Role of the STAR proteins in early mammalian development and pluripotency

Alessandro Dasti

TESI DOCTORAL UPF / YEAR 2019

Thesis supervisor

Dr. Gian Gaetano Tartaglia

Dr. Elias George Bechara El Halal

DEPARTMENT OF BIOINFORMATICS AND GENOMICS AT CENTER FOR GENOMIC REGULATION

Acknowledgments

There are so many people to acknowledge that I've met during this amazing experience that a whole thesis might not be enough. Jokes aside, I obviously have to be very grateful to Gian, one of the best bosses that a person could ask for. Thank you for having given me the great opportunity of doing the PhD in your lab. You haven't been merely a boss but much more. Thank you for having always treated all of us as PEERS, something that surely I'll miss once I'll be out of this lab, and thank you for all the things you've done for the whole group. It won't be easy to find another boss that takes so personally his employees' lives. Another gigantic thanks is obviously for Elias, since without you I would have never got the chance to known and come in the Tartaglia Lab. You have been (and you are and you will always be I hope) one of the most important persons of my Barcelonian experience. I don't have to tell you that but you already know that you are: my boss, my colleague, my friend, my confident, an elder brother, a guide, a life inspiration and a shoulder to lay on! You're one of the most generous persons I've ever met in my life: it's your strength and your weakness (just try to be more selective on who deserves your generosity and who does not!). Another person that has to be thanked for all the bullying she suffered coming from me is EMSIDE Fabrizio Maria Carla CRG (this is the simple way I saved you in my mobile agenda and you know that). You already known that our lives in the lab would have been much more

meaningless and boring without the other one. You agree right? I'm looking forward to see what will be your defense's video ☺
I also have to thank all the rest of the lab members, Alex for all the different and stimulating conversation we have had during these years (many of the things we talked about are either forbidden or indecent), Fernando for our *odi et amo* relationship, Nieves for her contributions, her support to the others as well as for your organizational skills that were always very helpful in this entropic lab. MaDGa first of all for many of the graphs present in this thesis as well as for our conversations about the psychology of the reletionships ahahaha. Natalia, for the cheat chats we have lately when in the morning, we are the only ones that are in the lab. The less new members as well as the newest, namely that character named Michele Mountains ahahahha. I wish you could have joined us before. Anyway, I'm happy I've met you and for sure I'll miss all the yogurtinos we had we Giulia! I want to thanks even the past members of the lab, in particular Mimma. I've to thank the people of my previous lab and my previous boss Thomas Graf. Some special thanks go surely to my favorite Chinese in town: Tian. You've enriched my CRG life both on the scientific and on the social aspect. A thanks goes as well to Greg and Fra, for all the crazy excursions, the ski trips, the jumps from the rocks and the beach volley matches! I want to thank as well Sergio, ex colleague and present friend that I hope to have back in Barcelona as soon as possible. I want to thank as well my favorite Sardinian and my favorite English: Andrea and Chris! I hope we´ll have more chances to hang out as we used to do more in the past! I want to thank as well las chicas del lab Gabaldon:

Ester, Ewa and the future mom Susana. Many thanks for being so patient and zen with us! I know, we were noisy neighbors. Of the CRG unrelated people I have to give a special thanks to Ignasi: moltes gracies per ser una persona tan polifacètica. Amb tu es pot parlar de tot: des de les txorrades mes tontes als temes més seriosos. Moltes gràcies per tots els teus consells i per haver-me ensenyat tan be el català! Gli ultimi ringrazimaenti vanno sicuramente alla famiglia, in particolar modo mia sorella che è stata sempre molto supportive e che sta diventando un punto di riferimento sempre più importante. Un altro grazie speciale va ad Oscar. Grazie per prenderti cura di mamma e grazie per tutto quello che fai per lei. Grazie per tutte le vongole veraci inviatemi direttamente dall'Italia. Ma il grazie sicuramente più importante e che non potrà mai ripagare tutti gli sforzi fatti va a MAMMA. Sei la persona piu' importante che ho. Sei una donna piccola ma dal cuore immenso ed hai una forza inestimabile. Grazie per tutti i sacrifici (e son stati tantissimi) che hai fatto per me e senza i quali io ora non sarei qui!

Abstract

The molecular basis that sustain embryonic stem cell (ESCs) pluripotency and their differentiation in all the types of somatic cells is subject of intensive research. Recently, the potential involvement of RNA Binding Proteins (RBPs) in this process started to gain high interest for many researchers. RBPs can shuttle between the cytoplasm and the nucleus to associate with nascent RNAs and regulate their fate at different post-transcriptional levels (splicing & maturation, nuclear export and transport, translation, stability, storage and degradation). The signal transduction and activation of RNA (STAR) is a family of RBPs that consists of 5 members that are evolutionally conserved and share a common structural domain. With this PhD project, we aim at understanding the role of two members of the STAR protein family, namely Sam68 and Quaking, in both mESCs pluripotency and differentiation in particular toward the cardiomyocyte lineage.

Preface

The stem cell biology field is gaining more and more interest for the relevance of these cells in development as well as for their potential applications in therapy. In order to explore all the possible therapeutic avenues that these cells can open, a deep knowledge of their biological properties is needed.

RNA binding proteins are proteins that, by binding and regulating all the different mechanism and steps of the RNA metabolism, can fine tune the gene expression, the stability of the RNAs, their translation as well as their splicing among other things. Because of this, it is not surprising that the study of the RBPs in the regulation of the stem cell biology, namely of their pluripotency and differentiation towards all the terminally differentiated cells that compose the mature being, has become of great interest for this field of biology.

The signal transduction and activation of RNA (STAR) protein family, is a family of RBPs composed by 5 members that are evolutionary conserved and share a common structural domain. These RBPs are involved in several developmental processes across different species indicating that they might be important actors for the early stages of the mammalian development.

Because of all the aforementioned reasons, we wanted to define and understand the role of two members of the STAR family, namely Sam68 and quaking, in both mouse embryonic stem cell pluripotency maintenance and in their differentiation. To achieve that, we made use of the more classical cell biology techniques along with the application of cutting-edge technologies for the study of the RNA.

In the first chapter of the results section, I make a description of the cellular characterization of the mESCs KO for either Sam68 or Quaking obtained through the CRISPR/Cas9 technology. In this section are described the results obtained by making use of more classical cellular and molecular biology techniques.

In the second section of the results, there is the description of the effects of the depletion of these two proteins in terms of gene expression, alternative splicing and translation regulation during the differentiation of the mESCs especially towards the cardiomyocyte lineage, obtained by using cutting-edge techniques such as RNA-sequencing.

In the third and last chapter of the results, I provide a functional significance, through a phenotypic characterization, for all the "molecular defects" unraveled by the previously described techniques that mESCs, depleted of either Sam68 or Quaking, show during the differentiation towards cardiomyocytes.

Resum

Els mecanismes moleculars que sostènen la pluripotència i la diferenciació de les cèl·lules mares embrionàries són un camp de intensa recerca. Recentement, el potencial rol de les proteïnes que lliguen els ARNs (RBPs) està prenent molt interés per a molts investigadors. Les RBPs es poden traslladar entre el nucli i citoplasma per associar-se amb els ARNs neixents regualnt els seus destins a diferents nivells post-transcriptionals (splicing i maduració, exportació nuclear i transport, tradució, regulació de la estabilitat i la degradació). La família de proteïnes STAR (signal and trasduction and activation of RNA) está formada per cinc proteïnes que estan conservades a nivell evolutiu i totes comparteixen el mateix domini estructural. En aquest projecte doctoral volem entendre el rol de dos membres d'aquesta familia de proteïnes, anomenades Sam68 i Quaking, en la pluripotència de les cèl·lules mares i en la seva diferenciació en particular cap a cardiomiocits.

Table of Contents

1 INTRODUCTION

1.1 First steps of development

As a result of fertilization, the zygote, which is considered as the earliest stage of the embryo, is formed and will give rise to both fetal and extraembryonic lineages. The zygote goes through eight mitotic cleavages establishing the blastomeres, morulae and, by day ~E3.5 in mouse (E5 in human), to the 64-cells preimplantation blastocyst[1]. The early blastocyst is formed by an external layer of cells called trophectoderm and an inner cell mass (ICM). The latter is composed by a bipotent progenitor. Indeed, at ~E4.5, the ICM segregates into the hypoblast, responsible of the extraembryonic tissues, and the epiblast that will generate all the terminally differentiated cells of the foetus[1–3]. Upon implantation, the blastocyst dramatically changes its shape to a more cylindrical structure called "egg cylinder" where the epiblast consists of 120 cells organized in a pseudostratified epithelium. At around day E6.5 the cells of the epiblast undergo lineage specification and gastrulation that in turn produces the three germ layers:

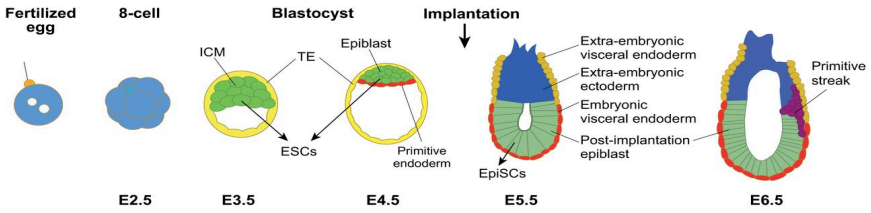


Figure 1 Early stages of mouse embryonic development and ESCs (image taken from "The pluripotent state in mouse and human", Davidson et al.[4])

the definitive endoderm (inner layer), the mesoderm (middle layer) and the definitive ectoderm (outer layer)[5]. Cells derived from the endoderm eventually form many of the internal linings of the body, including the lining of most of the gastrointestinal tract, the lungs, the liver, the pancreas and other glands that open into the gastrointestinal tract, and certain other organs, such as the upper urogenital tract and female vagina. The ectoderm, on the other hand, eventually forms certain "outer linings" of the body, including the epidermis and hair. It is also the precursor to mammary glands and the central and peripheral nervous systems. Cells derived from the mesoderm give rise to all other tissues of the body, including the dermis of the skin, the heart, the muscle system, the urogenital system, the bones, and the bone marrow (and therefore the blood). The mesoderm is the germ layer that distinguishes evolutionarily higher life-forms (i.e., those with bilateral symmetry) from lower life-forms (i.e., those with radial body symmetry). The mesoderm allows more highly evolved organisms to have an internal body cavity that houses and protects the organs, bathing them in fluids and supporting them with connective tissue.

1.2 Embryonic Stem Cells (ESCs)

The mouse pre-implantation epiblast cells represent the pluripotent embryonic stem cells (ESCs)[6–9]. These cells, can be derived from mouse embryos and cultured *in vitro* and have revealed to be a very important tool in order to study the early stages of mammalian

development. They have also been used for therapy approaches and disease modelling [10,11]. Historically, two forms of pluripotency have been maintained thus far *in vitro*, termed **naive** and **primed** state[12–14]. The most primitive state is mouse naive ESCs (nESCs), which resemble the *in vitro* "frozen in time" version of ~32 pluripotent cells embedded within ICM of the pre-implantation embryo (E3.5). These pluripotent cells are *in vivo* shielded by trophectoderm cells, together composing the blastocyst[2]. Under proper culture conditions with addition of selected inhibitors, ICM derived cells are *in vitro* "locked" in the reprogramming state[14,15]. One day later in development (E4.5) upon fertilization when the blastocyst matures, the mouse embryo reaches the uterine wall and implants into uterus[16,17] . At this stage, this early state of naïve pluripotency evolves into the primed pluripotent state of the egg cylinder epiblast resembling rosette formation of pluripotent cells centred in monolayer surrounding the cavity, the precursor of the egg cylinder. At this stage the peri-implantation rosette becomes responsive to the differentiation inducing signals of the gastrula[17]. The ESCs possess two main capacities: the pluripotency and the self-renewal[18,19]. The pluripotency is the ability to give rise to all the terminally differentiated cells, whereas the self-renewal, is the ability to proliferate indefinitely remaining in an undifferentiated and uncommitted state. Molecularly, the pluripotent state is sustained by the expression of the pluripotency associated transcription factors (TFs) such as Oct4, Nanog and Sox2[20–23]. Interestingly, these transcription factors not only sustain the pluripotent state, but are, as well, lineage specifiers that act by promoting the differentiation

towards a specific lineage while impeding the differentiation towards the other ones[24,25]. Therefore, the pluripotent and undifferentiated state of the ESCs is sustained by the co-expression of the different competing lineage specifying pluripotency TFs. As a result, when there is not an individual factor prevailing, the undifferentiated state is maintained, albeit very precariously. Conversely, when this fine equilibrium is disrupted by the overexpression of a lineage specifying TFs caused by the an external molecular signal (e.g. TGF-B), the cells take a commitment to differentiate into a specific lineage with the activation of a specific gene expression program that this process involves[26]. Because of all the aforementioned reasons, it is not surprising to consider the mESCs as a very powerful laboratory model to study the early events of the mammalian development *in vitro*.

1.3 Embryonic stem cells as an in vitro model of early mammalian development

A robust and efficient in vitro method to generate most terminally differentiated cell types derived from the three germ layers is through embryoid bodies (EBs)[27,28]. In this system, the ESCs are allowed to spontaneously differentiate as tissue-like spheroid in a suspension culture. The EB method has shown to be a quite faithful recapitulation of the early stages of mammalian development with the formation of a complex three-dimensional structure made of cells and extracellular matrix that drives the differentiation of the cells towards the three germ layers and their derivatives[29]. This method

consists in allowing the aggregation of the ESCs while withdrawing the cytokines used to maintain the undifferentiated state in culture (e.g. Leukemia inhibitory factor LIF). By doing so, the cells are allowed to spontaneously undertake a differentiation path giving rise to a bigger three-dimensional structure. To allow EB formation and avoid the aggregation of distinct EBs, these are cultured either in hanging drops or in suspension or in methylcellulose cultures[30,31].

1.4 Networks regulating ESC pluripotency

The mechanisms underlying self-renewal and pluripotency of the stem cells have been the subject of intensive studies in recent years[32], focusing on the transcriptional regulation of the pluripotency network governed by trinity of nuclear regulators, Oct4, Sox2, and Nanog[33]. Nanog, Oct4, or Sox2 deficient ES cells lose pluripotency and differentiate into extraembryonic lineage[21,34,35], suggesting that these factors are major regulators of the self-renewing state. Depletion of some factors that can compensate core pluripotency factors results in failed colony forming capability[36,37]. Current understanding of molecular, transcriptional and functional properties of ESCs makes them the best available approach to model the developmental "ground state" of pluripotency, very close to their *in vivo* counterparts[38].

1.5 Exit from pluripotency and early differentiation

To initiate differentiation, stem cells have to be forced out of self-renewal. Both transcriptional networks and epigenetic control of the pluripotent state have been the focus of several studies, and significant headway has also been made in understanding germ layer differentiation. Every pluripotent stem cells can be induced into endoderm, ectoderm and mesoderm, recapitulating the early events in embryogenesis. The process of differentiation is divided into two processes: exit from pluripotency and germ layer differentiation/lineage specification[39]. However, there is still a lack of knowledge about the process of exit from pluripotency, and the mechanisms by which the initial non-pluripotent stage is stabilized before epigenetic modifications generate the differentiation tracks. Betschinger and colleagues identified dual specificity phosphatases that regulates the exit from pluripotency via ERK signalling leading to nuclear clearance of TFE3 transcription factor[40].

1.6 Post-transcriptional control of pluripotency and differentiation

Recent advances in *omics* techniques shed new lights on the existence of another complex layer of gene expression regulation on the RNA level that can modulate pluripotency and lineage commitment. Indeed, in the earliest developmental stage, the zygotic genome is transcriptionally silent and development is guided exclusively by mRNAs and proteins from the oocyte. System wide studies revealed new roles of RNA binding proteins in several biological processes[41,42]. Another study using genomic and proteomic-approaches by Lu and colleagues suggested that post-transcriptional regulation may be responsible for a large proportion of protein level changes during ESC fate transition[43]. Many pluripotency factors, such as Oct4, Nanog, Sall4, Tcf3, Foxp1, Mbd2, and Yy2 have multiple isoforms that vary in expression, intracellular localization, stability, or function due to differences in their coding exons or untranslated regions resulting from highly regulated alternative splicing[44–48]. Hence the importance of RNA Binding Proteins in this process.

1.7 RNA Binding Proteins in general biological processes

Unlike the prokaryotes where transcription and translation are coupled events, in the eukaryotes these two processes are spatially and temporally separated allowing a series of processing and maturation of the nascent pre-mRNA to be carried out in the nucleus before it is exported to the cytoplasm. These mechanisms include the pre-mRNA splicing, nuclear export, transport, stability, polyadenylation and mRNA editing among others. Throughout these pre-mRNA processing mechanisms, the cell is able to fine tune the protein expression and diversify the cell proteome at several layers[49]. This dynamic process of mRNA maturation is carried out by RNA-binding proteins (RBPs) that form an interaction network involving all these different steps. To date, it is known that the human genome encodes more than 1500 RBPs[50]. It has been widely known that RBPs bind RNA sequences and/or structures via the classical RNA-binding domain. The most characterized RNA-binding domains (RBD) are: the RNA-recognition motif (RRM), the zinc finger domain, the K Homology (KH) domain and the Pumilio/FBF domain[51]. The RRM domain is the most represented. It is a 90 amino acids domain containing 2 conserved sequences called RNP1 and RNP2 disposed in a stranded β-sheet packed against 2 α-helices[52]. The zinc finger domain is composed by a β-hairpin and an α-helix pinned together by a zinc ion[53] whereas the KH domain is made of 70 amino acids that compose a β-sheet, made of three antiparallel β-strands packed against 3 α-helices[54]. However, recent advances in determining the

structures of the ribosome and spliceosome showed complex RNA-Protein interactions happening through "unconventional" RBDs. RBPs show a high degree of modularity, being formed by several RBDs, as well as other domains frequently needed in order to establish interactions with other proteins[51]. Usually, RBD binds a little stretch of 8-10 nucleotides and this binding is reinforced by the cooperation of other binding domains targeting the same RNA sequence[51,55]. The love story between RNAs and RBPs begins at the birth of the transcript where RBPs escort the RNAs covering or exposing different regions to help the mRNA progressing through the different stages of its life. RBPs orchestrate and regulate several coordinated post-transcriptional pathways leading to a normal development and cellular homeostasis. Recent studies about distinct ribonucleoparticles (RNPs) and long non-coding RNAs (lncRNAs) suggest that also RNAs regulate RBP functions[56]. Given the critical importance of RBPs in the dynamic gene expression system, it is not surprising that a deficiency in the normal function of RBPs will disrupt RNP organization and cause a number of clinical disorders[57,58].

1.8 RBPs in ESCs and normal embryonic development

The choice between pluripotency maintenance and differentiation is tightly regulated. To accomplish this, several layers of regulations are needed and RBPs play an important role in undertaking this choice.

For instance, MBNL regulates the alternative splicing of the transcription factor FOXP1, more specifically of the exon variant that codes for the forkhead domain. The inclusion of this exon variant leads to a FOXP1 isoform that is able to bind and transcribe pluripotency associated genes while repressing the differentiation-associated ones demonstrating that the alternative splicing of a single exon can have a major impact on the biology of the ESCs[46,44]. Furthermore, AS regulators are differentially expressed in pluripotent stem cells and somatic cells[59].They control the proper splicing of cell-state specific transcripts, and can rewire AS networks during cell fate transitions. Specifically, FOX2, SON, SFRS2, MYC, GCN5, ZCCHC24, and RBM47 facilitate pluripotency-specific AS of their target genes[59–62]. In contrast, MBNL1, MBNL2, RBM24, and SFRS11 promote differentiation-specific AS patterns for a large number of splicing events[63,64]. In addition to the alternative splicing, another important aspect of the RNA metabolism that has been shown to occur during early mouse development, mouse ESC differentiation and somatic cell reprogramming is the mRNA alternative polyadenylation. This process leads to the production of mRNAs with different untranslated regions, which can impact mRNA stability, translation, or intracellular localization[65]. In fact, it has been shown that ESCs favor proximal polyadenylation sites to the distal ones resulting in shorter 3'UTRs of the transcripts in ESCs compared to differentiated cells. A shorter 3'UTR means as well less regulatory sites for miRNAs and other RBPs [66,67]. One of the RBPs identified as important to generate an alternative polyadenylation (APA) profile pluripotency-specific is Fip1. Depletion of Fip1 leads

to differentiation of ESCs[68]. The control of the RNA stability is another mechanism implicated in the fine-tuning of the choice between pluripotency maintenance and differentiation. For instance, Larp7 binds Lin28 RNA, an important self-renewal factor. This interaction yields in a more stable form of Lin28 that ultimately inhibits the priming of the ESCs and therefore their differentiation[69]. Another class of RBPs that plays an important role in ESCs is represented by the "RNA modifiers" which are enzymes that introduce or read RNA modifications. In fact, the methylation of the adenosine into N^6-methyladenosine (m^6A) by the methyltransferase Mettl14 and Mettl3 is one of the major RNA modifications. These two enzymes work as a heterodimer to methylate the adenosines in position 6 leading to a stabilization of the modified transcripts[70]. This modification has been shown to fine-tune the levels of the pluripotency-related transcripts. Mettl3$^{-/-}$ cells have more self-renewal capacity when compared to their WT counterparts, and are unable to differentiate into cardiomyocyte or Tuj+ (*bona fide* neurons) cells and give rise to poorly differentiated teratomas when injected in nude mice. The mechanism underlying this phenotype is the increased stability of some pluripotency-related transcripts, such as Nanog, Jarif and Lin28 among others[71]. Last, but not the least important mechanism of RNA regulation operated by RBPs in ESCs is the control of the nuclear export. Two important nuclear-export factors Thoc2 and Thoc5 are involved in the nuclear export of pluripotency associated transcripts such as Nanog, Sox2, Klf4 and Essrb and therefore are involved as well in the control of their translation. The downregulation of Thoc5 seems to be critical in order

to shut down the pluripotency-associated gene expression program allowing the differentiation of the ESCs[72]. To conclude, it is becoming more evident that the RBPs play an important role in fine-tuning the ESCs biology by regulating all the different post-transcriptional events influencing the choice between pluripotency maintenance and differentiation. For this reason, we decided to explore the role of the STAR family members in this process.

