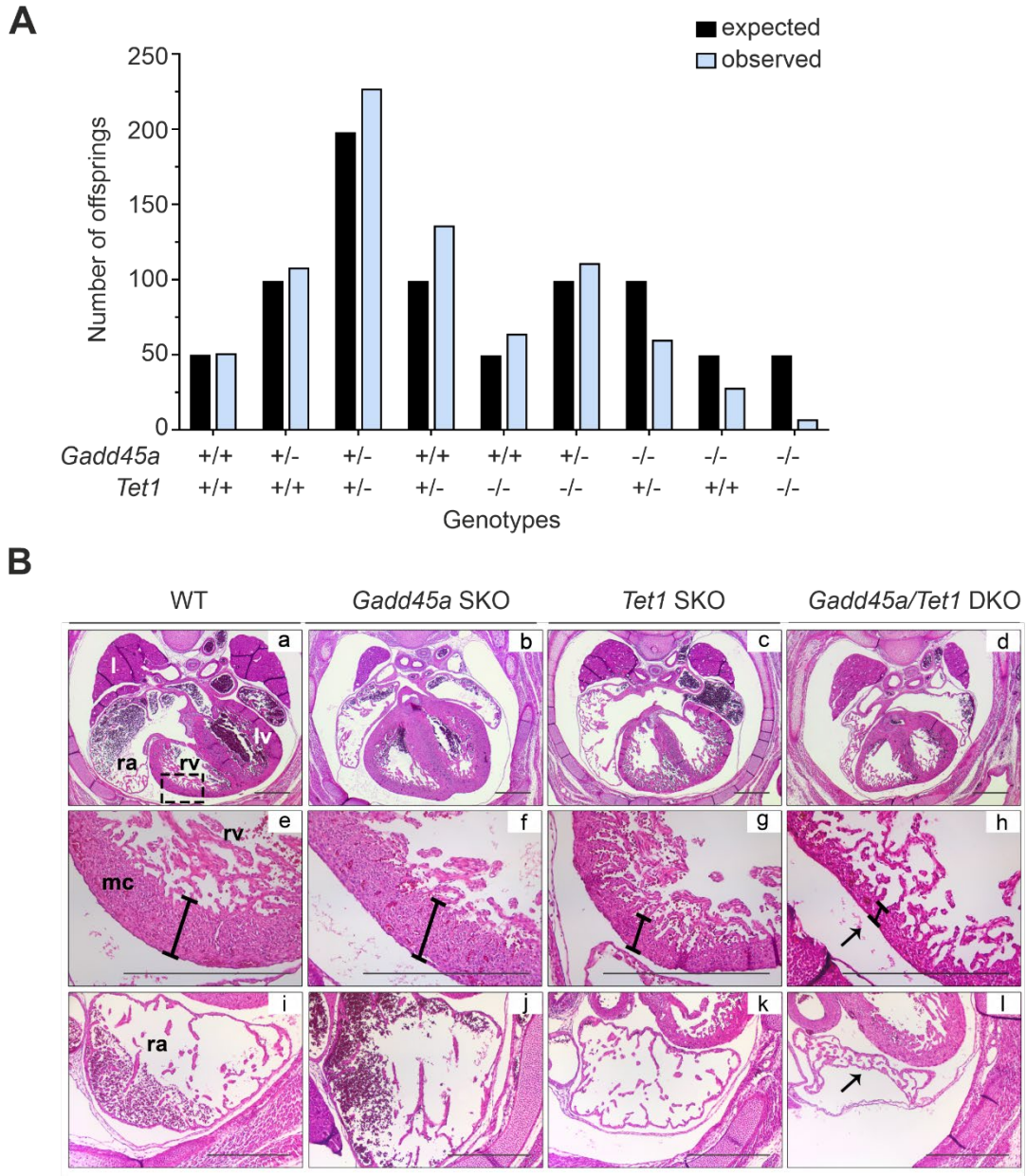

Figure 5-2: Analysis of *Gadd45a/Tet1* DKO mouse embryos. (A) Bar graph shows the expected vs. observed number of offspring from different *Gadd45a* and *Tet1* knockout genotypes. **(B)** (a-d) Representative H&E staining images of embryonic heart tissues collected at E15.5 stage from WT, *Gadd45a* SKO, *Tet1* SKO and *Gadd45a/Tet1* DKO mice (4x magnification) (e-h) Middle panel shows the myocardium (mc) thickness in right ventricle (rv) of WT, *Gadd45a* SKO, *Tet1* SKO and *Gadd45a/Tet1* DKO mice (20x magnification). (i-l) Bottom panel shows the right atrium (ra) of WT, *Gadd45a* SKO, *Tet1* SKO and *Gadd45a/Tet1* DKO mice (10x magnification).

5.2. Aim

Disruptions in epigenetic landscape affect transcription changes associated with normal heart development and correlate strongly with congenital heart defects (CHDs) and cardiovascular disorders (CVD), including dilated cardiomyopathy (DCM), ischemic cardiomyopathy (ICM), and heart failure (HF) [195,201,220–222]. In human cells, GADD45 proteins act as recruiters of TET1 and enhance their DNA demethylation activity [52,107]. However, the genome-wide prevalence and mechanism of how GADD45 proteins enhance TET-mediated DNA demethylation during embryonic development remains unclear. Furthermore, TET proteins regulate gene expression via catalytic (via DNA demethylation) and non-catalytic functions (via recruitment of with epigenetic modifiers such as PRC and SIN3a) [49,50,52,107,223]. It remains unclear whether GADD45 proteins play a role in balancing these contrasting TET protein functions. In part - II of my thesis, I aim to address these open questions using mESCs as the in-vitro system.

Unpublished studies in our laboratory showed synergistic roles for GADD45a and TET1 proteins during embryonic heart development. I explored the cause and the consequences of potential cardiac differentiation defects in *Gadd45* TKO (CDS) mESCs. I planned to perform RNA-seq in WT and *Gadd45* TKO (CDS) mESCs to elucidate transcription regulatory roles of GADD45 proteins. With [redacted] we planned to identify differential TET1 binding sites and differentially methylated regions (DMRs) between WT and *Gadd45* TKO (CDS) mESCs. Additionally, by cross-comparing with publically available *Tet1* knockout and *Tet1* catalytic mutant mESCs RNA-seq datasets, I planned to investigate GADD45 and TET1 functions in regulation of heart development-associated genes [224]. Due to lack of *Gadd45* TKO mice, I planned to establish an in-vitro cardiomyocyte differentiation assay using WT and *Gadd45* TKO (CDS) mESCs. I planned to perform gene expression analysis in these cardiac differentiated cells to identify any misregulated genes during cardiac differentiation defects in Gadd45 TKO (CDS) mESCs. I planned to perform scRNA-seq to elucidate the origin of potential cardiac differentiation defect in *Gadd45* TKO (CDS) mESCs.

5.3. Results

5.3.1. Characterization of *Gadd45* TKO (CDS) mESCs

I explored whether the embryonic heart development defects were due to an altered epigenetic landscape from depletion of GADD45 proteins. I hypothesized that GADD45 proteins regulate TET functions (catalytic or non-catalytic) during embryonic heart development. Given the redundancy of GADD45 proteins and lack of *Gadd45* TKO mice, we established a new set of *Gadd45* TKO mESCs to test our hypothesis (4 independent mESC clones, a kind gift from [redacted] and [redacted] this lab). This new set of *Gadd45* TKO mESCs harboured deletion of complete coding sequence (CDS) in all *Gadd45* genes, referred henceforth as *Gadd45* TKO (CDS) mESCs. This was in contrast to the previously published *Gadd45* TKO mESCs (also used in part-I of my thesis), which harbour only exon 1-2 deletions in *Gadd45* genes [77,225]. Importantly, the *Gadd45* TKO (CDS) mESCs harbour the *Gadd45* gene deletions similar to individual *Gadd45* SKO mice, ensuring that my in-vitro results from mESCs remain comparable with the in-vivo findings (Figure 5-2, 5-3 A) (Sanger sequencing data not shown).

Firstly, I characterized the newly generated WT and *Gadd45* TKO (CDS) mESCs by measuring the expression of *Gadd45a*,*b,g* genes using qPCR. The qPCR analysis showed no expression of *Gadd45a,b,g* genes in *Gadd45* TKO (CDS) mESCs (Figure 5-3 B). Next, I measured the expression of pluripotency genes (*Oct4*, *Sox2* and *Nanog*), to test whether GADD45 proteins regulate pluripotency in mESCs. There was no significant difference in expression of the pluripotency genes between WT and *Gadd45* TKO (CDS) mESCs (Figure 5-3 C).

Additionally, in collaboration with [redacted] (this lab), we measured the levels of DNA methylation (5mC) and its oxidized derivatives (5hmC, 5fC and 5caC) in WT and *Gadd45* TKO (CDS) mESCs (Figure 5-3 D). There was no significant difference in global 5mC levels between WT and *Gadd45* TKO (CDS) mESCs. On the contrary, all its oxidized derivatives (5hmC, 5fC and 5caC) showed a significant increase in *Gadd45* TKO (CDS) mESCs. These results suggest global changes in DNA demethylation activity in *Gadd45* TKO (CDS) mESCs.

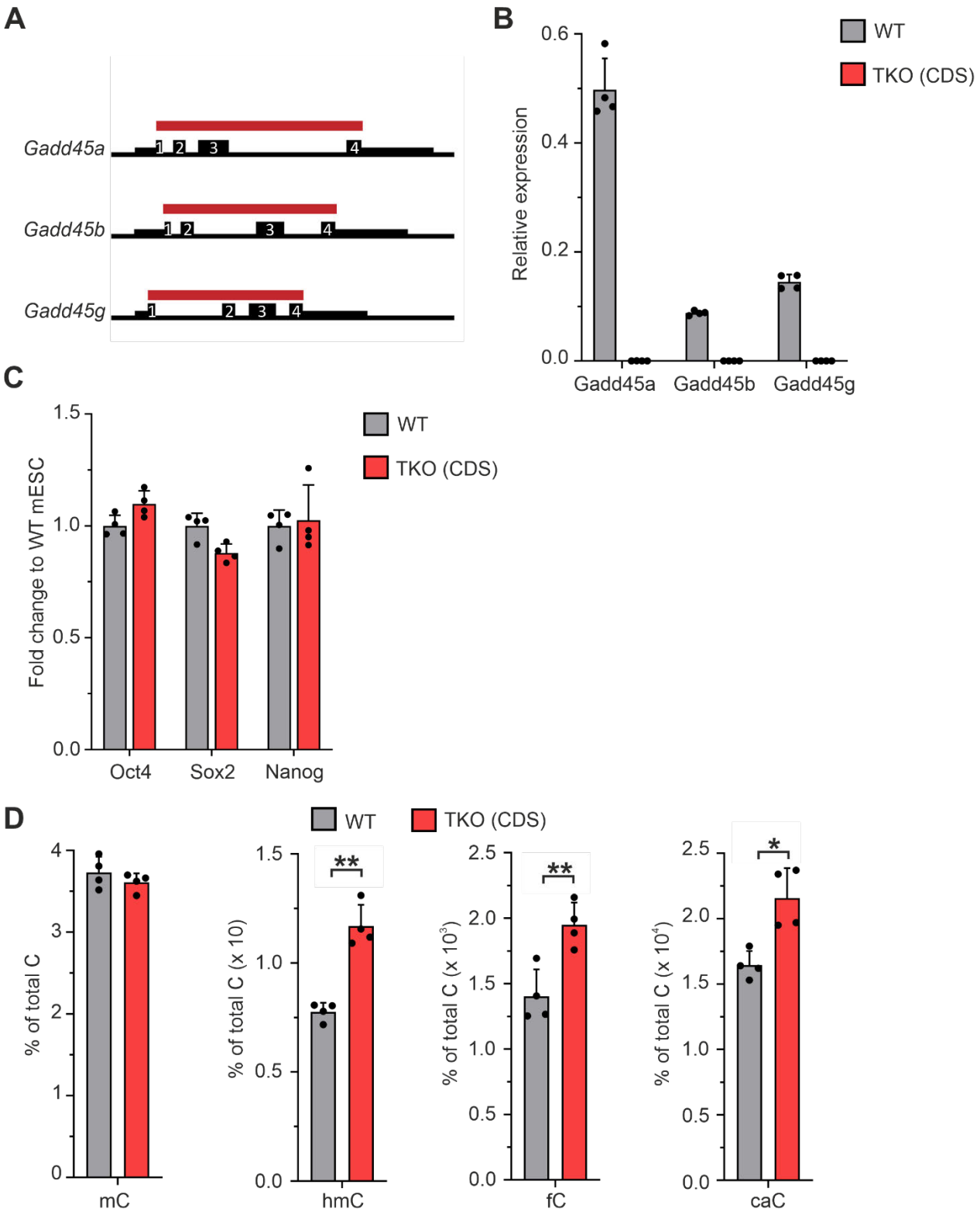

Figure 5-3: Characterization of *Gadd45* TKO (CDS) mESCs. (A) Schematic represents the deleted regions (red bars) in all three *Gadd45* genes in *Gadd45* TKO (CDS) mESCs. **(B)** Bar graph shows expression of *Gadd45* genes in WT and *Gadd45* TKO (CDS) mESCs. **(C)** Bar graph shows WT mESCs normalized expression of pluripotency genes (*Oct4, Sox2* and *Nanog*) in WT and *Gadd45* TKO (CDS) mESCs. **(D)** Bar graph shows the levels of modified cytosines (5mC, 5hmC, 5fC and 5caC) as percentage of total cytosines (C) in WT and *Gadd45* TKO (CDS) mESCs. Statistical significance was tested with two-tailed, unpaired Student's t-test (*: p-value <0.05; **: p-value <0.01; ****: p-value <0.0001). Relative expression was measured with respect to housekeeping genes, *Tbp* and *G6pd*.

5.3.2. *Gadd45* TKO (CDS) mESCs differentiate into all three germ layers

Next, I tested whether the *Gadd45* TKO (CDS) mESCs can efficiently differentiate into all three germ layers (endoderm, mesoderm and ectoderm), using embryoid body (EB) and retinoic acid (RA) treated EB (EB+RA) differentiation. qPCR analysis showed that WT and *Gadd45* TKO (CDS) mESCs differentiated similarly towards mesoderm (representative marker *Eomes*) and endoderm (representative marker *Hnf4a*) lineages in EB differentiation (Figure 5-4 A). Interestingly, *Gadd45* TKO (CDS) EB+RA samples showed increased expression of neuroectoderm markers (representative markers *Pax6, Hoxa2, Neurog1* and *Neurog2*) compared to WT EB+RA samples (Figure 5-4 A). These results showed that *Gadd45* TKO (CDS) mESCs can efficiently differentiate into all three germ layers during EB differentiation, however, have an increased bias towards neuroectoderm lineages.

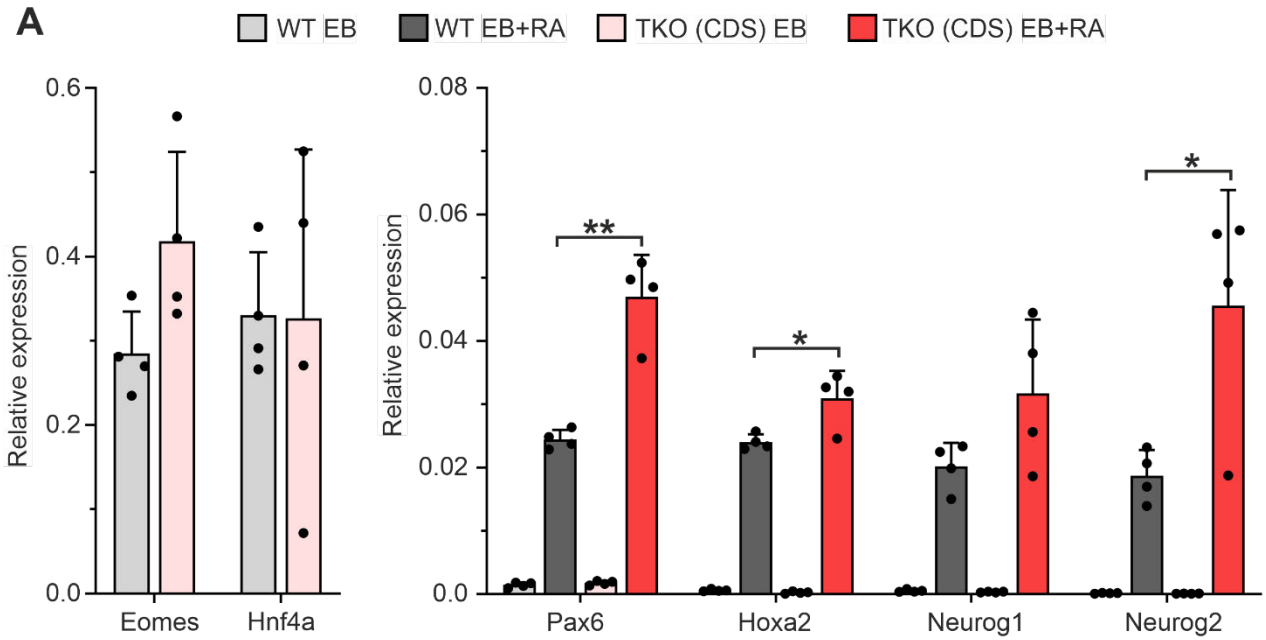

Figure 5-4: EB and EB+RA differentiation of WT and *Gadd45* TKO (CDS) mESCs. (A) Bar graph shows expression of representative marker genes for mesoderm (*Eomes*), endoderm (*Hnf4a*) and neuroectoderm (*Pax6, Hoxa2, Neurog1* and *Neurog2*) in EB and EB+RA differentiated WT and *Gadd45* TKO (CDS) mESCs. Statistical significance was tested with two-tailed, unpaired Student's t-test (*: p-value <0.05; **: p-value). Relative expression was measured with respect to housekeeping genes, *Tbp* and *G6pd*.

5.3.3. RNA-seq reveals reduced expression of heart development-associated genes

In order to understand the consequence of the deletion of *Gadd45* genes on gene transcription, I performed transcriptome analysis using RNA-seq in WT and *Gadd45* TKO (CDS) mESCs. The bioinformatics analysis was performed in collaboration with [redacted] (this lab).

RNA-seq analysis identified ~2000 differentially expressed genes (DEGs) between WT and *Gadd45* TKO (CDS) mESCs, with 1056 upregulated and 932 downregulated genes (FDR corrected p-value < 0.01) (Figure 5-5 A). It is important to note that the previous batch of *Gadd45*

TKO (with only exon1, 2 deletions) mESCs, had only 35 DEGs (FDR corrected p-value < 0.01) [77]. There was a marginal overlap of 11 DEGs between these two RNA-seq datasets (Figure 5-5 B). The 11 overlapping genes were related to MAPK signalling and did not contain any previously reported 2C (2-cell) like state-specific genes (data not shown) [77]. This dramatic difference in DEGs suggest that the *Gadd45* TKO mESCs (with only exon1, 2 deletions) were hypomorphs and *Gadd45* TKO (CDS) mESCs are likely complete functional *Gadd45* knockouts.

Gene Ontology (GO) analysis was performed to identify enriched biological processes among the upregulated (up DEGs) and downregulated (down DEGs) genes in *Gadd45* TKO (CDS) mESCs. The up DEGs were enriched primarily for broad metabolic terms related to ATP, amino acid and monosaccharide processes (Figure 5-5 C). Interestingly, the down DEGs were enriched for embryonic morphogenesis, heart development and blood vessel morphogenesis terms (Figure 5-5 D). I verified the downregulation of these heart development and blood vessel morphogenesis associated genes via qPCR (Figure 5-5 E). Furthermore, in collaboration with [redacted] and [redacted] (this lab), we performed assay for transposase-accessible chromatin with sequencing (ATAC-seq) to map open chromatin regions in WT and *Gadd45* TKO (CDS) mESCs. There were ~1000 differential ATAC peaks, of which ~130 overlapped with the RNA-seq DEGs from *Gadd45* TKO (CDS) mESCs (data not shown). Interestingly, GO analysis of genes associated with differential ATAC down peaks showed significant enrichment for cell migration, mesenchyme and outflow tract development related terms, in line with the RNA-seq down DEGs (data not shown).

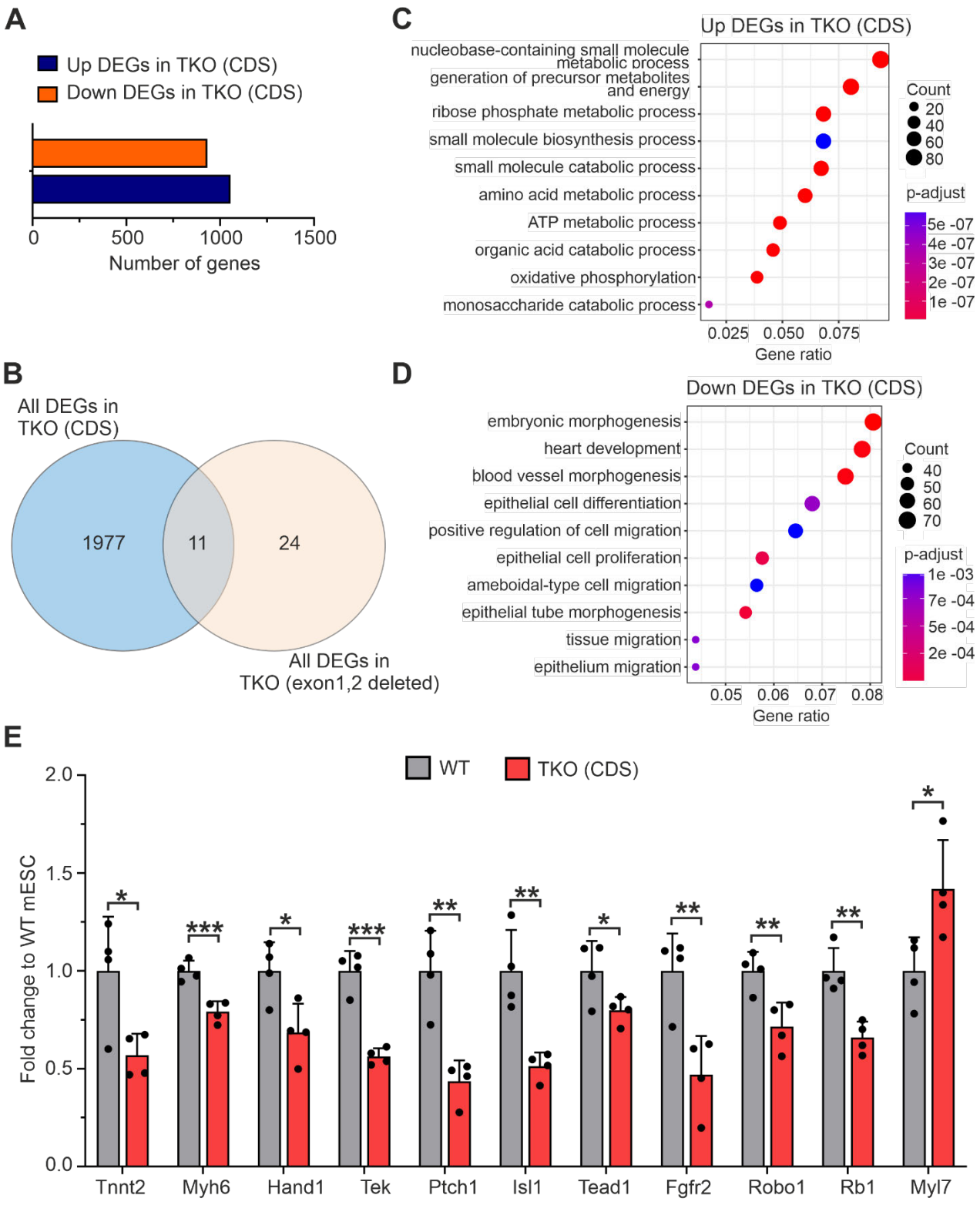

Figure 5-5: Analysis of RNA-seq from WT and *Gadd45* TKO (CDS) mESCs. (A) Bar graph shows the number of DEGs in *Gadd45* TKO (CDS) mESCs. **(B)** Venn diagram shows the overlap between DEGs of *Gadd45* TKO (CDS) mESCs and *Gadd45* TKO (with only exon1, 2 deletions) mESCs**. (C)** GO term analysis of up DEGs in *Gadd45* TKO (CDS) mESCs. **(D)** GO term analysis of down DEGs in *Gadd45* TKO (CDS) mESCs. **(E)** Bar graph showing the WT mESC normalized expression of heart development-associated genes in WT and *Gadd45* TKO (CDS) mESCs. The data is presented as fold change to WT mESCs. Statistical significance was tested with two-tailed, unpaired Student's t-test (*: p-value <0.05; **: p-value <0.01; ***: p-value <0.001). Relative expression was measured with respect to housekeeping genes, *Tbp* and *G6pd*.