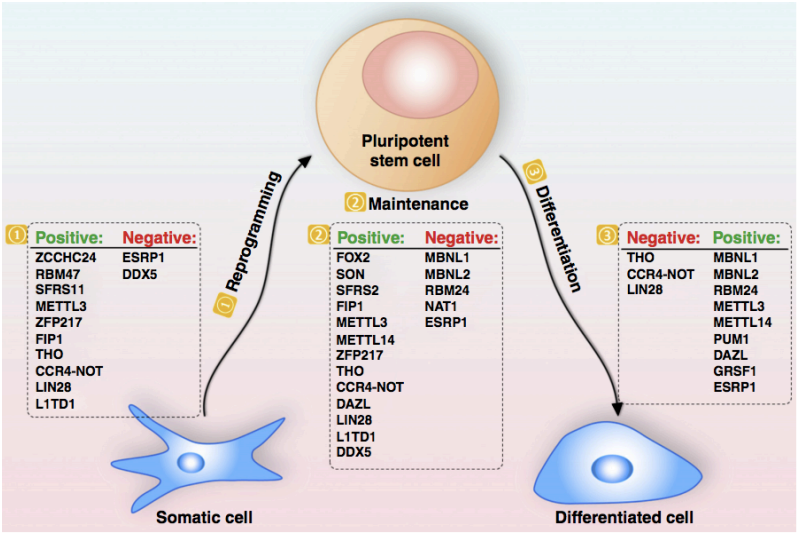


Figure 2 RBPs function in the fine tuning of the choice between pluripotency and differentiation in ESCs (adapted from Chen et al, 2017[73])

1.9 The Signal Transduction and Activation of RNA (STAR) protein family

The Signal Transduction and Activation of RNA (STAR) proteins constitute a family of RBPs composed by 5 different members that are evolutionary conserved in all the metazoa[74]. These are: Src associated in mitosis (SAM68 also known as KHDRBS1), quaking (QKI), Slm1 (know as well as KHDRBS2), Slm2/T-Star (called KHDRBS3) and Splicing Factor 1 (SF1). These proteins share, except SF1, a common structural domain called STAR domain[74] that consists of a central KH domain flanked by two domains, QUA1 at the C-terminal and QUA2 at the N-Terminal. While the QUA2 domain is involved in binding to RNA[75], QUA1 is necessary for the dimerization of the STAR proteins[76,77]. Conversely, SF1 does not contain the QUA1 domain. Notably, some of the STAR proteins can have different splicing isoforms (e.g. QKI has 4 different splicing isoforms[78,79]), and can be regulated by different post-translational modifications. The STAR proteins are important in all the processes of the RNA metabolism. They are involved in pre-mRNA splicing regulation, RNA localization, RNA transport, stability as well as regulation of the translation efficiency[80]. Furthermore, the members of this family take part in important developmental processes such as myelinogenesis and spermatogenesis and their alteration or deregulation can be associated to the onset of pathologies[81–87].

1.10 SAM68 at glance and its role in development

To date, SRC associated in mitosis (SAM68), a 443 amino acid long protein of 68kDa encoded by a gene with 9 exons, is the most studied and characterized protein of the family[88,89]. Through SELEX experiments it has been identified the UAAA motif as the one recognized by this RBP[90]. It is mainly a nuclear protein and it contains not only the STAR domain but a proline-rich region, an SH3, a WW-binding site, RGG boxes and, as well, a prominent string of tyrosines in the C-teminal part that flanks a bipartite nuclear localization signal[91,9293,94] . All these features are important for either post-translational modification or interactions with other proteins. As a consequence, the activity of SAM68 can be tightly modulated and linked to cell signaling pathways[94–96]. Furthermore, it has been shown that Sam68 regulates the gene expression through interactions with transcription factors or modulators[49,97]. Additionally, Sam68 has been shown to be involved in various processes of the RNA metabolism. Indeed, Sam68 regulates the alternative splicing of CD44, Bcl-x as well as the splicing of transcripts important during neurogenesis[91,96,98]. Moreover, it is involved in the regulation of the translation through the interaction with the translation factor eIF4F. This function is indispensable during male germ cell differentiation where Sam68 plays a key role by enhancing the translation of transcripts needed for the full differentiation of these cells[86,99]. Sam68 can also act in translation-independent mechanisms during male germ cells development. Indeed, Sam68 regulates the recruitment of the U1 small nuclear ribonucleoprotein (U1 snRNP)

on the alternative last exons (ALEs) present on male germ-cells specific transcripts. This recruitment impedes the recognition of the alternative polyadenylation site by the cleavage-polyadenylation complex resulting in the formation of the fully functional forms of these transcripts[100]. Because of all the aforementioned examples, it is not surprising that the main phenotypic feature of the mouse Sam68$^{-/-}$ is the male infertility. The KO mouse model was generated by the disruption of exon 4 and exon 5 that encode the KH domain[86,101]. Sam68 null mice show perinatal lethality but the offspring that survive does not exhibit major defects apart from the male infertility. Furthermore, these mice show motor-coordination defects and weight considerably less compared to their wild type littermates[87]. Interestingly, in the females null mice the onset of the mammary gland tumor is impeded[102]. As far as the neuronal development is concerned, it has been demonstrated that Sam68 plays a major role in the fine tuning of the switch between self-renewal and differentiation during the in vivo expansion of the neurogenic areas of the neocortex where the neuronal progenitor cells (NPCs) reside. More specifically, in NPCs where the level of SAM68 are high, Sam68 binds to a cryptic polyadenylation site (PAS) in the intron 7 of the aldehyde dehydrogenase 1A3 (ALDH1A3) preventing its recognition and premature transcript termination. This leads to the formation of a fully functional enzyme that is needed for the sustainment of the glycolytic metabolism that is essential for the self-renewal capacity of the cells. Conversely, upon differentiation, the levels of Sam68 decrease, the cryptic PAS is recognized leading to the formation of a truncated and non-functional form of the enzyme. This will result in

a metabolic switch, loss of the self-renewal capacity and differentiation of the NPCs[103]. Sam68 has also been shown to be involved in the adipogenic differentiation by stabilizing the long-non-conding RNA (lncRNA) Hotair, an important RNA for preadipocyte differentiation[104]. Sam68 regulates the adipogenesis not only through mRNA stabilization but also through the regulation of the alternative splicing of the ribosomal S6 kinase gene (S6K1) impeding the usage of the alternative exons 6a, 6b and 6c that lead to the formation of an anti-adipogenic form of the protein[105]. To summarize, Sam68 is involved in the differentiation of tissues of several embryonic origins by regulating the RNA metabolism.

1.11 Quaking and its role in development

Quaking is the second better characterized member of the STAR family. The pre-mRNA of QKI can give rise, through alternative splicing, to at least four different proteins called QKI5, QKI6, QKI7 and QKI7b, that differ in their C-terminal [78]. QKI5 has a nuclear localization signal therefore it is mainly nuclear but it can also shuttle between nucleus and cytosol[77]. Conversely, QKI6 and QKI7 reside exclusively in the cytoplasm. QKI recognizes and binds to RNAs through a bipartite sequence with a core (ACUAAY) and half site (UAAY) separated by 1 to 20 nucleotides[106,107]

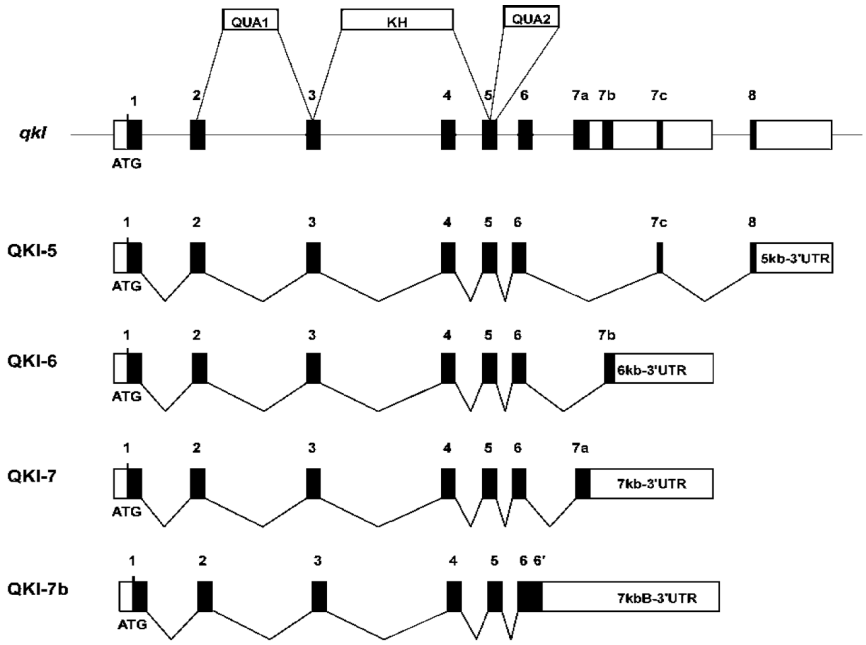


Figure 3 Alternative splicing of Quaking taken form Darbelli et al.[108]

The KO mouse model is embryonic lethal at embryonic day 10.5 due to cardiovascular defects suggesting a very important but still unknown role of this protein during early embryonic development. A natural homozygous mutant of QKI is viable (QKI^{v}), in which QKI expression is selectively attenuated in Schwann cells and oligodendrocytes[109–111]. Specifically, this mutant harbors 1 Megabase deletion in the promoter-enhancer region of the gene located less than 1000 bases 5' to QKI gene and includes not only the enhancer of QKI in glial cells but as well part of the Parkin 2 gene and the Parkin co-regulated gene (Pacrg)[78,112,113]. QKI^{v} mice show myelination defects in both central and peripheral nervous system. They only produce around 10% of non-functional myelin compared to normal brain. Consequently, the mutant mice develop severe

tremors and seizures around the 12th day after birth[114]. Indeed, among all the developmental processes in which QKI takes part, the myelination is probably the most well characterized[78,114–116]. During the nervous tissue development, QKI is expressed at high levels in the neuronal progenitor cell that gives rise to both the neurons and glial cells. QKI expression is shut down in neurons while it is maintained in the glial cells where it plays an important role in the specialization of these subset of nervous tissue cells[117]. More specifically, QKI-5 affects the pre-mRNA splicing of the myelin-associated glycoprotein (MAG) via an intronic element downstream of exon 12 and represses its inclusion[118]. Furthermore, QKI regulates various steps of the RNA metabolism of the transcript that encodes the myelin-basic protein (MBP), another protein important for a fully functional myelination. QKI regulates the nuclear retention of the MBP mRNA as well as its transport to the myelinating membranes. Moreover, through the binding to the 3'UTR of the MBP transcript in the cytoplasm, QKI-6 and QKI-7 play a role in the stabilization of this transcript[119–121]. Recently, it has been shown that QKI-5 binds to an intronic element and regulates the inclusion of the ninein exon 18 which is very important during the differentiation of the neuronal progenitor cells into mature neurons[122]. To date, the critical function of QKI has been elucidated not only in the neuronal tissue but in the vasculature development during embryogenesis as well. More specifically, QKI is required in the visceral endoderm for the expression of enzymes such as Aldh1a2 and Raldh2 that are important for the production of the retinoic acid (RA) and other soluble factors needed for the vasculogenesis[123]. Interestingly,

another important function of Quaking during development is the biogenesis of circular RNAs (circRNAs) during epithelial to mesenchyme transition (EMT). Indeed, as Simon J Conn and colleagues have demonstrated, QKI is responsible for the formation of around one third of all the circRNAs found during the EMT[124]. In cardiac tissues, QKI has been shown to regulate the biogenesis of the circRNA from the Titin and Fhod3 transcripts among others[125]. In a recent study in Zebrafish, homologs of QKI called QKIa and QKIc, were shown to be involved in muscle development. Both QKIa and QKIc bind and stabilize the mRNA of a splicing variant of the tropomyosin called tropomyosin 3.12, an isoform needed for the formation of the slow myofibrils[126]. To conclude, QKI appears to be indispensable for the embryonic development given its importance in the regulation of the RNA metabolism in several developmental processes.

As we previously described Sam68 and QKI are involved in several tissue specifications during development but what about the heart?

1.12 Cardiac Development

The heart is the first functional organ that forms during embryogenesis. It arises around day 7.5 in the mouse and at the third week in the human embryo. During gastrulation part of the mesoderm that emerges from the primitive streak receives cardiogenic signals that lead to pre-cardiac specification[127]. The pre-cardiac mesoderm is then composed by cardiac precursors that express the T-box

transcription factor Eomesodermin (Eomes). These cells will contribute specifically with the construction of the mature cardiovascular system[128] and are identified by the expression of two distinct markers Flk1 and PDGFRα[129,130]. The cells of the pre-cardiac mesoderm will then segregate in two major populations: the first heart field (FHF) that will give rise to the main organ namely to the left ventricle and part of the atria, and the secondary heart field (SHF) that will give rise to the arterial pole of the heart (namely the outflow tract and the right ventricle). The latter is formed by cells labelled by the expression of Isl1, Mef2c and Fgf10[131–133]. At this point emerges another class of cardiac progenitor from the dorsal mesenchymal protrusion (DMP) that will form the atrial septum (the wall that separates the two atria) and marked by the expression of the transcription factor Tbx18 and lacks the expression of Nkx2-5[134–136]. The FHF contains both myocardial and endocardial precursors. The endocardial cells delaminate to form two endocardial tubes surrounded by two separate myocardial tubes that eventually fuse and leads to the fusion of the endocardial tubes as well[137,138]. During this fusion process, the myocardium secrets a jelly matrix that will separate the endocardium form the myocardium. At his point the embryonic heart is no more than a linear tube. This linear tube will loop to allow the convergence of the inflow (venous) and outflow (arterial) tract[139]. As a consequence of the looping, the place where the PHF meets the SHF will become first area of septation. The septation is the process that leads to the formation of the four separated chambers of the mature heart. Concomitant to this process, some cells from the neural tube will migrate until the outflow tract

giving then origin to the conductive tissue of the organ with the formation of the anterior parasympathetic plexis that will innervate the heart tissue and regulates the heart rate[140]. TGF-β and Wnt signaling pathways are needed to fine-balance the expansion of the cardiac precursor and their differentiation[130,141–143]. All the signaling pathways in turns, lead to the activation of cardiac-specific transcription factors. These control and regulate all the aspects of the heart development, including the terminal differentiation of the cardiac cells, the identity of the cardiac chambers and the establishment of transcriptional gradients that defines the identity of each specific portion of the heart[144,145]. The core TFs that governs the cardiac development belong to the GATA family, MEF2 and NKX family among others[127,146–148]. Nkx2-5 is essential for the terminal differentiation of the myocardium, it is expressed by cells with a ventricular fate[149–151]. The GATA protein family, with its members GATA4, GATA5 and GATA6, represents another fundamental class of zinc finger TFs needed for cardiac development. Indeed, many of the cardiac-specific genes have sequences recognized by these TFs in their promoter-enhancer regions. Consequently, these TFs, cooperate with others such as Nkx2-5, Mef2a, Serum Response Factor (SRF) and Tbx5, to regulate the onset of the cardiac differentiation. Among the GATA family of TFs, GATA4 is undoubtedly the most important member. Indeed, it is needed not only in the early onset of cardiac differentiation but as well is indispensable for the atrioventricular valves formation and for the maintenance of the adult cardiac function along with its partner GATA6[154,155].

1.13 RBPs in the heart development

RBPs play different and crucial roles during all the steps of the heart development. The most important RBPs involved in the regulation of cardiac stage-specific alternative splicing events are: Hermes, RBM-24, RBM-20, RBM-15, RBFOX, MBNL1, ESRP, the SR proteins and CELF1. As mentioned above, an important family of RBPs involved in cardiac development is the one of the loosely related RRM-containing RBPs called RBM[156]. More specifically, RBM24 has been shown to take part in regulating the alternative splicing of muscle-related genes in the developing heart, favoring the cardiomyocyte differentiation of ESCs [63,157]. RBM24 levels increase during both mouse and human cardiac development and in the differentiated myocardium[158–162]. RBM24$^{-/-}$ mice die during embryogenesis due to severe cardiac defects caused by a complete loss of the sarcomeric structures while post-natal conditional Knock out mice exhibit dilated cardiomyopathy phenotype. All the phenotypic defects of the RBM24$^{-/-}$ mice are due to the misregulation of the sarcomeric AS events that result in disruptive sarcomeric structures[157,163]. More than 500 AS events are misregulated in the heart upon RBM24$^{-/-}$. Similarly to RBM24, RBM20 is initially expressed in the activated pre-cardiac cells and during cardiogenic differentiation of mESCs[164] and, like RBM24, it is involved in the regulation of the alternative splicing of cardiac-related genes[165–168]. Among the most known, important and well characterized splicing target of RBM20 there is the sarcomeric structural protein Titin which is a protein that is crucial for myofibril assembly, maintenance

and elasticity[168,169]. A splicing-independent example that shows the importance of the RBPs during cardiac development is given by the RBP called Hu antigen R/Elav like protein (HuR). Indeed, HuR stabilizes the mRNA of Mef2C that in turn promotes the expression of the sodium channel α subunit (SCN5α), an important sodium channel of the cardiac cells[170]. Importantly, little is known about the role of the STAR protein members in this process. QKI homolog, in Drosophila, *how*, was shown to be involved in the regulation of beating rate, the sarcomere organization, and expression of the sarcomeric proteins[171,172]. Similarly, some of the QKI mutant mice, show defect in the heart, particularly in the outflow tract although the heart tube formation and the contractile function appears to be normal[173].

2 Objective of the study

In the latest years, the research on stem cells has gone through an exponential growth. Mainly due to the fact that they represent a valuable research tool to investigate in depth the early stages of the mammalian development and their potential application in both disease modelling, drug discovery and therapy. Given the emerging importance that RBPs showed as master regulators in fine-tuning the balance between pluripotency and differentiation, it is crucial to understand their role in this process and the molecular mechanisms by which they achieve to maintain this balance. The RNA Binding Proteins family, STAR, has been shown to be involved in mammalian development. However, the role of two STAR members Sam68 and QKI in early development is still unknown. The objective of this study is to unravel the role Sam68 and QKI in both mouse embryonic stem cell pluripotency and mESCs differentiation. In order to achieve that, we first generated Sam68$^{-/-}$ and QKI$^{-/-}$ mESC lines by CRISPR/Cas9 technology. Then, we extensively characterized these cell lines on a cellular and molecular level, by using the most cutting-edge “omics” techniques.

3 Materials and Methods

3.1 Cell lines

Throughout this research project different human and mouse derived cell lines were used. In order to produce lentiviral vectors, the human embryonic kidney derived HEK293T cell line was used. This cell line was cultured in DMEM supplemented with 10% Fetal Bovine Serum (FBS), 1% Penicillin-Streptomycin (P/S) and 1% of L-Glutamine. The ES Rex1 reporter cell line was obtained by the Austin Smith laboratory. For maintenance culture the cells were plated at a density of 2 to $3x10^4$cells/cm^2 in 0,1% gelatin (ES-006-B, Millipore) precoated dishes and were passed every other day after dissociation with ACCUTASE™ cell detachment solution (SCR005, Millipore). These cells were maintained in 2i medium. The 2i medium consists of a 1:1 mix of Neurobasal medium (21103049, Gibco) and DMEM-F12 (12634010, Gibco) supplemented with both N2 (17502-048 Gibco) and B27 (17504-047 Gibco) factors, 1% of L-Glutamine (25030081, Gibco), 1% Penicillin-Streptomycin (15140122, Gibco), 1% Sodium Pyruvate (11360070, Gibco) and 1% of non-essential amino acids NEAA (11140068, Gibco), β-mercaptoethanol (31350010, Invitrogen), 1μM of PD0325901 (444968, Calbiochem), 3 μM of CHIR99021 (361571, Calbiochem) and leukemia inhibitory factor LIF (ESG-1106 Millipore). The E14 cell line was obtained by the Tissue Engineering unit of the Center for Genomic Regulation.

This cell line, is the one used for the generation of the Sam68 and Quaking KO lines, thanks to the efforts made by the CRG tissue engineering facility along with the facility of biomolecular screening and protein technology. Briefly, E14 WT cells were transfected with the bicistronic plasmid pSpCas9 PX459 codifying for both the Cas9 and the sgRNA targeting the first exon of either Sam68 or Quaking. Afterwards, the transfected cells were selected for the puromycin resistance. The Cas9 activity of each sgRNA was then verified by T7 exonuclease assay. After the puromycin selection the cells were FACS sorted in order to get colonies derived by a single clone. Consequently, the clones were singularly sequenced and the *bona fide* KO ones selected in order to carry out the research project.

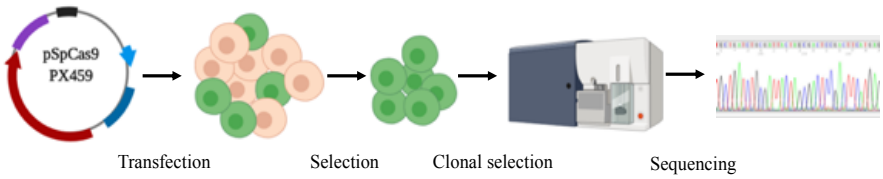


Figure 1 Generation of the KO cell line. Workflow of the generation of Sam68 and Quaking KO E14 cell lines through CRISPR-Cas9 technology.

For maintenance culturing, the cells were plated at a density of ~1,5X10^{4}cells/cm^2 on gelatin pre-coated culture dish and were passed every other day after dissociation with Trypsin-EDTA 0,05% (2530054, Gibco) and maintained in Knock-Out DMEM with 15% embryonic stem cells qualified FBS (16141079, Gibco), 1% P/S, 1% L-Glut, 1% Sodium Pyruvate, 1% NEAA and LIF. For differentiation medium, the amount of ES qualified FBS was reduced to 10%.

3.2 Lentiviral Production

Lentiviral vectors were produced by making use of the HEK293T packaging cell line. Briefly, 24 hours prior to transfection $9x10^6$ cells are plated in a 150mm diameter dish. A mix of 9μg of VSV-G envelope vector, 20μg of packaging Δ8.9 vector and 32μg of pLKO.1-shRNA mission vector (Sigma-Aldrich) is made and milli-Q water is then added up to a volume of 1125μl along with 125μl of 2,5M $CaCl_2$. The solution is then incubated on a rotating wheel at room temperature (RT) for 5 minutes. Afterwards, 1250μl of 2XHBS pH 7.4 are added dropwise while vortexing at maximum speed. The resulting solution is then incubated for 10 minutes at RT. The final solution is added to the cultured cells dropwise covering the whole plate surface. After 14-16 hours the medium is changed and collected twice with a syringe after further 24 and 48 hours. The medium containing the virus is then filtered through a 0,45μm filter and is put into a polipropilene 25X36 cm tube (Beckman Culture). The viral particles are then concentrated through ultracentrifugation for 2 hours and 20 minutes at 22'000 rpm at 22 degrees Celsius. The pellet obtained is then resuspended in ice-cold PBS in order to concentrate the lentiviral vector 1000 times.

3.3 Lentiviral spin-infection

For each infection were used 1,5X10^5 cell resuspended in 100μl of medium supplemented with polybrene at a concentration of 5 μg/mL. A variable amount (from 5 to 10μl) of lentiviral vector was used and the infection was performed by centrifugation at 1000G for 3 hours at 32 degrees Celsius. The cells were then resuspended and plated on a precoated gelatin dish. The medium was changed 24 hours after plating and the selection was performed with 10 μg/ml puromycin (P8833, Sigma-Aldrich) was initiated 48 hours after the infection.