5.3.4. Expression of heart development-associated genes is unaffected in *Gadd45* SKO (CDS) mESCs

In parallel to generation of *Gadd45* TKO (CDS) mESCs, we also generated *Gadd45* SKO (CDS) and matched WT parental mESCs (4 clones each, a kind gift from [redacted] this lab). As expected, *Gadd45a*, *Gadd45b*, *Gadd45g* genes were not expressed in the respective *Gadd45* SKO (CDS) mESCs (Figure 5-6 A). *Gadd45a* SKO (CDS) mESCs showed significantly lower expression of *Gadd45b* and *Gadd45g* genes compared to WT parental mESCs (Figure 5-6 A). Furthermore, I analysed for potential downregulation in expression of few representative heart development-associated genes in WT parental and *Gadd45* SKO (CDS) mESCs (directional or one-tailed hypothesis). Except for downregulation of *Rb1* in *Gadd45a* SKO (CDS) mESCs*,* there was no significant reduction in expression of heart development-associated genes in *Gadd45* SKO (CDS) mESCs (Figure 5-6 B). These results suggest that *Gadd45* genes exhibit functional redundancy in regulation of most heart development-associated genes, in line with absence of any embryonic heart development defects in *Gadd45a* SKO mice (Figure 5-2 B).

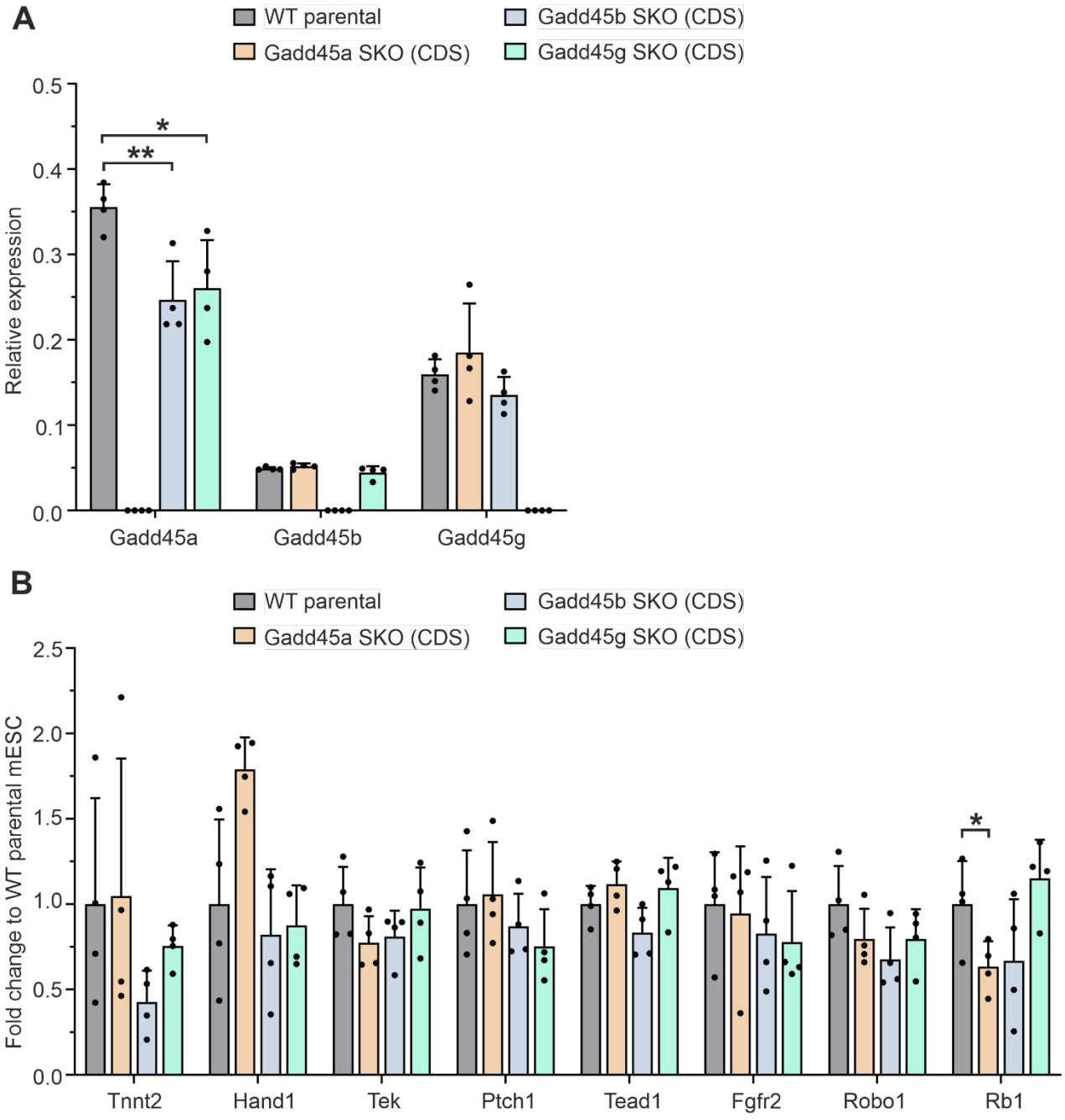

Figure 5-6: Analysis of heart development-associated genes in *Gadd45* SKO mESCs. (A) Bar graph shows the qPCR data for expression of *Gadd45* genes in WT parental and different *Gadd45* SKO (CDS) mESCs. **(B)** Bar graph shows the WT parental normalized expression of representative heart development-associated genes in WT parental and different *Gadd45* SKO (CDS) mESCs. The data is presented as fold change to WT parental mESCs. Statistical significance was tested with one-tailed, unpaired Student's t-test (*: p-value <0.05). Relative expression was measured with respect to housekeeping genes, *Tbp* and *G6pd*.

5.3.5. Reduced expression of heart development-associated genes is independent of signalling roles of GADD45 proteins

GADD45 proteins play essential roles in cellular signalling pathways such as p38 mitogen-activated protein kinase (p38), c-Jun N-terminal kinase (JNK) and extracellular-signal-regulated kinase (ERK) [226]. These signalling pathways also play essential roles during embryonic heart

development [227,228]. Therefore it was essential to determine whether the downregulation of heart development-associated genes was an indirect effect of altered cellular signalling in *Gadd45* TKO (CDS) mESCs. I used the compounds SB202190 (20 μM), JNK-IN-7 (1 μM) and PD0325901 (1 μM) to inhibit p38, JNK and ERK pathways respectively [229–231]. The p38 and JNK signalling is activated primarily during stress response (e.g. UV exposure, thermal stress), and therefore their activated phosphorylated forms are not detectable via western blot under normal growth conditions in mESCs (data not shown) [232,233]. I used published concentrations of SB202190 (20 μM) and JNK-IN-7 (1 μM), known to inhibit p38 and JNK signalling in other cell types [229–231]. In contrast, ERK signalling is highly active in mESCs and the phosphorylated ERK1/2 (phospho-ERK1/2) forms are detectable via western blot under normal growth conditions (Figure 5-7 A, C) [234]. 1 μM PD0325901 treated WT and *Gadd45* TKO (CDS) mESCs showed substantial reduction in levels of phospho-ERK1/2, confirming the inhibition of ERK signalling in these cells (Figure 5-7 A, C).

I treated WT mESCs with the above compounds and analysed the qPCR data for potential downregulation of representative heart development-associated genes (directional or one-tailed hypothesis). *Tek, Tnnt2* and *Myh6* showed significantly reduced expression upon inhibition of p38 pathway (Figure 5-7 B). However, the expression of these three genes was also downregulated upon p38 inhibition in *Gadd45* TKO (CDS) mESCs (Figure 5-7 D). Taken together, these results suggest that the p38 pathway regulates expression of heart development-associated genes via a GADD45-independent mechanism in mESCs. Moreover, inhibition of both JNK and ERK pathways did not show downregulated expression of these heart development-associated genes in WT and *Gadd45* TKO (CDS) mESCs (Figure 5-7 B, D). Therefore, I concluded that the downregulation of heart development-associated genes in *Gadd45* TKO (CDS) mESCs was independent of signalling functions of GADD45 proteins.

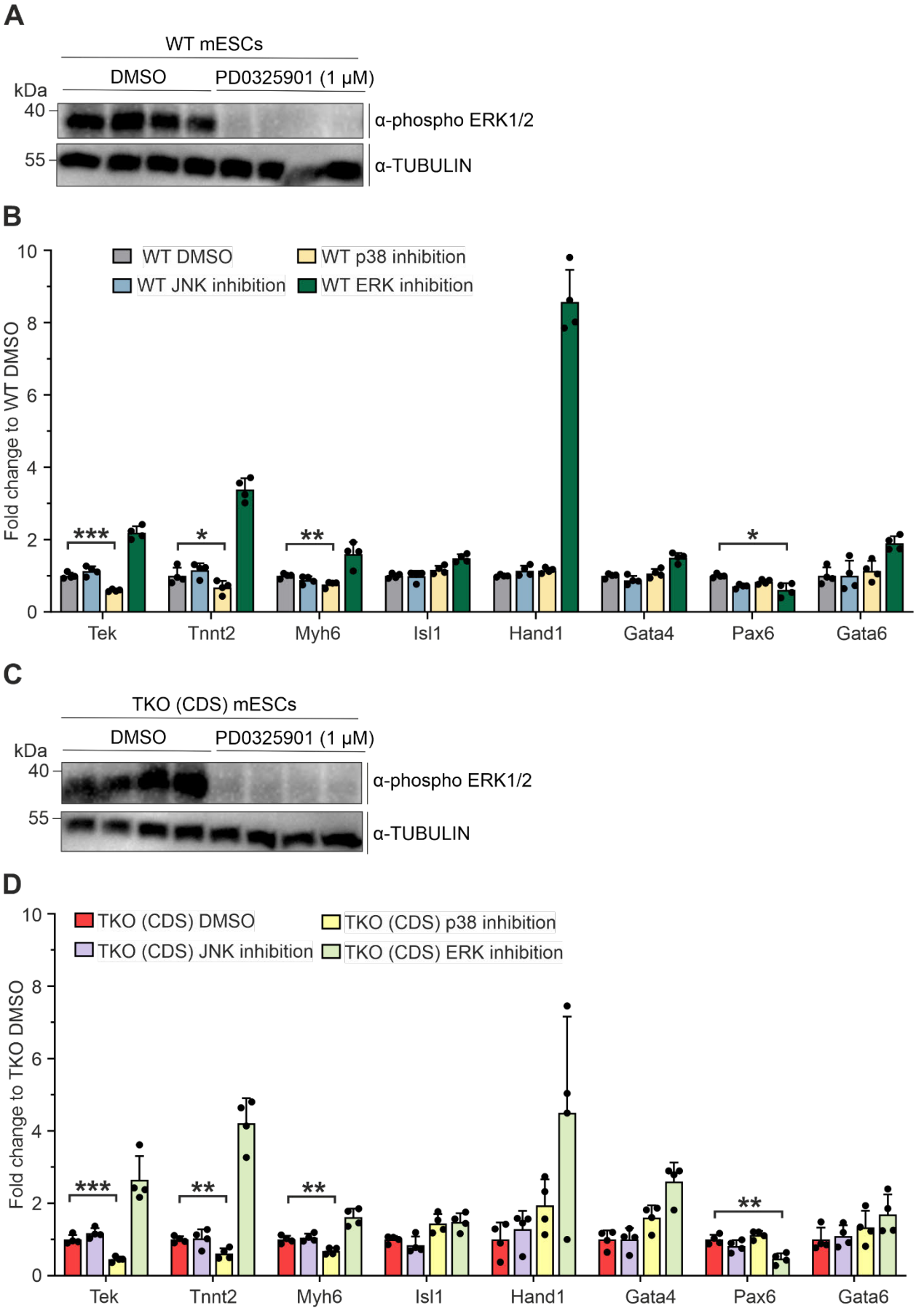

Figure 5-7: Inhibition of p38, JNK and ERK signalling pathways in WT and *Gadd45* TKO (CDS) mESCs.

(A) Western blot analysis of phospho-ERK1/2 in DMSO or 1 µM PD0325901 treated WT mESCs (4 independent clones). TUBULIN served as loading control. **(B)** Bar graph shows the qPCR data for representative heart development-associated genes upon inhibition of p38 (20 µM SB202190), JNK (1 µM JNK-IN-7) and ERK (1 µM PD0325901) pathway in WT mESC. The data is presented as fold change to DMSO-treated WT mESCs. **(C)** Western blot analysis of phospho-ERK1/2 in DMSO or 1 µM PD0325901 treated *Gadd45* TKO (CDS) mESCs (4 independent clones). TUBULIN served as loading control. **(D)** Bar graph shows the qPCR data for representative heart development-associated genes upon inhibition of p38, JNK and ERK pathway in *Gadd45* TKO (CDS) mESCs. The data is presented as fold change to DMSO treated *Gadd45* TKO (CDS) mESCs. Statistical significance was tested with one-tailed, unpaired Student's t-test (*: p-value <0.05; **: p-value <0.01; ***: p-value <0.001; ****: p-value <0.0001). Relative expression was measured with respect to housekeeping genes, *Tbp* and *G6pd*. Expected molecular weight for phospho-ERK1/2: 40kDa; TUBULIN: 55kDa.

5.3.6. Overexpression of *Gadd45a,b,g* rescues expression of heart development-associated genes in *Gadd45* TKO (CDS) mESCs

Next, I investigated whether overexpressing *Gadd45* genes could rescue the downregulated heart development-associated genes in *Gadd45* TKO (CDS) mESCs. I therefore performed transient transfection with pCS2+myc empty vector (EV) or combined pCS2+myc *Gadd45a*,b,g plasmids in WT and *Gadd45* TKO (CDS) mESCs (Figure 5-8 A). Both WT and *Gadd45* TKO (CDS) mESCs transfected with pCS2+myc Gadd45a,b,g plasmids showed significant induction of Gadd45 mRNAs (~1000-2000 fold) compared to pCS2+myc EV transfected controls (data not shown).

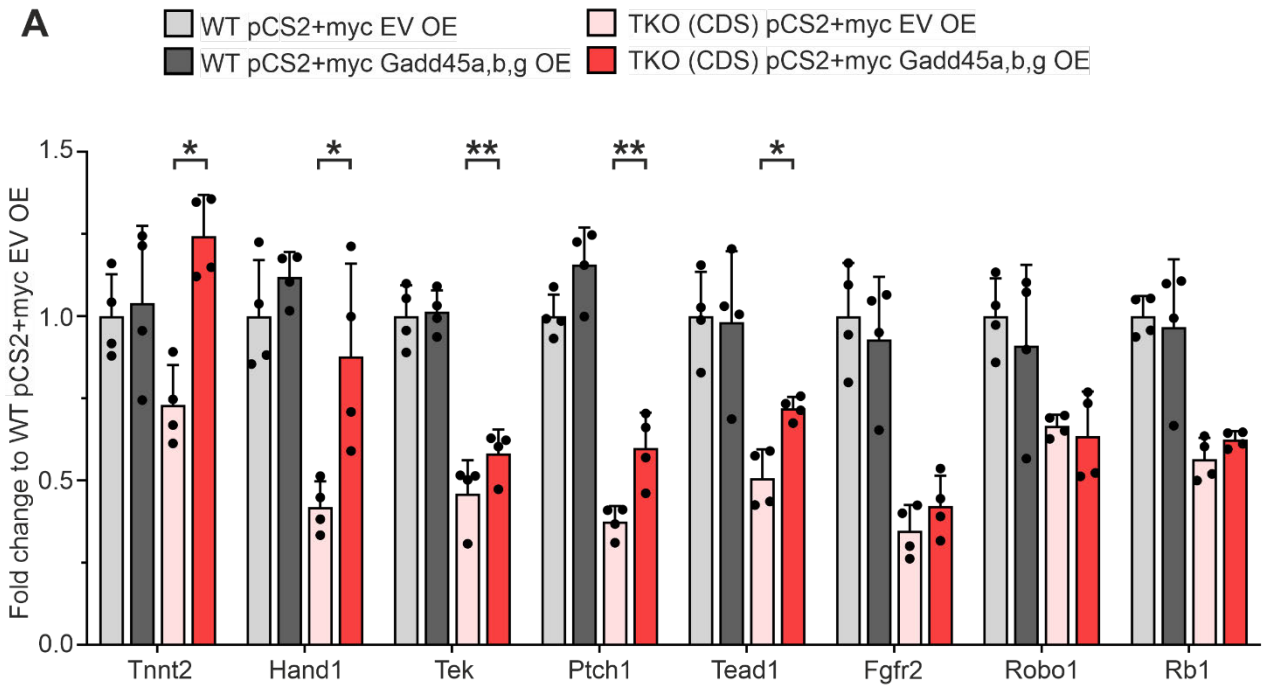

Figure 5-8: pCS2+myc *Gadd45a*,b,g OE in WT and *Gadd45* TKO (CDS) mESCs. (A) Bar graph shows the qPCR data for expression of heart development-associated genes upon pCS2+myc EV or pCS2+myc *Gadd45a*,*b,g* OE in WT and *Gadd45* TKO (CDS) mESCs. The data is presented as fold change to pCS2+myc EV transfected WT mESCs. Statistical significance was tested with two-tailed paired (comparing pCS2+myc EV OE vs pCS2+myc *Gadd45a,b,g* OE in respective genotypes) Student's t-test (*: p-value <0.05; **: p-value <0.01). Relative expression was measured with respect to housekeeping genes, *Tbp* and *G6pd*.

Overexpressing *Gadd45a*, *Gadd45b*, and *Gadd45g* in WT mESCs did not affect the expression of heart development-associated genes (Figure 5-8 A). In contrast, combined OE of *Gadd45a*, *Gadd45b*, and *Gadd45g* partially rescued the expression of downregulated genes associated with heart development (*Tnnt2, Hand1, Tek, Ptch1* and *Tead1*) in *Gadd45* TKO (CDS) mESCs (Figure 5-8 A). This further supports our hypothesis that GADD45 proteins regulate embryonic heart development-associated genes in mESCs.

5.3.7. Increased TET1 occupancy in *Gadd45* TKO (CDS) mESCs

Next, I investigated the mechanistic basis for the reduced expression of embryonic heart development-associated genes in *Gadd45* TKO (CDS) mESCs (Figure 5-5 D). It was still unclear whether GADD45 proteins guide TET1 to regulate local DNA methylation at heart development-associated genes in mESCs, like shown for TCF21-TARID locus [107]. In collaboration with [redacted] (this lab), we performed TET1 CUT&Tag sequencing followed by analysis of differential TET1 peaks between WT and *Gadd45* TKO (CDS) mESCs. Surprisingly, we observed a large number of TET1 peaks with increased occupancy (~6300 TET1-up peaks) as compared to TET1 peaks with reduced occupancy (~1500 TET1-down peaks) in *Gadd45* TKO (CDS) mESCs (Figure 5-9 A). GO analysis of genes associated with TET1-up peaks showed enrichment for cell cycle and chromatin remodelling related terms (data not shown). On the other hand, GO analysis of genes associated with TET1-down peaks showed enrichment for forebrain development and other neuronal differentiation related terms (data not shown).

These results are in line with my mass spectrometry result of DNA modification, showing increased levels of 5mC oxidation intermediates i.e. 5hmC, 5fC and 5caC (Figure 5-3 D). However, the substantially large number of TET1-up peaks was in contradiction to our proposed hypothesis that GADD45 proteins guide TET1 to specific loci in mESCs. Therefore, there was a need to consider alternate models for GADD45 and TET1-mediated gene regulation in mESCs. For instance, GADD45 proteins are primarily localized in the cytoplasm in mESCs and could potentially regulate the nuclear vs. cytoplasmic levels of TET1 in mESCs.

Alternatively, instead of acting as a recruiter of TET proteins, GADD45 proteins could be regulating the half-life of TET1-chromatin interaction to enhance TET1 DNA demethylation processivity [52]. In extension, to counteract any reduction in TET1 DNA demethylation activity due to absence of GADD45 proteins, cells might overcompensate by loading more TET1 proteins at its target sites in *Gadd45* TKO (CDS) mESCs. However, if the increased recruitment of TET1 proteins is not

regulated, it could also lead to increased repressive non-catalytic functions (via PRC recruitment) [224].

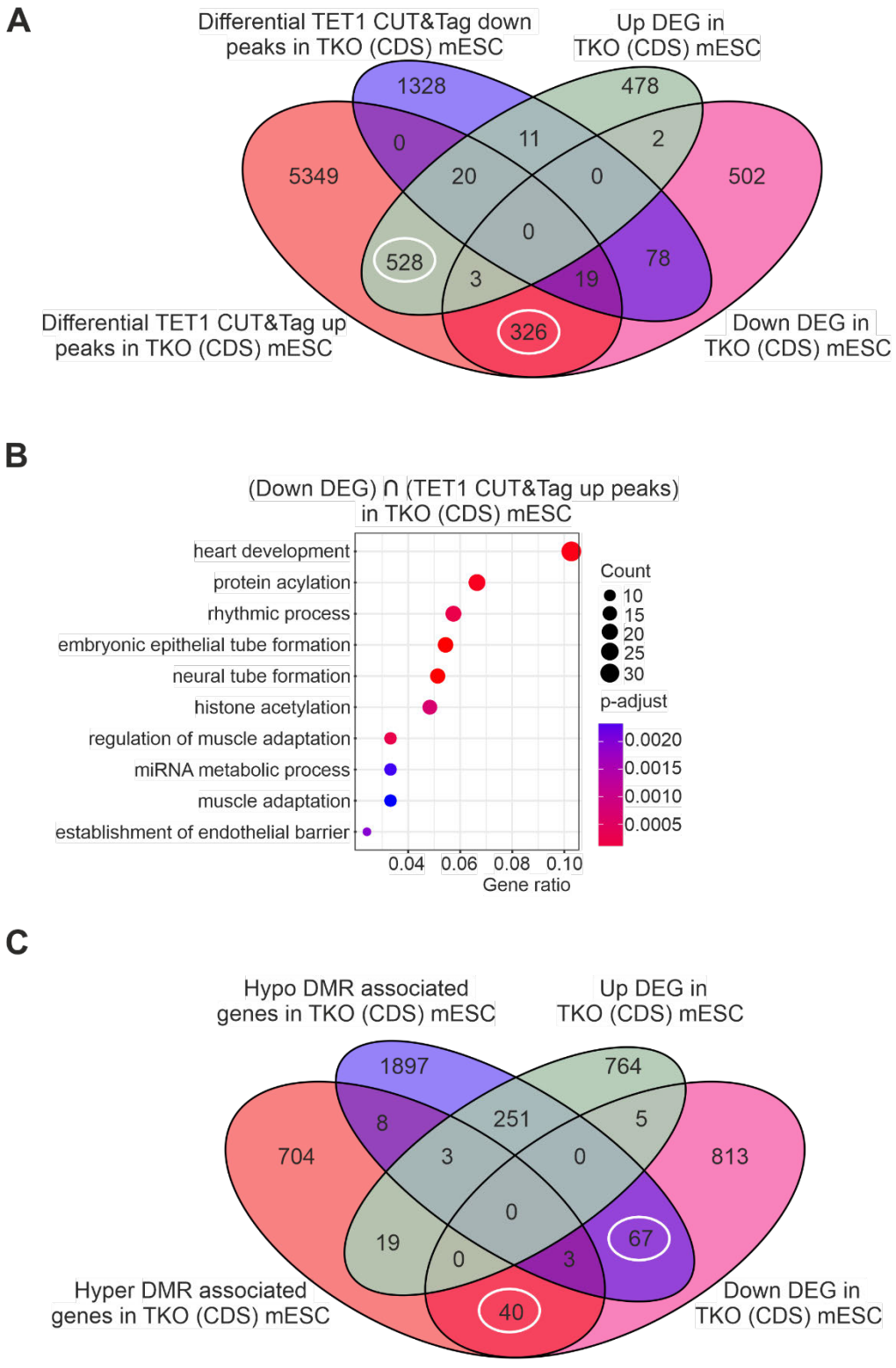

Figure 5-9: Analysis of differential TET1 CUT&Tag peaks and DMRs in *Gadd45* TKO (CDS) mESCs. (A) Venn diagram shows the overlap between differential TET1 CUT&Tag peaks (up and down) with DEGs from RNA-seq (up and down) of WT and *Gadd45* TKO (CDS) mESCs. White circles indicate the overlap

between TET1-up peaks with RNA-seq (up and down) DEGs **(B)** GO analysis of genes common between down DEGs and TET1-up CUT&Tag peaks (348 genes) in *Gadd45* TKO (CDS) mESCs. **(C)** Venn diagram shows the overlap between DMRs (hypo and hyper) and DEGs from RNA-seq of WT and *Gadd45* TKO (CDS) mESCs. The unexpected overlaps between differential TET1 peaks/DEGs/DMRs is explained in the description of results below.