3.4 Teratoma Assay

All the experimental protocols were performed in accordance with the recommendations for the proper care and use of laboratory animals [local (law 32/2007); European (EU directive n° 86/609, EU decree 2001- 486) regulations, and the Standards for Use of Laboratory Animals n° A5388-01 (NIH)] and were approved by the local ethical committee (CEEA-PRBB). For Teratoma formation, 100μl of solution containing 500.000 mouse ES cells and 1:15 matrigel were injected into both flank sites of SCID BEIGE mice. The teratoma formation was stopped after 3 weeks of development. The resulting tissue has been washed with PBS, fixed with 4%PFA, paraffin embedded, sectioned and haematoxylin/eosin staining was

performed. Sections were evaluated and analyzed for the presence of ectodermal, mesodermal and endodermal-derived tissue.

3.5 Cell Synchronization

In order to synchronize the cells, these were treated with 400nM of nocodazole (M1404, Sigma-Aldrich) for 14-16 hours. Once the synchronization was achieved, the cells were released by washing the cell plate three times with PBS.

3.6 Clonogenic Assay, Alkaline Phosphatase staining and colony area calculation

In order to set up the clonogenic assay, 100cells/ml of KO DMEM supplemented with LIF and 15% FBS were seeded in a gelatin precoated 6 well plate (2ml/well). The cells were left in culture for 10 days and the medium was changed every other day. At day 10 the cells were stained with the Alkaline phosphatase staining kit II (00-0055, Stemgent) according to the manufacture's indications. Cells were then counted and the area of the colonies was quantified using Image J software.

3.7 Proliferation Assay

To set up the proliferation assay $75x10^3$ cells/well were plated in a gelatin precoated 12 multiwell plate. Cells were then counted daily for the next three days. Each well was counted twice and the media, supplemented with LIF, was replaced any other day.

3.8 Embryoid Body assay and cardiomyocyte differentiation

To set up the Embryoid Body (EB) assay, the cells were seeded in hanging drops at a density of $2x10^4$cells/ml of KO-DMEM supplemented with 10% FBS ES tested. After three days of hanging drop culture, EBs were flushed and seeded on a low attachment plate. The medium was changed any other day until the day 10. In order to derive cardiomyocytes, at day three (D3) the cells were seeded on a gelatin precoated 96 multiwell plate and cultivated with the same medium for a maximum of 8 days more. The cells were then scored for the capacity of giving rise to beating *foci*.

3.9 RNA-sequencing

In order to perform the RNA-sequencing, total RNA was extracted with the Maxwell 16 LEV simplyRNA Cells Kit (AS1270, Promega) according to manufacturer instructions. The quality and quantity of

the extracted RNA was then verified using the Nanodrop. rRNA depletion and library preparation was done with the TruSeq stranded total RNA Library prer Human/Mouse/Rat (20020596, Illumina). The sequencing was performed using 2x125bp paired-ends reads on a HiSeq 2500 sequencer with HiSeq v4 chemistry. The reads were then demultiplexed and analyzed with the DESeq2 pipeline.

3.10 Retrotranscription

A variable amount of RNA (from 500ng up to 1μg) was retrotranscribed to cDNA using the NZY first-strand cDNA synthesis kit (MB125, NZYtech) according to manufacturer instructions. Residual RNA was then digested with the RNase H, a component of the kit.

3.11 Semiquantitave PCR

Semiquantitve Polymerase Chain Reaction (PCR) was performed by mixing the necessary amount of DNA with H_2O up to a volume of 10μL. 10μL of NZYtaq II 2X green master mix (MB358, NZYtech) were then added. To amplify the DNA a specific thermocyclator program was used according to the melting temperature of the primers as well as to the specificity of each DNA target. The products of the PCR reaction were then run on an agarose gel (from 0,5 up to

3% depending on the amplimer size) in TAE buffer. The run was always performed at 100V.

3.12 Real-Time PCR

The quantitative PCR was performed by using SYBR® master mix (4309155, Applied Biosystem) according to manufacturer instruction. The reaction was run on a Viia™ 7 system and the fold enrichment derived with the ΔΔCt method.

3.13 Total protein extraction, cytosol-nucleus fractionation, SDS-page and Western Blot

In order to extract proteins, an adequate volume of RIPA buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 1% Triton-X100, 0,1% sodium deoxycholate, 0,1% of sodium dodecilsulphate, 140mM of NaCl) supplemented with proteases inhibitor (1187358000111, Roche) was added to the cell pellet and kept on ice for 10 minutes pipetting at least once during the incubation. Afterwards, the cells are centrifuged at max speed (>13'000 rpm) on a tabletop centrifuge for 10 minutes at +4°C. The supernatant containing the proteins is then preserved and the concentration of protein measured spectrophotometrically as absorbance at 595nm in Bradford reagent. As far as the fractionate extraction is concerned, cells were scraped in PBS supplemented with 1mM EDTA, they were pelleted at full

speed and resuspended in an adequate volume of ice-cold NP40 lysis buffer (10mM Tris-HCl pH7.5, 0,15% NP40, 150mM NaCl) for 5 minutes. Then the sample was stratified adding 2,5 volumes of NP40 supplemented with 24% sucrose and centrifuged at 14'000rpm at +4°C for 10 minutes. The supernatant is conserved as cytosolic extract. The nuclear pellet was then washed with PBS/1mM EDTA, centrifuged and resuspended in glycerol buffer (20mM Trish-HCl pH7.9, 75mM NaCl, 0,5mM EDTA, 0,85mM DTT and 50% glycerol). To this, an equal volume of nuclei lysis buffer (10mM Hepes pH7.6, 1mM DTT, 7,5mM of $MgCl_2$, 0,2mM EDTA, 0,3M NaCl, 1M Urea and 1% NP40) was added. This was vortexed for 2 seconds twice and then centrifuged at max speed in a tabletop centrifuge. The supernatant was then taken as soluble nuclear fraction. An adequate volume of protein corresponding to a variable amount from 25µg up to 60µg is then mixed with LDS 4X (NP007, Invitrogen) and H_2O up to the desired volume. The mix is then loaded into a precast NUPAGE 4%-12% bis-tris gel (NP0321BOX, Invitrogen) immerged in MOPS buffer (NP0001, Invitrogen). The run is performed at 180V and once finished the gel is dry-blotted with the iBlot™ 2 NC transfer stacks (IB23002, Invitrogen) and its corresponding device using the P3 program for 7 minutes. The nitrocellulose is then blocked with 5% milk in T-TBS (TBS buffer with 0,1% Tween20) for 1h at RT shaking. Afterwards the membrane is incubated with the primary antibody in 2,5% milk in T-TBS for either 1h at RT or overnight at +4C always shaking. After the incubation with the primary antibody the membrane is washed 3 times for 5 minutes with T-TBS shaking and then incubated with the

corresponding secondary antibody in T-TBS with 2,5% milk 1 hour at RT. The membrane is washed 3 more times for 5 minutes with T-TBS. In order to reveal the signal the membrane is incubated with the Immobilon Classico Western HRP substrate (WBLUC0100, Millipore) and developed with the Amershan 600 (GE Healthcare and Life Sciences).

3.14 RNA-Immunoprecipitation (RIP)

An adequate volume (100µl per sample) of protein G Dynabeads (10004D, Life Technologies) were washed twice with 900µl of lysis buffer (150mM NaCl, 25mM Tris pH 7.4, 5mM EDTA, 1mM DTT, 0,5% NP40, protease inhibitor and 100U/ml of RNAase inhibitor SUPERase in AM2696, Invitrogen). Then, they were resuspended in 100µl of lysis buffer with 4µg of antibody or IgG for the negative control and incubated at +4°C on a rotating wheel. The cells were pelleted and resuspended in 1ml of lysis buffer on ice, then 5µl of turbo DNase (AM2238, Invitrogen) were added and were consequently incubated at 37°C in agitation at 400 rpm. The sample were then furtherly incubated on ice for 10 more minutes. The sample was centrifuged at 13000rpm for 10 minutes at +4°C. 10% of the supernatant was directly mixed with Qiazol (79306, Qiagen) for RNA extraction done according to the manufacturer's instructions. The remaining part of the lysate instead was mixed with the beads and incubated overnight at +4°C on a rotating wheel. After that, the beads were washed three times with wash buffer (150mM, 25mM

Tris pH 7.4, 5mM EDTA, 1mM DTT and 0,5% NP40), the last wash was then performed with PBS. 20% of the volume were then resuspended in NuPAGE LDS and H_2O and were used for SDS-WB analysis while the rest was mixed with 1mL of Qiazol and the extracted RNA was retrotranscribed and analyzed with either semi-quantitative PCR or qPCR.

3.15 Ribosome Profiling

For each experiment of ribosome profiling we used 3 plates of 100mm at full confluence. Cells were treated with cycloheximide (01810, Sigma Aldrich) at a concentration of 100 μg/mL and were incubated for 15 minutes at 37°C. The media was withdrawn and the cells washed with ice cold PBS supplemented with cycloheximide, scraped and pelleted. The cells were then lysed with lysis buffer (20mM Tris-HCl pH7,4, 150mM of NaCl, 5mM of $MgCl_2$, 1mM DTT, 100 μg/mL, 1% of Triton and 25U/mL of Turbo DNase I [AM 2238, Invitrogen] in RNase free water) and glass beads by vortexing 5 seconds for 3 times. The lysate was then left 10 minutes on ice and consequently centrifuged at 10'000 rpm for 10 minutes. To proceed we took an amount of sample corresponding to 8-10 OD_{260nm}. In order to digest the unprotected RNA, RNAse I (AM2294, Invitrogen) at a concentration of 100U/μl was added and incubated at 25°C for 15 minutes. 10μl of SUPERase In (AM 2694, Invitrogen) at 20U/μl were added and the sample loaded on a sucrose cushion at 34% w/v in a polycarbonate ultracentrifuge tube (349622, Beckman Culture).

The sample was then centrifuged at 70'000 rpm for 4 hours at +4°C. The supernatant was discarded and the pellet resuspended in the resuspension buffer of the Maxwell kit (AS1270, Promega) of RNA-extraction. The RNA was then extracted according to the manufacturer instructions and resuspended in 40µl of RNAse free water. In order to select the fragments of interest we pre-run a 15% TBE-Urea polyacrylamide gel (EC6885BOX, Invitrogen) at 200V for 15 minutes in TBE buffer. 5µg of each sample were then mixed with the TBE-Urea sample buffer. The RNA was then denatured at 80°C for 90 seconds and loaded onto the gel were they were run at 200V for 65 minutes. The gel was then stained for 5 minutes with SYBR green (S32717, Invitrogen) and fragments between 28nt and 32nt excised. The bands were transferred into a RNase-free tube, resuspended in 360µl of RNase free water and the RNA extracted by crushing the gel with a pestel. Then it was incubated at 70°C for 10 minutes in order to melt the gel. Once melted the sample was loaded into a spin column with filter made of 2 pieces of whatmann paper and centrifuged at max speed for 2 minutes. Consequently, the RNA was precipitated by adding 40µl of 3M sodium acetate, 1µl of glycoblue (00585590, Invitrogen) and 500µl of isopropanol. The precipitation was done either for 1 hour in dry ice or overnight at -80°C. Afterward the samples was centrifuged at 20'000g for 30 minutes at 4°C. The supernatant was discarded and the pellet washed with ethanol at 70% in RNase free water. It was then furtherly centrifuged at 20'000g for 2 minutes and then air dried for 10 minutes. Eventually, it was resuspended in 26µl of RNase free water. The ribosomal RNA was then removed by making use of the Illumina

Ribo-zero rRNA removal kit (20020596, Illumina) according to manufacturer's instructions. The RNA left was then precipitated by adding 300µl of isopropanol 20µl of sodium acetate, 90µl of H_2O and precipitated for 1 hour in dry ice. Then it was centrifuged again at 20'000g for 30 minutes at +4°C, the supernatant discarded and the pellet washed with 500µl of ethanol 70%, centrifuged again at 20'000g for 5 minutes at 4°C and the pellet was air-dried and resuspended in 10µl of TRIS-HCl pH8 10mM. 33µl of water were added and the samples were incubated at 80 °C for 90 seconds. In order to dephosphorylate the RNA, 5µl of T4 PNK buffer, 1µl of SUPERase IN and 1µl of T4 PNK enzyme (M0202L, New England Biolabs) were added and the mix incubated at 37°C for 1 hour. The enzyme was then inactivated by rising the temperature at 70°C. The RNA was then precipitated and washed as previously stated and resuspended in 8,5µl of water. In order to add the linker, 1,5µl of preadenylated linker were previously denaturated (0.5µg/µl) at 80 °C for 90 seconds and then cooled down at RT. Then it was ligated by adding 2µl of T4 RNA truncated ligase buffer, 6µl of PEG 8000, 1µl of SUPERase In (20 U/µl) and 1µl of T4 RNA truncated ligase enzyme (200 U/µl, M024L, New England Biolabs), incubated for 2 hours and a half at RT. Then the RNA was precipitated and washed as previously described and resuspended in 10µl of Tris-HCl pH8 10mM. Then the samples were run on a 15% TBE gel as previously described, the RNA was excised, extracted and precipitated as previously described and resuspended in 10µl of TRIS HCl pH 8 10 mM. In order to convert it to cDNA 2µl of reverse transcription

primer at 1.25μM were added along with 4μl of 5x first strand buffer, 1μl of dNTPs mix (10 mM), 1μl of 1M DTT , 1μl of SUPERase In (20 U/μl), 1μl of Superscript III enzyme (200 U/μl, 18080093, Invitrogen) and retrotranscribed at 50°C for 30 minutes. The cDNA was then precipitated and washed as previously described and resuspended in 10μl of water. The cDNA was then electrophoretically run, excised and precipitated as previously described. The cDNA is then circularized adding 2μl of circligase buffer, 1mM of ATP and 50mM of $MnCl_2$ and 100 units of CircLigase I (CL4111K, Epicentre). The sample was then incubated 1 hour at 60°C in a thermal cycler and then heat-inactivate the enzyme for 10 min at 80°C and the rRNA depleted once again. After precipitation and wash the cDNA was resuspended in 5μl of TRIS HCl pH 8 10 mM and was PCR amplified in 100μl volume containing 20μl of Phusion HF buffer, 2μl of dNTPs mix (10 mM), 0.5μl of FW primer (100 μM), 0.5μl of RV barcoded primer (100 μM) and 71μl of RNase free water, 1μl of phusion polymerase (2 U/μl, M05305, New England Biolabs). The PCR was performed with the following program: 1 cycle at 98°C for 30 sec, 14 cycles at 98°C for 10 sec, 65°C for 10 sec, 72°C for 5 sec. Eventually, the samples were run on an 8% (wt/vol) polyacrylamide non-denaturing gel for 40 min at 180V. The gel was then stained and a band at ~180nt was excised, extracted and purified as previously described. Then the libraries obtained were quality controlled and sequenced using single-end 50bp sequencing on a HiSeq 2500 sequencer with HiSeq v4 chemistry. The bioinformatics analysis was performed by the bioinformatics core facility of the CRG. Briefly, the adaptor was

trimmed off the raw sequences (fastq files) with skewer (version 0.2.2). The quality of both raw and trimmed reads was assessed with the FastQC tool. The reads that aligned (using bowtie2 version 3.2.0) to the rRNAs or the tRNAs (coordinates from the UCSC table browser), were removed. All the remaining reads were size selected: only reads ranging from 22nt to 36nt were considered for further analysis, to capture the ribosome protected fragments (approximately 30bp). These were then aligned to the Gencode M14 version of the genome (mm10/GRCm38) using the STAR mapper (version 2.5.3a). Read coverage around the TSS was assessed, for each selected read length separated, in order to define an offset/shift value of ribosome position (the shift are then applied to the read mapping position to obtain the approximate position of the translated codon). The reads were finally sorted out according to their alignment to either the coding sequence (CDS) or the 5' or 3' untranslated regions.

3.16 EBs inclusion and Immunofluorescence

EBs at D10 were collected and transferred into a 15ml conical tube. The medium was removed and the EBs fixed with 4% paraformaldehyde (PFA) for 30 minutes at room temperature. The PFA was removed and the EBs washed with PBS for 5 minutes. The EBs were then incubated for 30 minutes with solutions made of PBS with increasing concentration of sucrose (10, 20 and 30%). After that, the solution was carefully removed and the EBs were included in

OCT and frozen. The tissue was then cut at the microtome and the put on a glass slide. In order to stain it, first the slide was put at +37°C for 30 minutes then it was washed 3 times in PBS. Afterwards, the slide was incubated with blocking solution (5% BSA, 0,1% Triton-X100). The slide was then incubated in a humid chamber overnight at +4°C with blocking solution and the primary antibody. Then, it was washed 3 times for 5 minutes with PBS and incubated with a specific secondary antibody always in blocking solution for 1h at RT. Eventually, the slide was washed 3 more times in PBS for 5 minutes and it the slide was covered with a coverslip attached with the mounting solution containing DAPI to stain the nuclei.

4 RESULTS

4.1 CELLULAR CHARACTERIZATION

4.1.1 Generation of *bona fide* Sam68 and Quaking knock-out mouse embryonic stem cell lines

The general workflow of the generation of the KO cell lines is described in the material and methods section. Briefly, in the Sam68$^{-/-}$ cell line, a 16 nucleotides deletion was generated right after the start codon in the first exon. This mutation completely disrupts the reading frame and abolishes the production of the protein. On the other hand, in the QKI$^{-/-}$ cell line a 2nt deletion was produced leading to a premature stop codon. The genomic sequencing of the Sam68 and QKI loci is shown in figure 1.

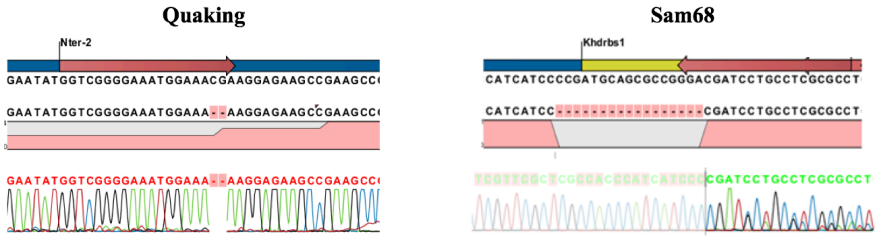


Figure 1 Generation of QKI and Sam68 KO lines. Sequencing of the CRISPR-Cas9 generated mESCs KO for Quaking and Sam68. Each cell line, shows the homozygous nucleotide depletions in QKI and Sam68 respectively.

Given the possible redundancy in function of the two members of the STAR family, we wanted to investigate whether the complete abolishment of one of these two proteins would lead to the

deregulation of the other one. Notably, the KO of any of the two proteins does not affect the expression levels of the other one, suggesting that there is not compensation between those two proteins.

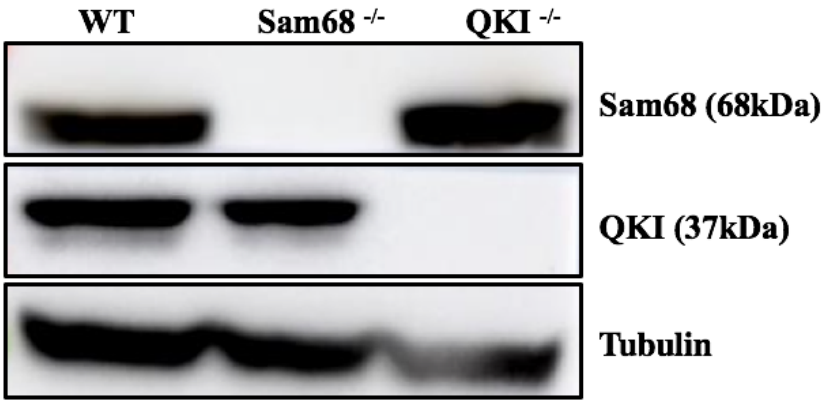


Figure 2 Sam68 and Quaking KO cell lines. Western blot showing the complete depletion of Sam68 and Quaking at protein levels in their respective KO cell lines. It also shows the absence of compensation mechanism between the two STAR family members.

4.1.2 Sam68, Quaking and stem cells features

To understand the role of Sam68 and QKI in pluripotency we first analyzed the phenotype of the generated cell lines by observing the ESCs colonies under the phase-contrast microscopy. As shown in figure 3, both Sam68$^{-/-}$ and QKI$^{-/-}$ ESCs colonies do not exhibit morphological differences compared to their WT counterparts when kept and maintained in normal serum condition. In fact, during normal culture maintenance, KO colonies form the typical mESCs

domed-shape morphology with defined borders as the WT ones and do not show any apparent anomaly.

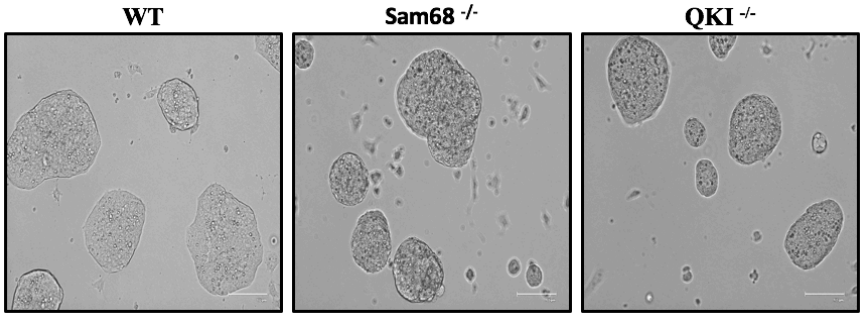


Figure 3 WT and KO mESCs colonies. The KO colonies maintain the typical visible stem cells features as the WT ones.

4.1.3 Sam68, Quaking and pluripotency exit

In order to undertake a differentiation path, a stem cell has to exit from the pluripotent state. This means that the cell has to both dynamically orchestrate a transcriptional program that leads to a terminally differentiated state and most importantly, to disrupt all the transcriptional circuitry that sustains the pluripotent state. The exit from pluripotency has been widely studied at the levels of transcription and epigenetics through several high-throughput screenings revealing master transcription factors and identifying dual specificity phosphatases in regulating this exit via ERK signalling leading to nuclear clearance of TFE3 transcription factor[40]. However, involvement of RBPs as another layer of regulation and the molecular mechanisms ruling this process are still poorly understood. In order to assess the roles of both Sam68 and Quaking in the exiting from the

pluripotent state we made use of a cell line Rex1-GFP (RG1)[174]. This cell line was generated by knocking in of a destabilized form of the green fluorescent protein (GFP) with a half-life of 2 hours in a monoallelic way under the promoter of the pluripotency-associated factor Rex1 (gene name Zfp42).

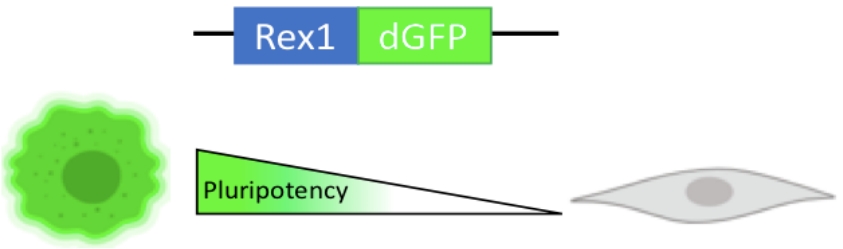


Figure 4 Rex1-GFP (RG1) cells. The figure shows the genetic construct that characterizes this cell line. The cartoon shows how GFP expression is lost upon loss of pluripotency

In pluripotency conditions, Rex1-GFP is expressed. However, when the cells start to differentiate, they gradually loose the expression of the pluripotency factor Rex1 and consequently of the GFP. During the first steps of this experiment, the cells are maintained in conditions that ensure the pluripotency maintenance. We depleted either Sam68 or QKI via lentiviral infection of short-hairpin RNAs in the Rex1-GFP cells. The depletion levels of each protein are shown in figure 5A. To test the role of Sam68 and QKI in the exit from pluripotency we allowed the cells to differentiate, using a differentiation-permissive medium and monitoring the dynamics of the GFP expression by time course flow cytometry. We calculated the percentage of cells exiting from pluripotency by the ratio of cells that lost the expression of GFP over the total number of cells in each

condition (Figure 5B). As shown in the graphs, the depletion of Sam68 or QKI did not affect the expression of Rex1-GFP compared to the control cells in differentiation-permissive medium. We concluded that depletion of Sam68 or Quaking does not interfere with the disruption of the pluripotency molecular circuitry.

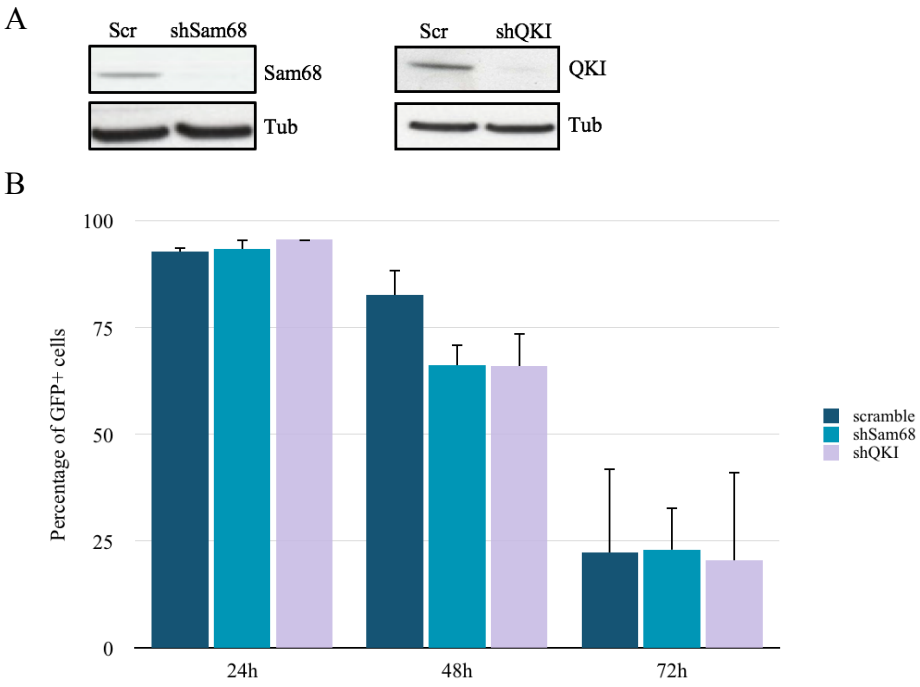


Figure 5 Dynamics of the pluripotency circuitry disruption. A- Western blot showing the depletion levels of either Sam68 or QKI in Rex1-GFP cell lines. B- The graph shows the dynamic of the loss of the GFP expression, and therefore of the pluripotency, when the cells are cultured in differentiation-permissive medium.

4.1.4 Sam68, Quaking and pluripotency *in vivo*

The teratoma formation assay is a gold standard for the assessment of the pluripotency capacity *in vivo* of the ESCs. A teratoma is a non-malignant tumor derived by the stem cells injected in a nude mouse. The resulting tumor is composed by a disorganized mixture of cells more or less differentiated that derive from all the three germ-layers if the cells of origin are truly pluripotent[175]. We implanted Sam68$^{-/-}$ or QKI$^{-/-}$ ESCs into different immune-compromised mice and allowed them to proliferate and differentiate to form a teratoma. After three weeks, we removed the tumors that had reached sufficient size and subjected them to hematoxylin-eosin staining (figure 6). The histological sections show cells derived from either ectoderm, mesoderm and endoderm in both teratomas derived from Sam68$^{-/-}$ or QKI$^{-/-}$. This strongly suggests that the depletion of either Sam68 or Quaking does not alter the pluripotency capacity of the mESCs *in vivo*. It is important to point out that the high variability in tissue formation in this assay has been widely reported[176].

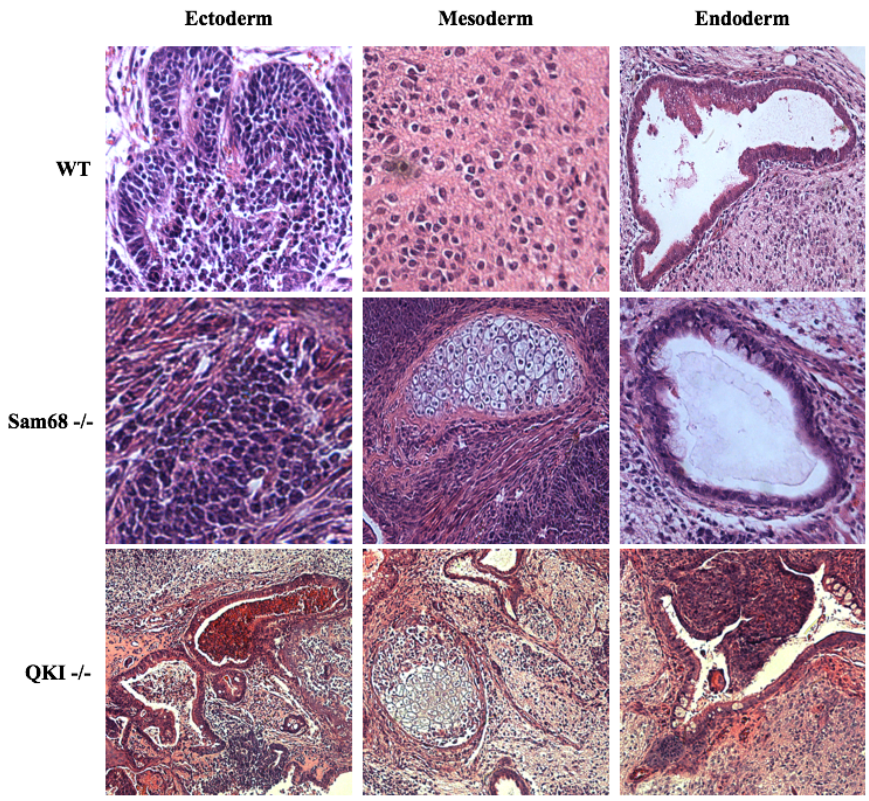


Figure 6 Teratoma formation assay. Teratomas derived from either WT E14 mESCs or the KO counterparts. All the teratomas contain cells derived from all the germ layers.

4.1.5 Sam68, Quaking and proliferation of the mESCs

The role of Sam68 and Quaking in the regulation of the cell cycle has been shown in different cellular contexts[80150], but not in ESCs. We therefore used our Sam68$^{-/-}$ and QKI$^{-/-}$ ESCs and synchronized them using nocodazole, a drug that promotes the arrest in the G2/M phase. After a 16 hours treatment, a time-window that ensures complete arrest in that cell cycle phase, we released the cells and maintained them in normal medium. Consequently, we counted the cells every

24 hours for 3 days and plotted the results in figure 7. As the graph shows, mESCs depleted of either Sam68 or QKI proliferate significantly less in the first divisions compared to the WT counterparts. This reduced growth rate lasted until 72 hours upon QKI depletion but not upon Sam68 depletion suggesting that Sam68 is very important for the first cell divisions and that a compensatory mechanism is activated to help the cells overcome the loss of Sam68. Both results go on line with previous findings showing that the absence of Sam68 is associated with a prolonged G2-M phase progression in a fibroblast chicken cell line, leading to a slower cell cycle[177], and that Quaking directly binds to and control the expression of mRNAs encoding critical cell cycle factors such as cyclin E in worms, cdc25 in flies, and p27 in mammals[178].

Taken together, these data demonstrate that Sam68 and Quaking are involved in the control of the cell cycle in ESCs and further studies are needed to decipher the molecular mechanism underlying this proliferative defect.

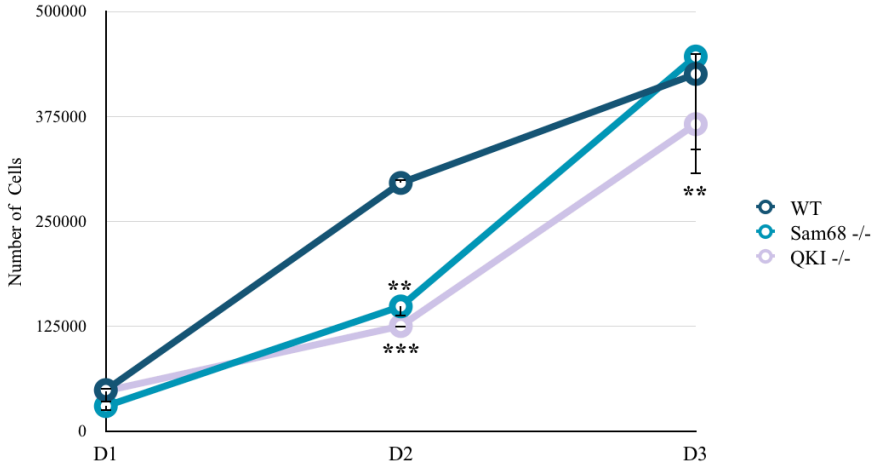


Figure 7 Proliferation assay of WT, Sam68$^{-/-}$ and QKI$^{-/-}$ mESCs. Each time-point is the average of three independent replicas.

4.1.6 Sam68, Quaking and self-renewal

The defect in proliferation upon Sam68 or Quaking depletion lead us to investigate their role in the self-renewal of the undifferentiated mESCs, a property linked as well to the proliferation. Previous reports have demonstrated the role of Sam68 in the self-renewal of the neuronal progenitor cells (NPCs) by promoting the glycolytic metabolism. More specifically, in proliferating NPCs, where the levels of Sam68 are high, Sam68 masks an alternative polyadenylation site (aPAS) of the mRNA of the Aldehyde dehydrogenase 1A3 (Aldh1a3) leading to the production of a fully functional enzyme that sustains the glycolytic metabolism and therefore the proliferation. Conversely, when the levels of Sam68 decrease, alternative polyadenylation occurs and a non-functional form of the enzyme is produced. As a consequence the glycolytic metabolism decreases thus promoting differentiation of the NPCs to neurons[103]. In order to assess whether the absence of either Sam68 or Quaking affects the self-renewal capacity of the mESCs, we set up a clonogenic assay. Unlike the proliferation assay, in the clonogenic assay the cells are plated at a clonal density allowing them to be less exposed and dependent on the cell-cell contacts and/or to the activity of paracrine signals. Then, we synchronized the cells, plated them a clonal density and kept them in maintenance media for 10 days. We, then, revealed the colonies with the alkaline phosphatase staining. We repeated the experiment in three different biological replicates. We counted the colonies and represented the results in figure 8. The depletion of either proteins led to a dramatic effect on both the density

and the size of the colonies. As shown in figure 8, QKI$^{-/-}$ and Sam68$^{-/-}$ ESCs give a significantly lower number of colonies suggesting that indeed, they do exert a role in the self-renewal capacity of the mESCs. Furthermore, we measured the size of the colonies using ImageJ software and represented the average size in each condition in figure 9. We observed that the colonies derived from either Sam68$^{-/-}$ or QKI$^{-/-}$ ESCs have a significantly lower area when compared to the WT cells. This corroborate the phenotype observed in the proliferation assay. Indeed, the area of a colony is directly dependent on the proliferative capacity of the cells during the first divisions after plating. These results demonstrate that both Sam68 and QKI affect the self-renewal and the proliferative capacity of the mESCs. This leaves us with some open questions: Is it a conserved mechanism by which Sam68 regulate the self-renewal capacity of undifferentiated cells? What about Quaking?

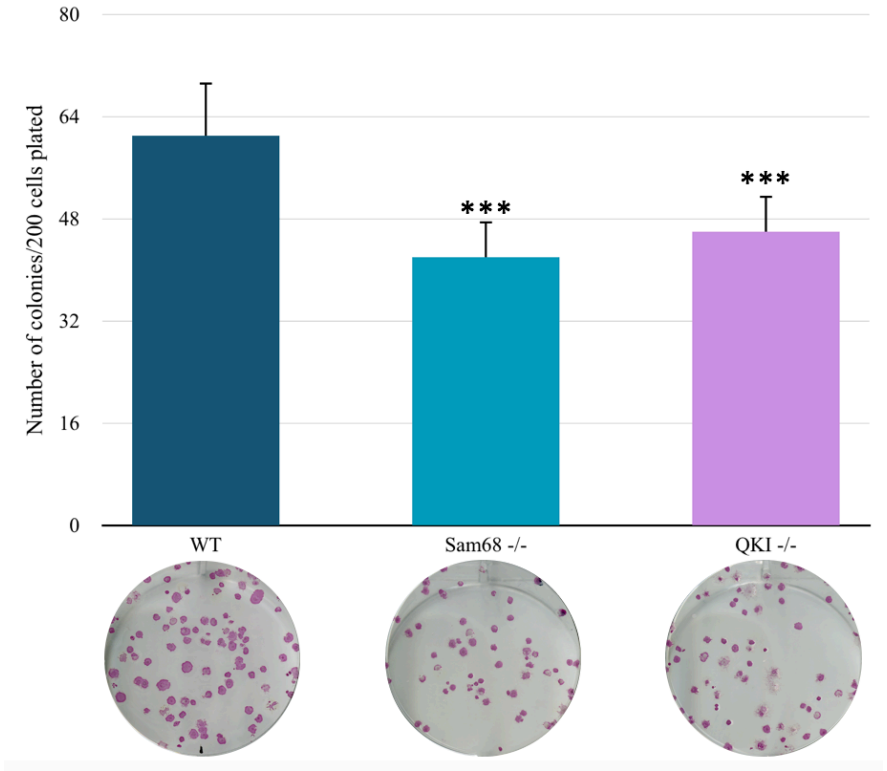


Figure 8 Clonogenic assay. In the upper panel: the graph shows the number of colonies obtained after plating 200 cells per well. The lower panel shows pictures of stained colonies.

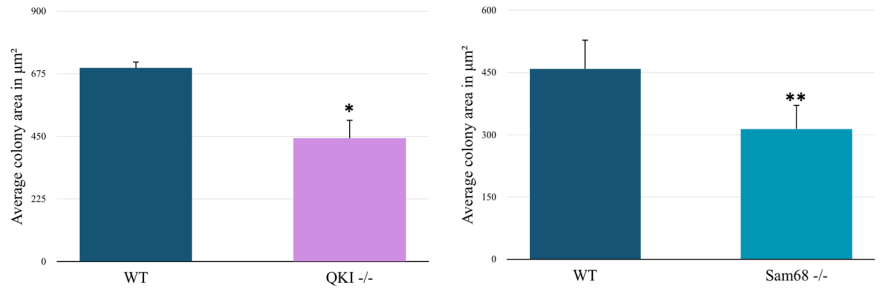


Figure 9 Area of the colonies of the clonogenic assay. The bar plot on the left shows the average area of the colonies in square micron obtained by both WT and QKI$^{-/-}$ mESCs (left) and WT and Sam68$^{-/-}$ mESCs (right).

4.1.7 Sam68 and Quaking cellular localization

Given that both Sam68 and QKI are known regulators of the RNA metabolism depending on their specific cellular localization, we wanted to check the distribution of the two STAR family members in the different cellular compartments of both undifferentiated mESCs and in differentiation conditions. To do so, we performed a fractioned protein extraction of the nucleus and the cytosol in both pluripotency and EBs at day 10 of differentiation. We have noticed that the expression of both proteins is downregulated during mESCs EBs differentiation (figure 10). The expression levels of these two proteins have been shown to decrease during neuronal differentiation[103,179]. Interestingly, while Sam68 seems to be enriched in the nucleic fraction, suggesting a more predominant role in the AS regulation and/or in the nuclear export, QKI seems to be more enriched in the cytosolic one. This in turn suggests a more prominent role of the latter in the mechanism that regulates the stability and the translation efficiency of its target transcripts.

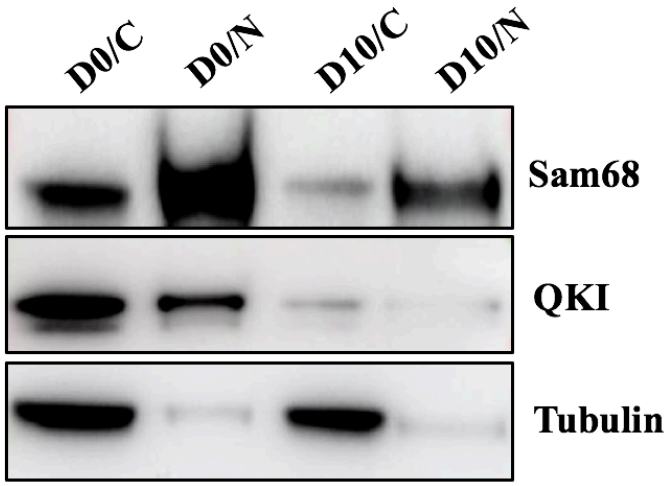


Figure 10 Nucleus-Cytoplasm fractioned WB. WB showing the level of Sam68 and QKI protein expression in the nucleus and cytosol of mESCs and EB at D10 of differentiation.

4.2 MOLECULAR CHARACTERIZATION

Given the importance of the RBPs in regulating different processes involved in the regulation of the RNA metabolism, we wanted to investigate whether the depletion of either Sam68 or QKI could have an impact on the RNA regulation in mESCs both in pluripotency conditions and during embryonic bodies differentiation. To achieve that, we decided to perform total RNA-sequencing experiments in different conditions. We chose to have a sequencing depth that would allow us to detect alterations in the AS events as well. On the other hand, we carried out the ribosome profiling to find out how the depletion of either proteins can affect the translation of the transcripts on a global scale.

4.2.1 Sam68, Quaking and cardiac related genes

We performed pair-end RNA-sequencing on WT and either $Sam68^{-/-}$ or $QKI^{-/-}$ lines in pluripotency condition (Day 0), in the early stages of differentiation, corresponding to the day 3 of EBs assay (D3), and in the later stages of differentiation corresponding to the day 10 of EBs differentiation (D10). We summarized the sequencing results in table 1.

D0	WT 1st	WT 2nd	WT 3rd	S 1st	S 2nd	S 3rd	Q 1st	Q 2nd	Q 3rd
raw reads	91.354.108	91.955.523	92.016.607	98.398.678	96.557.007	94.962.196	87.689.985	90.686.885	89.701.699
uniquely mapped	77.311.819	77.364.572	77.381.822	87.962.741	85.993.115	84.677.269	74.450.307	76.907.448	76.336.882
% of uniquely mapped	84,63%	84,13%	84,10%	89,39%	89,06%	89,17%	84,90%	84,81%	85,10%

D3	WT 1st	WT 2nd	WT 3rd	S 1st	S 2nd	S 3rd	Q 1st	Q 2nd	Q 3rd
raw reads	89.364.875	90.340.215	88.786.891	99.027.997	95.070.113	96.325.674	92.152.304	92.579.852	92.958.574
uniquely mapped	75.655.555	77.988.120	75.396.095	87.880.704	86.105.057	87.196.977	78.196.435	79.638.629	78.907.359
% of uniquely mapped	84,66%	86,33%	84,92%	88,74%	90,57%	90,52%	84,86%	86,02%	84,88%

D10	WT 1st	WT 2nd	WT 3rd	S 1st	S 2nd	S 3rd	Q 1st	Q 2nd	Q 3rd
raw reads	90.528.355	92.349.099	92.407.825	96.372.761	96.991.949	94.603.099	93.340.179	90.441.821	92.158.892
uniquely mapped	77.183.270	78.782.912	79.472.962	86.028.802	86.857.141	83.647.467	80.402.538	75.944.468	78.897.948
% of uniquely mapped	85,26%	85,31%	86,00%	89,27%	89,55%	88,42%	86,14%	83,97%	85,61%

Table 1 RNA sequencing results for the three different conditions and the three time points. The raw reads are represented in the first lanes for each time point and replicate. The number and percentage of uniquely mapped reads after adaptor trimming and ribosomal and transfer RNA elimination are shown in lane 2 and three respectively.

We multiplexed the samples and used the Illumina Hi-Seq 2500 to generate more than 800 million reads per lane. Triplicate Samples from each condition and time point were pooled in the same lane.

The downstream analysis of the results was performed using the DESeq2 pipeline for RNA-Seq data. Briefly, it includes generating FASTQ-format files containing raw read sequences and aligning these reads to the annotated mouse genome GRCm38/mm10, and quantifying the expression of the genes. It is worth pointing out that RNA-Seq analysis presents unique computational challenges not encountered in other sequencing-based analyses and requires specific consideration to the biases inherent in expression data. Mapping RNA-Seq reads to the genome is considerably more challenging than mapping DNA sequencing reads because many reads map across splice junctions. DESeq2 pipeline resolves this problem by supplementing the reference genome with sequences derived from exon–exon splice junctions acquired from known gene annotations. DESeq2 maps reads with a “splicing-aware” way recognizing the difference between a read aligning across an exon–intron boundary

and a read with a short insertion. It takes into account the alignment yield, base wise accuracy, mismatch and gap placement, and exon junction discovery. The number of uniquely mapped reads per sample is very similar among the different replicas and conditions and allowed us to perform the downstream statistical analysis.

First, we used the principal component analysis (PCA) to compare the different biological replicas and conditions. In general, PCA helps to capture the most important features of a dataset by identifying the most relevant directions of variance in the data allowing a “friendly” visualization of the data. Both figure 10 and 11 show high reproducibility of the different replicas, excluding the technical variability for the different experiments. As expected, both the WT and the KO samples cluster according to the time-point, meaning that the difference during differentiation of the same sample is higher than the difference between KO and WT at the same time-point. This is not surprising given that when a stem cell undergoes differentiation its gene expression program changes dramatically.