We therefore overlapped our DEGs and TET1 CUT&Tag datasets, to identify correlation between DEGs and differential TET1 occupancy sites in *Gadd45* TKO (CDS) mESCs. The analysis revealed a substantial overlap between the up (551 genes) and down (348 genes) DEGs with TET1-up peaks in *Gadd45* TKO (CDS) mESCs (Figure 5-9 A). On the other hand, there was a poor overlap between TET1-down peaks and RNA-seq DEGs (up or down) in *Gadd45* TKO (CDS) mESCs.

The overlap was performed by intersecting genomic coordinates of differential TET1 peaks and DEGs (3kb upstream of TSS + gene body). Such an approach of overlapping genomic loci can lead to unexpected results, for instance, 5 overlapping genes between up and down DEGs (Figure 5-9 A). Such instances arise when up and down DEGs share their genomic coordinates, for example, anti-sense orientated genes on opposite strands. Similarly, unexpected overlap between TET1-up and TET1-down peaks, is due to the genomic coordinates of one DEG overlapping with two separate up and down TET1 peaks.

Interestingly, GO analysis for the 348 genes with reduced expression and increased TET1 occupancy in *Gadd45* TKO (CDS) mESCs showed enrichment of heart development related terms (Figure 5-9 B). These counterintuitive results suggest that increased TET1 occupancy in absence of GADD45 proteins correlates with transcription repression of heart development-associated genes. To resolve these counterintuitive result, in collaboration with [redacted] (this lab), we performed DNA methylation analysis, using Infinium Mouse Methylation Array, in WT and *Gadd45* TKO (CDS) mESCs. On a genome-wide level, there was a substantial number of sites that lost DNA methylation (~2200 genes associated with hypo DMRs) in *Gadd45* TKO (CDS) mESCs (Figure 5-9 C). In contrast, there were fewer sites with increased DNA methylation (~800 genes associated with hyper DMRs), interestingly enriched in gene promoters in *Gadd45* TKO (CDS) mESCs (data not shown). It is worth noting that the genome-wide changes in DNA methylation observed in *Gadd45* TKO (CDS) mESCs (global hypomethylation and promoter specific hypermethylation) mimic ageing associated changes in DNA methylation [235,236].

Furthermore, I closely analysed whether these differential DNA methylation associated genes overlapped with DEGs in *Gadd45* TKO (CDS) mESCs. The unexpected overlap between hypo and hyper DMR associated genes stems from few genes being proximal to DMRs changing in opposite directions. There was a poor overlap between down DEGs (containing the heart development-associated genes) and DMRs in *Gadd45* TKO (CDS) mESCs (Figure 5-9 C). The few down DEGs that overlapped with hyper DMR (43 genes) and hypo DMR (70 genes) did not contain heart

development-associated genes, indicating that the downregulation of heart development-associated genes in *Gadd45* TKO (CDS) mESCs was via a DNA methylation independent mechanism. The DNA methylation independent downregulation of heart development-associated genes taken together with increased TET1 occupancy at those genes, hints towards increase in non-catalytic TET1 activity (potentially via recruitment of PRC) in absence of GADD45 proteins in mESCs (Figure 5-9) [224].

5.3.8. ChIP Enrichment Analysis (ChEA) on TET1 regulated DEGs from *Gadd45* TKO (CDS) mESCs

I further explored whether GADD45 proteins play a role in balancing catalytic vs. non-catalytic functions of TET1. I performed ChEA on the subset of 348 genes with reduced expression and increased TET1 occupancy in *Gadd45* TKO (CDS) mESCs (containing heart development-associated genes) (Figure 5-9 A). ChEA computes over-representation of transcription factor targets from the public ChIP-X database [237]. Interestingly, these 348 genes showed significant enrichment for mESC specific binding sites of many PRC sub-units (SUZ12, MTF2 and EZH2) (Figure 5-10 A).

Next, I used publicly available RNA-seq dataset from *Tet1* KO (exon 4 deleted) and *Tet1* catalytic mutant (exon 10 point mutations H1652Y and D1654A) mESCs [224]. In their analysis, Chrysanthou *et. al.,* also identified genes regulated exclusively via non-catalytic TET1 function [224]. They did so by identifying DEGs unique to *Tet1* KO mESCs (both catalytic and non-catalytic functions lost), excluding DEGs common with *Tet1* catalytic mutant mESCs (only catalytic functions lost).

There was a significant overlap between DEGs from (a) *Gadd45* TKO (CDS) and *Tet1* KO mESCs and (b) *Gadd45* TKO (CDS) and DEGs unique to *Tet1* KO mESCs (FDR corrected p-value < 0.05) (Figure 5-10 B, D). Furthermore, I performed ChEA on the 387 and 323 overlapping DEGs and again observed an enrichment for mESC specific binding sites of many PRC sub-units (SUZ12, MTF2, RNF2 and EZH2) (Figure 5-10 C, E). These results further strengthen the hypothesis that GADD45 proteins play a role in transcription regulation at TET1 and PRC occupied sites in mESCs. It will be important to perform CUT&Tag for H3K27me3 or SUZ12 to analyse for potential increase in TET1-mediated PRC recruitment at heart development-associated genes in *Gadd45* TKO (CDS) mESCs.

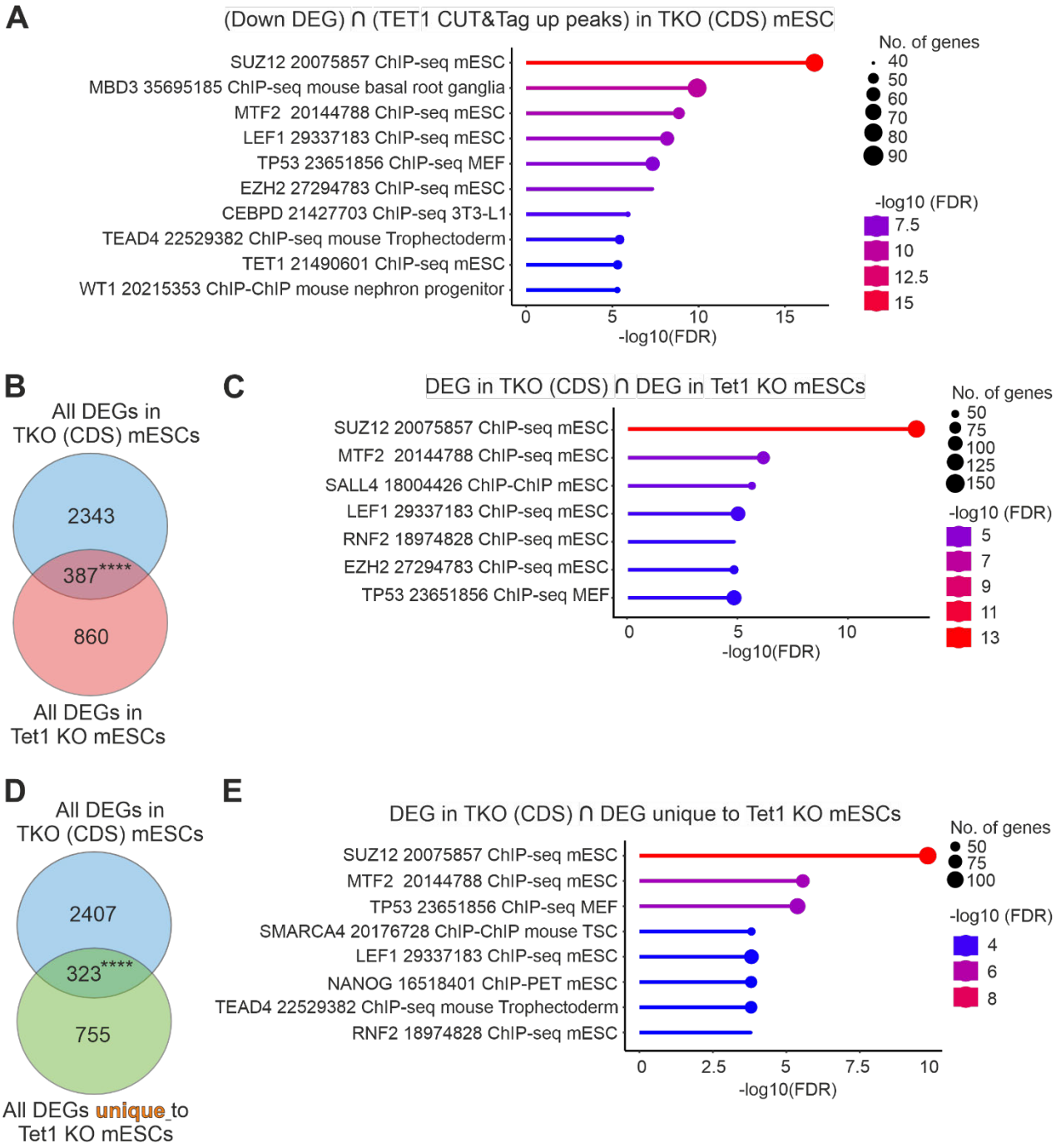

Figure 5-10: ChEA on DEGs from *Gadd45* TKO (CDS) and *Tet1* KO mESCs. (A) ChEA on 348 genes overlapping between down DEGs and TET1 CUT&Tag up peaks in *Gadd45* TKO (CDS) mESCs. The Y-axis labels provide information on the name of ChEA enriched protein, PubMed ID of used dataset, assay type and cell type, while the X-axis presents the log transformed FDR values. **(B)** Venn diagram showing overlap between all DEGs from *Gadd45* TKO (CDS) and all DEGs from *Tet1* KO mESCs. **(C)** ChEA on the 387 overlapping genes between all DEGs from *Gadd45* TKO (CDS) and all DEGs from *Tet1* KO mESCs. **(D)** Venn diagram showing overlap between all DEGs from *Gadd45* TKO (CDS) and all DEGs unique to *Tet1* KO mESCs. **(E)** ChEA on the 323 overlapping genes between all DEGs from *Gadd45* TKO (CDS) and all DEGs unique to *Tet1* KO mESCs. Fisher exact test was used to compute significance of the Venn diagram overlaps; ****: p-value < 0.0001.

5.3.9. *Gadd45* TKO (CDS) mESCs show altered cardiac differentiation

The mESC transcriptome analysis revealed a role for GADD45 proteins in regulation of heart development-associated genes (Figure 5-5 D). However, it was unclear whether this downregulated expression in mESC stage translates to defective cardiac differentiation in *Gadd45* TKO (CDS) mESCs. Since, we do not have access to *Gadd45* TKO mice, I needed to establish an in-vitro cardiac differentiation protocol. By modulating culture media conditions, mESCs can differentiate along defined lineages, producing diverse cell types, including cardiomyocytes [80]. I established the in-vitro cardiac differentiation protocol combining two published chemically defined media strategies [238,239]. By inhibiting BMP4 (2 µM Dorsomorphin treatment from day -3 to day 2) and Wnt (1 µM XAV939 day 3 to day 5) signalling in EBs at specific time points, I envisaged to mimic the in-vivo differentiation scenario to achieve cardiomyocyte differentiation of mESCs (Figure 5-11 A) [175–177]. With this protocol I would also be able to isolate specific cardiac differentiation lineages for cell-type specific molecular analysis.

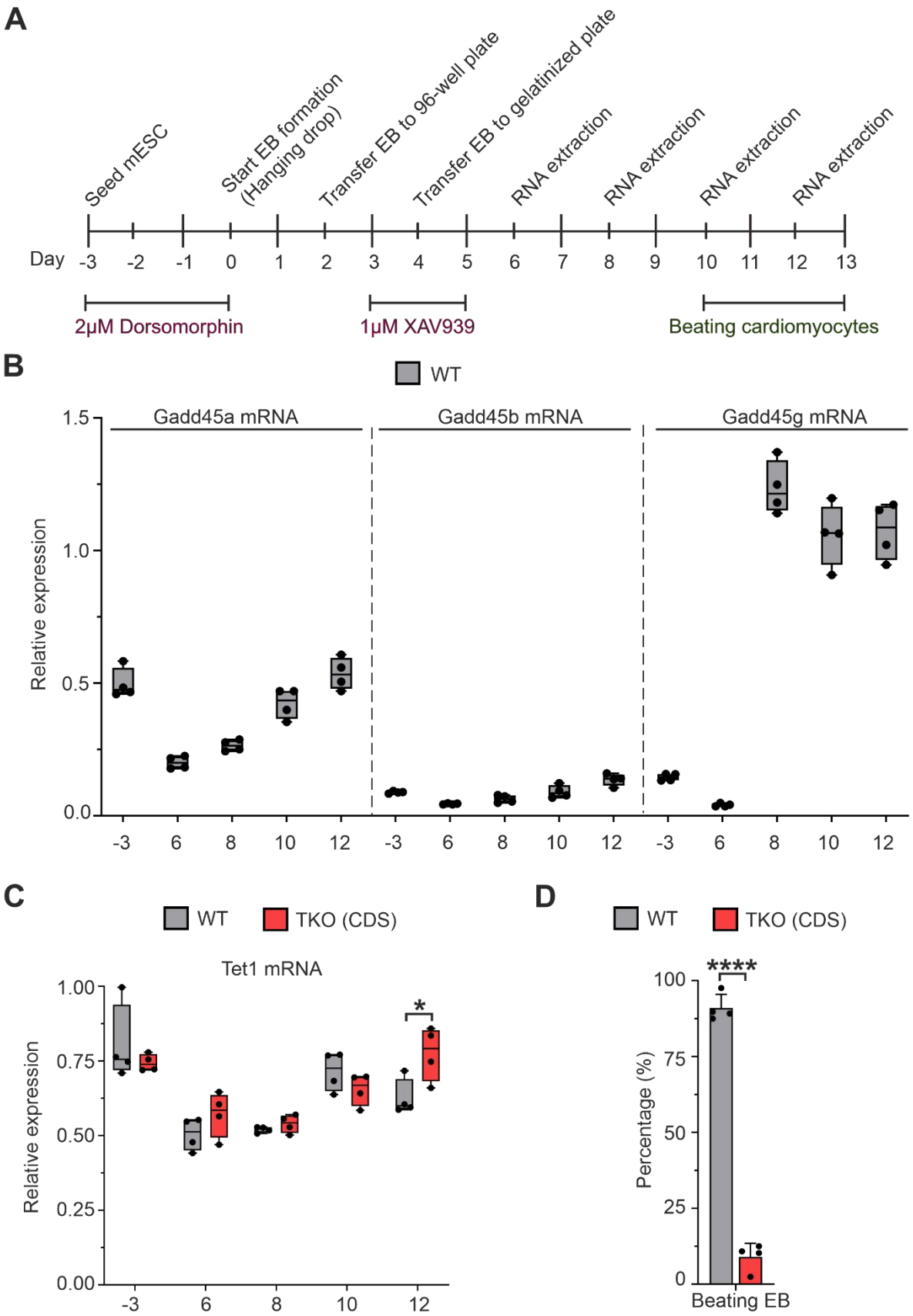

Figure 5-11: In-vitro cardiac differentiation of WT and *Gadd45* TKO (CDS) mESCs. (A) Schematic represents the in-vitro cardiac differentiation timeline with information about inhibitor concentration and time point of treatment. **(B)** Box plot shows the qPCR data for expression of *Gadd45* genes from in-vitro cardiac differentiated WT mESCs. X-axis represents the day of cardiac differentiation (-3, 6, 8, 10 and 12). **(C)** Box

plot shows the qPCR data for expression of *Tet1* from in-vitro cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. X-axis represents the day of cardiac differentiation (-3, 6, 8, 10 and 12). **(D)** Bar graph shows the percentage of beating EB from in-vitro cardiac differentiated WT mESCs at day 10 of differentiation (~120 EB per genotype were analysed). Statistical significance was tested with two-tailed, unpaired Student's t-test (*: p-value <0.05; ****: p-value <0.0001). Relative expression was measured with respect to housekeeping genes, *Tbp and G6pd*.

I performed qPCR analysis to analyse expression of genes at different stage of cardiac differentiation [day -3 (undifferentiated mESCs), day 6, day 8, day 10 and day 12] in cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. WT cells showed the lowest expression of *Gadd45a,b,g* genes at day 6 of cardiac differentiation, followed by an increase in expression in the following days (day 8-12) (Figure 5-11 B). Interestingly, *Gadd45g* showed the highest increase in expression during cardiac differentiation and was the highest expressed *Gadd45* isoform in cardiac differentiated cells, in line with the in-vivo data for expression of *Gadd45* isoforms in heart tissue [240]. Furthermore, I measured expression of *Tet1* gene during the cardiac differentiation timeline, which showed similar levels of expression during day 6-10 of the cardiac differentiation in WT and *Gadd45* TKO (CDS) mESCs. However, there was significantly higher *Tet1* expression in day 12 cardiac differentiated *Gadd45* TKO (CDS) mESCs (Figure 5-11 C). Notably, *Gadd45* and *Tet1* genes showed similar pattern of expression during cardiac differentiation in WT mESCs (lowest at day 6 followed by increase in expression as cells differentiate to cardiomyocytes).

Lastly, the cardiac-differentiated EB from WT mESCs exhibited spontaneous beating regions at day 10 of differentiation, similar to the rhythmic contractions observed in cardiomyocytes within an embryonic heart in-vivo. Interestingly, the number of beating EBs was significantly reduced in cardiac-differentiated *Gadd45* TKO (CDS) mESCs (Figure 5-11 D). This strongly supports our hypothesis that the GADD45 proteins regulate cardiac differentiation during embryonic development.

I investigated the transcriptome of these cardiac differentiated cells to identify the basis for this reduced number of beating EBs in *Gadd45* TKO (CDS) mESCs. I performed RNA-seq in cardiac-differentiated WT and *Gadd45* TKO (CDS) mESCs at day 6, 8, 10 and 12 of differentiation and data was analysed in collaboration with [redacted] (this lab). There were different numbers of up and down DEGs at the respective time points of cardiac differentiation, with the highest number of DEGs observed at day 12 of cardiac differentiation (Figure 5-12 A). GO analysis for DEGs at day 6 of cardiac differentiation showed no significantly enriched terms. GO analysis for genes downregulated in *Gadd45* TKO (CDS) at day 8 of cardiac differentiation showed significant enrichment of angiogenesis and epithelial cell differentiation related terms (Figure 5-12 B). Expectedly, GO analysis for genes downregulated in *Gadd45* TKO (CDS) at day 10 and 12 of cardiac differentiation showed significant enrichment of cardiac muscle tissue and circulatory system development related terms (Figure 5-12 C, D).

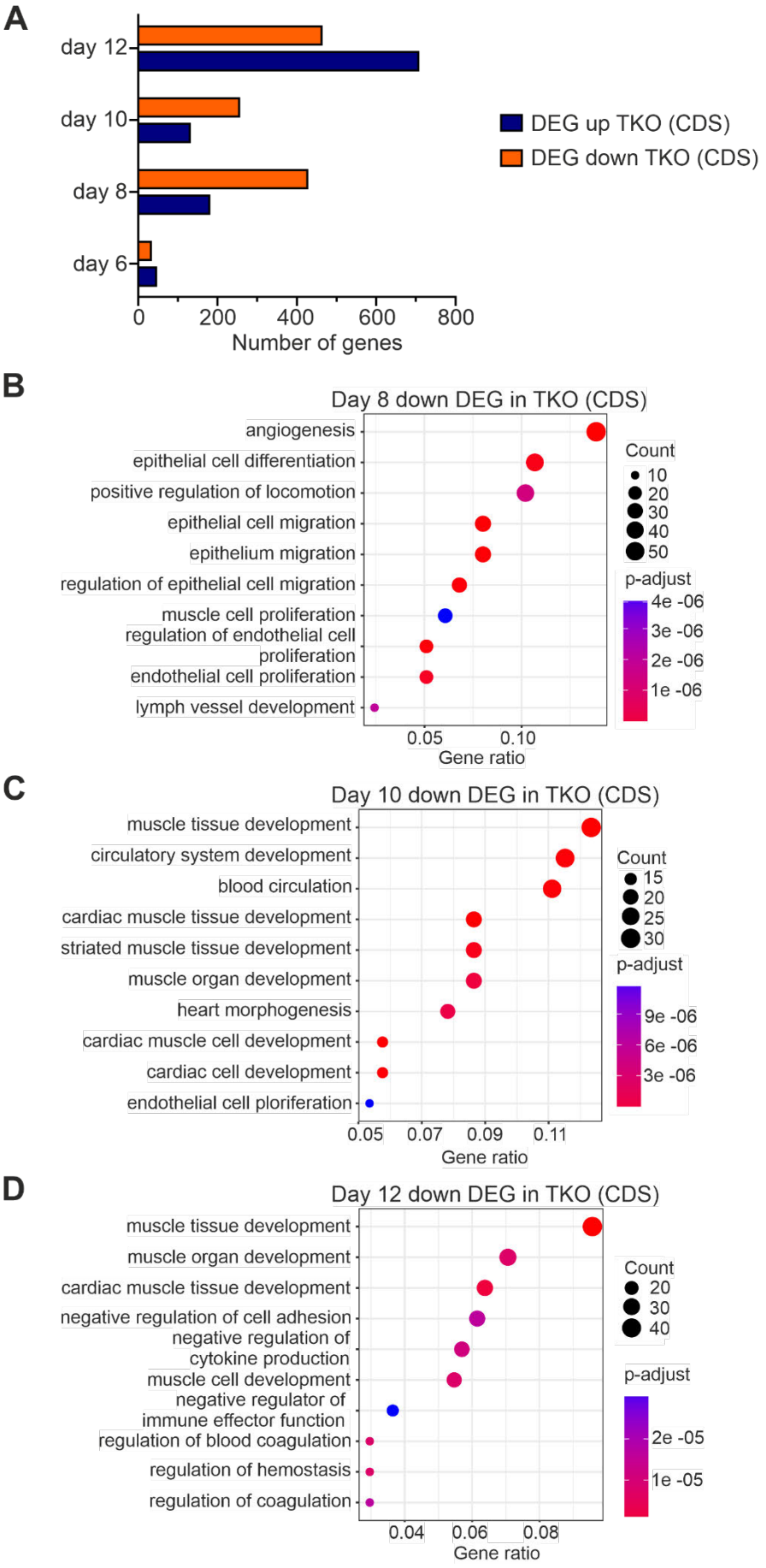
A
day 12
day 10
day 8
day 6
0
200
400
600
800
Number of genes
DEG up TKO (CDS)
DEG down TKO (CDS)
B
Day 8 down DEG in TKO (CDS)
angiogenesis
epithelial cell differentiation
positive regulation of locomotion
epithelial cell migration
epithelium migration
regulation of epithelial cell migration
muscle cell proliferation
regulation of endothelial cell proliferation
endothelial cell proliferation
lymph vessel development
0.05
0.10
Gene ratio
Count
10
20
30
40
50
p-adjust
4e -06
3e -06
2e -06
1e -06
C
Day 10 down DEG in TKO (CDS)
muscle tissue development
circulatory system development
blood circulation
cardiac muscle tissue development
striated muscle tissue development
muscle organ development
heart morphogenesis
cardiac muscle cell development
cardiac cell development
endothelial cell ploriferation
0.05
0.07
0.09
0.11
Gene ratio
Count
15
20
25
30
p-adjust
9e -06
6e -06
3e -06
D
Day 12 down DEG in TKO (CDS)
muscle tissue development
muscle organ development
cardiac muscle tissue development
negative regulation of cell adhesion
negative regulation of cytokine production
muscle cell development
negative regulator of immune effector function
regulation of blood coagulation
regulation of hemostasis
regulation of coagulation
0.04
0.06
0.08
Gene ratio
Count
20
30
40
p-adjust
2e -05
1e -05

Figure 5-12: Analysis of RNA-seq data from in-vitro cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. (A) Bar graph shows the number of DEGs from in-vitro cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs at day 6, 8, 10, and 12 of differentiation. **(B), (C) and (D)** GO analysis differentially downregulated genes (down DEGs) in in-vitro cardiac differentiated *Gadd45* TKO (CDS) mESCs at day 8, 10 and 12 of differentiation respectively.