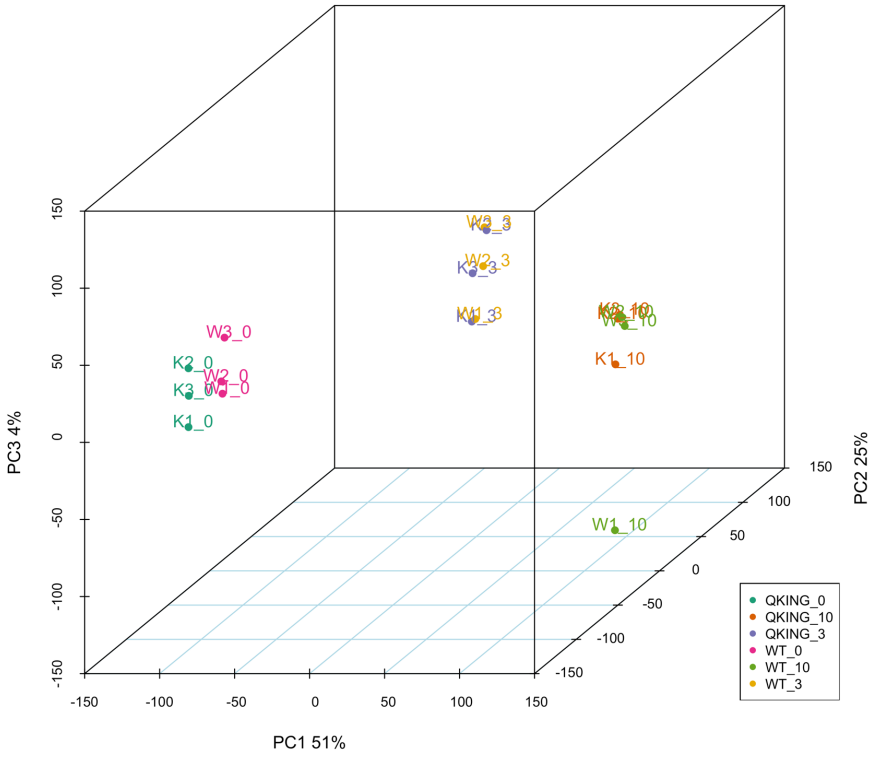


Figure 10 Principal Component Analysis of QKI$^{-/-}$ RNA-sequencing. The PCA shows a high degree of reproducibility. The biological triplicates of each condition and time point cluster together.

Of note, Sam68$^{-/-}$ leads to major differences in RNA expression if compared to QKI$^{-/-}$.

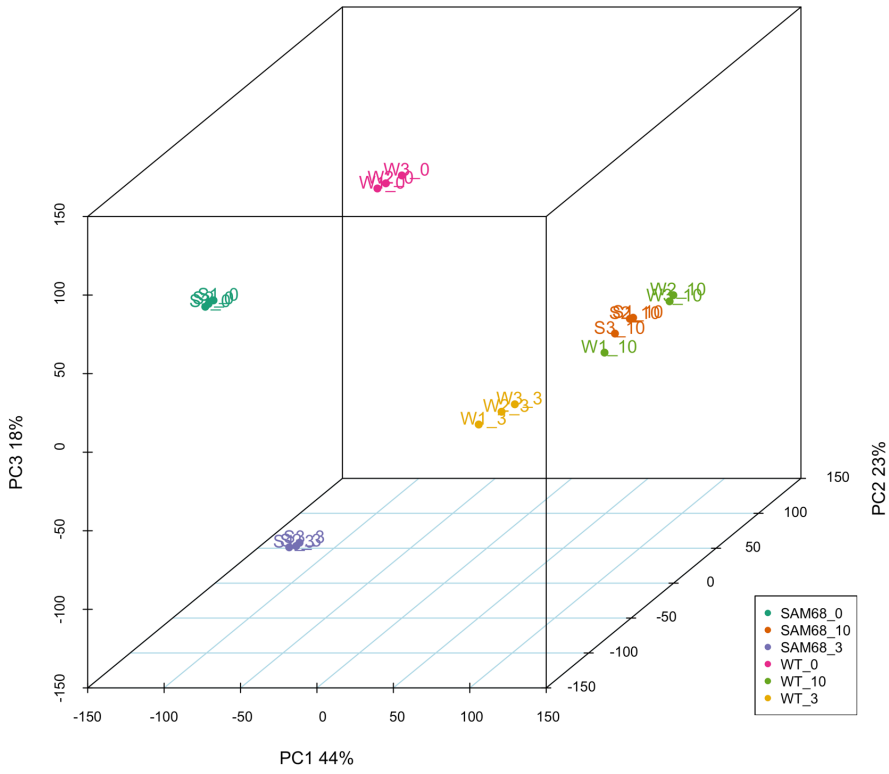


Figure 11 Principal Component Analysis of Sam68$^{-/-}$ cells. The PCA shows a high degree of reproducibility. Either WT or Sam68$^{-/-}$ biological triplicates cluster together at each time point.

These major differences in the PCA can be partially explained by the fact that upon Sam68 depletion, the number of differentially expressed genes is at least 2 folds higher when compared to the QKI$^{-/-}$ cell line, this can be due to the different levels of expression of the two proteins (see figure 10). Moreover, the depth that we reached in the RNA-seq experiments, allowed us to detect the differential expression of non-coding RNAs, which, in general, show lower levels of expression if compared to the protein coding transcripts. Indeed, as shown in table 2, we detected the differential expression of long non-coding RNAs (lncRNAs), microRNAs (miRNAs) as

well as small nucleolar RNAs (snRNAs), antisense transcripts and other non-coding RNAs. Similar to protein-coding RNAs, non-coding RNAs result to be more differentially expressed upon Sam68 depletion, suggesting a broader role of this protein at regulating the expression of these RNAs.

	Q D0	S D0	Q D3	S D3	Q D10	S D10
Protein	1246	2621	256	2900	991	3659
lncRNA	44	275	10	280	19	148
miRNA	10	152	5	169	1	14
Others	79	3159	15	3272	46	1649

Table 2 Different classes of differentially expressed RNAs. This table shows the differentially expressed RNAs divided in different classes for both STAR proteins in different time-points. Only genes with ± 1,5 fold change with an adjusted p value < 0.01 are represented.

We therefore checked in detail the effects of the depletion of each protein on the protein coding genes. We only considered the transcripts that showed a fold change > ±1.5 and an adjusted p value < 0.01. While $QKI^{-/-}$ affected 1246 coding transcripts at day 0, 76% of which are downregulated, $Sam68^{-/-}$ showed approximately 2 folds more regulation with 2621 affected transcripts of which 47% are downregulated. Whereas, at day 3, these numbers were 256 and 2900 respectively and 991 and 3659 at day 10 (figure 12 A). Interestingly, as shown in figure 12 B, the overlap between regulated targets for the two proteins in each time point was rather limited revealing distinct regulatory functions of these highly similar factors.

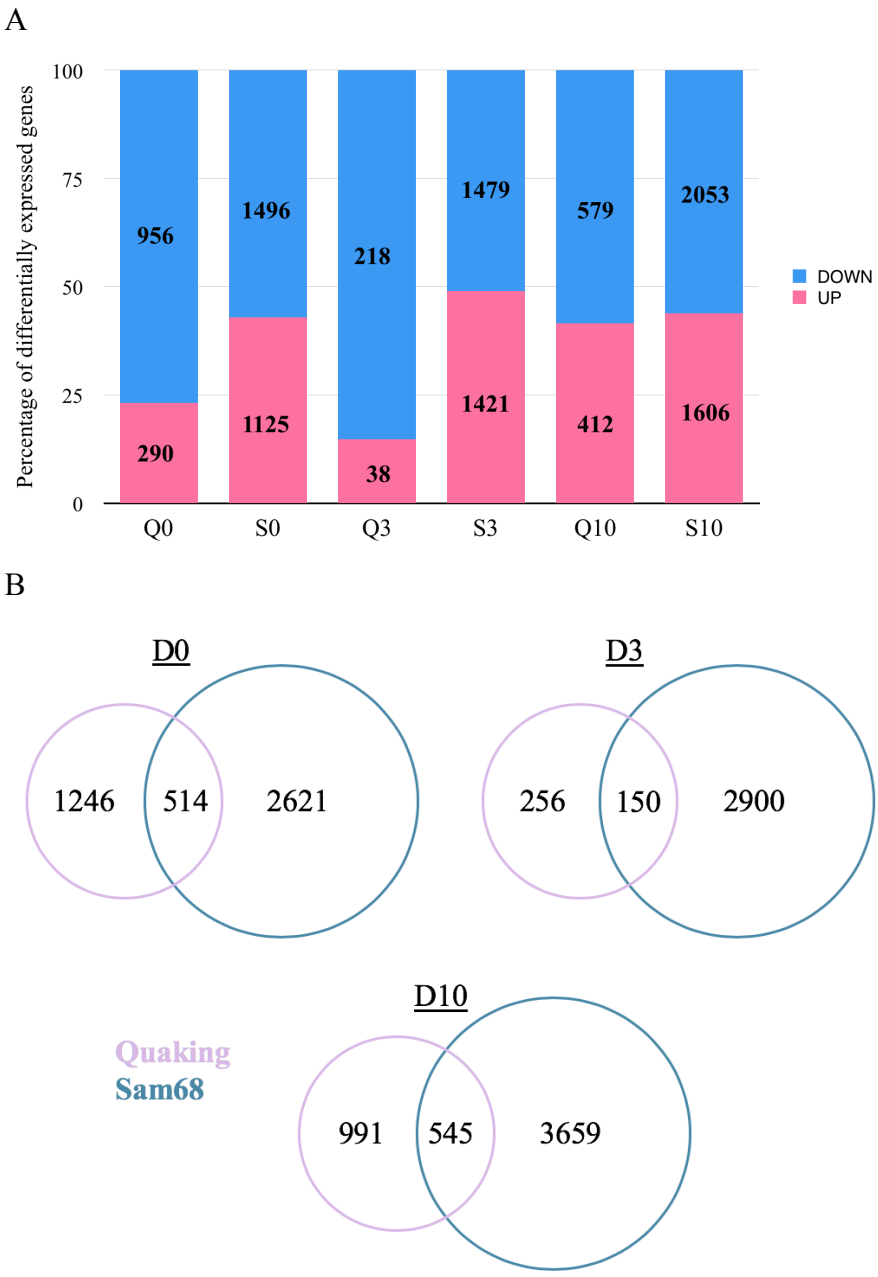
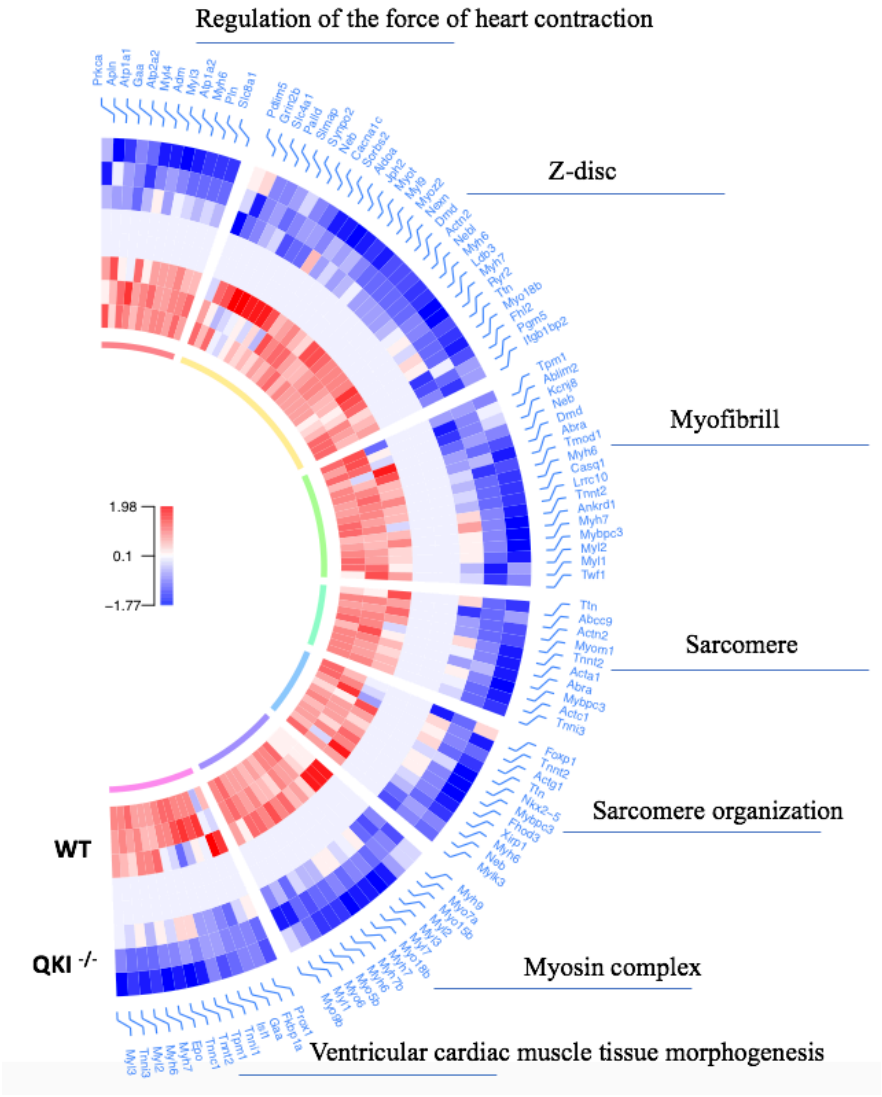


Figure 12 RNA-sequencing results. A: bar plot showing the number and the percentage of the upregulated and downregulated genes for each time point. B Venn diagrams representing the overlap of misregulated genes at each time point of both Sam68$^{-/-}$ and QKI$^{-/-}$ cells. In both cases only genes with adjusted p value <0,01 and fold change of ±1,5 are represented.

It was previously shown that regulation of Sam68 or Quaking expression affects transcripts involved in development of the nervous system or of the male germ cells[86,119]. To understand which pathways are regulated by Sam68 and Quaking proteins during very early development we performed a pre-ranked gene-set enrichment analysis (GSEA) with a selected gene set database related with mouse pathways on the differentially expressed genes at D10 of embryoid bodies differentiation. This analysis revealed that, in addition to genes involved in neurogenesis and muscle development, the most striking differences are pathways related to the cardiac differentiation, a process in which the regulatory function of these two proteins is mostly unknown to date. We selected those GO ontology gene sets related with cardiac development and we plotted the significant gene sets in a circus plot. Notably, as shown in figure 13, while the majority of the cardiac-related genes are downregulated at D10 in the $QKI^{-/-}$, they are generally upregulated in the $Sam68^{-/-}$ cell line suggesting different and perhaps opposite functions of the two STAR family members in the development of this tissue. Therefore we decided to focus our effort to understand the role of Sam68 and Quaking proteins in cardiac development.


Regulation of the force of heart contraction
Z-disc
Myofibrill
Sarcomere
Sarcomere organization
Myosin complex
Ventricular cardiac muscle tissue morphogenesis
WT
QKI -/-
1.98
0.1
−1.77

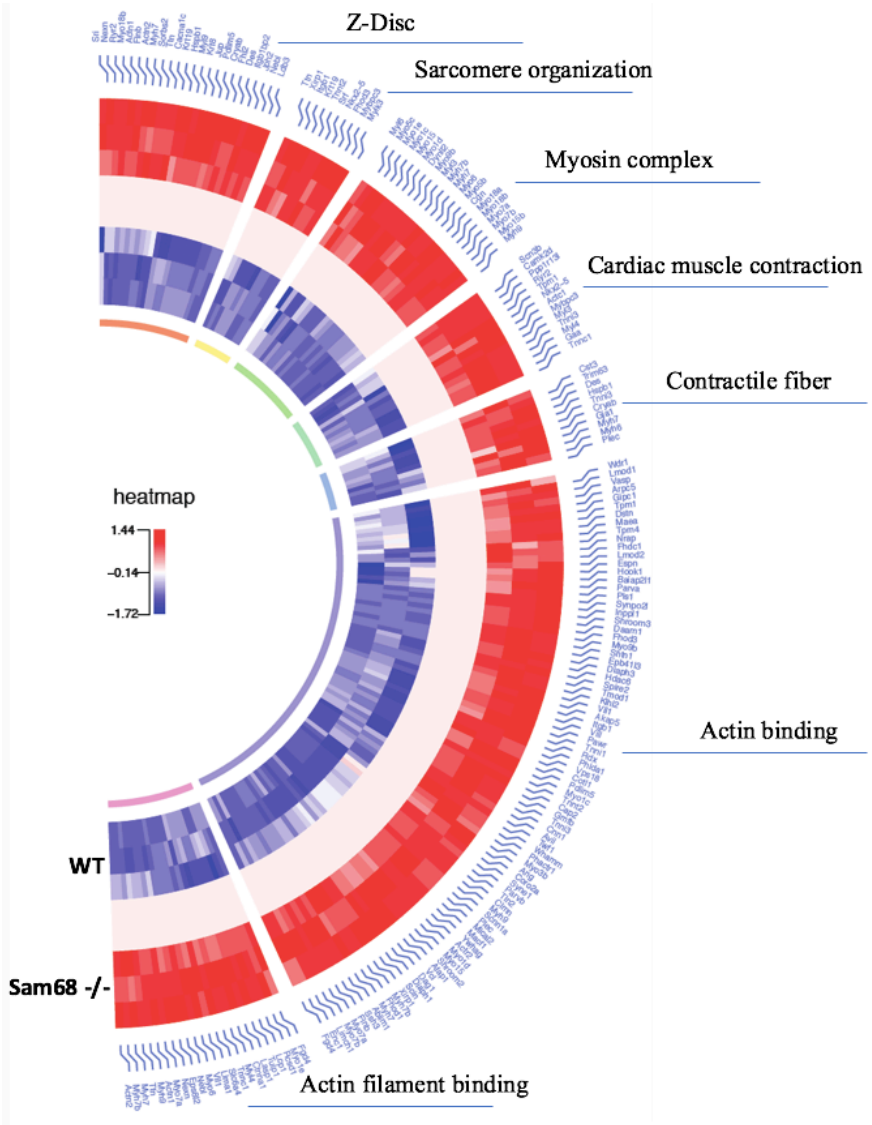


Figure 13 Changes in the transcriptome upon QKI or Sam68 depletion. Differentially expressed RNAs involved in the cardiovascular system development at D10 of EBs differentiation. RNAs belonging to this category appear downregulated for QKI$^{-/-}$ cells (upper plot) and upregulated for Sam68$^{-/-}$ (lower plot). Gene ontology classes are indicated.

4.2.2 Quaking and the regulation of cardiac related transcripts

The deregulation of cardiac related genes that we observed at D10 is most probably the consequence of an earlier alteration of major transcription factors involved in this process. For this reason, we tested and validated the RNA-sequencing results via real-time PCR. First, we checked the mRNA expression levels of the cardiogenic TF core (GATA4, GATA6 Mef2c and Nkx2-5[180]) at D3 of EB differentiation. Real-time PCR validation shows a significant and dramatic downregulation of both GATA4 and GATA6 but not of Mef2c or Nkx2-5 upon quaking depletion, perhaps due to their very low levels of expression at this timepoint.

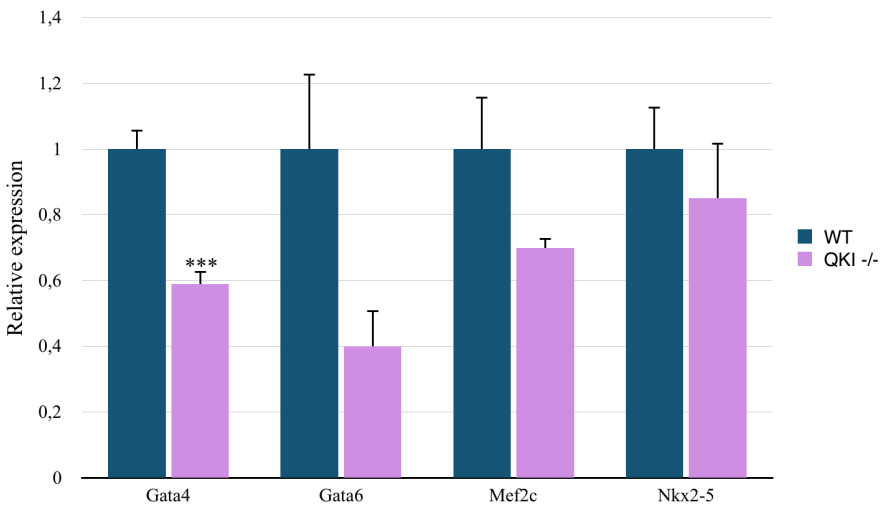


Figure 14 RNA-sequencing shows downregulation of the cardiogenic TFs. Real-time PCR of WT and QKI$^{-/-}$ mESCs at D3 of EBs differentiation shows severe and significant downregulation of GATA4 and other cardiogenic TFs.

This downregulation, could explain the considerable downregulation at D10 of EBs differentiation of functional and structural cardiac-related transcripts as shown in the graph in figure 15. More specifically, we detected a sharp downregulation of structural proteins that are part of the sarcomeric structure such as actin alpha 2 (Actn2)[181], myosin binding protein H-like (Mybphl)[182], nebulin (Nebl)[183], two different forms of heart-specific myosin heavy chain protein (Myh6 and Myh7)[184], cardiac actin (Actc1)[185], different components of the tropomyosin binding complex such as the tropomodulin (Tmod1)[157], two forms of cardiac troponin itself (Tnni3 and Tnnc1)[187,188] and a skeletal muscle troponin expressed in the cardiac tissue as well such as troponin T3 (Tnnt3)[189]. Moreover, upon depletion of QKI we observed the downregulation of two important cardiac-specific splicing regulators RBM 24[63] and RBM 20[164] and two important ion channels: the sodium voltage-gated channel alpha subunit 5 (SCN5a)[190] and the ryanodine-receptor Ryr2 that is important for the coordination of the calcium-induced calcium release, mechanism that in turn is needed to regulate the cardiomyocyte contraction. Furthermore, this receptor is involved as well in the embryonic heart development[166]. All these data, taken together, suggest a role of Quaking as a master regulator of different layers during the mESCs differentiation toward the cardiomyocyte lineage.

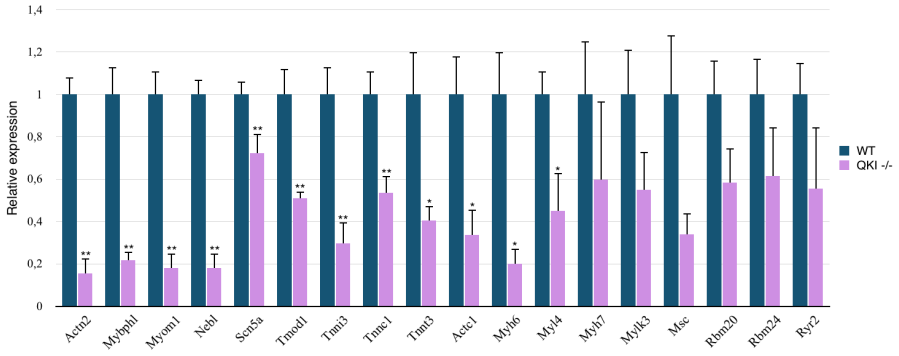


Figure 15 RNA-sequencing validation. Real-time PCR shows severe downregulation of mRNAs encoding for structural, developmental and functional cardiac specific proteins at D10 of mESC EBs differentiation.

4.2.3 Sam68 and cardiac related transcripts

We carried out the validation via real-time PCR of the RNA-sequencing results of Sam68$^{-/-}$. We were able to detect a statistically significant upregulation of the following transcription factors at D10 of EBs differentiation: GATA4, Nkx2-5 and the pace-maker cells specific TF Tbx18[193]. Conversely, the cardiac progenitor transcription factor Isl1, a LIM homeodomain transcription factor[132], is significantly downregulated upon depletion of Sam68. These data, suggest a general impairment in the establishment of the cardiomyocyte gene expression program and therefore of the cardiomyocyte identity[132,194]. As a consequence, there is a general upregulation of transcripts that encoding the cardiac functional and structural proteins, such as the myosins (Myl2 and Myl4)[195,196] and the cardiac-specific actin (Actc1), and important ion channels, such as Ryr2 and the calcium channel subunit CACNB2[197]. However, only the upregulation of troponin T2 (Tnnt2) and of Myl4 are statistically

significant upon validation (this is due to the higher heterogeneity of the Sam68 KO EBs that leads to higher inter-replicas variability compared to the Quaking KO samples). On the other hand, genes encoding for component of the cardiac extracellular matrix like the collagen type V alpha 1 chain (Col5a2)[198] or the laminin subunit alpha 1 (Lama1)[199] are significantly down and up regulated respectively in Sam68$^{-/-}$.

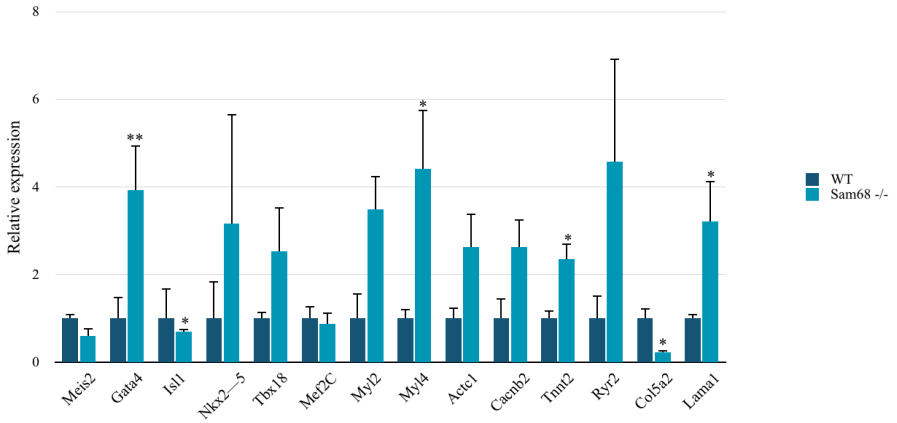


Figure 16 RNA-sequencing validation. Real-time PCR shows deregulation of mRNAs coding for structural, developmental and functional cardiac specific proteins at D10 of mESC EBs differentiation.

4.2.4 Sam68, Quaking and AS regulation of cardiac genes

Several studies have demonstrated the roles of both Sam68 and Quaking in regulating the alternative splicing of many transcripts involved in several biological processes[98,105,118,118]On the other hand, we have seen that quaking regulates the expression levels of splicing factors such as RBM20 and RBM24. This has lead us to expect that

the depletion of either protein will have a major impact on the splicing outcome. Therefore, we analyzed the RNA-seq data to detect the changes of the AS events. In general, as shown in figure 17, the majority of the AS deregulated events that happen upon depletion of either QKI or Sam68 belong the categories of the intron retention, alternative 3' or 5' splice sites and the cassette exons.

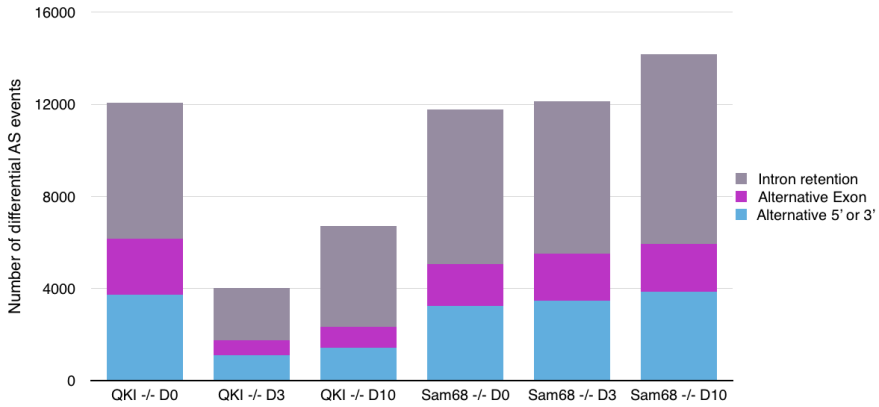


Figure 17 Global changes in AS upon Sam68 or QKI depletion. Changes in AS events at D0, D3, and D10 of EBs differentiation upon Sam68 or QKI depletion. Only intron retention, alternative exon and alternative 3' and 5' splice sites are considered.

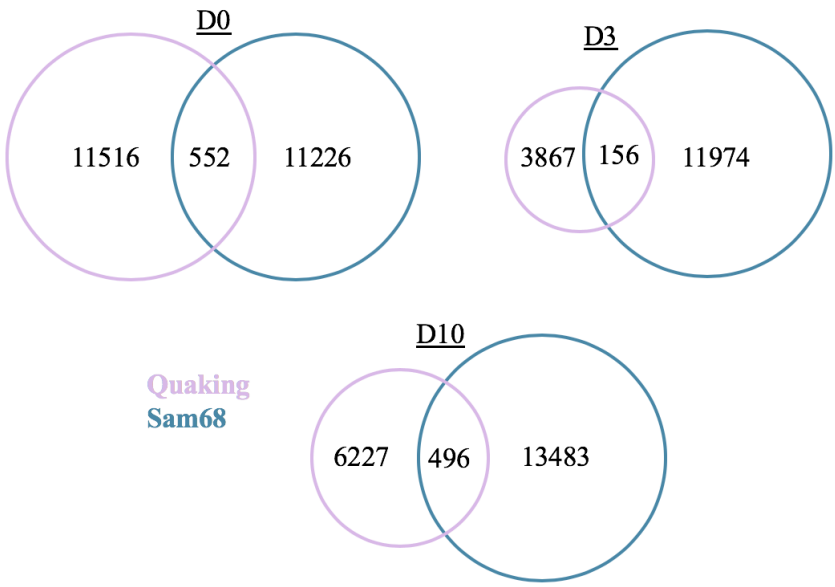


Figure 18 Venn diagrams of the alternative splicing. Venn diagrams showing the overlaps of the common altered AS events upon QKI or Sam68 KO at different time-points. Only events with an adjusted p value<0,01 and with a DPSI >25% or <-25% are considered.

Despite the large amount of deregulated alternative splicing events by each protein and at each time point, our analysis showed less than 5% overlap between the two knock-out lines, indicating divergent roles of the two proteins in the regulation of this process in mESCs. We decided to focus on the analysis of the cassette-exons AS events. for the QKI$^{-/-}$ cell lines and we detected and validated several aberrant AS events that occur on transcripts that encodes for cardiac-muscle related proteins. More specifically, we found aberrant inclusion of exons of sarcomeric structural proteins such as nebulin (Nebl), a cardiac-specific myofibrils assembly protein[200], titin and myo18a, a recently described component of the A-band of the ventricular cells[201]. Interestingly, different aberrant AS events happen as well on

the mRNA of the formin homology 2 domain containing 3 (Fhod3), an important protein for the formation of the sarcomeric actin that is important for the maintenance of the mouse heart[202,203] and whose mutations are linked to the development of hypertrophic cardiomyopathies[204,205]. Furthermore, the depletion of quaking affects as well the AS of ion channels important for the electrophysiological proprieties of the heart such as the potassium voltage-gated channel subfamily D member 3 (Kdcn3), important for the repolarization of the cardiac cells[206], the ryanodine receptor Ryr2, crucial for the calcium-induced calcium release that in turn is essential for the contractile activity of the cardiomyocytes, or the sarcolemma membrane-associated protein (Slmap) always important for the excitation-contraction coupling and that has even pathological relevance[207–209]. Last but not least, some cardiac developmentally-related transcripts are affected in their AS in QKI$^{-/-}$. Among these we validated the aberrant AS pattern for the transcript that encodes the stromal interaction molecule 1, Stim1, recently described to be involved in the developmental growth of the ventricular cardiomyocytes and in the generation of the cardiac rhythm[210–212] . Another developmentally-related aberrant AS event is the one of the adhesion G-protein coupled receptor 126, Adgrg6, thought to be involved during the first phases of the heart development and in the trabeculation of the organ[213,214].

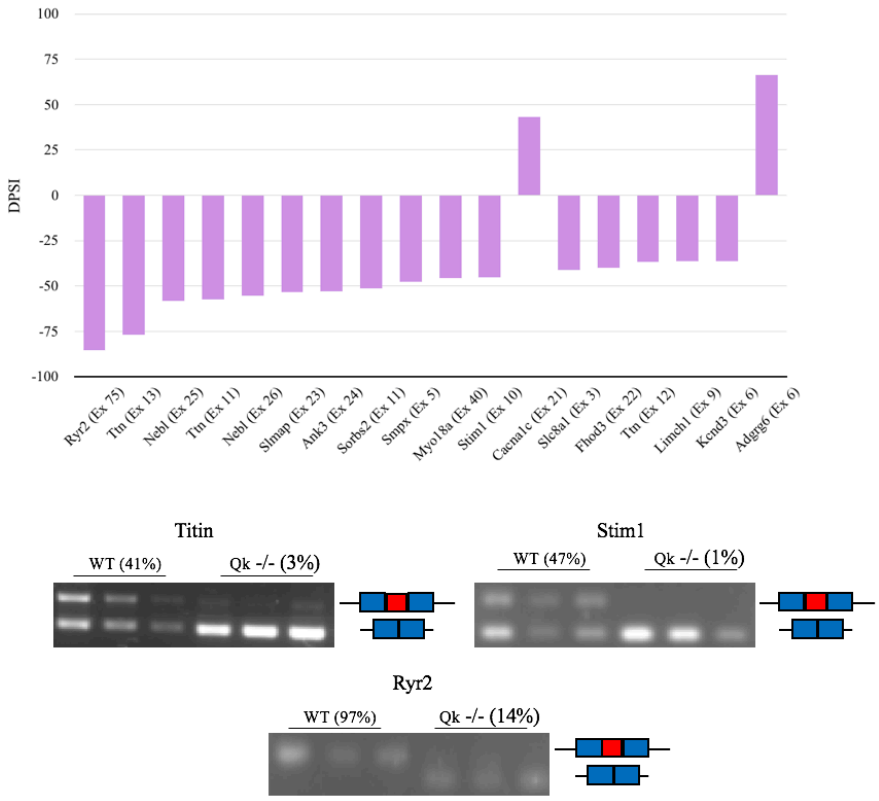


Figure 19 AS changes in QKI$^{-/-}$ EBs. The bar plot shows the delta percentage of splice inclusion (DPSI) of cardiac-related transcripts at D10 of EB differentiation in QKI$^{-/-}$ cells. A value inferior to 0 means less inclusion. The affected exons are indicated by their numbers. In the bottom part of the figure some validations performed with semiquantitative-PCR.

To sum up, the depletion of QKI in differentiating embryonic bodies disrupts the correct AS scenario of cardiac-related transcripts that encodes developmentally-related proteins as well as for structural and functional proteins important for the correct functioning of the cardiomyocytes and, consequently, of the whole organ. With respect to Sam68$^{-/-}$ cells, at D10 of EBs differentiation we also observed altered splicing of cardiac-related genes. In detail, as shown in figure 20, upon Sam68$^{-/-}$ there is more inclusion of three exons of the gene

the calcium channel, voltage dependent, L type alpha 1C subunit (CACNA1C) transcript. Those exons are normally skipped in the WT condition. Conversely, other three exons of the same transcript are more skipped in Sam68$^{-/-}$ cells demonstrating that Sam68 is a general regulator of the splicing of this important gene. All the exons whose AS is misregulated encodes for the transmembrane domain of this calcium channel subunit[215]. More in detail, this subunit is of particular interest for both the electrophysiological properties of the cardiomyocytes and for its involvement in the more common cardiac pathologies. Specifically, this subunit is responsible for the inward current of Ca^{2+} that leads to the depolarization of the cardiomyocytes that in turn leads to the calcium-induced calcium release that ultimately is responsible for the sarcomere contraction. Beside this, it has as well a clinical relevance being associated to the long and short QT syndromes, two of the most common causes of sudden death[216–219] . It is important to point out that exon 21 of this transcript is regulated by both Sam68 and Quaking. In fact, depletion of either QKI or Sam68 leads to more inclusion of this exon in the final transcript. However, Sam68 and Quaking show antagonistic effect on exon 6 of the Adgrg6 transcript, where Sam68 seems to induce more inclusion of this exon contrary to quaking that leads to more skipping. Another important gene whose AS is altered in Sam68$^{-/-}$ is the myosin binding protein C cardiac (MYBPC3). This gene encodes a component of the sarcomere where it resides at the level of the A band. It is important for the regulation of the sarcomere shortening and its stabilization. Moreover, similarly to CACNA1C, mutations in this gene are associated to the most common cardiomyopathies[220,221].

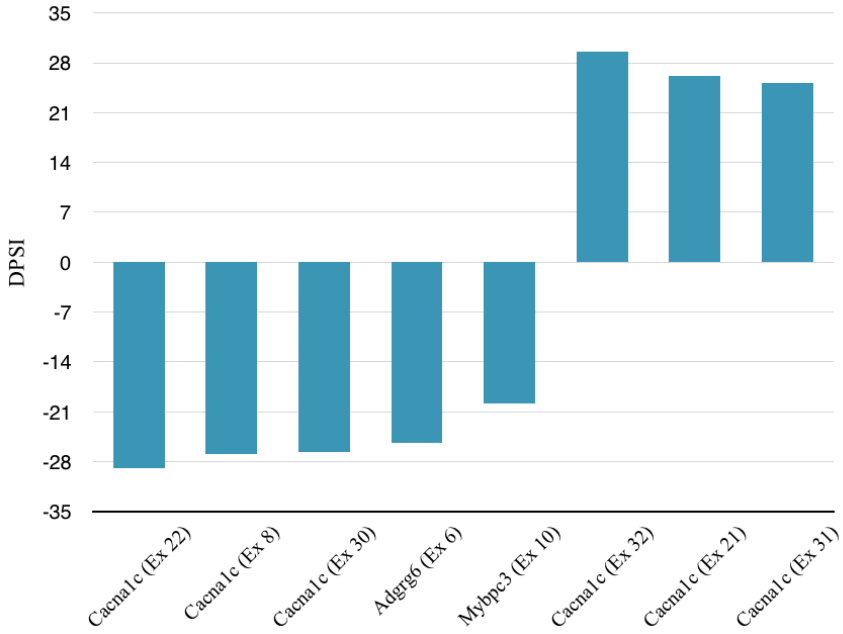


Figure 20 AS changes in Sam68$^{-/-}$ EBs. The bar plot shows the delta percentage of splice inclusion (DPSI) of cardiac-related transcripts at D10 of EB differentiation in Sam68$^{-/-}$ cells compared to WT cells. A value inferior to 0 means less inclusion. Affected exons are indicated by their numbers.

To summarize this part, both Sam68 and Quaking regulate the AS of cardiac-related transcripts. It is important to notice that the absence of Sam68 severely affects the AS of several cassette-exon of an important regulator of the calcium currents and therefore of the cardiomyocytes contraction. All these data, taken together, demonstrate that both Sam68 and QKI are important regulators of AS events that occur on transcripts coding for important structural, functional and developmentally-related cardiac genes.

4.2.5 Sam68 and the translation of cardiac-related transcripts

Given the known role of Sam68 as translation regulator during the differentiation of the male germ cells[86], we decided to check whether this role is conserved in other cell differentiation models. Therefore, we performed ribosome profiling on both WT and Sam68$^{-/-}$ EBs at day 0 and day 10 of differentiation. The ribosome profiling is a powerful technique that provides a snapshot of the actively translated transcripts in a cell. The experiment was performed in triplicates for each time point and condition. After library preparation and deep sequencing, the reads were processed and aligned to the mouse transcriptome. The results are shown in table 3.

D0	W 1st	W 2nd	W 3rd	Q 1st	Q 2nd	Q 3rd	S 1st	S 2nd	S 3rd
total reads	51.751.215	48.335.477	88.017.098	27.530.280	8.921.129	19.814.970	31.336.363	56.918.783	47.049.664
trimmed	39.566.044	33.242.569	65.211.232	19.200.180	7.432.807	14.946.690	21.164.077	35.024.567	14.994.943
uniquely mapped	9.789.905	10.429.460	18.262.100	5.826.386	2.498.599	4.535.399	6.137.155	10.976.210	8.303.106
reads in the CDS	6.897.461	6.749.066	13.841.191	3.622.572	1.691.537	3.176.711	5.039.971	8.686.689	6.053.143
reads in the 5' UTR	215.685	214.072	393.008	102.234	47.027	87.738	160.167	304.803	229.948
reads in the 3' UTR	123.327	131.786	254.515	61.386	31.143	59.635	113.195	202.043	124.940

D10	W 1st	W 2nd	W 3rd	Q 1st	Q 2nd	Q 3rd	S 1st	S 2nd	S 3rd
total reads	57.497.710	44.835.313	37.544.622	23.428.712	39.112.459	24.578.708	58.235.812	62.686.865	39.507.908
trimmed	41.038.476	37.908.362	10.347.433	8.591.659	17.848.490	14.254.139	40.414.873	35.024.567	14.994.943
uniquely mapped	8.698.012	6.702.534	2.663.602	1.780.820	4.161.156	4.022.730	10.404.552	8.317.715	4.060.286
reads in the CDS	6.325.529	4.467.914	648.457	1.024.074	2.562.358	2.965.745	7.485.046	3.189.227	1.251.987
reads in the 5' UTR	204.030	133.367	20.903	36.191	115.180	111.525	259.370	79.884	42.665
reads in the 3' UTR	127.052	83.807	6.476	18.272	67.527	59.799	159.613	37.879	18.915

Table 3 Summary of the Ribo-Seq results. The reads obtained from the deep sequencing for each replicate and time point for WT, QKI$^{-/-}$ and Sam68$^{-/-}$ are indicated before and after being processed and aligned to the mouse transcriptome.

First, we observed that more than 80% of the uniquely aligned reads fall in CDS and less than 5% of the reads map to 5' and 3' UTRs regions indicating the high quality of the data. We then calculated the Pearson correlation to verify the reproducibility of the different replicas. As shown by the values listed in the table 4, the R^2 values are higher than 0.9 for all the pairwise comparisons attesting for the

high reproducibility among the different replicas of the same sample. Of note, the correlation is lower at D10 compared to D0 perhaps reflecting the lower homogeneity of the differentiated samples compared to when then cells are still fully pluripotent.

	W0.	W0…
W0.	-	0,98
W0..	0,976	0,979

	Q0.	Q0..
Q0.	-	0,962
Q0..	0,96	0,967

	S0.	S0…
S0.	-	0,975
S0..	0,972	0,966

	W10.	W10…
W10.	-	0,911
W10..	0,973	0,93

	Q10.	Q10…
Q10.	-	0,965
Q10..	0,937	0,935

	S10.	S10…
S10.	-	0,956
S10..	0,967	0,958

Table 4 Pearson correlation. The tables show the values of the Pearson correlations among the different replicas. The values show a high degree of reproducibility.

Our RNA-seq results revealed a new role of Sam68 and QKI in the heart development, we therefore focused our Ribosome profiling analysis on transcripts involved in this pathway. Upon Sam68 depletion, we observed a general deregulation of the translation efficiency of several transcripts that are involved in the cardiac development or that sustain the cardiomyocytes functionality. Some of these transcripts are more efficiently translated while others are less efficiently translated in Sam68$^{-/-}$ embryonic bodies compared to the WT ones. We could not find the same trend for QKI, this could mean that the regulation of heart development via QKI occurs through other molecular mechanisms. Among the transcripts that are more efficiently translated in Sam68$^{-/-}$ during EB differentiation we mainly found TFs, ion channels and structural proteins involved in the organogenesis of the heart. For instance, we detected a sharp increase in the translation of the channel Kcnj8, a potassium channel

expressed in both ventricular and atrial cardiomyocytes[222] , another increase in translation of Nkx2-5 or of Osr1, transcription factors involved in the cardiac morphogenesis[223] and the transcripts that encode the structural protein myosin heavy chain, cardiac type 6 (Myh6)[184]. To sum up, for these specific transcripts, Sam68 could act as a repressor of translation. Conversely, other mRNAs are less efficiently translated upon Sam68 depletion, suggesting that this STAR family member acts, directly or indirectly, as an enhancer of their translation. Among these we found Sox17, a known cardiac-mesoderm specifier[224]; the Indian hedgehog mediator, that, along with Sonic hedgehog, is indispensable for the embryonic development of the heart[225], and components of the cardiac extracellular matrix, such as the glycoprotein Lama1. Last but not least, among the transcripts that are less translated upon Sam68 depletion in differentiating EBs we found GATA4, TF crucial for the embryonic development of the heart and that together with Mef2c, Tbx5 and Nkx2-5 constitute the cardiogenic TFs core. We show the results of the Ribo-Seq experiment on GATA4 transcript in figure 21.

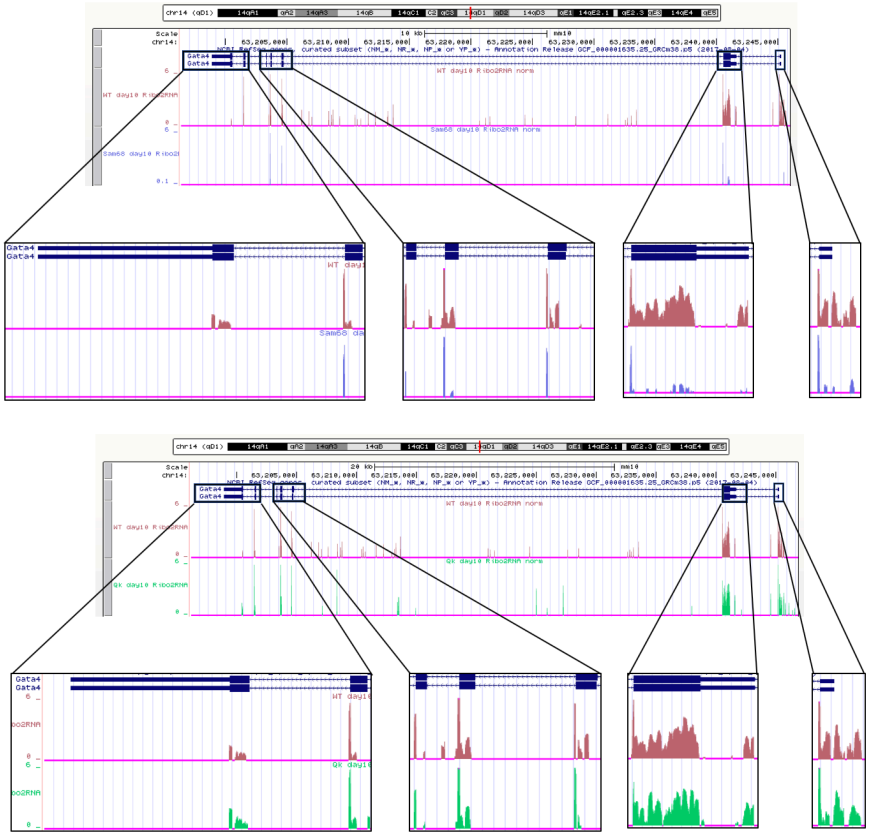


Figure 21 Genome browser captures of Ribo-seq experiment for both Sam68 (upper panel, in blue) and QKI (lower panel, green) on GATA4 mRNA. The amplifications of the CDS parts clearly show a decrease of ribosome reads upon depletion of Sam68 but not upon depletion of QKI. The 5'UTR of the gene is at the right part of the panels and the WT track is shown in brown.