I validated the RNA-seq results with an independent cardiac differentiation of WT and *Gadd45* TKO (CDS) mESCs followed by qPCR (Figure 5-13). The expression of *T* gene, a mesodermal marker gene, showed no differences between cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs, suggesting no global mesoderm differentiation defects (Figure 5-13 A). Interestingly, starting from day 8 of cardiac differentiation there was a significantly reduced expression of *Gata4* gene, a cardiac progenitor marker, in *Gadd45* TKO (CDS) (Figure 5-13 B). Furthermore, starting from day 8 of cardiac differentiation there was a significantly reduced expression of *Tek* gene, an endothelial cell marker, in *Gadd45* TKO (CDS) (Figure 5-13 C). Lastly, starting from day 8 of cardiac differentiation there was a significantly reduced expression of *Myh6, Actc1* and *Tnnt2* genes, a set of cardiomyocyte markers, in *Gadd45* TKO (CDS) (Figure 5-13 D-F). These representative examples validate my RNA-seq results and indicate defects in activation of embryonic heart development-associated genes, leading to impaired cardiac differentiation in *Gadd45* TKO (CDS) mESCs.

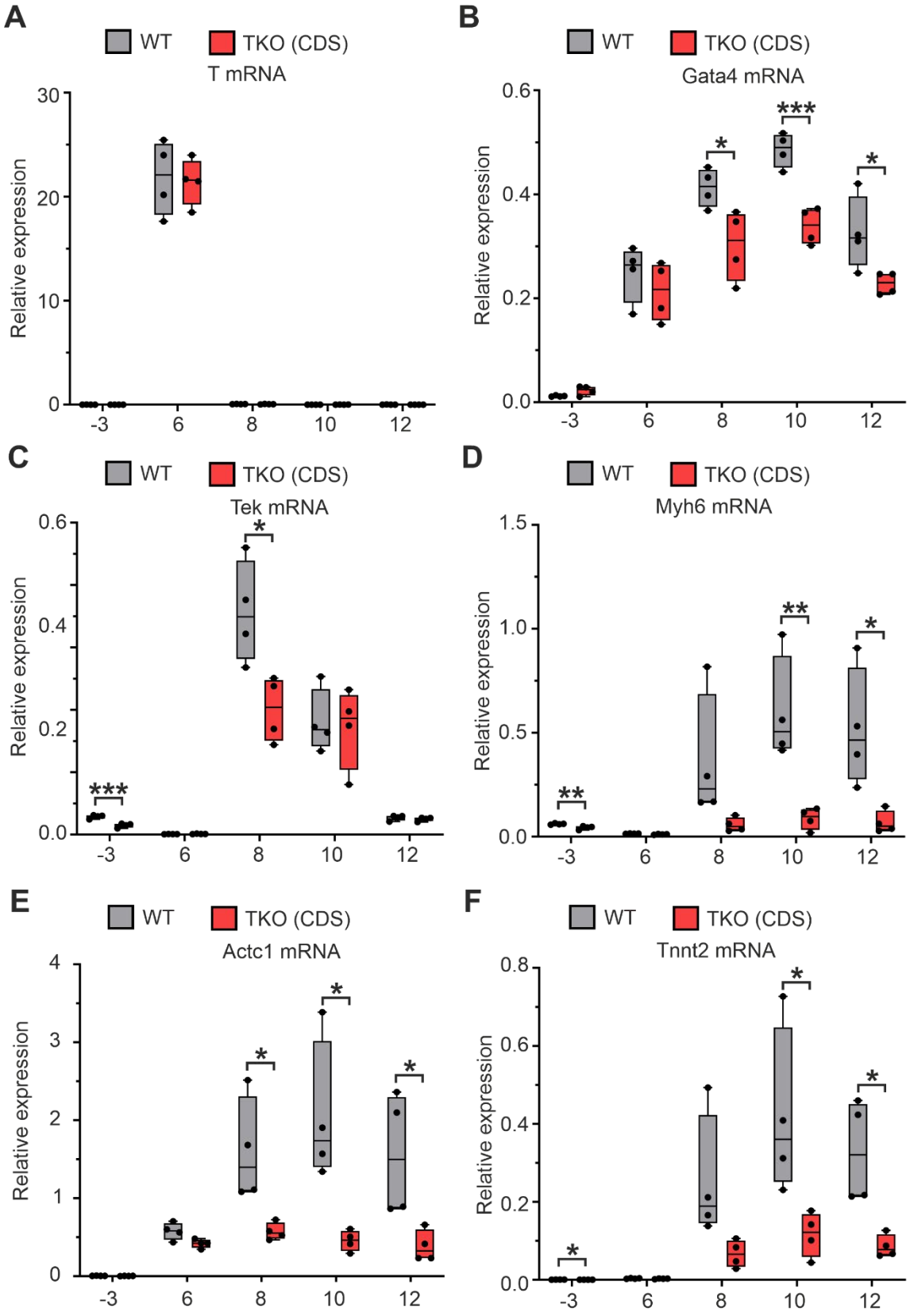

Figure 5-13: Validation of RNA-seq data from in-vitro cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. (A) – (F) Box plot shows the qPCR data for expression of *T, Gata4, Tek, Myh6, Actc1* and *Tnnt2* genes from in-vitro cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. X-axis represents the day of cardiac differentiation (-3, 6, 8, 10 and 12). Statistical significance was tested with two-tailed, unpaired Student's t-test (*: p-value <0.05; **: p-value <0.01; ***: p-value <0.001). Relative expression was measured with respect to housekeeping genes, *Tbp* and *G6pd*.

5.3.10. *Gadd45* TKO (CDS) mESCs show impaired differentiation towards mesenchymal stem cells (MSCs).

I performed basic cell-type annotation analysis using all the DEGs in day 12 of cardiac differentiation using the online tool ENRICHR [241–243]. Interestingly, this crude analysis showed enrichment of MSCs along with the expected cardiac muscle cells (data not shown). MSCs are multipotent stem cells which originate from the mesoderm during embryonic development [244]. They play essential roles in maintaining tissue homeostasis and facilitating regeneration post injury [245,246]. MSCs are particularly significant in the treatment of CVD due to their unique characteristics, such as their capacity to differentiate into cardiovascular cells and ability to promote the formation of new blood vessels [245,246]. Furthermore, *Dlk1* (a MSC marker gene) expression was significantly lower in cardiac differentiated *Gadd45* TKO (CDS) mESCs at all differentiation time points, indicative of an impaired MSC differentiation (Figure 5-14 A).

To analyse for potential MSC differentiation defects, I differentiated WT and *Gadd45* TKO (CDS) mESCs towards MSC lineage using retinoic acid treatment based published protocol (Figure 5-14 B) [247]. The WT mESCs differentiated towards the MSC lineage and showed significant increase in expression of MSC marker genes (*Dlk1, Cd44/ Pgp-1* and *Cd105/ Endoglin*) (Figure 5-14 C) [248–250]. However, *Gadd45* TKO (CDS) mESCs, showed significantly lower expression of *Dlk1* and *Cd105* genes (and trend for reduced expression of *Cd44* gene) compared to WT. This further supports the hypothesis that *Gadd45* TKO (CDS) mESCs have a MSC differentiation defect (Figure 5-14 C). This potentially in part could contribute to cardiac differentiation defects in *Gadd45* TKO (CDS) mESCs (Figure 5-12).

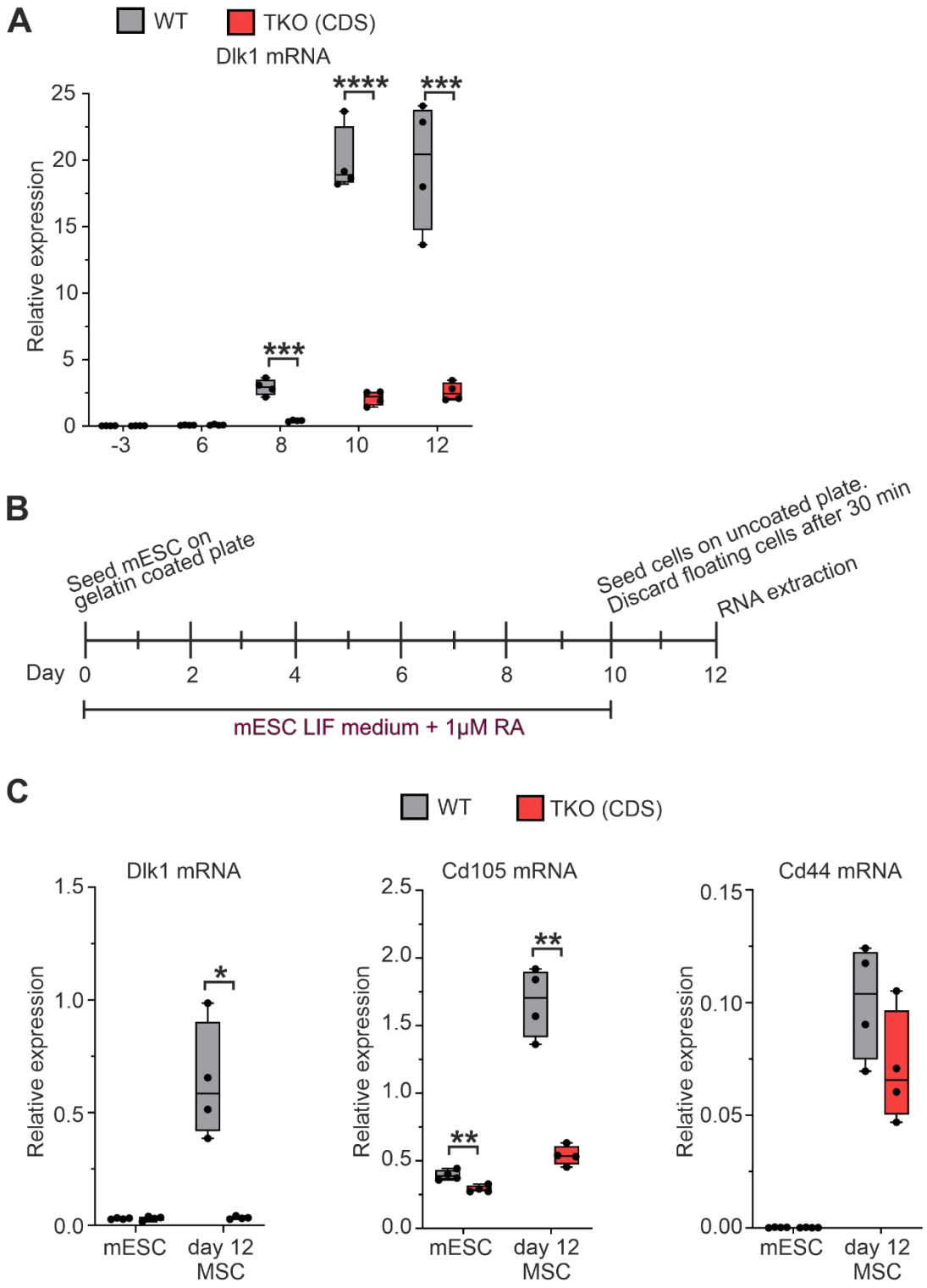

Figure 5-14: MSC differentiation of WT and *Gadd45* TKO (CDS) mESCs. (A) Box plot shows the qPCR data for expression of *Dlk1* gene from in-vitro cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. X-axis represents the day of cardiac differentiation (-3, 6, 8, 10 and 12). **(B)** Schematic represents the mESC to MSC differentiation timeline using retinoic acid treatment and MSC cell attachment properties for enriching MSC. **(C)** Box plot shows the qPCR data for expression of *Dlk1*, *Cd105* and *Cd44* genes from MSC differentiated WT and *Gadd45* TKO (CDS) mESCs. X-axis represents the differentiation status: mESC or day 12 MSC differentiated cells. Statistical significance was tested with two-tailed, unpaired Student's t-test (*: p-value <0.05; **: p-value <0.01; ***: p-value <0.001; ****: p-value < 0.0001). Relative expression was measured with respect to housekeeping genes, *Tbp* and *G6pd*

5.3.11. Pilot scRNA-seq for cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs

The bulk RNA-seq revealed an impaired cardiac differentiation in *Gadd45* TKO (CDS) mESCs (Figure 5-12). Furthermore, *Gadd45* TKO (CDS) mESCs showed impaired MSC differentiation (Figure 5-14). However, the bulk RNA-seq and qPCR analysis is limited due to cell type heterogeneity, making it difficult to conclude whether the reduced expression of cardiac marker genes is due to reduced number of cardiomyocytes or reduced expression in individual cardiomyocytes. Furthermore, the origin of this cardiac differentiation defect due to any prior impaired lineage commitment in *Gadd45* TKO (CDS) remains unclear. Therefore, I need to perform scRNA-seq in multiple time points of cardiac differentiation from WT and *Gadd45* TKO (CDS) mESCs.

Owing to their large cell size (>70 µm) cardiomyocytes pose challenges to conventional microfluidics and cell sorting based scRNA-seq approaches [251]. Therefore, I implemented a pilot scRNA-seq in singlets using the split-pool barcoding approach to overcome the cell size limitations, in day 12 cardiac differentiated WT (~5000 cells) and *Gadd45* TKO (CDS) mESCs (~6000 cells) [252]. The pilot sequencing was performed in singlets to test for suitability of this scRNA-seq approach and potentially to identify trends in gene expression and cell number without statistical analysis.

The scRNA-seq data quality was good with ~70% of reads mapping to protein coding genes, <4% mitochondrial reads and >75% normalized read coverage across gene body (data not shown). We identified ~2500-3000 genes in both cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. In collaboration with [redacted] (Bioinformatics core facility at IMB), we analysed the scRNA-seq data using Igraph-Louvain unsupervised clustering approach with k=20 (k represents number of nearest neighbours when creating clusters) and identified 10 cell clusters (Figure 5-15 A) [253]. For example, cluster 8 was associated with high expression of angiogenesis and vasculature development related genes, while cluster 9 was associated with high expression of known cardiomyocyte markers (*Tnnt2, Myh6 and Actc1*) (Figure 5-15 B). This suggests that the split-pool barcoding approach overcomes the cell size limitations and is suitable for scRNA-seq analysis of cardiac differentiated mESCs.

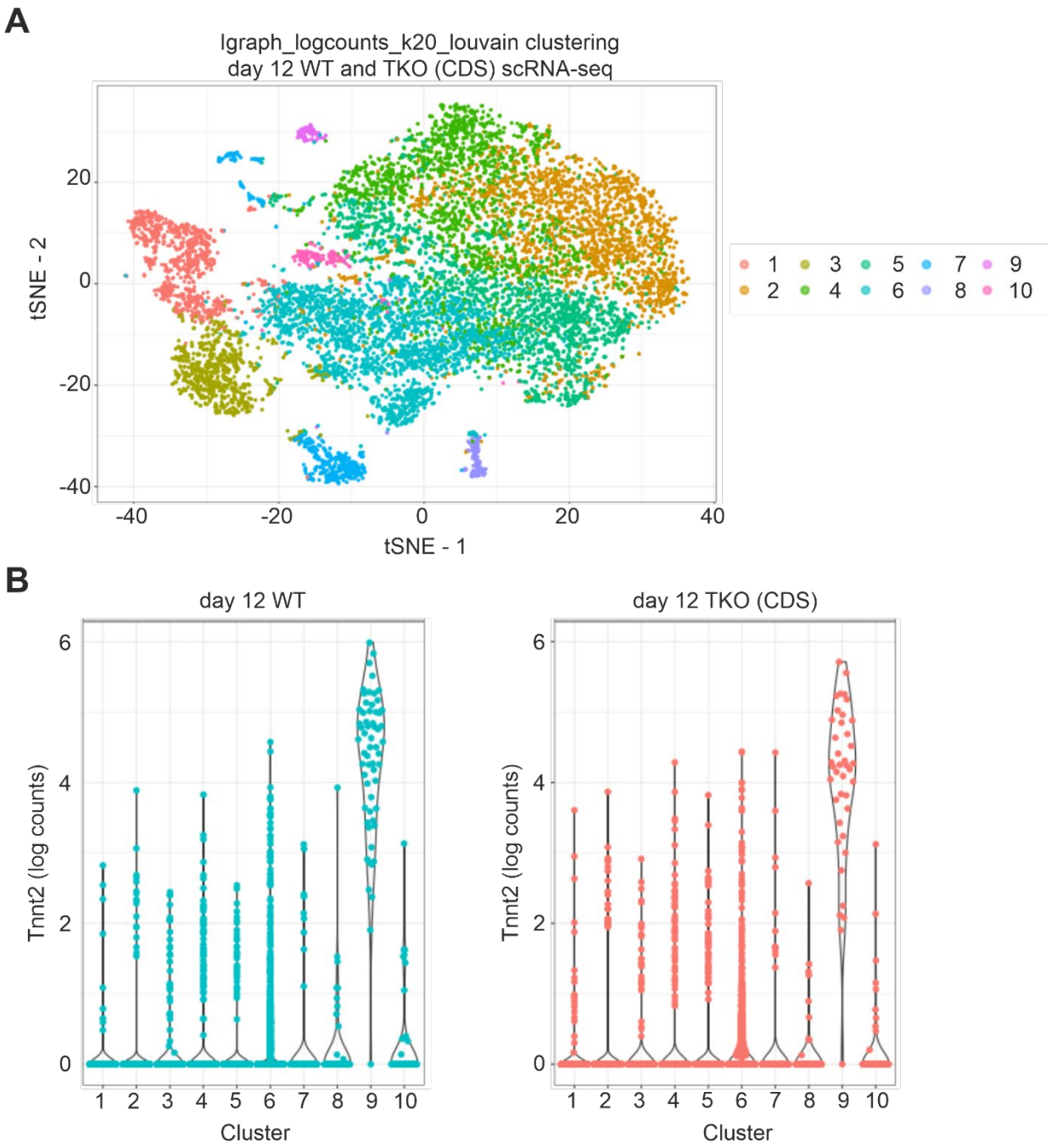

Figure 5-15: Pilot scRNA-seq in day 12 cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. (A) Igraph-Louvain unsupervised clustering represented using t-Distributed Stochastic Neighbour Embedding (tSNE). The parameter k20 indicates the number of nearest neighbours used in the graph-based clustering algorithm. Each point in the tSNE plot represents a single cell, with colours denoting distinct clusters identified from day 12 cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. **(B)** Violin plots represent the expression of *Tnnt2* gene (as log counts) in individual clusters (1 to 10) identified by Igraph-Louvain unsupervised clustering. Each point represents a single cell in day 12 cardiac differentiated WT (blue) and *Gadd45* TKO (CDS) (pink) mESCs. Cluster 9 represents the high cardiomyocyte marker (*Tnnt2*) expressing cells, where there are more cells in day 12 cardiac differentiated WT (60 blue dots) compared to *Gadd45* TKO (CDS) (39 pink dots) mESCs.

We could not perform any statistical analysis in this pilot scRNA-seq, since there were no replicates in our experimental design. Therefore, I analysed for potential trends in expression of cardiac marker genes by plotting their expression across the 10 clusters and manually counted the number of cells in cluster 9 (cardiomyocyte annotated cluster) in day 12 cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs. In line with our bulk RNA-seq data (Figure 5-12 D), scRNA-seq data showed a trend of reduced *Tnnt2* expression (log counts) in day 12 cardiac differentiated *Gadd45* TKO (CDS) mESCs (Figure 5-15 B). Additionally, there were 60 vs. 39 cells with high *Tnnt2* expression in cluster 9 from day 12 cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs, respectively (Figure 5-15 B). The lower number of cardiac marker expressing cells was not due to low starting number of cells from cardiac differentiated *Gadd45* TKO (CDS) mESCs (scRNA-seq analysis performed with 5000 WT cells vs. 6000 *Gadd45* TKO (CDS) cells). These preliminary findings hint that not only lower number of cardiomyocytes arise during cardiac differentiation of *Gadd45* TKO (CDS) mESCs, but also the individual cardiomyocytes have reduced expression of cardiac marker genes. This reduced expression of genes essential for sarcomere formation in individual cardiomyocytes, aligns with the significantly lower number of beating EB in cardiac differentiated *Gadd45* TKO (CDS) mESCs (Figure 5-11 D).

5.4. Discussion

The development of the mouse embryonic heart is a highly coordinated process involving the specification, differentiation, and morphogenesis of various cell populations [139,140,147]. Key TFs and signalling pathways guide these processes, and disruptions often lead to CHDs [192–195]. The precise control over gene expression (including expression of TFs) during embryonic heart development is facilitated by the dynamic and reversible nature of epigenetic modifications, such as DNA methylation and histone modifications [28,192,193]. Despite these insights, our mechanistic understanding of the factors modulating the epigenetic landscape during embryonic heart development remains limited. The unpublished results from our lab in *Gadd45a/Tet1* DKO mice hint at synergistic interplay between GADD45a and TET1 proteins in regulating embryonic heart development. In part - II of my thesis, I investigated the underlying epigenetic mechanism of how GADD45 proteins regulate TET1 functions to regulate gene transcription during embryonic heart development.

RNA-seq analysis revealed significant downregulation of heart development-associated genes in *Gadd45* TKO (CDS) mESCs (Figure 5-5 D, E). Importantly, *Gadd45* genes exhibit functional redundancy in regulating most heart development-associated genes in mESCs, in line with absence of embryonic heart development defects in *Gadd45a* SKO mice (Figure 5-2 B). Interestingly, the downregulation of heart development-associated genes was independent of GADD45 signalling functions in mESCs (Figure 5-7). Furthermore, the partial rescue of heart development-associated gene expression following combined overexpression of Gadd45a, b, g in *Gadd45* TKO (CDS) mESCs, reinforces the hypothesis that GADD45 proteins are critical regulators of these heart development-associated genes in mESCs (Figure 5-8).

In human cells, GADD45 proteins were shown to recruit TET1, enhancing their DNA demethylation activity [106,107]. However, the genome-wide prevalence and whether GADD45 proteins regulate embryonic heart development via a similar mechanism was unknown. One of the most striking result from our work was the increased TET1 occupancy at ~6300 genomic loci in *Gadd45* TKO (CDS) mESCs, which overlap substantially with the down DEGs (Figure 5-9 A). This genome-wide increase in TET1 occupancy aligns the mass spectrometry results showing elevated levels of 5mC oxidation intermediates (5hmC, 5fC, and 5caC) (Figure 5-3 D). These results were in contradiction with the proposed hypothesis of GADD45 acting as a recruiter of TET1 to regulate transcription in mESCs.

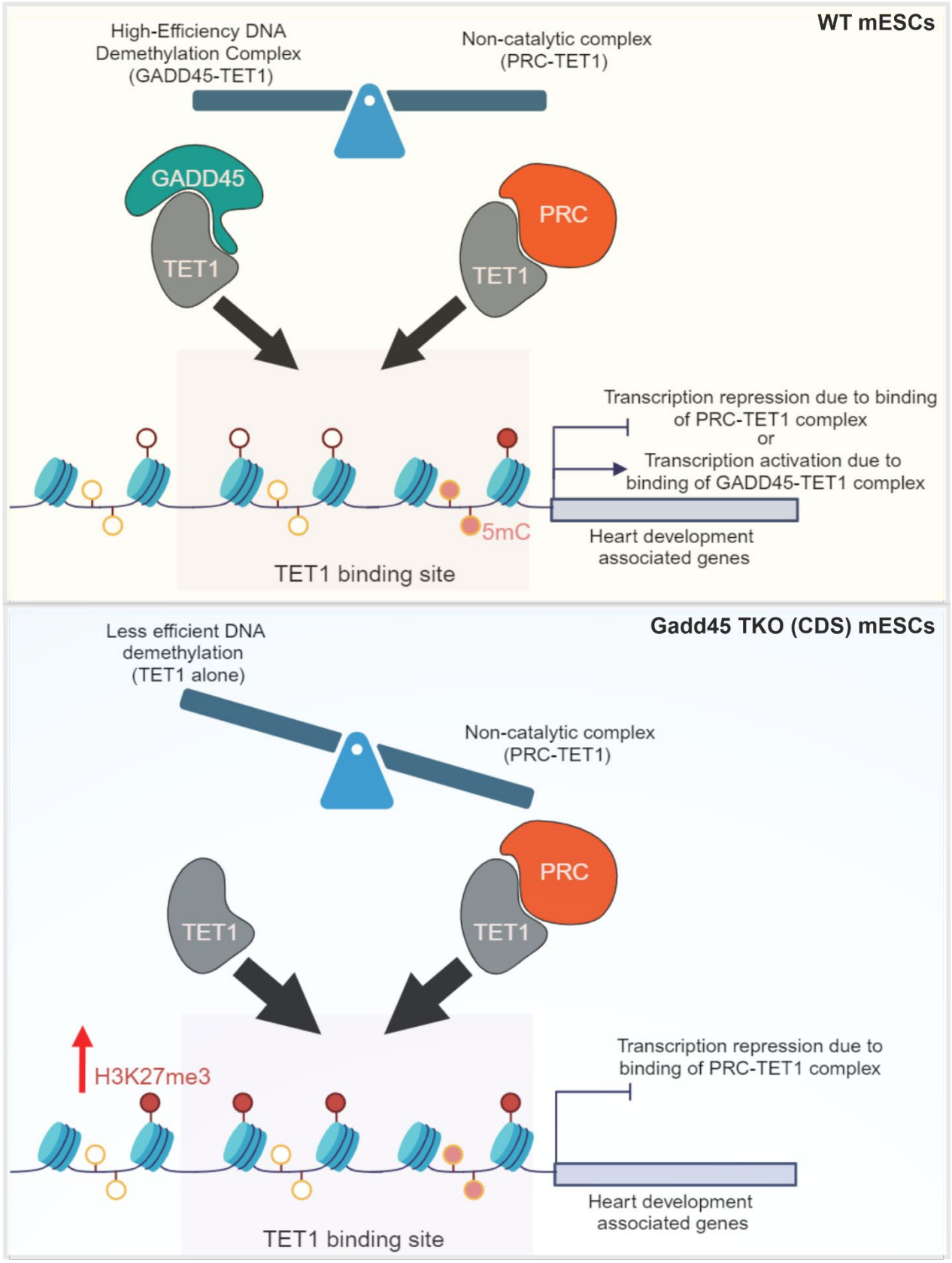

Figure 5-16: Proposed model of GADD45 regulating heart development-associated genes by balancing TET1 catalytic and non-catalytic functions. The high efficiency DNA demethylation complex (GADD45-TET1) and the repressive non-catalytic complex (PRC-TET1) regulate transcription through dynamic changes in 5mC and H3K27me3 levels in promoters of heart development-associated genes in WT

mESCs (top panel). The loss of GADD45 proteins and increased TET1 recruitment in *Gadd45* TKO (CDS) mESCs shifts the balance towards repressive non-catalytic TET1 functions leading to reduced transcription (bottom panel). The thickness of TET1 arrows symbolize increased recruitment to target loci.

It is known that GADD45 proteins enhance DNA demethylation catalytic activity of TET1 (high efficiency DNA demethylation complex) in comparison to DNA demethylation by TET1 alone (Figure 5-16) [52]. The increased TET1 recruitment could therefore be a compensatory cellular response to maintain DNA methylation dynamics, in the absence of GADD45 (Figure 5-9 A, 5-16). I hypothesize that this compensatory increase in TET1 occupancy creates imbalance between catalytic and non-catalytic TET functions. The increased TET1 binding at its target sites in absence of GADD45 could inadvertently lead to increased recruitment of PRC-TET1 complex, disrupting the underlying epigenetic landscape (Figure 5-16) [224]. Moreover, our data shows that the downregulation of heart development-associated genes was independent of DNA methylation changes despite increased TET1 occupancy (Figure 5-9 A, C). This observation aligns with increased non-catalytic TET1 activity via recruitment of PRC at these heart development-associated genes. I therefore adapted my hypothesis to test whether GADD45 proteins regulate transcription by suppressing the non-catalytic TET1 functions during embryonic heart development. The hypothesis based on these findings is summarized in a model (Figure 5-16).

In support of my hypothesis, ChEA of the downregulated genes with increased TET1 occupancy showed enrichment of sites known to have PRC protein binding in mESCs (Figure 5-10 A). Furthermore, the DEGs from *Gadd45* TKO (CDS) mESCs overlap significantly with DEGs from *Tet1* KO mESCs, with the overlapping genes enriched for PRC protein binding sites (Figure 5-10 B-E) [224]. Further studies, such as CUT&Tag for H3K27me3 or SUZ12, will be necessary to elucidate the potential increase in TET1-mediated PRC recruitment at heart development-associated genes in *Gadd45* TKO (CDS) mESCs. In addition, it is important to map GADD45 binding sites in mESCs to test whether GADD45-TET1 bound sites show lower PRC occupancy in WT mESCs and whether these sites are “hijacked” by TET1-PRC complex in *Gadd45* TKO (CDS) mESCs.

Despite the global increase in TET1 occupancy and major transcriptome changes, *Gadd45* TKO (CDS) mESCs demonstrated the ability to differentiate into all three germ layers during EB differentiation, albeit with a bias towards neuroectoderm lineages (Figure 5-4). The bias towards ectoderm differentiation could be linked to altered expression of cell migration regulating genes, as cells migrating through the primitive streak differentiate into mesoderm and endoderm, whereas cells not ingressing comprise the ectoderm [139,143]. Interestingly, genes positively regulating cell migration are also downregulated in *Gadd45* TKO (CDS) mESCs (Figure 5-5 D). However, this hypothesis needs further investigation. These results suggest that while GADD45 proteins are not essential for maintaining pluripotency but they may influence lineage specification (Figure 5-4, 5-

3 C). This was further demonstrated by the impaired cardiac differentiation and reduced number of beating EBs from *Gadd45* TKO (CDS) mESCs (Figure 5-11 D, 5-12, 5-13). Additionally, MSCs are multipotent stem cells required for tissue homeostasis and are known to differentiate towards cardiomyocytes during regeneration post-injury [246]. I observed MSC differentiation defect during cardiac and direct MSC differentiation of *Gadd45* TKO (CDS) mESCs (Figure 5-14). These results are in line with increased *Gadd45* mRNA levels during cardiomyocyte re-differentiation post ischemic injury, potentially extending the role of GADD45 proteins during cardiomyocyte regeneration and homeostasis [217].

Therefore, I took a more generalized approach to study early cardiac lineage commitment defects and implemented the split-pool barcoding scRNA-seq approach in cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs (Figure 5-15). The preliminary results show a trend for reduced expression of cardiac marker genes and reduced number of cardiomyocytes in cardiac differentiated *Gadd45* TKO (CDS) mESCs (Figure 5-15 B). These results will be followed up by performing scRNA-seq in biological replicates and at multiple cardiac differentiation time points in WT and *Gadd45* TKO (CDS) mESCs. This will help us map the earliest cardiac lineage commitment defect during cardiac differentiation in *Gadd45* TKO (CDS) mESCs and identify differential epigenetic changes due to imbalanced TET functions in differentiating *Gadd45* TKO (CDS) mESCs. For instance, single cell multiome (RNA-seq and mapping H3K27me3 sites) analysis will help understand lineage-specific epigenetic alterations in cardiac differentiated WT and *Gadd45* TKO (CDS) mESCs.

Lastly, it is important to note the substantial differences observed between *Gadd45* TKO (CDS) mESCs (used in part - II of my thesis) and previously published *Gadd45* TKO (with only exon1, 2 deletions) mESCs [77]. There were only 35 DEGs in *Gadd45* TKO (with only exon1, 2 deletions) mESCs in contrast to ~2000 DEGs in *Gadd45* TKO (CDS) mESCs (Figure 5-5 A, B). The poor overlap of 11 DEGs did not contain the 2c-like genes shown to be deregulated in *Gadd45* TKO (with only exon1, 2 deletions) mESCs (Figure 5-5 B) [77]. Unlike the global DNA hypomethylation observed in *Gadd45* TKO (CDS) mESCs, the *Gadd45* TKO (with only exon1, 2 deletions) mESCs showed a bias towards DNA hypermethylation (~7000 sites). Moreover, the increased bias towards neuroectoderm lineages and impaired expression of heart development associated genes was not observed in *Gadd45* TKO (exon 1, 2 deletions) mESCs [77]. Interestingly, shorter Gadd45 transcript variants (exon 2 skipped form) are known to be expressed in cells [225]. Despite lacking start codons, the exon 1, 2 deleted Gadd45 transcripts, which are highly expressed in *Gadd45* TKO (with only exon1, 2 deletions) mESCs, could get translated into GADD45 proteins with exon 1, 2 deletions via non-AUG start codon [254]. Therefore, it is plausible that the conclusions drawn in Schüle *et. al.* (2019) stem from the absence of full-length GADD45 proteins or the presence of GADD45 proteins

lacking exons 1 and 2. In extension, my novel findings from *Gadd45* TKO (CDS) mESCs likely reflect the genuine defects linked to complete *Gadd45* knockout in mESCs.

In summary, my work highlights the importance of GADD45-TET1 interaction (direct or indirect) and furthers our understanding in embryonic heart development. My data also points toward a novel mechanism wherein the GADD45-TET1 interaction (directly or indirectly) limits the binding of repressive PRC-TET1 complex to chromatin and promotes timely activation of heart development genes during normal embryonic heart development.

6. Material and Methods

6.1. Material

6.1.1. Equipment

-150°C freezer (Sanyo); -20°C freezer (Liebherr); -80°C freezer (Sanyo); agarose gel chambers (Bio-Rad); 2100 Bioanalyzer (Agilent technologies); bacterial incubators (Thermo Scientific); bacterial shaker (Infors); balances (Sartorius); Bioruptor pico (Diagenode); cell counter (Bio-Rad); cell culture dishes and flasks (TPP); cell culture incubators (Thermo Scientific); centrifuges (Heraeus); ChemiDoc XR+ System (Bio-Rad); ChemiDoc XRS+ System (Bio-Rad); Criterion TGX Precast Midi Protein Gels (Bio-Rad); cryo tubes (Greiner Bio-One); Cryo-Safe Cooler (Belart); DNA LoBind 1.5 ml Tubes (Eppendorf); DynaMag-2 Magnet rack (Invitrogen); E-Gel electrophoresis system (Invitrogen); fridges (Liebherr); heating blocks (Eppendorf); Infinite M200 plate reader (Tecan); laminar flow hoods (Dometic); LightCycler 480 (Roche); magnetic stirrer (Heidolph); MaXtract High Density, 2mL (Qiagen); microcentrifuges (Heraeus); microscope (Leica); microwave oven (Sharp); multichannel pipettes (Sartorius); multidispenser pipette (Eppendorf); Nanodrop 2000 spectrophotometer (Thermo Scientific); NextSeq500 (Illumina); NextSeq2000 (Illumina); orbital shaker (Neolab); PAGE midigel chambers (Bio-Rad); PCR thermocyclers (Biometra); pH meter (Mettler Toledo); pipet boy (Integra); pipettes (Eppendorf); power supplies (Bio-Rad); Qubit (Thermo Fisher); rotator (Neolab); silicone sealing mats (nerbe); SpeedVac concentrator (Eppendorf); Falcon (Sarstedt); Trans-Blot Turbo (Bio-Rad); Trans-Blot Turbo Midi PVDF (Bio-Rad); tubes with cell-strainer cap (Falcon); ultrapure water purification system (Millipore); ultrapure water purification system (Sartorius); UV crossliner (Stratagene); vortexer (Scientific industries); water baths (Neolab);

6.1.2. Chemicals and pre-made buffers

0.05 % Trypsin-EDTA (1x) (Gibco); 0.1 % gelatin in ultrapure water (Millipore); 0.25 % Trypsin-EDTA (1x) (Gibco); agarose (Biozym); ammonium acetate (Sigma-Aldrich); ampicillin (Sigma-Aldrich); ATP (NEB); bicinchoninic acid (BCA) (Sigma-Aldrich); bovine serum albumin (BSA) (Sigma-Aldrich); BSA, Molecular Biology Grade (NEB); buffer EB (Qiagen); calcium chloride (Sigma-Aldrich); chloroform-isoamylalcohol (Roth); copper sulphate (Sigma-Aldrich); Cutsmart buffer (NEB); dimethyl adipimidate (DMA) (Sigma-Aldrich); dimethyl formamide (DMF) (Sigma-Aldrich); dimethyl sulfoxide (DMSO) (Sigma-Aldrich); dithiothreitol (DTT) (Sigma-Aldrich); DMEM high glucose, pyruvate, no glutamine (Gibco); DMEM, high glucose, no glutamine, no pyruvate (Gibco); Dorsomorphin (Sigma); DPBS, no calcium, no magnesium (Gibco); EDTA (Sigma-Aldrich); ES-grade FBS (PAN Biotech); ethanol (Sigma-Aldrich); ethidium bromide (Roth); fetal

bovine serum (FBS) (Lonza); 16% Formaldehyde (w/v) methanol-free (Thermo Scientific); Gelred Nucleic Acid Stain (Sigma); glycerol (Sigma-Aldrich); glycogen, RNA grade (Thermo Scientific); hydrochloric acid (Sigma-Aldrich); isopropanol (Sigma-Aldrich); L-Cysteine (Sigma-Aldrich); L-Glutamine 100x 200 mM (Gibco); L-Glycine (Sigma-Aldrich); Lipofectamine RNAiMAX (Invitrogen); lithium chloride (Sigma-Aldrich); lysis buffer AL (Qiagen); magnesium chloride (Sigma-Aldrich); MEM Non-essential amino acids (NEAA) 100x (Gibco); methanol (Sigma-Aldrich); β-mercaptoethanol (Sigma-Aldrich); nuclease-free water (Qiagen); NuPAGE LDS Sample Buffer (4X) (Invitrogen); OptiMEM (Gibco); paraformaldehyde (Sigma-Aldrich); PBS (Gibco); penicillin/streptomycin 10,000 U/ml (PAN); phenol-chloroform-isoamylalcohol (PCI) (Roth); potassium chloride (Sigma-Aldrich); Protease inhibitor cocktail tablets EDTA-free (Roche); QIAzol lysis reagent (Qiagen); random primers (Invitrogen); restore™ Western Blot Stripping Buffer (Thermo Scientific); retinoic acid (Sigma-Aldrich);Roti C/I (Roth); skim milk powder (Sigma-Aldrich); SOC medium (Thermo Scientific); sodium acetate (Sigma-Aldrich); sodium bicarbonate (Sigma-Aldrich); sodium chloride (Sigma-Aldrich); sodium deoxycholate (Sigma-Aldrich); sodium dodecyl sulfate (Sigma-Aldrich); sodium hydroxide (Sigma-Aldrich); sodium pyruvate 100 mM (Gibco); spermidine (Sigma-Aldrich); SuperSignal West Femto/Pico (Thermo Fisher Scientific); Tris-HCl pH 8 1M (Sigma-Aldrich); Triton X-100 (Sigma-Aldrich); Tween-20 (Sigma-Aldrich); NP-40 (Sigma-Aldrich); XAV939 (StemCell Technologies); XL1-blue chemically competent bacteria (homemade).

6.1.3. Enzymes

AluI (NEB); BsrGI-HF (NEB); DNAse I (Sigma); EcoRI-HF (NEB); HindIII-HF (NEB); Nuclease S1 (Invitrogen); NEBNext® Ultra™ II Q5® Master Mix (NEB); pA-Tn5 (IMB PPCF); Phusion High-Fidelity DNA Polymerase (NEB); ProteinaseK (Qiagen); Q5 Polymerase (NEB); RNAse free DNAse set (Qiagen); RNaseA (Qiagen); RNase H (NEB); RNase III (IMB PPCF); SspI-HF (NEB); Superscript II reverse transcriptase (IMB PPCF); XbaI (NEB); KLD enzyme mix (NEB)

6.1.4. Kits

Blood & Cell Culture DNA Midi kit (Qiagen); ChIP DNA clean and concentrator (Zymo Research); LightCycler 480 Probes Master (Roche); Evercode™ Cell Fixation v2 (Parse Biosciences, SKU: ECF2101 and ECF2001); Evercode™ WT Mini v2 (Parse Biosciences, Version 2.3-UM0021); MEGAscript SP6 (Fisher Scientific); MEGAscript™ T7 Transcription Kit (Thermo Scientific); NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific); PCR Mycoplasma Test Kit I/C (Promokine); Protein-A dynabeads (Invitrogen); Protein-G dynabeads (Invitrogen); Qiaprep Maxiprep kit (Qiagen); Qiaprep Midiprep kit (Qiagen); Qiaprep Miniprep kit (Qiagen); Qiaquick Gel extraction kit (Qiagen); QIAquick PCR purification kit (Qiagen); Qubit dsDNA HS Assay Kit

(Thermo Fisher Scientific); RNeasy Mini Kit (Qiagen); Stranded Total RNA Prep Ligation with Ribo-Zero Plus Kit (Illumina); NuGen Ovation Ultralow System V2 (Tecan).

6.1.5. Cell culture media and buffers

Table 6-1: Composition of buffers and solutions used in this study

Solution	Components
mESC medium	DMEM, 15% Pansera ESC-grade FBS, 2mM L-Glutamine, 50U/ml Penicilin/Streptomycin, 1mM sodium pyruvate, 1x NEAA, 100µM β-mercaptoethanol, 20ng/ml Leukemia Inhibitory Factor (LIF).
mESC differentiation medium	DMEM, 15% ESC-grade FBS, 2mM L-Glutamine, 50U/ml Penicilin/Streptomycin, 1mM sodium pyruvate, 1x NEAA, 100µM β-mercaptoethanol.
dNTP mix	10 mM dATP, 10 mM dCTP, 10 mM dGTP, 10 mM dTTP
Gelred (10 000x)	Invitrogen
Luria broth (LB)	for 1 l total volume in water: 10 g bactotryptone, 5 g yeast extract, 10 g NaCl pH 7.0, autoclaved
NuPAGE 4x +DTT	4x NuPAGE LDS sample buffer including 400 mM DTT
PBS	140 mM NaCl, 2.7 mM KCl, 1.5 mM KH_2PO4, 8.1 mM Na_2HPO4, pH 7.4, autoclaved
PBS-0.1 % Tween (PBS-T)	1x TBS, 0.1 % Tween-20
PBS-T 1% BSA	1x PBS, 1% BSA
PBS-T 5% BSA	1x PBS, 5% BSA
SDS-Running buffer 10x	for 1 litre total volume: 10 g SDS, 30.3 g Tris, 144.1 g glycine

TBE (10x)	1 M Trizma base, 1 M boric acid, 20 mM EDTA, autoclaved
TBS (20x)	3 M NaCl, 53.7 mM KCl, 839 M Trizma hydrochloride, 160 mM Trizma base pH 7.4, autoclaved
TBST	1x TBS, 0.1 % Tween-20
Western blot blocking buffer	5% milk in 1x TBST
TE buffer	10 mM Tris-HCl pH 8.0, 1 mM EDTA, autoclaved
NE buffer CUT&Tag	20 mM HEPES-KOH pH 7.6, 10 mM KCl, 0.1% Tween-20, 0.5 mM Spermidine, 20% Glycerol, 1X Protease inhibitor Cocktail (PIC)
Binding buffer CUT&Tag	20 mM HEPES-KOH pH 7.6, 10 mM KCl, 1 mM $MgCl_2$, 10 mM CaCl2
Blocking buffer CUT&Tag	20 mM HEPES-KOH pH 7.6, 150 mM NaCl, 0.5 mM Spermidine, 0.1% BSA, 2 mM EDTA, 1X PIC
Wash buffer CUT&Tag	20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 0.1% BSA, 1X PIC
pA-Tn5 buffer	2 µl of pA-Tn5 diluted in 20 mM HEPES-KOH pH 7.5, 0.5 mM Spermidine, 0.1% BSA, 1X PIC 300mM NaCl, 0.025% Tween 20
Tagmentation buffer	20 mM Tris-HCl pH 8.3, 300 mM NaCl, 10mM $MgCl_2$, 0.8 mM ATP, 10% dimethyl formamide (DMF)
Stop buffer CUT&Tag	20 mM EDTA, 20 mM EGTA, 0.03% SDS
Dimethyl adipimate (DMA) cross-linker	200mM stock solution prepared by dissolving 50mg DMA powder in 500µl PBS and 500µl DMSO
ChIP Lysis Buffer A	100mM Tris-HCl pH 8, 10mM DTT in ultra-pure water

ChIP Lysis Buffer B	10mM HEPES-KOH pH 7.5, 10mM EDTA, 0.5mM EGTA, 0.5% Triton X-100 in ultra-pure water
ChIP Lysis Buffer C	10mM HEPES-KOH pH 7.5, 10mM EDTA, 0.5mM EGTA, 200mM NaCl in ultra-pure water
ChIP Lysis Buffer D	50mM Tris-HCl pH 8.0, 10mM EDTA, 0.75% SDS in ultra-pure water
ChIP Buffer	15 mM Tris–HCl pH 8.0, 180 mM NaCl, 1.2 mM EDTA, 1.2% Triton X-100 in ultra-pure water
ChIP Wash Low Salt Buffer	20mM Tris–HCl pH 8.0, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100 in ultra-pure water
ChIP Wash High Salt Buffer	20mM Tris–HCl pH 8.0, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100 in ultra-pure water
ChIP Wash LiCl Buffer	10mM Tris–HCl pH 8.0, 250mM LiCl, 1mM EDTA, 1% sodium deoxycholate, 1% NP-40 in ultra-pure water
Papain activation buffer	1.1 mM EDTA, 0.067 mM mercaptoethanol and 5.5 mM cysteine-HCl
DnaseI solution for scRNA-seq cell preparation	2% Pansera ESC-grade FBS, 1x DnaseI buffer (Qiagen) diluted in 1x PBS buffer w/o magnesium and calcium

6.1.6. Antibodies

Table 6-2: Antibodies used in this study

Antibody	Supplier	Catalog number
Alexa Fluor 488 Goat anti-mouse	Invitrogen	A-11001
Alexa Fluor 488 Goat anti-rabbit	Invitrogen	A-11008
Alexa Fluor 546 Goat anti-rabbit	Invitrogen	A-11010

Alexa Fluor 633 Goat anti-Mouse	Invitrogen	A-21050
Alexa Fluor 633 Goat anti-Rabbit	Invitrogen	A-21071
CASPASE-3 antibody	Cell Signaling Technology	9662S
GAPDH antibody	Thermo Scientific	PA1-987
GFP antibody	Sigma-Aldrich	SAB4301138
Goat anti-rabbit-HRP	Dianova	111-035-144
Goat-anti-mouse-HRP	Dianova	115-035-146
Guinea Pig anti-Rabbit IgG	Antibodies online	ABIN101961
HA tag antibody	Abcam	ab9110
HA tag antibody	Cell Signaling Technology	3724S
HA tag antibody	Biolegend	901501
HA Tag CUTANA™ antibody	Epicypher	13-2010
HISTONE H3 antibody	Abcam	ab1791
HSF1 antibody	Cell Signaling Technology	4356S
HSP40 antibody	Cell Signaling Technology	4871S
HSP90α antibody	Cell Signaling Technology	8165S
MEKK4 antibody	Invitrogen	PA5-52459
p38 MAPK antibody	Cell Signaling Technology	8690S

PARP1 antibody	Cell Signaling Technology	9542S
Phospho ERK1/2	Cell Signaling Technology	9101S
Phospho serine-326 HSF1 antibody	Abcam	ab115702
Rabbit anti-Mouse IgG	Thermo Scientific	31190
S9.6 antibody	Homemade IMB PPCF	-
TATA binding protein (TBP) antibody	Abcam	ab818
TET1 antibody [N1]	Biozol	GTX125888
Cardiac Troponin T Antibody	Bioss	BSS-BS-0648R
TUBULIN antibody	Sigma	T5168-.2ML

6.1.7. Oligonucleotide sequences

Table 6-3: UPL assays used for RT-qPCR (mouse)

Gene	**Sequence forward (5'→3')**	**Sequence reverse (5'→3')**	**UPL Probe**
Actc1 mRNA	CAGAGCTGTCTTCCCGTCCATC	CAGAGTCAGGATACCTCGCTTG	#123
Cd105 mRNA	CAGCCAAAGTGTGGCAATCAGG	GCTACTCAGGACAAGATGGTCG	#95
Cd44 mRNA	CGGAACCACAGCCTCCTTTCAA	TGCCATCCGTTCTGAAACCACG	#35
Dlk1 mRNA	CCCTGCGTGATCAATGGT	CACAGAAGTTGCCTGAGAAGC	#4
Dnajb1 mRNA	TTTTCGACCGCTATGGAGAG	TAGCACCACCACTGCTTCCT	#45
Dnajb1 promoter	GTCCGACGACGAGATCAAG	TCTTGTCCGGGTGGTAGC	#70

Eomes mRNA	ACCGGCACCAAACTGAGA	AAGCTCAAGAAAGGAAACATGC	#9
Fgfr2 mRNA	AGACCACAAATGGGCGACT	CCACATTAACACCCCGAAGGA	#92
G6pd mRNA	AAGCAGAGTGAGCCCTTC	CATAGGAATTACGGGCAAAGA	#78
Gadd45a mRNA	GCTGCCAAGCTGCTCAAC	TCGTCGTCTTCGTCAGCA	#98
Gadd45b mRNA	CGGCCAAACTGATGAATGT	ATCTGCAGAGCGATATCATCC	#79
Gadd45g mRNA	GTCCGCCAAAGTCCTGAAT	GCTATGTCGCCCTCATCTTC	#71
Gata6 mRNA	GGTCTCTACAGCAAGATGAATGG	TGGCACAGGACAGTCCAAG	#40
Hand1 mRNA	CAAGCGGAAAAGGGAGTTG	GTGCGCCCTTTAATCCTCTT	#51
Hnf4a mRNA	CCAAGAGGTCCATGGTGTTTA	CCGAGGGACGATGTAGTCAT	#68
Hoxa2 mRNA	GAAGGCGGCCAAGAAAAC	CATCAGCTATTTCCAGGGATTC	#70
Hsp90aa1 mRNA	GTGTTCATTCAGCCACGATG	CATTAACTGGGCAATTTCTGC	#25
Hsp90aa1 pre-mRNA	CAGCCACGGTGAGTGATG	CTGAGGGAACTCGGATGGT	#32
Hsp90ab1 mRNA	CGCAAGAACATCGTCAAGAA	GGAATCTTCATGAATTCCAAGC	#1
Hsp90ab1 pre-mRNA	GTAAGGCCGGGGTCTCTC	GAGGCTTTGGGACAGATGAG	#75
Hsp90ab1 promoter	CCAGAACTCAGGGGACTGAA	TGTGCGATTTGTAGGGAACA	#10
Hspa1a mRNA	AGGCCTCTGCTGGCTCTC	TGCAGGACAAACTAAGGAGTGA	#94
Hspa1a pre-mRNA	CGCCTACTTCAACGACTCTCA	GATCCGCAGCACGTTTAGA	#49