To summarize, in differentiating mESCs, Sam68 exerts translational control on cardiac-associated transcripts behaving either as a translational repressor or enhancer (Figure 22).

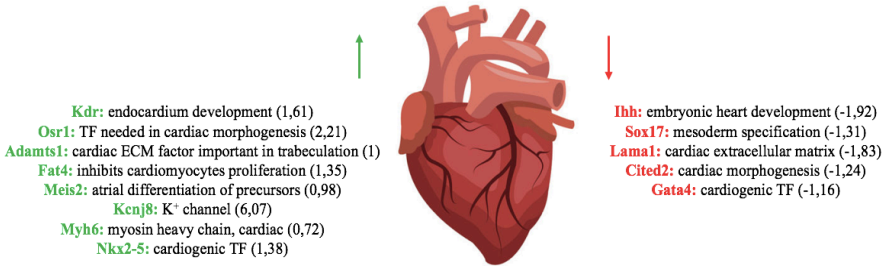


Figure 22 Ribosome profiling. Cardiac-related transcripts whose translation is controlled by Sam68. On the left side, transcripts that are more actively translated upon Sam68 KO. On the right, transcripts that are less actively translated upon Sam68 KO. In parenthesis the $\log_2$ fold change is represented.

4.3 FUNCTIONAL CHARACTERIZATION

4.3.1 Sam68, Quaking and cardiomyocytes differentiation

Our omics data shed the light on the involvement of both Sam68 and Quaking in the cardiomyocyte's differentiation of the mESCs. Therefore, we decided to set up an EB-derived cardiomyocyte differentiation assay. Briefly, in this EB protocol variant, at D3 of hanging drop differentiation, each EB is singularly placed in a well of a 96 wells plate and is allowed to adhere to the dish surface. This in turn, favors the differentiation of the mESCs specifically towards the cardiomyocyte lineage. More specifically, after 8 days of replating the EBs and upon the adhesion of the EBs to the pre-coated gelatin dishes, they give rise to beating foci, *bona fide* cardiomyocytes. This approach has allowed us to assess the physiological output in the absence of any of these two STAR members in this mammalian model of differentiation. Indeed, depletion of either Sam68 or QKI led to a dramatic and statistically significant decrease in the number of beating *foci* compared to WT EBs as presented in figure 23. Of note, cells depleted of Quaking show higher degree of impairment at generating contractile *foci* and compared to Sam68$^{-/-}$. This experiment allowed us to conclude that both proteins are involved in cardiomyocyte differentiation of the mESCs.

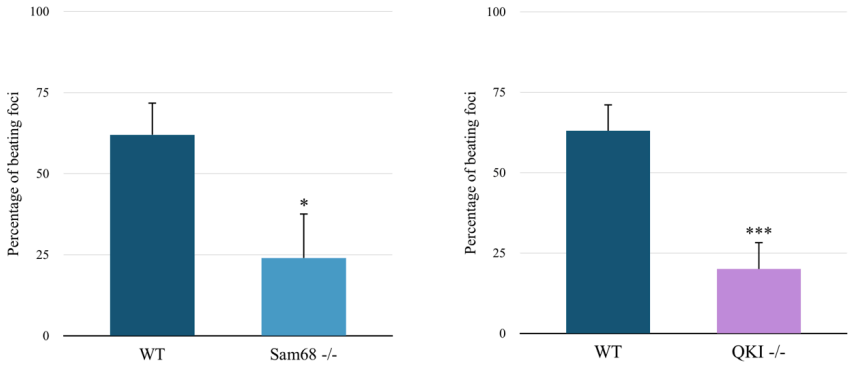


Figure 23 Cardiomyocyte differentiation of mESCs deprived of either Sam68 or QKI. The bar plots represent the percentage of beating *foci* for WT and either Sam68$^{-/-}$ or QKI$^{-/-}$ cells. Each plot is the average of three independent experiments.

4.3.2 Sam68 and heart beating

Surprisingly, while performing this differentiation assay we noticed that not only Sam68$^{-/-}$ mESCs give rise to less beating *foci* when compared to the WT counterparts, but the beating activity shows some abnormalities. Thus, we decided to assess whether the absence of Sam68 is linked to an abnormal beating activity. To achieve this, we counted and scored the *foci* beating activity per minute. As shown in figure 24, Sam68$^{-/-}$ cells have a significantly higher beating rate compared to the WT. It is known that among the components that regulate the beating activity of a cardiomyocyte there are the ion channels and the consequent ion currents that they generate. As previously described, the depletion of Sam68 in differentiating EBs leads to the dramatic upregulation of both the ryanodine receptor 2 Ryr2 and the calcium voltage gated channel subunit CACNB2, beside leading to a complete aberrant AS of the other calcium channel subunit CACNA1C. These molecular defects might be the reason

underlying the beating impairment, nonetheless more investigations are needed in order to define the link between Sam68 and the cardiomyocyte's ion channels.

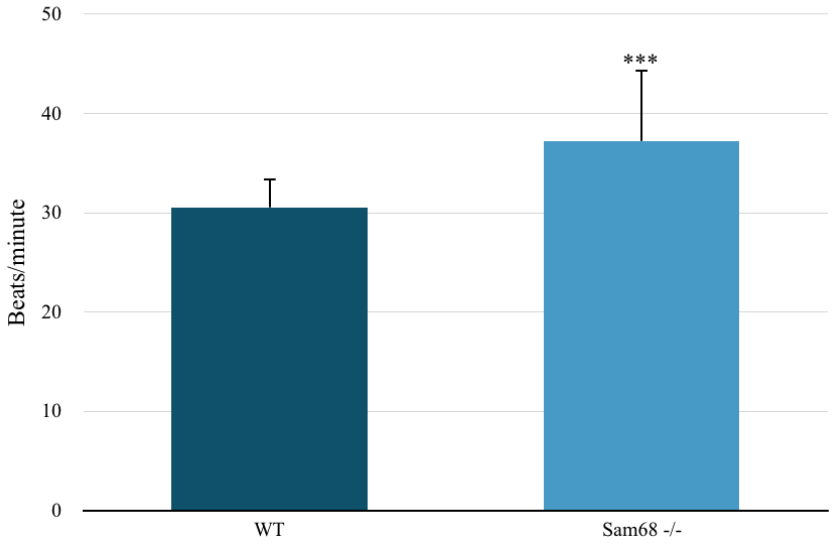


Figure 24 Beating activity. In this bar plot the beating activity is shown as beat per minute. Each bar is the average of three independent replicas.

4.3.3 Sam68 and GATA4

Our data show that both Sam68 and Quaking are involved in cardiac development at very early stages and, to achieve this, they act on several layers: regulation of master transcription factors, alternative splicing factors and alternative splicing events. To better understand the molecular mechanisms ruling this process via Sam68 we focused our study on GATA4 (that showed opposite trends in transcription and translation levels) and checked whether its deregulation can be due to a direct binding of Sam68. We first performed a bioinformatic

analysis using *cat*RAPID, a tool developed by our laboratory. *cat*RAPID predicts the interactions between pairs of proteins and RNAs[226].

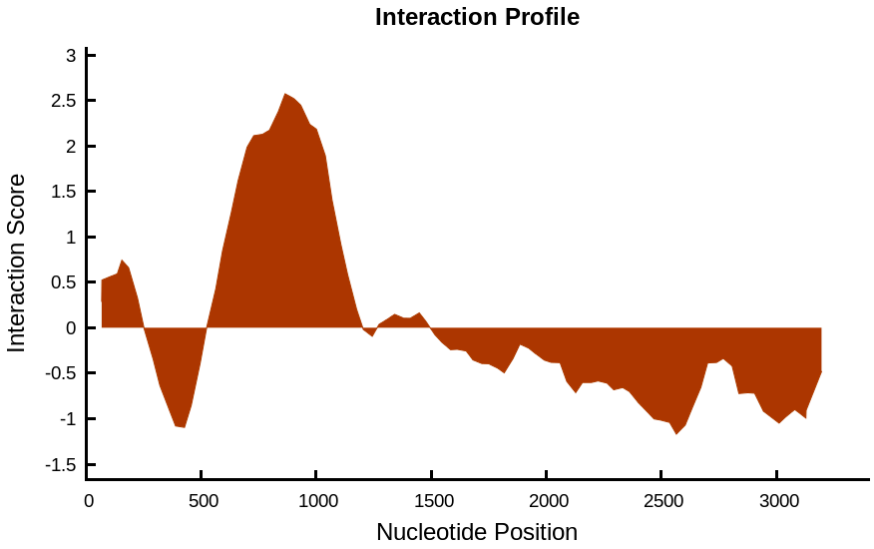


Figure 25 *cat*RAPID interaction profile. Profile of prediction of the interaction between GATA4 mRNA and Sam68 protein.

*cat*RAPID predicted a high interaction score between nucleotides 500 and 1000 of the GATA4 mRNA and Sam68 protein (figure 25). To further confirm this prediction, we experimentally tested the interaction propensity by performing a RNA immunoprecipitation experiment in differentiating EBs. We performed RIP followed by both semi-quantitative PCR and real-time PCR. We designed 3 different pairs of primers to detect GATA4 mRNA. The results of the RIP are represented in figure 26. The specific enrichment of the GATA4 mRNA pulled down with Sam68 is validated by either pair of primers in both semi quantitative and real time PCR techniques. compared to the IgG negative control and the unbound RPO mRNA.

To our surprise and unlike Sam68, Quaking does not bind GATA4 mRNA as shown in both the semi-quantitative and real-time PCR experiments. These results confirm both the direct and indirect regulation of GATA4 by Sam68 and QKI respectively.

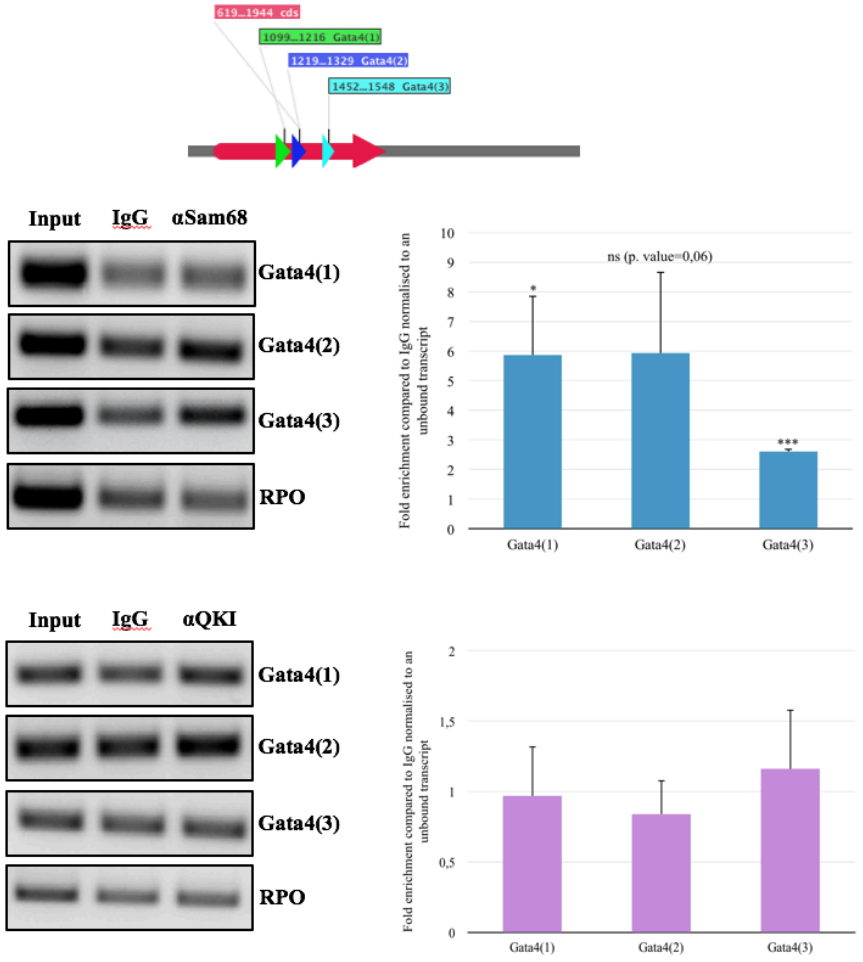


Figure 26 RIP of Sam68 and QKI to detect GATA4 binding. The cartoon shows the sites of primer annealing on the sequence of GATA4 mRNA. The left panels: semi-quantitative PCR for both Sam68 (upper panel) and QKI (lower panel). The enrichment of the GATA4 transcript in the samples precipitated with an antibody directed against Sam68 compared to either the negative control (IgG) and the house-keeping transcript RPO is validated by the three primer pairs. Conversely, GATA4 did not show enrichment when precipitated with QKI. The right panels: real-time PCR showing the enrichment of the GATA4 transcript normalized to a non-bound RNA (RPO) in Sam68.

4.3.4 Functional significance of the interaction between Sam68 and GATA4 mRNA

Our experiments showed that Sam68 binds to GATA4 mRNA and that upon depletion of Sam68, the is an increase of the mRNA levels of GATA4. One would expect this increase to be accompanied by an increase of the GATA4 protein levels, however we observed less ribosomes associated to GATA4 mRNA in the absence of Sam68. So, we asked the question: what happens to GATA4 protein levels? First, to rule out the effect of Sam68 on the stability of the GATA4 mRNA we treated the cells with actinomycin D that blocks the transcription and measured the decay of the GATA4 mRNA at different time points during 12 hours. We could not detect any difference of GATA4 mRNA levels between Sam68$^{-/-}$ and WT cells (data not shown). We therefore checked the GATA4 translation efficiency in the absence of Sam68. To achieve that, we performed both western blot and immunofluorescence in EB at D10. As shown in figure 27, both the WB and the IF show significant downregulation of the GATA4 protein in cells depleted of Sam68 expression. More in detail, by IF we detected that the expression of GATA4 protein is almost abolished in Sam68$^{-/-}$ cells compared to the WT cells. This result indicates that Sam68 directly regulates the translation of GATA4 mRNA. All these data, along with previous evidences, demonstrate a role of Sam68 in the regulation of the differentiation towards the cardiac lineage of mESCs probably mainly due, but not only, to the regulatory action that this member of the STAR family has on GATA4, a master TF for the cardiomyocyte development.

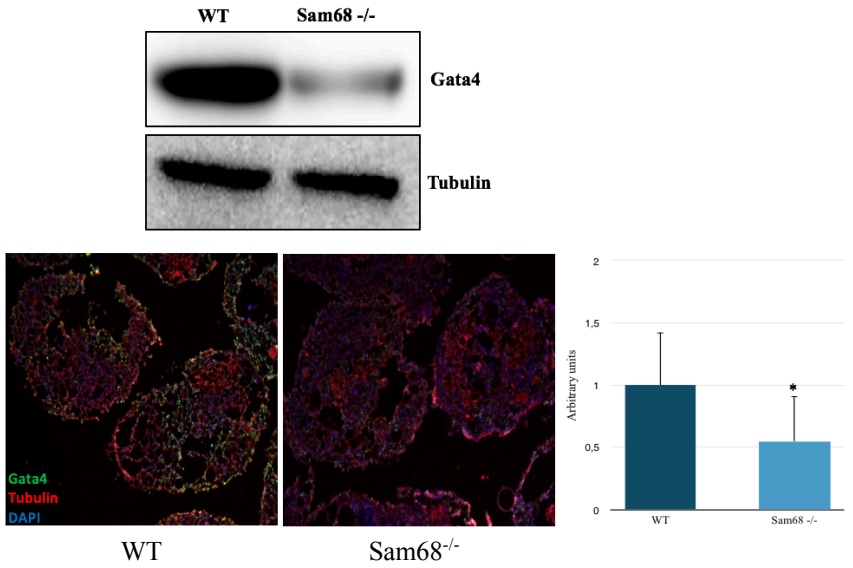


Figure 27 GATA4 protein expression. Upper panel: WB of GATA4 in both WT and Sam68$^{-/-}$ at D10 of EBs differentiation. Lower panel: on the left IF of GATA4 and on the right its quantification.

5 Review: RNA-centric approaches to study RNA-Protein Interaction in vitro and in silico

The RNA binding proteins exert crucial functions and take part to all regulative steps that bridge the transcription to the protein expression. To date, it is known that more than 1500 genes codify for RBPs and their misregualtion can even have pathological outcomes. Given that, it is not surprising the huge effort that the researches are committed to in order to understand which are the intrinsic and extrinsic determinants that determine the binding of an RBP to the RNA. In this review we critically analyze the main *in vitro* and *in silico* techniques used in order to study the determinants of these important interactions. This review is focused particularly on the intrinsic factors such as the RNA secondary structure, and on the intrinsic determinants that can have an impact on the binding such as expression, post-translational and post-transcriptional modifications.

Dasti A, Cid-Samper F, Bechara E, Tartaglia GG. RNA-centric approaches to study RNA-protein interactions in vitro and in silico. Methods (San Diego, Calif). 2020;178:11–8. DOI: 10.1016/j.ymeth.2019.09.011

6 Whsc1 links pluripotency exit with mesendoderm specification

The mechanism that controls the loss of pluripotency as well as those that regulate the differentiation of the embryonic stem cell is a topic of immense investigation. In order to differentiate, a stem cell has to disrupt the transcriptional factor circuitry that allows the sustainment of the pluripotency and activate the expression of lineage instructive transcription factors that ultimately will instruct the cell on the differentiative path it has to undertake. Beside the transcription factors and the non-coding RNAs, the chromatin related factors (CFRs) play a crucial role in this pluripotency disruption and cell fate decisions. Indeed, the CRFs are factors that, by adding chemical modifications on the histone or on nucleotides, can alter the chromatin state. This results in a more or less chromatin compaction that allows or not the gene expression. In fact, when an ES differentiate its chromatin state dramatically changes. Along the differentiation there is an increase in chromatin compaction and consequent silencing of pluripotency associated genes. In this research article, we describe how a chromatin modifier such as Whsc1, more specifically a methyltransferase, links pluripotency exit to mesendoderm specification. Indeed, our work demonstrates that Whsc1, thanks to its methyltransferase independent activity, is able to promote both the exit from pluripotency and the activation of the mesendoderm regulators.

Tian TV, Di Stefano B, Stik G, Vila-Casadesús M, Sardina JL, Vidal E, et al. Whsc1 links pluripotency exit with mesendoderm specification. Nature cell biology. 2019;21(7):824–34. DOI: 10.1038/s41556-019-0342-1

7 DISCUSSION

7.1 RNA Binding Proteins: central players in the development process?

Recently, the most recurrent questions in the stem cell field are: what is fine tuning the balance between pluripotency and differentiation? what are the external and internal cues that are responsible for the decision of a differentiation path that a stem cell undertakes towards a specific lineage? Extensive studies on transcription factors (TFs) and chromatin remodelers have been carried out during the last decades in order to understand this switch[227,228].More recently, the emerging roles of RNA binding proteins in this process started to be revealed[229,230] .For instance, Mbnl1 impedes the inclusion of exon 16b of the TF FOXP1 thus not allowing the expression of pluripotency genes[229]. Generally speaking, RNA binding proteins are involved in the regulation of the RNA metabolism, therefore, they have a crucial role in the regulation of gene expression that in turn is responsible of a specific cell identity[159,231]. They act by bridging gene expression with protein expression, and because of the importance of the gene and protein expression in the biology of a cell, it is clear that RBPs can have a crucial role in the stem cell biology. RBPs are classified according to their structural domains[51]. A specific class of RBPs is represented by the signal transduction and activation of RNA (STAR) protein family that consists of 5 members that are all evolutionary conserved and share a common structural domain called

STAR domain[74]. To date, it is known that these proteins play various important roles in several developmental processes across different species, indicating that they might be important during the early and late stages of the mammalian development[86,105,179]. Because of all the aforementioned reasons, we decided to investigate whether two members of the STAR protein family, namely Sam68 and Quaking, have a role in both pluripotency and differentiation of the mouse embryonic stem cells.

7.2 Sam68, QKI and pluripotency: any link?

In order to understand the link between either QKI or Sam68 and pluripotency, we made use of both basic and cutting-edge molecular and cellular biology techniques. More specifically, first we generated *bona fide* Sam68 and Quaking Knock-out lines through the CRISPR-Cas9 genome editing technology. Consequently, we decided to deeply characterize the two cell lines for their features in both pluripotency conditions and during mESCs differentiation. First, we observed that the KO lines do have the same visible features of the wild type counterparts (dome shape and net borders). Second, we confirmed that the depletion of Sam68 does not affect the expression levels of QKI and *vice versa* suggesting that the two proteins exert different functions and lack redundancy in this cellular system. We then tested whether the KO of these STAR family members affect the pluripotency features of the mESCs. More in detail, we assessed whether the KO of either Sam68 or QKI would affect both the

pluripotency itself, described as capacity of giving rise to terminally differentiated cells derived from all the 3 germ layers; and its disruption, mechanism needed for a cell in order to undertake any differentiation path[26,40]. As far as the pluripotency is concerned, we demonstrated that the absence of either Sam68 or QKI does not affect the pluripotency of the mESCs. Indeed, the teratomas extracted from the nude mice are composed of terminally differentiated cells derived from all the three germ layers. Similarly, we have found that both Sam68 and Quaking do not participate in the disruption of the core pluripotency maintenance TFs network. **All these data, taken together, show that both Sam68 and QKI do not alter the maintenance or the disruption of the pluripotency in mESCs** unlike other RBPs, such as Thoc5 whose downregulation is necessary for the disruption of the pluripotency transcription factor circuitry[229].