Hspa1b promoter	CCAGGAATCTTCCAGCAGTT	CCCACAACTCCGATTACTCAA	#22
Hsph1 mRNA	CAACCTCAAGAAGCCAGTGA	AGTTCAAGCCCACAATCTGC	#42
Hsph1 pre-mRNA	CTTCCCGAGGGAGATGGT	GAAGCTGAGTGGGGAGTCG	#97
Isl1 mRNA	ATGATGGTGGTTTACAGGCTAAC	TCGATGCTACTTCACTGCCAG	#134
Kdr mRNA	TGCCTACCTCACCTGTTTCC	GTCTGGCTGTCATCTGGGA	#18
Myh6 mRNA	CCAGCCATCTCCTCTGTTAGGT	GCTGGAAGATGAGTGCTCAGAG	#1
Nanog mRNA	GCCTCCAGCAGATGCAAG	GGTTTTGAAACCAGGTCTTAACC	#25
Nestin mRNA	TCCCTTAGTCTGGAAGTGGCTA	GGTGTCTGCAAGCGAGAGTT	#67
Oct4 mRNA	AATGCCGTGAAGTTGGAGAA	CCTTCTGCAGGGCTTTCAT	#95
Pax3 mRNA	GCCCACGTCTATTCCACAA	GAATAGTGCTTTGGTGTACAGTGC	#69
Pax6 mRNA	GTTCCCTGTCCTGTGGACTC	ACCGCCCTTGGTTAAAGTCT	#78
Ptch1 mRNA	GATGCTTGCAGGGTCCGAAT	GGCTGGAGACACCTCAGGAC	#162
Rhox2a mRNA	GGGAGTGAGTGAAGCCACAG	GCCTTTACAGCCTCTTATAACTCCT	#91
Sox2 mRNA	GACGTCGTAGCGGTGCAT	ACGGCAGCTACAGCATGA	#68
T mRNA	GCTCTAAGGAACCACCGGTCATC	ATGGACTGCAGCATGGACAG	#23
Tbp mRNA	CCAGGAAATAATTCTGGCTCA	GGGGAGCTGTGATGTGAAGT	#97
Tek mRNA	TGTCACGAGGTCAAGAAGTGT	ACACCATAGGACCAGACATCAC	#158
Tfap2b mRNA	CAGTGACCTGCACTCCAGAA	GCCAGCAGATCCGTAAATTC	#49

Tnnt2 mRNA	GATGCACTTTGGAGGGTACA	CCTTGGCCTTCTCTCTCAGTT	#53

6.1.8. Plasmids

Table 6-4: Plasmids used in this study

Insert name	Supplier	Catalog number
dRNaseH1-EGFP-dCas9 (dRED) (OE)	Addgene	139836
Mouse *Tnnt2* cDNA ORF Clone, N-GFPSpark® tag	SinoBiological (Biozol)	MG51609-ANG
pCS2 + myc-Mm*Gadd45a*	[redacted] (unpublished, this lab)	-
pCS2 + myc-Mm*Gadd45b*	[redacted] (unpublished, this lab)	-
pCS2 + myc-MmG*add45g*	[redacted] (this lab)	Reference [255]
RNASEH1-GFP (OE)	Addgene	108699
Xenopus laevis cTnI (WISH probe)	Gift from Prof. Michael Kühl lab, Ulm, Germany	Reference [177]
Xenopus laevis Islet1 (WISH probe)	Gift from Prof. Michael Kühl lab, Ulm, Germany	Reference [177]
Xenopus laevis MHC-α (WISH probe)	Gift from Prof. Michael Kühl lab, Ulm, Germany	Reference [177]
Xenopus laevis Nkx2-5 (WISH probe)	Gift from Prof. Michael Kühl lab, Ulm, Germany	Reference [177]
Xenopus laevis Tbx5 (WISH probe)	Gift from Prof. Michael Kühl lab, Ulm, Germany	Reference [177]

6.1.9. Software

Table 6-5: Softwares used in this study

Name	Version
Agilent Mass Profiler Professional	V13.0
Agilent MassHunter Qualitatitve Analysis	B.06.00
Agilent MassHunter Quantitative Analysis	B.05.02
Agilent MassHunter Quantitative Analysis	B.09.00
BioRad Image Lab	5.2.1
Fiji Image J	V5
GraphPad Prism	10.2.2
LAS X Leica software	1.4.6
LightCycler 480 software	1.5.1.62
Tecan i-control	2.0.10.0

6.2. Methods

6.2.1. Molecular biology

General molecular biology methods including preparation of chemically competent XL1-blue E. coli bacteria, plasmid amplification in E. coli, spectrophotometric quantification of DNA and RNA, restriction digests, DNA ligations, PCR, and agarose gel electrophoresis were carried out as previously described [256]. All oligonucleotides used in this study were synthesized by Integrated DNA Technologies (IDT). Plasmid DNA was sequenced by StarSEQ.

6.2.2. Plasmid maintenance and purification

For long-term storage, bacterial culture was mixed with 33% glycerol and stored at -80°C. To recover the plasmids, the frozen glycerol stock was scraped with a pipette tip, which was then transferred to 5 ml LB medium containing antibiotic for overnight incubation at 37°C. For plasmid preparation, QIAGEN Mini and Midi Kits were used according to manufacturer's protocol. DNA amount and purity were estimated on a Nanodrop 2000 spectrophotometer.

6.2.3. Cloning of plasmids

Plasmids generated for this study were cloned using the KLD Enzyme Mix from New England BioLabs (NEB), according to the manufacturer's recommendations. In brief, PCR primers were designed using NCBI primer design, PCRs were performed with Q5 polymerase. PCR-amplified inserts were purified using QIAquick Gel extraction kit, and ligated with the backbone vector using the KLD Enzyme Mix. Plasmids were then amplified in XL1-blue competent E.coli, purified and sequenced to confirm correct insertion.

6.2.4. Genomic DNA extraction and LC-MS/MS analysis

Genomic DNA was isolated with a DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's instructions after treatment of lysed cells with RNase A (final concentration 1 mg/ml) for 10 min in RT. DNA was subsequently ethanol precipitated with ammonium acetate. Michael Musheev carried out quantification of 5mC, 5hmC, 5fC and 5caC as previously described [77].

6.2.5. RNA isolation and cDNA synthesis

RNA was isolated using the QIAGEN Qiazol reagent according to manufacturer's protocols. RNA pellet was resuspended in nuclease free water. Residual DNA was digested prior to complementary DNA (cDNA) synthesis using Amplification Grade DNaseI (Cat. No. AMPD1,

Sigma), according to manufacturer's protocols. cDNA was generated using reverse transcriptase provided by IMB protein production CF according to supplemented protocols. 1 µg RNA (DnaseI treated) was mixed with 2 µl 100 µM dN6 primer and 1 µl 10 mM dNTPs in a final volume of 12 µl. The mix was denatured 5 min at 65°C and cooled on ice for 2 min. 4 µl 5x First Strand Synthesis buffer, 2 µl 0.1 M DTT, 1 µl Ribolock and 0.2 µl reverse transcriptase were added and samples were incubated for 10 min at 25°C, 90 min 42°C and 5 min 72°C. The resulting cDNA was diluted 1:6 in nuclease free water for qPCR.

6.2.6. Real-time quantitative PCR

Real-time quantitative PCR (RT-qPCR) was conducted in a 384-well format using the Roche LightCycler480 System. It is based on detection of a specific PCR product using short hydrolysis probes. Gene-specific primers and corresponding probes were determined using the Roche Universal ProbeLibrary Assay Design Centre or primer3 (open source) & probeBAGEL (custom software provided by M. Misak). 5 µl diluted cDNA or diluted ChIP DNA sample was mixed with 5.5 µl 2x ProbesMaster, 0.12 µl probe, and 0.12 µl primer mixture containing forward and reverse primer at 50 mM each. The following settings were used for RT-qPCR:

	PCR step	temperature	time	Ramp rate
1	Initial Denaturation	95°C	10 min	4.8°C/s
2	Denaturation	95°C	10 s	4.8°C/s
3	Annealing	60°C	15 s	2.5°C/s
4	Elongation (signal acquisition per cycle)	72°C	10 s	4.8°C/s
	Go to step 2 and repeat 49 cycles			
5	Cooling	40°C	30 s	2.5°C/s

Cp values and expression levels were calculated using LightCycler 480 software release 1.5.1.62. All samples were measured in technical replicates and normalized to loading controls.

6.2.7. Western blot

For whole cell lysates, cultured cell were lysed in RIPA lysis buffer and scraped off from the plate. Cells were incubated on ice, and were intermittently pipetted / vortexed to achieve cell lysis. Samples were cleared by centrifugation for 10 min at 21,000 g at 4°C. Cell lysates were mixed with 4x NuPAGE sample buffer, containing 100 mM DTT and incubated for 10 min at 95°C. Samples were separated on a precast SDS-PAGE gel by applying 90V for 2 hours. For blotting, Trans-Blot Turbo Midi 0.2 µm PVDF Transfer membrane was used according to manufacturer's instructions. Briefly, the blotting setup was assembled as followed (from bottom to top): 2 whatman papers, PVDF membrane, PAGE gel, 2 whatman papers. The transfer was performed using the Turbo function for midi gels in the Trans-Blot Turbo transfer system for 7 min. Blots were blocked using 5% milk in 1x TBST for 60 min at room temperature. Primary antibody was diluted 1:1000 in western blot blocking buffer and blot was incubated overnight at 4°C. Blot was washed three times using TBST buffer. Secondary antibody was diluted 1:3000 in western blot blocking buffer and blot was incubated for 60 min at room temperature. Blot was washed three times using TBST buffer. Signals were developed with SuperSignal West Pico or Femto Chemiluminescent Substrate and analysed using a ChemiDoc with Image Lab software. Quantification of band intensity was performed using Image Lab Software.

6.2.8. BCA Assay

Protein lysate concentration was measured using the Bicinchoninic acid (BCA) assay. Briefly, a standard curve was prepared using known concentrations of a BSA protein standards. Working reagent was prepared by mixing BCA and copper sulphate solution in a 50:1 ratio. Protein lysates and standards were diluted 1:1 in DPBS buffer and 25µl was dispensed into a 96-well microplate. Working reagent (200 µl) was added to each well and the microplate was covered and incubated at 37°C for 30 minutes to allow colour development. After incubation, the absorbance of each well was measured at 562 nm using Tecan microplate reader. A standard curve was generated by plotting the absorbance values of the standards against their respective concentrations. The protein concentration of the samples was determined by interpolation from the standard curve. The final protein concentration was corrected for dilution factors and expressed as µg/µl.

6.2.9. Nuclear-Cytoplasmic fractionation

The nuclear-cytoplasmic protein fraction was performed using NE-PER Nuclear and Cytoplasmic Extraction Kit (Cat no. 78833). Briefly, cells were harvested by gently scrapping in cold 1 ml PBS and centrifuged at 1000 x g for 3 min in a cold centrifuge. Washed cells were re-suspended in PBS and centrifuged again at the same speed. To the cell pellet, 300 µl of ice-cold CER I (with protease inhibitor cocktail) was added. The pellet was fully suspended by gentle pipetting and incubated on

ice for 10 minutes. 16.5 µl of ice-cold CER II was added to the tube followed by gentle pipetting and incubation on ice for 1 minute. The tube was centrifuged at ~16,000x g for 5 minutes in a cold centrifuge. The supernatant (cytoplasmic extract) was immediately transferred to a clean pre-chilled tube and stored on ice. The pellet (nuclear fraction) was washed with 1 ml of cold PBS without resuspension and centrifuged. The nuclear pellet was suspended in 150 µl ice-cold NER and vortexed for 15 seconds. Samples were vortexed every ~15 minutes for a total of 40 minutes. The tube was centrifuged at ~16,000x g for 10 minutes in a cold centrifuge. The supernatant (nuclear extract) was transferred to a clean pre-chilled tube and stored at -20°C until further use. Respective protein lysate concentration was measured using BCA assay and equal amount of proteins from each fraction were run on western blot.

6.2.10. mESC cell culture

mESCs were cultured on pre-coated gelatinized (0.1% Gelatin, Millipore) plates in mESC medium at 37°C 5% CO_2. mESC medium was changed every day and cells were passaged every other day. For passaging, cells were washed with PBS and incubated 3 min with 2 ml of 0.05% trypsin/EDTA at 37°C. Trypsin was quenched by adding mESC medium and colonies were dissociated into single cells by gentle titruation. The cells were pelleted 3 min at 300xg, resuspended in fresh culture medium and split 1:8 to a new pre-coated gelatinized plate.

1 x 10^6 mESCs were seeded in a 10 cm dish. Next day the mESCs were treated with DMSO or respective inhibitors [SB202190 (20 µM), JNK-IN-7 (1 µM) and PD0325901 (1 µM)] for 24 hrs followed by protein and RNA extraction.

6.2.11. Cell freezing

To freeze cells from a culture dish they were dissociated as described before. After quenching and removing trypsin mESC were resuspended in 6 ml PANsera +10% DMSO. The suspension was mixed and distributed to four 1.5-ml-freezing-vials. Vials were slowly frozen in an isopropanol-based freezing container or styrofoam box at -80°C. To thaw cells they were resuspended in 10 ml pre-warmed medium. DMSO containing medium was removed by centrifugation at 300xg for 3 min and all cells were plated on a culture dish

6.2.12. Plasmid transfection

For harvesting after 48 h 1.8 x10^6 mESC were plated in culture medium w/o P/S on a 10-cm-dish. ~5 µg DNA was prepared in 250 µl OptiMEM and 15 µl Lipofectamine 2000 (Thermo-Fischer) (1 part DNA: 3 part Lipofectamine ratio) was prepared in 250 µl OptiMEM. Both solutions incubated for 5 min and mixed together by carefully pipetting up and down. After 20 min incubation at room

temperature 500 µl transfection solution was added dropwise to the cells and distributed by gentle swirling.

6.2.13. In-vitro embryoid body (EB) differentiation

For EB differentiation $3.5x10^6$ cells were plated on non-adherent 10 cm bacterial dishes (Greiner) in 15 ml mESC differentiation medium. Every other day, EBs were collected in a 50 ml falcon and settled down by gravity. EBs were resuspended in fresh mESC differentiation medium and transferred on a new non-adherent 10 cm bacterial dish. For retinoic acid treatment of EBs, 5µM retinoic acid (final concentration) was added to mESC differentiation medium on day 4 and day 6 of differentiation.

6.2.14. Cardiomyocyte differentiation protocol

0.8×10^6 mESCs were seeded in 10cm dish with 2 µM Dorsomorphin containing standard mESC media (day -3). The 2 µM Dorsomorphin containing standard mESC media was changed every day for the next 72 hrs. Dorsomorphin-treated mESCs were dissociated with trypsin and washed with standard media without LIF. The cell pellet was resuspended in 2 µM Dorsomorphin containing mESC differentiation media. The cell suspension was processed for hanging drop EB formation using 1000 cells in 25µl volume (day 0). The drops were pipetted in the lid of a 15cm cell culture plate. 15ml of DPBS was used to maintain humidity. The EB hanging drops were not disturbed for next 2 days. On day 2 of differentiation, individual 25µl EB containing drop was transferred to an individual well of 96 well ultra-low binding plate. 180µl of mESC differentiation media without Dorsomorphin was added to each well. On day 3 of differentiation, 22µl of mESC differentiation media containing 10 µM of XAV939 was added to each well, to achieve a final XAV939 concentration of 1uM. On day 4 of differentiation, the EBs were transferred to gelatin-coated 10cm dish (RNA-seq, scRNA-seq, qPCR assays) or 48-well plates for beating EB counting analysis. The cells are still in mESC differentiation media containing 1 µM of XAV939. On day 5 of differentiation, media was changed with mESC differentiation media without XAV939. From here on, mESC differentiation media was changed every 2nd day, and EBs start to show beating movements around 10-12 days of differentiation.

6.2.15. MSC differentiation protocol

mESCs were seeded at low density of ~0.5×10^6 per 10 cm cell culture dish and treated with 1 µM RA. The medium was refreshed every second day over the 10 day differentiation period. The differentiated cells formed a compact monolayer, which was then trypsinized and disaggregated. The resulting single-cell suspension was reseeded in an uncoated cell culture dish, allowing selection of MSCs which attach within ~30 minutes of seeding. The floating cells were removed by

changing the medium. This process was repeated one more time, yielding cells that exhibited typical MSC (fibroblast-like) morphology.

6.2.16. Fluorescence-activated cell sorting (FACS)

Cell sorting was conducted by the Cytometry CF of the IMB. RNASEH1-GFP transfected mESC were detached and dissociated as described before and resuspended in 2 ml mESC medium. Sorting was conducted in a Becton Dickinson FACSAria III SORP equipped with a 100 µm nozzle. All measured events were gated by forward scatter-area (FSC-A) and side scatter area (SSC-A) to exclude cell debris. Cell doublets were detected by measuring SSC-A towards SSC width. GFP positive cells were detected after excitation with a blue laser (BL 488 nm), passing a 530/30 band pass filter. Cells were gated by area of fluorescence signal relative to FSC-A.

6.2.17. Immunofluorescence assay

Cells growing on coverslips were washed briefly using 1× PBS (5 min, twice at room temperature) followed by fixation with 4% paraformaldehyde (PFA, Sigma, 158127) prepared in 1× PBS (pH 7.4) for 12 min. 250mM Glycine was added to quench any residual formaldehyde and incubated for 5 min. Cells were permeabilized and blocked simultaneously in 2% BSA (Sigma, A2153) and 0.5% Triton X-100 solution (prepared in 1× PBS) for 60 min. Primary antibodies (Table 6-2) were diluted in 0.5% BSA and 0.1% Triton X-100 solution (prepared in 1× PBS). Cells were incubated overnight at 4°C in primary antibody solution. Next day, cells were washed thrice using 0.1% Triton X-100 solution prepared in 1x PBS. Secondary antibodies (Table 6-2) were diluted in 0.5% BSA and 0.1% Triton X-100 solution (prepared in 1× PBS). Cells were incubated in secondary antibody for 60 min at room temperature. After three washes with 0.1% Triton X-100 solution prepared in 1x PBS, cells were counterstained with 0.05 µg/ml 4′,6-diamidino-2-phenylindole (DAPI) solution for 2 min at room temperature, washed in 1× PBS and mounted in Slowfade Gold Antifade (Invitrogen, S36937). Coverslips were stored at 4°C until they were imaged.

6.2.18. Chromatin Immunoprecipitation

Briefly, mESCs were fixed in 1% formaldehyde for 10 min and quenched in 200 mM glycine for 5 min. Cells were scraped on ice in cold 1x PBS and centrifuged at 4000 rpm. The supernatant was discarded, cell pellet was snap frozen in liquid nitrogen and stored at -80°C. Nuclei were isolated by resuspending the pellet in series of ChIP Lysis buffers. First, the cell pellet was resuspended in 1ml ice cold ChIP Lysis buffer A and incubated on ice for 15 min. Cells were centrifuged for 5 min at 4000 rpm at 4°C. The cell pellet was resuspended in 1ml ChIP Lysis buffer B and incubated on ice for 15 min followed by centrifugation. Cell pellet was resuspended in 1ml ChIP Lysis buffer C

and incubated on ice for 5 min, followed by centrifugation. Lastly, the nuclei were resuspended in 300 µl ChIP Lysis buffer D and containing 0.75% SDS. Chromatin was sheared in a Bioruptor Pico (Diagenode, Seraing, Belgium) to an average length of 200–500 bp. Chromatin was diluted five times in ChIP buffer, pre-cleared and incubated overnight at 4°C with anti-HA antibodies. Immune complexes were captured on Protein A/G magnetic beads (GE Healthcare, Chicago, IL, USA), serially washed using (a) ChIP Wash Low Salt Buffer, (b) ChIP Wash High Salt Buffer, (c) ChIP Wash LiCl Buffer and (d) TE buffer at room temperature. DNA was eluted in ChIP elution buffer (0.1 M $NaHCO_3$, 1% SDS). After decrosslinking at 65°C for 6 h, DNA was purified with the ChIP DNA Clean and Concentrator Kit (Zymo Research). Immunoprecipitated and input DNA was analysed by qPCR (Table 6-3).

6.2.19. Cleavage Under Targets and Tagmentation (CUT&Tag)

mESCs were grown in 10cm dish for 48hrs. Media was discarded and cells were cross-linked in the plate with 2 mM DMA for 45 min at RT. Cells were washed with 1xPBS and harvested using trypsin method. Cells were washed twice with 1xPBS and required number of cells for the experiment were transferred to new tube (250k cells per sample. Cells were washed once with wash buffer. 250k cells were resuspended in 300 µl NE buffer and incubated in ice for 10 min. Meanwhile, 15 µl BioMag®Plus ConcanavalinA beads per 250k cells were added to 1 mL of Binding Buffer. Beads were washed twice in Binding Buffer and finally resuspended in 150 µl of Binding Buffer for each sample. The 150 µl beads were added to the 300 µl of NE buffer with cells and incubated at RT for 15min in rotating wheel. The samples were pulse spun and placed on magnet to discard the liquid. Samples were resuspended in 1 mL Blocking Buffer and incubated at RT for 10 minutes. The samples were pulse spun and placed on magnet to discard the liquid.

Samples were resuspended in 100 µl of Wash Buffer with 0.1% Tween 20 and 1 µg of primary antibodies (anti-HA or IgG). The samples were incubated rotating at RT for 2h. 500 µl of wash buffer was added to sample, placed on magnet and all liquid was discarded. Sample was washed twice with 1 mL Wash Buffer. Samples were resuspended in 100 µl of Wash Buffer with 0.1% Tween 20 and 1 µg of secondary antibodies (guinea pig anti-rabbit or rabbit anti-mouse). The samples were incubated rotating at RT for 1h. 500 µl of wash buffer was added to sample, placed on magnet and all liquid was discarded. Sample was washed twice with 1 mL Wash Buffer. Samples were resuspended in 100 µl pA-Tn5 buffer and incubated rotating at RT for 1h. 500 µl of wash buffer was added to sample, placed on magnet and all liquid was discarded. Samples were washed twice with 1 mL Wash Buffer. Samples were resuspended in 100 µl of Tagmentation buffer, equilibrated on ice-water for 2min and incubated for 1h at 37°C. 20 µl of stop buffer and 5µl of Proteinase K (20mg/ul) was added to each sample and incubated overnight at 37°C at 1400rpm.

100 µl of TE buffer was added to the samples and DNA was extracted using ChIP DNA Clean and Concentrator Kit (Zymo Research). 14 µl of the eluted DNA was used for library preparation using NEBNext® Ultra™ II Q5® Master Mix with 1.2 µl of 10 µM adapter primers.

	PCR step	temperature	time
1	Hot-start activation	72°C	5 min
2	Denaturation	98°C	30 s
3	Denaturation	98°C	10 s
4	Annealing and Elongation	68°C	8 s
	Go to step 3 and repeat 14 cycles		
5	Cooling	4°C	pause

Samples were cleaned up using AMPure beads by adding 1.1 volume (33µl AMPure beads) to each sample. Samples were incubated for 5 min and washed twice with 200 µl 80% ethanol, and eluted in 30 µl nuclease-free water.

6.2.20. Single cell RNA-seq (scRNA-seq)

WT and *Gadd45* TKO mESCs were differentiated towards cardiomyocyte lineage using the above mentioned protocol. On day 12 of differentiation, cells were dissociated using Papain. Briefly, 5x Papain stock solution (125U/ml) was diluted to 1x solution using Papain activation buffer. The diluted Papain was incubated at 37°C for 30min to activate the enzymatic activity of Papain. After this incubation the solution was diluted 1:1 using 1x PBS (magnesium and calcium free) to prepare the working Papain solution concentration. The differentiated cells were washed with 1x PBS (magnesium and calcium free) and incubate with 3ml of working Papain solution for 10min at 37°C. The cells were liberated from the plate by gentle titruation using a wide-bore 1ml pipette and inactivate Papain using 10 ml of mESC differentiation medium. Cells were spun down at 500xg for 4 min at 4°C. Wash the cells once with 10 ml of mESC differentiation medium. Cells were then incubated with DnaseI solution at 37°C for 15 min to get rid of any DNA from dead cells (10µl DnaseI enzyme per ml of buffer). Cells were washed once with 10 ml of 2% Pansera ESC-grade FBS diluted in 1x PBS buffer w/o magnesium and calcium (this prevents the cell clumping). Lastly,

the cells were passed through a 200 µm cell strainer (pluriselect, SKU 43-50200-03) to remove any cell clumps.

4 x 10^6 single cells were fixed following Evercode™ Cell Fixation v2 (SKU: ECF2101 and ECF2001). NGS library prep was performed with Evercode™ WT Mini v2 (Version 2.3 – UM0021) starting with 20,000 projected barcoded cells, as per manufacturer's instructions. Libraries were amplified with a total of 22 PCR cycles. Post PCR purification steps were performed to exclude residual primer and adapter dimers. Libraries were profiled in a High Sensitivity DNA on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 4.0 Fluorometer (Invitrogen by Thermo Fisher Scientific). Samples were pooled in equimolar ratio and sequenced on a NextSeq2000 P2 (200 cycles) flow cell using paired-end mode: Read1: 66 cycles, Read2: 86 cycles and 2x8 cycles for the Unique Dual Index Read.

6.2.21. Library preparation

All the steps of library preparation and sequencing were performed by IMB Genomics core facility. RNA-seq library prep was performed with Illumina`s Stranded Total RNA Prep Ligation with Ribo-Zero Plus Kit following Stranded Total RNA Prep Ligation with Ribo-Zero Plus Reference Guide (April 2021, Document # 1000000124514 v02), as per manufacturer's instructions. Libraries were prepared with equal starting amount of total RNA and amplified in 9 PCR cycles. Two post PCR purification steps were performed to exclude residual primer and adapter dimers. Libraries were profiled in a High Sensitivity DNA Chip on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). Samples were pooled in equimolar ratio and sequenced using a NextSeq500 High output or NextSeq2000 P3 flow cells. RNA-seq was performed in single-end mode.

ChIP-seq library preparation was performed using Tecan`s NuGen Ovation Ultralow System V2 (M01379 v5.1), as per manufacturer's instructions. Libraries were prepared with equal starting amount of ChIP-DNA and amplified for 14-16 PCR Cycles. Libraries were profiled in a High Sensitivity DNA Chip on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 4.0 Fluorometer (Invitrogen by Thermo Fisher Scientific). Samples were pooled in equimolar ratio and purified with AMPure XP Beads (0.8X) to get rid of the Primer- and Adapter-Dimer. The pool was sequenced using a NextSeq500 High output or NextSeq2000 P3 flow cells. ChIP-seq was performed in paired-end mode.

6.2.22. Sequencing data analysis

Bioinformatics analysis was performed in collaboration with [redacted] and [redacted] (this lab) using the standard IMB NGS pipelines with default settings

(https://gitlab.rlp.net/imbforge/NGSpipe2go). The raw read quality was analysed using FastQC (v0.11.9). The raw reads alignment to mouse reference genome (mm10) was performed using STAR (v2.7.10) for RNA-seq and bowtie2 (v2.4.5) for ChIP-seq/ CUT&Tag datasets. Uniquely mapped reads were used to perform differential expression analysis in RNA-seq data using DESeq2 (cut-off FDR p-value < 0.01). For ChIP-seq and CUT&Tag datasets, peak calling was performed using MACS2 (v2.1.2) followed by edgeR based Diffbind tool for calling differential peaks.

The Bioconductor package clusterProfiler was used to find enriched GO terms under the ontology "biological processes". The list of all genes expressed in respective RNA-seq samples were used as background list for GO term enrichment analysis. Overlaps between RNA-seq datasets were identified using ENSEMBL IDs if available, and gene symbols if not. Overlaps between RNA-seq and CUT&Tag/ DMRs was performed by intersecting genomic coordinates of differential CUT&Tag/ DMR sites and DEGs (3kb upstream of TSS + gene body). Statistical significance of overlaps was determined using Fisher exact test. Additionally, I performed some basic analysis of RNA-seq, ChIP-seq and ChEA (2022 database) using the open source galaxy platform (usegaalxy.eu) and shinyGO 0.80.

The Illumina Mouse Methylation array was analysed by [redacted] using default settings of Rnbeads package.

The scRNA-seq data was first processed using split-pipe software (https://fredhutch.github.io/easybuild-life-sciences/updates/2022-02-09-Parse-splitpipe/). The pre-processed data was then analysed by [redacted] (IMB's Bioinformatics Core Facility) using a custom pipeline. The read quality was assessed using FastQC, followed by read mapping using STAR. Cells passing the QC criteria of: total counts > 2000, total features > 500, mitochondrial counts < 30%, were used for downstream clustering and visualization analysis. Normalized log2-expression values were computed for each endogenous gene based on the appropriate size factors from cran package. Clustering was performed using Louvain algorithm based cluster method with number of nearest neighbours "k=20" to consider during graph construction (igraph_logcounts_k20_louvain). The marker genes per cluster were further annotated with GO enrichment results.

7. References

1. Chazaud, C. & Yamanaka, Y. Lineage specification in the mouse preimplantation embryo. *Development* **143**, 1063–1074 (2016).

2. Aiken, C. E. M., Swoboda, P. P. L., Skepper, J. N. & Johnson, M. H. The direct measurement of embryogenic volume and nucleo-cytoplasmic ratio during mouse pre-implantation development. *Reproduction* **128**, 527–535 (2004).

3. Johnson, M. H. & Ziomek, C. A. The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71–80 (1981).

4. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Reviews Genetics 2013 14:3* **14**, 204–220 (2013).

5. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **128**, 635–638 (2007).

6. Toranõ, E. G., Garciá, M. G., Fernández-Morera, J. L., Ninõ-Garciá, P. & Fernández, A. F. The Impact of External Factors on the Epigenome: In Utero and over Lifetime. *Biomed Res Int* **2016**, (2016).

7. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics 2016 17:8* **17**, 487–500 (2016).

8. Brookes, E. & Shi, Y. Diverse epigenetic mechanisms of human disease. *Annu Rev Genet* **48**, 237–268 (2014).

9. Talbert, P. B. & Henikoff, S. The Yin and Yang of Histone Marks in Transcription. *Annu Rev Genomics Hum Genet* **22**, 147–170 (2021).

10. Patel, D. J. & Wang, Z. Readout of epigenetic modifications. *Annu Rev Biochem* **82**, 81–118 (2013).

11. Cedar, H., Solage, A., Glaser, G. & Razin, A. Direct detection of methylated cytosine in DNA by use of the restriction enzyme Mspl. *Nucleic Acids Res* **6**, 2125–2132 (1979).

12. 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).

13. 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).

14. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics 2012 13:7* **13**, 484–492 (2012).

15. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature 2013 502:7472* **502**, 472–479 (2013).

16. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science (1979)* **324**, 929–930 (2009).

17. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (1979)* **324**, 930–935 (2009).

18. He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (1979)* **333**, 1303–1307 (2011).

19. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES. *Journal of Biological Chemistry* **286**, 35334–35338 (2011).

20. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature 2012 484:7394* **484**, 339–344 (2012).

21. Song, C. X. *et al.* Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell* **153**, 678–691 (2013).

22. Meng, W. Y., Wang, Z. X., Zhang, Y., Hou, Y. & Xue, J. H. Epigenetic marks or not? The discovery of novel DNA modifications in eukaryotes. *Journal of Biological Chemistry* **300**, 106791 (2024).

23. Hon, G. C. *et al.* 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. *Mol Cell* **56**, 286–297 (2014).

24. Shi, D. Q., Ali, I., Tang, J. & Yang, W. C. New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Front Genet* **8**, (2017).

25. Song, C. X. *et al.* Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell* **153**, 678–691 (2013).

26. Wang, H., Yang, Y., Liu, J. & Qian, L. Direct cell reprogramming: approaches, mechanisms and progress. *Nat Rev Mol Cell Biol* **22**, 410–424 (2021).

27. Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nature Biotechnology 2010 28:10* **28**, 1079–1088 (2010).

28. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science (1979)* **293**, 1089–1093 (2001).

29. Shen, L. *et al.* Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. *Cell Stem Cell* **15**, 459–471 (2014).

30. Mulholland, C. B. *et al.* Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. *Nature Communications 2020 11:1* **11**, 1–24 (2020).

31. Guo, F. *et al.* Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell* **15**, 447–459 (2014).

32. Song, Y. *et al.* Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs. *Mol Cell* **75**, 905-920.e6 (2019).

33. Bogdanović, O. & Lister, R. DNA methylation and the preservation of cell identity. *Curr Opin Genet Dev* **46**, 9–14 (2017).

34. Seisenberger, S. *et al.* The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Mol Cell* **48**, 849–862 (2012).

35. Guo, F. *et al.* The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* **161**, 1437–1452 (2015).

36. Aref-Eshghi, E. *et al.* Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *The American Journal of Human Genetics* **106**, 356–370 (2020).

37. Serra-Juhé, C. *et al.* DNA methylation abnormalities in congenital heart disease. *Epigenetics* **10**, 167–177 (2015).

38. Liu, J. *et al.* Association between placental DNA methylation and fetal congenital heart disease. *Molecular Genetics and Genomics* **298**, 243–251 (2023).

39. Chang, S. *et al.* DNA methylation abnormalities of imprinted genes in congenital heart disease: a pilot study. *BMC Med Genomics* **14**, 1–11 (2021).

40. Andrews, S. *et al.* Mechanisms and function of de novo DNA methylation in placental development reveals an essential role for DNMT3B. *Nature Communications 2023 14:1* **14**, 1–12 (2023).

41. 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).

42. Dahlet, T. *et al.* Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. *Nature Communications 2020 11:1* **11**, 1–14 (2020).

43. Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature 2012 492:7429* **492**, 443–447 (2012).

44. 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).

45. Kang, J. *et al.* Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proc Natl Acad Sci U S A* **112**, E4236–E4245 (2015).

46. Dai, H. Q. *et al.* TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. *Nature 2016 538:7626* **538**, 528–532 (2016).

47. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev* **28**, 2103–2119 (2014).

48. Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature 2013 495:7441* **495**, 370–374 (2013).

49. Neri, F. *et al.* TET1 is controlled by pluripotency-associated factors in ESCs and downmodulated by PRC2 in differentiated cells and tissues. *Nucleic Acids Res* **43**, 6814–6826 (2015).

50. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature 2011 473:7347* **473**, 389–393 (2011).

51. Zhu, F. *et al.* Sin3a–Tet1 interaction activates gene transcription and is required for embryonic stem cell pluripotency. *Nucleic Acids Res* **46**, 6026–6040 (2018).

52. Kienhöfer, S. *et al.* GADD45a physically and functionally interacts with TET1. *Differentiation* **90**, 59–68 (2015).

53. Barreto, G. *et al.* Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature 2006 445:7128* **445**, 671–675 (2007).

54. Tamura, R. E. *et al.* GADD45 proteins: central players in tumorigenesis. *Curr Mol Med* **12**, 634–651 (2012).

55. Suzuki, M. *et al.* Molecular cloning, expression, and mapping of a novel human cDNA, GRP17, highly homologous to human gadd45 and murine MyD118. *Journal of Human Genetics 1999 44:5* **44**, 300–303 (1999).

56. Abdollahi, A., Lord, K. A., Hoffman-Liebermann, B. & Liebermann, D. A. Sequence and expression of a cDNA encoding MyD118: a novel myeloid differentiation primary response gene induced by multiple cytokines. *Oncogene* **6**, 165–167 (1991).

57. Fornace, A. J., Alamo, I. & Hollander, M. C. DNA damage-inducible transcripts in mammalian cells. *Proceedings of the National Academy of Sciences* **85**, 8800–8804 (1988).

58. Kovalsky, O., Lung, F. D. T., Roller, P. P. & Fornace, A. J. Oligomerization of Human Gadd45a Protein. *Journal of Biological Chemistry* **276**, 39330–39339 (2001).

59. Schrag, J. D., Jiralerspong, S., Banville, M., Jaramillo, M. L. & O'Connor-McCourt, M. D. The crystal structure and dimerization interface of GADD45gamma. *Proc Natl Acad Sci U S A* **105**, 6566–6571 (2008).

60. 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).

61. Papa, S. *et al.* Insights into the structural basis of the GADD45β-mediated inactivation of the JNK kinase, MKK7/JNKK2. *Journal of Biological Chemistry* **282**, 19029–19041 (2007).

62. Smith, M. L. *et al.* Interaction of the p53-Regulated Protein Gadd45 with Proliferating Cell Nuclear Antigen. *Science (1979)* **266**, 1376–1380 (1994).

63. Jin, S. *et al.* GADD45-induced cell cycle G2-M arrest associates with altered subcellular distribution of cyclin B1 and is independent of p38 kinase activity. *Oncogene 2002 21:57* **21**, 8696–8704 (2002).

64. Wang, X. W. *et al.* GADD45 induction of a G2/M cell cycle checkpoint. *Proc Natl Acad Sci U S A* **96**, 3706–3711 (1999).

65. Li, Z. *et al.* Gadd45a promotes DNA demethylation through TDG. *Nucleic Acids Res* **43**, 3986–3997 (2015).

66. Schäfer, A. Gadd45 Proteins: Key Players of Repair-Mediated DNA Demethylation. *Adv Exp Med Biol* **793**, 35–50 (2013).

67. Engel, N. *et al.* Conserved DNA methylation in Gadd45a-/- mice. *Epigenetics* **4**, 98–99 (2009).

68. 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).

69. 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).

70. Schäfer, A. *et al.* Impaired DNA demethylation of C/EBP sites causes premature aging. *Genes Dev* **32**, 742–762 (2018).

71. Patel, K., Murray, M. G. & Whelan, K. A. Roles for GADD45 in Development and Cancer. *Adv Exp Med Biol* **1360**, 23–39 (2022).

72. Leach, P. T. *et al.* Gadd45b knockout mice exhibit selective deficits in hippocampus-dependent long-term memory. *Learning & Memory* **19**, 319–324 (2012).

73. 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).

74. Ijiri, K. *et al.* A Novel Role for GADD45β as a Mediator of MMP-13 Gene Expression during Chondrocyte Terminal Differentiation. *Journal of Biological Chemistry* **280**, 38544–38555 (2005).

75. Salvador, J. M. *et al.* Mice Lacking the p53-Effector Gene Gadd45a Develop a Lupus-Like Syndrome. *Immunity* **16**, 499–508 (2002).

76. Chathuranga, W. A. G. *et al.* Gadd45β is critical for regulation of type I interferon signaling by facilitating G3BP-mediated stress granule formation. *Cell Rep* **42**, 113358 (2023).

77. Schüle, K. M. *et al.* GADD45 promotes locus-specific DNA demethylation and 2C cycling in embryonic stem cells. *Genes Dev* **33**, 782–798 (2019).

78. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature 1981 292:5819* **292**, 154–156 (1981).

79. Ghimire, S. *et al.* Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. *Scientific Reports 2018 8:1* **8**, 1–11 (2018).

80. Veenvliet, J. V., Lenne, P. F., Turner, D. A., Nachman, I. & Trivedi, V. Sculpting with stem cells: how models of embryo development take shape. *Development (Cambridge)* **148**, (2021).

81. Sanz, L. A. *et al.* Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol Cell* **63**, 167–178 (2016).

82. Hegazy, Y. A., Fernando, C. M. & Tran, E. J. The balancing act of R-loop biology: The good, the bad, and the ugly. *Journal of Biological Chemistry* **295**, 905–913 (2020).

83. Niehrs, C. & Luke, B. Regulatory R-loops as facilitators of gene expression and genome stability. *Nature Reviews Molecular Cell Biology 2020 21:3* **21**, 167–178 (2020).

84. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Mol Cell* **45**, 814–825 (2012).

85. Sze, S. *et al.* TERRA R-loops connect and protect sister telomeres in mitosis. *Cell Rep* **42**, 113235 (2023).

86. Pacesa, M. *et al.* R-loop formation and conformational activation mechanisms of Cas9. *Nature 2022 609:7925* **609**, 191–196 (2022).

87. Lee, C. Y. *et al.* R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nature Communications 2020 11:1* **11**, 1–15 (2020).

88. R-Loops. **2528**, (2022).

89. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* **18**, 1618–1629 (2004).

90. J, C. *et al.* Systematic Evaluation of Different R-loop Mapping Methods: Achieving Consensus, Resolving Discrepancies and Uncovering Distinct Types of RNA:DNA Hybrids. (2022) doi:10.1101/2022.02.18.480986.

91. Dumelie, J. G. & Jaffrey, S. R. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife* **6**, (2017).

92. Wahba, L., Costantino, L., Tan, F. J., Zimmer, A. & Koshland, D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev* **30**, 1327–1338 (2016).

93. Stork, C. T. *et al.* Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *Elife* **5**, (2016).

94. Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics 2015 16:10* **16**, 583–597 (2015).

95. Brickner, J. R., Garzon, J. L. & Cimprich, K. A. Walking a tightrope: The complex balancing act of R-loops in genome stability. *Mol Cell* **82**, 2267–2297 (2022).

96. Cristini, A. *et al.* RNase H2, mutated in Aicardi-Goutières syndrome, resolves co-transcriptional R-loops to prevent DNA breaks and inflammation. *Nature Communications 2022 13:1* **13**, 1–14 (2022).

97. Groh, M., Lufino, M. M. P., Wade-Martins, R. & Gromak, N. R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLoS Genet* **10**, e1004318 (2014).

98. Liu, Z. *et al.* San1 deficiency leads to cardiomyopathy due to excessive R-loop-associated DNA damage and cardiomyocyte hypoplasia. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1867**, 166237 (2021).

99. Tan-Wong, S. M., Dhir, S. & Proudfoot, N. J. R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol Cell* **76**, 600-616.e6 (2019).

100. Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J. & Fazzio, T. G. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nature Structural & Molecular Biology 2015 22:12* **22**, 999–1007 (2015).

101. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature 2014 516:7531* **516**, 436–439 (2014).

102. Ma, T. S. *et al.* Hypoxia-induced transcriptional stress is mediated by ROS-induced R-loops. *Nucleic Acids Res* **51**, 11584–11599 (2023).

103. Yan, Q., Shields, E. J., Bonasio, R. & Sarma, K. Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach. *Cell Rep* **29**, 1369-1380.e5 (2019).

104. Li, Y. *et al.* R-loops coordinate with SOX2 in regulating reprogramming to pluripotency. *Sci Adv* **6**, 777–787 (2020).

105. Grunseich, C. *et al.* Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Mol Cell* **69**, 426-437.e7 (2018).

106. 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).

107. Arab, K. *et al.* GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nature Genetics 2019 51:2* **51**, 217–223 (2019).

108. Sabino, J. C. *et al.* Epigenetic reprogramming by TET enzymes impacts co-transcriptional R-loops. *Elife* **11**, (2022).

109. Sabino, J. C. *et al.* Epigenetic reprogramming by TET enzymes impacts co-transcriptional R-loops. *Elife* **11**, (2022).

110. Rodríguez-Aguilera, J. R. *et al.* Genome-wide 5-hydroxymethylcytosine (5hmC) emerges at early stage of in vitro differentiation of a putative hepatocyte progenitor. *Scientific Reports 2020 10:1* **10**, 1–16 (2020).

111. Xu, X. *et al.* 5hmC modification regulates R-loop accumulation in response to stress. *Front Psychiatry* **14**, 1198502 (2023).

112. Sarge, K. D., Murphy, S. P. & Morimoto, R. I. Activation of Heat Shock Gene Transcription by Heat Shock Factor 1 Involves Oligomerization, Acquisition of DNA-Binding Activity, and Nuclear Localization and Can Occur in the Absence of Stress. *Mol Cell Biol* **13**, 1392–1407 (1993).

113. Richter, K., Haslbeck, M. & Buchner, J. The Heat Shock Response: Life on the Verge of Death. *Mol Cell* **40**, 253–266 (2010).

114. Richter, K., Haslbeck, M. & Buchner, J. The Heat Shock Response: Life on the Verge of Death. *Mol Cell* **40**, 253–266 (2010).

115. Fujimoto, M. *et al.* HSF1 phosphorylation establishes an active chromatin state via the TRRAP–TIP60 complex and promotes tumorigenesis. *Nature Communications 2022 13:1* **13**, 1–18 (2022).

116. Bierkamp, C. *et al.* Lack of maternal Heat Shock Factor 1 results in multiple cellular and developmental defects, including mitochondrial damage and altered redox homeostasis, and leads to reduced survival of mammalian oocytes and embryos. *Dev Biol* **339**, 338–353 (2010).

117. Xiao, X. Z. *et al.* HSF1 is required for extra-embryonic development, postnatal growth and protection during inflammatory responses in mice. *EMBO J* **18**, 5943–5952 (1999).

118. Du, J. *et al.* Mechanically sensitive HSF1 is a key regulator of left-right symmetry breaking in zebrafish embryos. *iScience* **26**, 107864 (2023).

119. Metchat, A. *et al.* Mammalian Heat Shock Factor 1 Is Essential for Oocyte Meiosis and Directly Regulates Hsp90α Expression. *Journal of Biological Chemistry* **284**, 9521–9528 (2009).

120. Miller, D. J. & Fort, P. E. Heat shock proteins regulatory role in neurodevelopment. *Front Neurosci* **12**, 412741 (2018).

121. Christians, E. S., Zhou, Q., Renard, J. P. & Benjamin, I. J. Heat shock proteins in mammalian development. *Semin Cell Dev Biol* **14**, 283–290 (2003).

122. Jing, R., Duncan, C. B. & Duncan, S. A. A small-molecule screen reveals that HSP90β promotes the conversion of induced pluripotent stem cell-derived endoderm to a hepatic fate and regulates HNF4A turnover. *Development (Cambridge)* **144**, 1764–1774 (2017).

123. Bradley, E., Bieberich, E., Mivechi, N. F., Tangpisuthipongsa, D. & Wang, G. Regulation of Embryonic Stem Cell Pluripotency by Heat Shock Protein 90. *Stem Cells* **30**, 1624–1633 (2012).

124. Nagaraju, G. P. *et al.* Epigenetic effects of inhibition of heat shock protein 90 (HSP90) in human pancreatic and colon cancer. *Cancer Lett* **402**, 110–116 (2017).

125. Cramer, T., Rosenberg, T., Kisliouk, T. & Meiri, N. Early-life epigenetic changes along the corticotropin-releasing hormone (CRH) gene influence resilience or vulnerability to heat stress later in life. *Molecular Psychiatry 2018 24:7* **24**, 1013–1026 (2018).

126. Rosenberg, T. *et al.* Embryonic Heat Conditioning Induces TET-Dependent Cross-Tolerance to Hypothalamic Inflammation Later in Life. *Front Genet* **11**, 536336 (2020).

127. Ueda, T., Kohama, Y., Kuge, A., Kido, E. & Sakurai, H. GADD45 family proteins suppress JNK signaling by targeting MKK7. *Arch Biochem Biophys* **635**, 1–7 (2017).

128. Mak, S. K. & Kültz, D. Gadd45 Proteins Induce G2/M Arrest and Modulate Apoptosis in Kidney Cells Exposed to Hyperosmotic Stress. *Journal of Biological Chemistry* **279**, 39075–39084 (2004).

129. 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).

130. Gao, H. *et al.* B23 Regulates GADD45a Nuclear Translocation and Contributes to GADD45a-induced Cell Cycle G2-M Arrest. *Journal of Biological Chemistry* **280**, 10988–10996 (2005).

131. Wu, C. Heat shock transcription factors: Structure and regulation. *Annu Rev Cell Dev Biol* **11**, 441–469 (1995).

132. Bullard, S. A. *et al.* Gadd45a Protein Promotes Skeletal Muscle Atrophy by Forming a Complex with the Protein Kinase MEKK4. *Journal of Biological Chemistry* **291**, 17496–17509 (2016).

133. Prinsloo, E., Setati, M. M., Longshaw, V. M. & Blatch, G. L. Chaperoning stem cells: a role for heat shock proteins in the modulation of stem cell self-renewal and differentiation? *BioEssays* **31**, 370–377 (2009).

134. Fuse, T., Yamada, K., Asai, K., Kato, T. & Nakanishi, M. Heat Shock-Mediated Cell Cycle Arrest Is Accompanied by Induction of p21 CKI. *Biochem Biophys Res Commun* **225**, 759–763 (1996).

135. Abraham, K. J. *et al.* Nucleolar RNA polymerase II drives ribosome biogenesis. *Nature 2020 585:7824* **585**, 298–302 (2020).

136. Ibrahim, N. O. *et al.* Induction of the Hypoxia-Inducible Factor System by Low Levels of Heat Shock Protein 90 Inhibitors. *Cancer Res* **65**, 11094–11100 (2005).

137. Chen, Y. *et al.* Hsp90 Chaperone Inhibitor 17-AAG Attenuates Aβ-Induced Synaptic Toxicity and Memory Impairment. *Journal of Neuroscience* **34**, 2464–2470 (2014).

138. Zhang, W. &, Wang, M. & Zhang, Y. Tn5 transposase-based epigenomic profiling methods are prone to open chromatin bias. *bioRxiv* 2021.07.09.451758 (2021) doi:10.1101/2021.07.09.451758.

139. Vincent, S. D. & Buckingham, M. E. How to Make a Heart: The Origin and Regulation of Cardiac Progenitor Cells. *Curr Top Dev Biol* **90**, 1–41 (2010).

140. Buckingham, M., Meilhac, S. & Zaffran, S. Building the mammalian heart from two sources of myocardial cells. *Nature Reviews Genetics 2005 6:11* **6**, 826–835 (2005).

141. Tam, P. P. L., Parameswaran, M., Kinder, S. J. & Weinberger, R. P. The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. *Development* **124**, 1631–1642 (1997).

142. Ye, D., Xie, H., Hu, B. & Lin, F. Endoderm convergence controls subduction of the myocardial precursors during heart-tube formation. *Development (Cambridge)* **142**, 2928–2940 (2015).

143. Fukui, H. *et al.* S1P-Yap1 Signaling Regulates Endoderm Formation Required for Cardiac Precursor Cell Migration in Zebrafish. *Dev Cell* **31**, 128–136 (2014).

144. Abu-Issa, R. & Kirby, M. L. Heart field: From mesoderm to heart tube. *Annu Rev Cell Dev Biol* **23**, 45–68 (2007).

145. Brade, T., Pane, L. S., Moretti, A., Chien, K. R. & Laugwitz, K. L. Embryonic Heart Progenitors and Cardiogenesis. *Cold Spring Harb Perspect Med* **3**, a013847 (2013).

146. Zaffran, S. & Frasch, M. Early Signals in Cardiac Development. *Circ Res* **91**, 457–469 (2002).

147. Kelly, R. G. & Buckingham, M. E. The anterior heart-forming field: voyage to the arterial pole of the heart. *Trends in Genetics* **18**, 210–216 (2002).

148. Harvey, R. P. Patterning the vertebrate heart. *Nature Reviews Genetics 2002 3:7* **3**, 544–556 (2002).

149. Bruneau, B. G. The developmental genetics of congenital heart disease. *Nature 2008 451:7181* **451**, 943–948 (2008).

150. Savolainen, S. M., Foley, J. F. & Elmore, S. A. Histology atlas of the developing mouse heart with emphasis on E11.5 to E18.5. *Toxicol Pathol* **37**, 395–414 (2009).

151. D'Amato, G., Luxán, G. & de la Pompa, J. L. Notch signalling in ventricular chamber development and cardiomyopathy. *FEBS J* **283**, 4223–4237 (2016).

152. Crocini, C. & Gotthardt, M. Cardiac sarcomere mechanics in health and disease. *Biophysical Reviews 2021 13:5* **13**, 637–652 (2021).

153. Ehler, E. Actin-associated proteins and cardiomyopathy—the 'unknown' beyond troponin and tropomyosin. *Biophys Rev* **10**, 1121–1128 (2018).

154. Barrick, S. K. & Greenberg, M. J. Cardiac myosin contraction and mechanotransduction in health and disease. *Journal of Biological Chemistry* **297**, 101297 (2021).

155. Sitbon, Y. H., Yadav, S., Kazmierczak, K. & Szczesna-Cordary, D. Insights into myosin regulatory and essential light chains: a focus on their roles in cardiac and skeletal muscle function, development and disease. *J Muscle Res Cell Motil* **41**, 313–327 (2020).

156. Farah, C. S. & Reinach, F. C. The troponin complex and regulation of muscle contraction. *The FASEB Journal* **9**, 755–767 (1995).

157. Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nature Medicine 2000 6:4* **6**, 389–395 (2000).

158. Streef, T. J. & Smits, A. M. Epicardial Contribution to the Developing and Injured Heart: Exploring the Cellular Composition of the Epicardium. *Front Cardiovasc Med* **8**, 750243 (2021).

159. Hall, C., Gehmlich, K., Denning, C. & Pavlovic, D. Complex Relationship Between Cardiac Fibroblasts and Cardiomyocytes in Health and Disease. *J Am Heart Assoc* **10**, 1–15 (2021).

160. Psarras, S., Beis, D., Nikouli, S., Tsikitis, M. & Capetanaki, Y. Three in a Box: Understanding Cardiomyocyte, Fibroblast, and Innate Immune Cell Interactions to Orchestrate Cardiac Repair Processes. *Front Cardiovasc Med* **6**, 437135 (2019).

161. George, R. M., Maldonado-Velez, G. & Firulli, A. B. The heart of the neural crest: Cardiac neural crest cells in development and regeneration. *Development (Cambridge)* **147**, (2021).

162. Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science (1979)* **313**, 1922–1927 (2006).

163. Prall, O. W. J. *et al.* An Nkx2-5/Bmp2/Smad1 Negative Feedback Loop Controls Heart Progenitor Specification and Proliferation. *Cell* **128**, 947–959 (2007).

164. Canac, R. *et al.* Deciphering Transcriptional Networks during Human Cardiac Development. *Cells* **11**, 3915 (2022).

165. Harvey, R. P. NK-2 homeobox genes and heart development. *Dev Biol* **178**, 203–216 (1996).

166. Tanaka, M., Chen, Z., Bartunkova, S., Yamasaki, N. & Izumo, S. The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. *Development* **126**, 1269–1280 (1999).

167. Watt, A. J., Battle, M. A., Li, J. & Duncan, S. A. GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proc Natl Acad Sci U S A* **101**, 12573–12578 (2004).

168. Kuo, C. T. *et al.* GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev* **11**, 1048–1060 (1997).

169. Srivastava, D., Cserjesi, P. & Olson, E. N. A Subclass of bHLH Proteins Required for Cardiac Morphogenesis. *Science (1979)* **270**, 1995–1999 (1995).

170. McFadden, D. G. *et al.* The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development* **132**, 189–201 (2005).

171. Hatcher, C. J. *et al.* TBX5 Transcription Factor Regulates Cell Proliferation during Cardiogenesis. *Dev Biol* **230**, 177–188 (2001).

172. Moskowitz, I. P. G. *et al.* The T-Box transcription factor Tbx5 is required for the patterning and maturation of the murine cardiac conduction system. *Development* **131**, 4107–4116 (2004).

173. Liu, R. *et al.* Tead1 is required for perinatal cardiomyocyte proliferation. *PLoS One* **14**, e0212017 (2019).

174. Chen, Z., Friedrich, G. A. & Soriano, P. Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev* **8**, 2293–2301 (1994).

175. Yuasa, S. & Fukuda, K. Multiple roles for BMP signaling in cardiac development. *Drug Discov Today Ther Strateg* **5**, 209–214 (2008).

176. Naito, A. T. *et al.* Developmental stage-specific biphasic roles of Wnt/β-catenin signaling in cardiomyogenesis and hematopoiesis. *Proc Natl Acad Sci U S A* **103**, 19812–19817 (2006).

177. Guo, Y. *et al.* The Wnt inhibitor Dkk1 is required for maintaining the normal cardiac differentiation program in Xenopus laevis. *Dev Biol* **449**, 1–13 (2019).

178. Hatzistergos, K. E. *et al.* Tumor Suppressors RB1 and CDKN2a Cooperatively Regulate Cell-Cycle Progression and Differentiation During Cardiomyocyte Development and Repair. *Circ Res* **124**, 1184–1197 (2019).

179. Zhao, J. & Mommersteeg, M. T. M. Slit–Robo signalling in heart development. *Cardiovasc Res* **114**, 794–804 (2018).

180. Mommersteeg, M. T. M., Yeh, M. L., Parnavelas, J. G. & Andrews, W. D. Disrupted Slit-Robo signalling results in membranous ventricular septum defects and bicuspid aortic valves. *Cardiovasc Res* **106**, 55–66 (2015).

181. Del Monte, G., Grego-Bessa, J., González-Rajal, A., Bolós, V. & De La Pompa, J. L. Monitoring Notch1 activity in development: Evidence for a feedback regulatory loop. *Developmental Dynamics* **236**, 2594–2614 (2007).

182. Molkentin, J. D., Lin, Q., Duncan, S. A. & Olson, E. N. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* **11**, 1061–1072 (1997).

183. McElhinney, D. B., Geiger, E., Blinder, J., Benson, D. W. & Goldmuntz, E. NKX2.5mutations in patients with congenital heart disease. *J Am Coll Cardiol* **42**, 1650–1655 (2003).

184. Firulli, A. B., McFadden, D. G., Lin, Q., Srivastava, D. & Olson, E. N. Heart and extra-embryonic mesodermal defects in mouse embryos lacking the bHLH transcription factor Hand1. *Nature Genetics 1998 18:3* **18**, 266–270 (1998).

185. Bruneau, B. G. *et al.* Chamber-Specific Cardiac Expression of Tbx5 and Heart Defects in Holt–Oram Syndrome. *Dev Biol* **211**, 100–108 (1999).

186. Zhang, H. & Bradley, A. Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development* **122**, 2977–2986 (1996).

187. Takada, S. *et al.* Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev* **8**, 174–189 (1994).

188. Horitani, K. & Shiojima, I. Wnt signaling in cardiac development and heart diseases. *In Vitro Cell Dev Biol Anim* **60**, 482–488 (2024).

189. High, F. A. *et al.* Murine Jagged1/Notch signaling in the second heart field orchestrates Fgf8 expression and tissue-tissue interactions during outflow tract development. *J Clin Invest* **119**, 1986–1996 (2009).

190. Garg, V. *et al.* Mutations in NOTCH1 cause aortic valve disease. *Nature 2005 437:7056* **437**, 270–274 (2005).

191. Freeman, S. B. *et al.* Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project. *Genetics in Medicine* **10**, 173–180 (2008).

192. Chang, C. P. & Bruneau, B. G. Epigenetics and cardiovascular development. *Annu Rev Physiol* **74**, 41–68 (2012).

193. Akerberg, B. N. & Pu, W. T. Genetic and Epigenetic Control of Heart Development. *Cold Spring Harb Perspect Biol* **12**, a036756 (2020).

194. Davis, K. *et al.* The role of demethylases in cardiac development and disease. *J Mol Cell Cardiol* **158**, 89–100 (2021).

195. Wang, Z., Zhao, Y. T. & Zhao, T. C. Histone deacetylases in modulating cardiac disease and their clinical translational and therapeutic implications. *Exp Biol Med* **246**, 213–225 (2021).

196. Gilsbach, R. *et al.* Distinct epigenetic programs regulate cardiac myocyte development and disease in the human heart in vivo. *Nature Communications 2018 9:1* **9**, 1–14 (2018).

197. Gilsbach, R. *et al.* Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease. *Nature Communications 2014 5:1* **5**, 1–13 (2014).

198. Gorsi, B., Mosbruger, T. L., Smith, M., Hill, J. T. & Yost, H. J. Nkx2.5-dependent alterations of the embryonic heart DNA methylome identify novel cis-regulatory elements in cardiac development. doi:10.1101/186395.

199. Grau, J., Schmidt, F. & Schulz, M. H. Widespread effects of DNA methylation and intra-motif dependencies revealed by novel transcription factor binding models. *Nucleic Acids Res* **51**, e95–e95 (2023).

200. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science (1979)* **356**, (2017).

201. Fang, S. *et al.* Tet inactivation disrupts YY1 binding and long-range chromatin interactions during embryonic heart development. *Nature Communications 2019 10:1* **10**, 1–18 (2019).

202. Oda, M. *et al.* DNA Methylation Restricts Lineage-specific Functions of Transcription Factor Gata4 during Embryonic Stem Cell Differentiation. *PLoS Genet* **9**, e1003574 (2013).

203. Chamberlain, A. A. *et al.* DNA methylation is developmentally regulated for genes essential for cardiogenesis. *J Am Heart Assoc* **3**, (2014).

204. Gorsi, B., Mosbruger, T. L., Smith, M., Hill, J. T. & Yost, H. J. Nkx2.5-dependent alterations of the embryonic heart DNA methylome identify novel cis-regulatory elements in cardiac development. *bioRxiv* 186395 (2017) doi:10.1101/186395.

205. Fang, X. *et al.* Knockdown of DNA methyltransferase 3a alters gene expression and inhibits function of embryonic cardiomyocytes. *The FASEB Journal* **30**, 3238–3255 (2016).

206. Dindot, S. V., Person, R., Strivens, M., Garcia, R. & Beaudet, A. L. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res* **19**, 1374–1383 (2009).

207. Park, K. S. *et al.* Cardiac pathologies in mouse loss of imprinting models are due to misexpression of h19 long noncoding rna. *Elife* **10**, (2021).

208. McKean, D. M. *et al.* Loss of RNA expression and allele-specific expression associated with congenital heart disease. *Nature Communications 2016 7:1* **7**, 1–9 (2016).

209. Dill, T. L. & Naya, F. J. A Hearty Dose of Noncoding RNAs: The Imprinted DLK1-DIO3 Locus in Cardiac Development and Disease. *Journal of Cardiovascular Development and Disease 2018, Vol. 5, Page 37* **5**, 37 (2018).

210. Dill, T. L. & Naya, F. J. A Hearty Dose of Noncoding RNAs: The Imprinted DLK1-DIO3 Locus in Cardiac Development and Disease. *Journal of Cardiovascular Development and Disease 2018, Vol. 5, Page 37* **5**, 37 (2018).

211. He, A. *et al.* Polycomb Repressive Complex 2 Regulates Normal Development of the Mouse Heart. *Circ Res* **110**, 406–415 (2012).

212. Sim, C. B. *et al.* Dynamic changes in the cardiac methylome during postnatal development. *The FASEB Journal* **29**, 1329–1343 (2015).

213. Lan, Y. *et al.* Stage-specific regulation of DNA methylation by TET enzymes during human cardiac differentiation. *Cell Rep* **37**, 110095 (2021).

214. Kim, M. Y. *et al.* Gadd45β is a novel mediator of cardiomyocyte apoptosis induced by ischaemia/hypoxia. *Cardiovasc Res* **87**, 119–126 (2010).

215. Lucas, A. *et al.* Gadd45γ regulates cardiomyocyte death and post-myocardial infarction left ventricular remodelling. *Cardiovasc Res* **108**, 254–267 (2015).

216. Simões, M. A. *et al.* Immunoexpression of GADD45β in the myocardium of newborns experiencing perinatal hypoxia. *Pathol Res Pract* **213**, 222–226 (2017).

217. Shakked, A. *et al.* Redifferentiated cardiomyocytes retain residual dedifferentiation signatures and are protected against ischemic injury. *Nature Cardiovascular Research 2023 2:4* **2**, 383–398 (2023).

218. Wang, J. *et al.* GADD45B inhibits MKK7-induced cardiac hypertrophy and the polymorphisms of GADD45B is associated with inter-ventricular septum hypertrophy. *Biochem Biophys Res Commun* **372**, 623–628 (2008).

219. Vijaya, M. *et al.* Differential gene expression profiles during embryonic heart development in diabetic mice pregnancy. *Gene* **516**, 218–227 (2013).

220. Krolevets, M. *et al.* DNA methylation and cardiovascular disease in humans: a systematic review and database of known CpG methylation sites. *Clinical Epigenetics 2023 15:1* **15**, 1–16 (2023).

221. Cao, J. *et al.* The role of DNA methylation in syndromic and non-syndromic congenital heart disease. *Clinical Epigenetics 2021 13:1* **13**, 1–18 (2021).

222. Pepin, M. E. *et al.* Genome-wide DNA methylation encodes cardiac transcriptional reprogramming in human ischemic heart failure. *Laboratory Investigation* **99**, 371–386 (2019).

223. Stolz, P. *et al.* TET1 regulates gene expression and repression of endogenous retroviruses independent of DNA demethylation. *Nucleic Acids Res* **50**, 8491–8511 (2022).

224. Chrysanthou, S. *et al.* The DNA dioxygenase Tet1 regulates H3K27 modification and embryonic stem cell biology independent of its catalytic activity. *Nucleic Acids Res* **50**, 3169–3189 (2022).

225. Carbonell, C. *et al.* Functional Network Analysis Reveals the Relevance of SKIIP in the Regulation of Alternative Splicing by p38 SAPK. *CellReports* **27**, 847-859.e6 (2019).

226. 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).

227. Rose, B. A., Force, T. & Wang, Y. Mitogen-activated protein kinase signaling in the heart: Angels versus demons in a heart-breaking tale. *Physiol Rev* **90**, 1507–1546 (2010).

228. Wang, Y. Mitogen-Activated Protein Kinases in Heart Development and Diseases. *Circulation* **116**, 1413–1423 (2007).

229. Zhang, T. *et al.* Discovery of Potent and Selective Covalent Inhibitors of JNK. *Chem Biol* **19**, 140–154 (2012).

230. DAVIES, S. P., REDDY, H., CAIVANO, M. & COHEN, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochemical Journal* **351**, 95–105 (2000).

231. Barrett, S. D. *et al.* The discovery of the benzhydroxamate MEK inhibitors CI-1040 and PD 0325901. *Bioorg Med Chem Lett* **18**, 6501–6504 (2008).

232. Zarubin, T. & Han, J. Activation and signaling of the p38 MAP kinase pathway. *Cell Research 2005 15:1* **15**, 11–18 (2005).

233. Johnson, G. L. & Nakamura, K. The c-jun kinase/stress-activated pathway: Regulation, function and role in human disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1773**, 1341–1348 (2007).

234. Li, J. *et al.* MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. *Differentiation* **75**, 299–307 (2007).

235. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics 2018 19:6* **19**, 371–384 (2018).

236. Field, A. E. *et al.* Molecular Cell Review DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. (2018) doi:10.1016/j.molcel.2018.08.008.

237. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).

238. Wang, H., Hao, J. & Hong, C. C. Cardiac induction of embryonic stem cells by a small molecule inhibitor of Wnt/β-catenin signaling. *ACS Chem Biol* **6**, 192–197 (2011).

239. Hao, J. *et al.* Dorsomorphin, a Selective Small Molecule Inhibitor of BMP Signaling, Promotes Cardiomyogenesis in Embryonic Stem Cells. *PLoS One* **3**, e2904 (2008).

240. Karlsson, M. *et al.* A single–cell type transcriptomics map of human tissues. *Sci Adv* **7**, (2021).

241. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 1–14 (2013).

242. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–W97 (2016).

243. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90 (2021).

244. Liu, P. *et al.* Mesenchymal stem cells: Emerging concepts and recent advances in their roles in organismal homeostasis and therapy. *Front Cell Infect Microbiol* **13**, 1131218 (2023).

245. White, I. A., Sanina, C., Balkan, W. & Hare, J. M. Mesenchymal stem cells in cardiology. *Methods in Molecular Biology* **1416**, 55–87 (2016).

246. Guo, Y., Yu, Y., Hu, S., Chen, Y. & Shen, Z. The therapeutic potential of mesenchymal stem cells for cardiovascular diseases. *Cell Death & Disease 2020 11:5* **11**, 1–10 (2020).

247. D'Angelo, W., Chen, B., Gurung, C. & Guo, Y. L. Characterization of embryonic stem cell-differentiated fibroblasts as mesenchymal stem cells with robust expansion capacity and attenuated innate immunity. *Stem Cell Res Ther* **9**, 1–12 (2018).

248. Lin, C. S., Xin, Z. C., Dai, J. & Lue, T. F. Commonly used mesenchymal stem cell markers and tracking labels: Limitations and challenges. *Histol Histopathol* **28**, 1109–1116 (2013).

249. Ramos, T. L. *et al.* MSC surface markers (CD44, CD73, and CD90) can identify human MSC-derived extracellular vesicles by conventional flow cytometry. *Cell Communication and Signaling* **14**, 1–14 (2016).

250. Kluth, S. M. *et al.* DLK-1 as a Marker to Distinguish Unrestricted Somatic Stem Cells and Mesenchymal Stromal Cells in Cord Blood. *https://home.liebertpub.com/scd* **19**, 1471–1483 (2010).

251. Grancharova, T. *et al.* A comprehensive analysis of gene expression changes in a high replicate and open-source dataset of differentiating hiPSC-derived cardiomyocytes. *Scientific Reports 2021 11:1* **11**, 1–21 (2021).

252. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science (1979)* **360**, 176–182 (2018).

253. Seth, S., Mallik, S., Bhadra, T. & Zhao, Z. Dimensionality Reduction and Louvain Agglomerative Hierarchical Clustering for Cluster-Specified Frequent Biomarker Discovery in Single-Cell Sequencing Data. *Front Genet* **13**, 828479 (2022).

254. Kearse, M. G. & Wilusz, J. E. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev* **31**, 1717–1731 (2017).

255. 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).

256. Molecular Cloning This is a free sample of content from Molecular Cloning: A Laboratory Manual, 4th edition. Click here for more information or to buy the book. (2012).

8. List of abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylytosine
5mC	5-methylcytosine
AM-AR	Active modification - active removal
AM-PD	Active modification - passive dilution
BER	Base excision repair
BMP	Bone morphogenetic protein
Bp	Base pair
BSA	Bovine serum albumin
cDNA	Copy DNA
CDS	Coding sequence
CHD	Congenital heart defects
ChIP	Chromatin immunoprecipitation
CpG	Cytosine-Guanine dinucleotide
CRISPR	Clustered regularly interspaced short palindromic repeats
Ctrl	Control
CUT&Tag	Cleavage Under Targets and Tagmentation
CVD	Cardiovascular disorders
DEG	Differentially expressed gene

DKO	Double knockout
DMEM	Dulbecco´s modified Eagle´s medium
DMR	Differentially methylated region
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Nucleoside triphosphate
DTT	Dithiothreitol
E	Embryonic day
EB	Embryoid body
EDTA	Ethylenediamine tetraacetic acid
Eomes	Eomesodermin
EPI	Epiblast
ERK	Extracellular-signal regulated kinases
EV	Empty vector
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FC	Fold change
FDR	False discovery rate
Gadd45a	*Gadd45a* Growth arrest and DNA-damage inducible protein 45 alpha
Gadd45b	*Gadd45a* Growth arrest and DNA-damage inducible protein 45 beta

Gadd45g	*Gadd45a* Growth arrest and DNA-damage inducible protein 45 gamma
GATA4	GATA binding protein 4
GFP	Green fluorescent protein
GO	Gene ontology
gRNA	Guide ribonucleic acid
HAND1	Heart and neural crest derivatives expressed 1
HAND2	Heart and neural crest derivatives expressed 2
Hr	Hour
HSF1	Heat shock factor 1
HSP	Heat shock protein
HSR	Heat shock response
ICM	Inner cell mass
ING1	Inhibitor of growth protein 1
JNK	c-Jun N-terminal kinases
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LIF	Leukemia inhibitory factor
MEF	Mouse embryonic fibroblast
MEKK4	Mitogen-activated protein kinase kinase kinase 4
mESCs	Mouse embryonic stem cells
Min	Minutes
mRNA	Messenger RNA

NEAA	Non-essential amino acid solution
NER	Nucleotide excision repair
NKX2-5	NK2 homeobox 5
NP-40	Nonidet P40
Oct4	Octamer-binding protein 4
OE	Overexpression
PAGE	Polyacrylamide gel electrophoresis
Pax3	Paired box 3
Pax6	Paired box 6
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Primitive endoderm
PGCs	Primordial germ cells
qPCR	Quantitative polymerase chain reaction
RA	Retinoic acid
RIPA	Radio immunoprecipitation assay
RNA	Ribonucleic acid
rpm	Rounds per minute
sc	Single cell
SDS	Sodium dodecyl sulfate
seq	Sequencing

SKO	Single knockout
Sox2	SRY-box 2
TAF12	TATA-box binding protein associated factor 12
TARID	TCF21 antisense RNA inducing promoter demethylation
TBP	TATA binding protein
TBS	Tris buffered saline
TBX5	T-box transcription factor 5
TDG	Thymine DNA glycosylase
TE	Trophectoderm
TET	Ten-eleven translocation
TF	Transcription factor
TKO	Triple knockout
TSS	Transcription start site
UPL	Universal probe library
Wnt	Wingless-type MMTV integration site family member
WT	Wild-type
YY1	Ying-Yang 1

9. Acknowledgements

10. Lebenslauf