7.3 Sam68, QKI and self-renewal: The hows & whys.

Given that both proteins have been reported to regulate cell proliferation[97,111] we assessed their functions in the proliferation of mESCs. We demonstrated that both Sam68$^{-/-}$ and QKI$^{-/-}$ cell lines show a decreased rate of proliferation compared to the WT indicating that the two proteins are involved in the regulation of the mESCs proliferation. This finding is in line with previous discoveries that show a prolonged G2-M phase in Sam68 depleted chicken fibrobalsts[177]. Of note, QKI$^{-/-}$ cells are even more impaired compared to the Sam68$^{-/-}$ suggesting that Quaking might have a more important

role in the regulation of cell cycle in these cells. Indeed, it was shown that depletion of QKI affects the proliferation of colon cancer cells by regulation the expression of cyclin D1[232]. Furthermore, given that proliferation and self-renewal are somehow linked phenomena, we tested whether the defect in proliferation had an ultimate effect on the self-renewal capacity of the cells. Moreover, it has been already shown that the absence of Sam68 in neuronal progenitor cells is linked to a reduced self-renewal of these undifferentiated cells[103]. Our results show that both QKI$^{-/-}$ and Sam68$^{-/-}$ mESCs have a reduced self-renewal capacity generating smaller and fewer colonies compared to the WT. This corroborates the role of these proteins in the maintenance of this cellular feature. **Collectively, these data, demonstrate that both Quaking and Sam68 are important for the control of the self-renewal and the regulation of the proliferation of the mESCs.**

7.4 What are the molecular mechanism hidden behind these features?

Given the known role of both family members in development, especially of the nervous system[98,98,122], we wanted to verify whether they have a role during the unbiased differentiation of mESCs towards others cell lineages. Our RNA-sequencing at early and late stages of embryoid bodies (EBs) differentiation in both KO lines showed deregulation of genes involved in the differentiation of the nervous tissue, as previously known. Strikingly, it revealed novel

pathways that were previously undescribed. The majority of the statically significant differentially expressed genes are either involved in the differentiation of the cardiovascular system, more specifically of the heart, or codes for important functional and structural proteins of this tissue. More surprisingly, the majority of the differentially expressed genes that belong to this category show an opposite trend for the two KO lines, while downregulated in QKI$^{-/-}$ they showed an upregulation in Sam68$^{-/-}$. **These data, suggest not only that both proteins have a role in the development of the heart but that probably they exert completely different, or perhaps opposite, functions in this biological context.** QKI acts early on the core TFs network already at D3 of EBs differentiation. Although at this stage of differentiation these TFs are lowly expressed, we observed a dramatic downregulation of GATA4 and GATA6[233]. This downregulation expanded and got amplified at D10 where many genes that code for structural, functional and cardiac-specific splicing regulators (Rbm24 and Rbm20) showed a dramatic decrease in their expression levels. All these defects in the early stages of differentiation, might be a further reason of the lethality of the QKI$^{-/-}$ embryos. Conversely, in the case of Sam68$^{-/-}$ cells, while there was an absence of deregulation of genes involved in the cardiovascular system development at D3 of EBs differentiation, we observed a major defect at D10 of EBs differentiation. Indeed, there is a dramatic upregulation of the mRNA levels of the transcription factor GATA4 as well as of two others important cardiac TFs, Nkx2-5 and Tbx18, known to be crucial for the formation of the sinoatrial node the specialized myocardial structure that initiates the electric

impulse that ultimately stimulates the contraction of the organ[193]. To sum up, knocking out Sam68 or QKI leads to a misregulation of a circuit of genes involved in the development of the cardiovascular system. This phenotype might also explain the high perinatal mortality of Sam68 KO mice although this has to be proven.

7.5 Are Sam68 and QKI master regulators orchestrating the heart development?

Given their known role at regulating alternative splicing, particularly during the development of the nervous system[98,122,179], and given that QKI affects the level of expression of other splicing regulators, we decided to analyze the global changes in the AS landscape of EBs differentiation of mESCs upon depletion of either protein. Surprisingly, the majority of the altered events in the KO cells falls in the category of the intron retention. This might be not surprising given the recently discovered role of Sam68 at promoting the splicing of introns by interacting with U1 snRNP[234], and with the known role of QKI at generating circRNAs always by binding to intronic sequences[124]. We decide to focus specifically on the altered AS events linked to the cardiac development. We observed more alteration of cardiac-related AS events for the QKI$^{-/-}$ cell line than for the Sam68 $^{-/-}$. This difference can partially be explained by the mild downregulation of Rbm20 and Rbm24 in the QKI$^{-/-}$ line. However, the overlap between the observed altered AS events and the ones that have been previously described to be regulated by the two Rbm

proteins, is very limited[63,159,167]. This suggests the existence of a Rbm independent mechanism of splicing operated by QKI. We have shown that the depletion of QKI causes aberrant alternative splicing for important structural sarcomeric proteins such as nebulette[183] and titin[168,169], important ion channels, such as the potassium channel Kdcn3[206] or the calcium channel crucial for the calcium-induced calcium release mechanism that allow the coupling excitation-contraction Ryr2[192], and last but not least of important developmental proteins such as Stim1[211] and Adgrg6[213,214]. Conversely, fewer cardiac-related AS events are altered in Sam68$^{-/-}$ suggesting that the regulation of heart development by these two proteins could happen at different and independent post transcriptional levels. Nevertheless, we found that the CACNA1C channel subunit that is crucial for the inward calcium currents that in turn are important for the cardiomyocyte depolarization and their consequent contractions, showed altered pattern of AS[216,217,219]. CACNA1C has three alternatively spliced exons that are more included and three others that are more excluded in the mature transcript in the Sam68$^{-/-}$ line compared to the WT .Interestingly, the exon 21 of the CACNA1C transcript, that encodes a transmembrane domain[215], is target of both Sam68 and QKI. Abrogating their expression causes an increase in inclusion of this exon suggesting that both proteins are needed in order to splice it out. Importantly, mutations in this gene and its altered alternative splicing are associated to the development of cardiomyopathies[216,218,219]. Additionally, depletion of Sam68 causes as well less inclusion of exon10 of the MYBPC3 transcript, an important sarcomeric structural protein whose mutations are as well

linked to the development of cardiomyopathies[220,221]. **These data demonstrate that both Sam68 and Quaking exert a role in the differentiation of the mESCs towards the cardiomyocyte lineage. They achieve this at different layers by controlling the gene expression, where they show almost an opposite effect, as well as the AS of cardiac-related transcripts.**

7.6 What happens at the protein level?

Given the known role of Sam68 in translation regulation during the differentiation of the male germ cells[86], we decided to study the global translational activity in EBs at day 10 of differentiation through the ribosome profiling. The ribosome profiling, is a powerful technique developed by the Ingolia laboratory, used in order to have a snapshot of the translational activity of a cell in a given time[235]. Ribosome profiling demonstrated that Sam68 affects the translation of important cardiac-related mRNAs both positively and negatively, suggesting that it can act as both translational enhancer or repressor. Interestingly, we demonstrated that the absence of Sam68 in EBs at day 10 of differentiation, causes a decreased translation of the cardiogenic TF GATA4 mRNA. **These data demonstrate that Sam68 has a role in the cardiac differentiation of the mESCs even through mechanisms of translational regulation.**

7.7 If heart it is, Heart will it be?

Given all the results we have shown in the involvement of these two STAR family members in the cardiogenic differentiation of the mESCs, we decided to carry out a cardiomyocyte differentiation assay. We allowed the differentiation of both WT and Sam68$^{-/-}$ or QKI$^{-/-}$ KO cells into cardiomyocytes and scored for the presence of beating *foci*. Interestingly, both knock out lines give rise to a statistically significant lower number of beating *foci*. More specifically, while for the WT line almost the 70% of the formed foci showed a beating activity, this percentage dropped to 20% in the case of the KO lines. This result is similar to the beating EBs obtained from GATA4$^{-/-}$ embryonic stem cells[236]. Moreover, the beating *foci* generated in the QKI$^{-/-}$ line are less defined and smaller, whereas the ones generated by Sam68$^{-/-}$ showed faster beating activity compared to the WT. Potentially, this beating defect, might be due to the altered alternative splicing of the calcium channel subunit CACNA1C that, along with the severe upregulation of the other two calcium channel Ryr2 and the CACNB2 subunit, could lead to a general deregulation of the intracellular calcium currents and consequently of the contraction. Although this has to be demonstrated, this hypothesis is at least supported by several independent researches that show a link between the CACNA1C mutations and the arrhythmic activity of the heart[237–239].**All these data, strongly support the hypothesis that both STAR family members have important roles during the cardiomyocyte differentiation of the mESCs.**

7.8 STARs, GATA4 & hearts, who is doing what?

Our RNA-seq results showed deregulation of the mRNA of GATA4 in opposite sense in both Sam68$^{-/-}$ and QKI$^{-/-}$. On the other hand, the ribosome profiling revealed its decreased translation in the Sam68$^{-/-}$ cells. Given the major defects in the generation of *bona fide* cardiomyocytes, we wanted to test the hypothesis that the two STAR family members could regulate the GATA4 transcripts levels directly by binding to it. As a consequence, we performed RNA immunoprecipitation in EBs at day 10 of differentiation and we found out that Sam68, unlike Quaking, binds directly to GATA4 mRNA. Last but not least, we wanted to verify whether the decreased translational activity of GATA4 in Sam68 led to less protein production as well. Coherently with the ribosome profiling findings, there is a dramatic reduction of the cardiogenic transcription factor GATA4 upon Sam68 KO detected by both WB and IF. **These data demonstrate that Sam68 directly binds to GATA4 mRNA and regulates the amount of GATA4 protein level most probably through a regulation of its translational activity.** However, there is still an open question: why, despite the increase of GATA4 mRNA levels there is a decrease in its protein levels there? A possible hypothesis that could explain the divergent effects of Sam68 at the mRNA and protein level of GATA4 is a mechanism in which Sam68 plays as an enhancer of the translation of GATA4 maintaining its proper translation and protein levels. In the absence of Sam68, GATA4 mRNA is poorly translated into protein; the cell senses the low levels of GATA4 protein and tries to restore them by

upregulating the transcription of its gene. Alternatively another appealing hypothesis might be based on the function of Sam68 in the nuclear export of the RNAs, a function already known in the case of the export of the HIV transcripts into the cytosol of the infected cells[243]. In this scenario, Sam68 might enhance the nuclear export of GATA4 transcript. In presence of Sam68, GATA4 mRNA is properly exported to the cytosol and consequently translated and its protein levels. Conversely, in the absence of Sam68, GATA4 might be retained in the nucleus, thus less associated to actively translated ribosomes leading to low protein production. With the data collected so far none of the two hypotheses can be ruled out and actually both molecular mechanisms could coexist as Sam68 is present both in the nuclear and cytoplasmic compartments. Further researches are needed in order to fully confirm or rule out these two appealing scenarios of GATA4 mRNA and protein regulation orchestrated by Sam68 during the differentiation of mESCs into fully mature cardiomyocytes. To date, this is the first evidence of post-transcriptional regulation of GATA4 mRNA by a specific RNA binding protein. On the other hand, this regulation is GATA4 specific, in fact the other GATA transcription factors do not show any deregulation of their expression levels upon Sam68 depletion.

8 CONCLUSIONS

The work carried out during my PhD studies at the Center for Genomic Regulation (CRG) of Barcelona, has been compiled under the form a thesis entitled “The role of the STAR proteins in early mammalian development and pluripotency”. The thesis presents my personal contribution to the understanding of the role of two members of the STAR protein family, namely Sam68 and Quaking, in the regulation of both pluripotency and differentiation of the mouse embryonic stem cells specifically towards the cardiomyocyte lineage.

I can summarize the major findings as follows:

1. Sam68 and Quaking do not affect the *in vivo* pluripotent capacity of the mESCs

2. Both Sam68 and Quaking are not involved in the disruption of the core pluripotency transcription factor network

3. Sam68 and Quaking do positively control both the self-renewal capacity and the proliferation of mESCs

4. Sam68 and Quaking are involved in the differentiation of the mESCs towards cardiomyocyte by regulating both the expression of cardiac-related transcripts as well as their alternative splicing

5. Sam68 is involved in the control of the beating activity of the cardiomyocytes. This is perhaps due to the regulation of the alternative splicing of the calcium channel CACNA1C and of the expression of the ryanodine receptor Ryr2

6. Sam68 controls, directly or indirectly, the translational activity of several cardiac-related transcripts

7. Sam68 binds GATA4 mRNA and influences its translation either enhancing it directly or indirectly or by influencing its nuclear export

9 FUTURE PERSPECTIVES

The findings of this study about the role of Sam68 and Quaking in the differentiation of the mESCs towards the cardiomyocyte lineage open up extensive avenues of research to understand the mechanism that sustain the defects seen during the differentiation of the mESCs into cardiomyocytes in either Sam68$^{-/-}$ or QKI$^{-/-}$ cells. First, it would be interesting to perform another cutting-edge technique used to study the interactions between RBPs and the RNAs, namely the cross-linking followed by the immunoprecipitation of the RBP of interest (CLIP). This technique, developed for the first time in 2005 and subsequently modified in order to increase its resolution, unravels, with nucleotide resolution in one of its variants, the binding sites of an RBP on the RNAs[244]. Consequently, through this technique, several information can be retrieved. For instance, it makes possible to discover to most favorable motifs of binding to the RNA and the binding position of the RBP on its target RNA, revealing the mechanism of regulation of that RBP on that specific RNA. For example, a binding that occurs at the 3' or 5' UTR most probably affects the translation and the stability of the target RNAs, while a binding in proximity of the exon-intron border can indicate a role in the splicing regulation of that specific transcript. Additionally, we recently performed Sam68 immunoprecipitation followed by mass-spectrometry on EBs at D10 of differentiation in order to find out Sam68 interactors and its macromolecular complexes. Surprisingly, we have found a significant interaction between Sam68 and Interleukin enhancer-binding factor 2 and 3 (Ilf2 and Ilf3), two

factors that are involved in the biogenesis of the circular RNAs[245], circRNA, a recently discovered class of RNAs with several functions of regulation of the RNAs and protein expression[246]. Of note, QKI was recently identified as an important actor in the biogenesis of the circRNA during the epithelial-mesenchymal transition[124]. Moreover, during the differentiation of human embryonic stem cells (hESCs) into cardiomyocyte, several circRNAs are differentially expressed in different time points of differentiation, suggesting that this specific class of RNAs might have an important role during this differentiation process[224]. Therefore, another appealing hypothesis that could explain the defects of the differentiation of the mESCs into cardiomyocytes could be based on the possible role that either Sam68 or QKI might have in the generation of specific circRNAs.

Appendix

qPCR primers

RPO Fw
TTCATTGTGGGAGCAGAG

RPO Rv
CAGCAGTTTCTCCAGAGC

Gata4 Fw
CCCTACCCAGCCTACATGG

Gata4 Rv
ACATATCGAGATTGGGGTGTCT

Mef2c Fw
ATCCCGATGCAGACGATTCAG

Mef2c Rv
AACAGCACACAATCTTTGCCT

Nkx2-5 Fw
GACGTAGCCTGGTGTCTCG

Nkx2-5 Rv
GTGTGGAATCCGTCGAAAGT

Actn2 Fw
TGGCACCCAGATCGAGAAC

Actn2 Rv
GTGGAACCGCATTTTTCCCC

Mybphl1 Fw
CACTGGAGATAGCCTCATGCT

Mybphl1 Rv
GTGCTCTTCAATAGAGGGCAG

Myom1 Fw
TCTCTACACGATCCGAGTGC

Myom1 Rv
GCTCCTTCGATCTCTGCATC

Nebl Fw
TACAACCCTCTGGGAAGTGC

Nebl Rv
GAGGGTGGCGTTTCCTTTAT

Scn5a Fw
ATGGCAAACTTCCTGTTACCTC

Scn5a Rv

CCACGGGCTTGTTTTTCAGC
Tmod1 Fw
TGAGCTAGATGAACTAGACCCTG

Tmod1 Rv
CGGTCCTTAAATTCCTTCGCTTG

Tnni3 Fw
TCTGCCAACTACCGAGCCTAT

Tnni3 Rv
CTCTTCTGCCTCTCGTTCCAT

Tnnc1 Fw
GCGGTAGAACAGTTGACAGAG

Tnnc1 Rv
CCAGCTCCTTGGTGCTGAT

Tnnt3 Fw
GGAACGCCAGAACAGATTGG

Tnnt3 Rv
TGGAGGACAGAGCCTTTTTCTT

Actc1 Fw
CTGGATTCTGGCGATGGTGTA

Actc1 Rv
CGGACAATTTCACGTTCAGCA

Myh6 Fw
GCCCAGTACCTCCGAAAGTC

Myh6 Rv
GCCTTAACATACTCCTCCTTGTC

Myh7 Fw
ACTGTCAACACTAAGAGGGTCA

Myh7 Rv
TTGGATGATTTGATCTTTCCAGGG

Myl4 Fw
ACCCAAGCCTGAAGAAGAGATGA

Myl4 Rv
CCCTCCACGAAGTCCTCATA

Mylk3 Fw
ACCATGTACTGACTACAGGAGG

Mylk3 Rv
CCACTGTTCGCACAGGTATGT

Msc Fw
GCCTGGCTTCCAGCTACATC

Msc Rv
CACGTCAGGTTCACAGGGTG

Rbm20 Fw
GGCCAAAACAAGCCCGATATT

Rbm20 Rv
CCCTGTCTGAGGTAGGCTCT

Rbm24 Fw
TTTTGCCTTTGGCGTTCAACA

Rbm24 Rv
GCTGCACATGGGGAATGAC

Ryr2 Fw
TCAAACCACGAACACATTGAGG

Ryr2 Rv
AGGCGGTAAAACATGATGTCAG

Meis2 Fw
CAGGGTGGTCCAATGGGAATG

Meis2 Rv

GGGGGTCCATGTCTTAACTGAG

Isl1 Fw
CAGTCCCAGAGTCATCCGAGT

Isl1 Rv
TGGGTTAGCAGTTTTGTCGTT

Tbx18 Fw
GTACCTGGCTTGGCACGAC

Tbx18 Rv
GCATTGCTGGAAACATGCG

Myl2 Fw
ATCGACAAGAATGACCTAAGGGA

Myl2 Rv
ATTTTTCACGTTCACTCGTCCT

CACNB2 Fw
ACTAGAGAACATGAGGCTACAGC

CACNB2 Rv
GCACTATGTCACCCAAACTGGAT

Tnnt2 Fw
CAGAGGAGGCCAACGTAGAAG

Tnnt2 Rv
CTCCATCGGGGATCTTGGGT

Ryr2 Fw
TCAAACCACGAACACATTGAGG

Ryr2 Rv
AGGCGGTAAAACATGATGTCAG

Col5a2 Fw
TTGGAAACCTTCTCCATGTCAGA

Col5a2 Rv
TCCCCAGTGGGTGTTATAGGA

Lama1 Fw
CAGCGCCAATGCTACCTGT

Lama1 Rv
GGATTCGTACTGTTACCGTCACA

Alternative splicing

Nebl Fw
ATCCTGTGACTGAACGGGTG

Nebl Rv
GCTGGTCTGATGCATGTACCC

Limch1 Fw
GGCGCCTCTGAGAAAGAAGAA

Limch1 Rv
CTCCTCACACCGCATGTCAAA

Titin Fw
GCAACCAAAGCCAAAGAGCAA

Titin Rv
GCTGCTCTGGGACCTTTGTG

Myo18a Fw
GGAAGCTGCTAACCAAAGCCT

Myo18a Rv
TGAGTCCACATCTGAGTCCCC

Ryr2 Fw
CCTCAGACCCAGAGAGGACAG

Ryr2 Rv
AGTAGTTTGTGCCACACAGCT

Stim1 Fw
CTCTCAACATCGACCCCAGCT

Stim1 Rv
TCTGAGATCCCAGGCCAAGC

Kcdn3 Fw
GCAAGACCACCTCACTCATCG

Kcdn3 Rv
GGACTTCTGGTGGAGGGGTAG

Ank3 Fw
ACCATCACGGAGAAGCACAAA

Ank3 Rv
GGGAGTCATCACCTAGCTCCT

Slmap Fw
AAGACAACCTGAAGCTGCTGC

Slmap Rv
GCTTCTAGAGGGAGGACGGTG

Smpx Fw
CCAGGACCTGTTGTCAACTTGT

Smpx Rv
CCCTTGGAAAAACACGTCAACG

Sorbs2 Fw
CTGTCGACAGGCCAAAGGAC

Sorbs2 Rv
CTGTCGACAGGCCAAAGGAC

Gpr126 Fw
CTACCTGATCCAGCTTCCTGC

Gpr126 Rv
CCATTCTGCTACTTTGGTCTGCA

Cacn1c.1 (Mmu EX0008751) Fw
CATCTCTCTGGCTGCTGAGGA

Cacn1c.1 (Mmu EX0008751) Rv
AGATGAGGGACACGCTAACCA

Cacn1c.2 (Mmu EX0008750) Fw
TTCCGGCAGAAGAGGATCCTT

Cacn1c.2 (Mmu EX0008750) Rv
TCTGGAAATCTCCTCGGGCTT

Cacn1c.3 (Mmu EX0008746) Fw
ACTATGGCCAGAGCTGCCTC

Cacn1c.3 (Mmu EX0008746) Rv
CATAGAGGGAGAGCATTGGGT

MYBPC3 Fw
GCCATGAAGATGCTGGGACTC

MYBPC3 Rv
CTCAGGATCTCCCACACGTCT

Primers for RIP

GATA4.2 Fw
CACCCCAATCTCGATATGTTTGA

GATA4.2 R
GCACAGGTAGTGTCCCGTC

GATA4.3 Fw
CACGCTGTGGCGTCGTAAT

GATA4.3 Rv
CTGGTTTGAATCCCCTCCTTC

Antibodies

Anti-Sam68: C-20 sc-333 Santa Cruz

Anti-Sam68: H4 sc-514468 Santa Cruz

Anti-Quaking: N-20 sc-103851 Santa Cruz

Anti-panQuaking: N147/6 MABN624 Millipore

Anti-GATA4: produced in house by the protein technologies unit of the CRG

Anti-betaTubulin: ab6046 abcam

Normal Rabbit IgG: sc-3888 Santa Cruz

Protein-G-HRP conjugated: ab97046 abcam

Alexafluor 488 donkey anti-mouse: A21202 life technologies

Alexafluor plus555 goat anti-rabbit: A32732 life technologies

Goat anti-rabbit-HRP conjugated: P0448 Dako

gRNA sequences for the generation of the KO lines

Sam68 sgRNA: GGCGCGAGGCAGGATCGTCC

Quaking sgRNA: GGTCGGGGAAATGGAAACGA

Oligonucleotides for the Ribosome profiling libraries

Preadenylated linker: 1/5rApp/CTGTAGGCACCATCAAT/3ddC/

Reverse transcription primer:
5′(Phos)AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA
GATCTCGGTGGTCGC(SpC18)CACTCA(SpC18)TTCAGACGT
GTGCTCTTCCGATCTATTGATGG TGCCTACAG-3′

Forward library primer: 5′-AATGATACGGCGACCACCGAG
ATCTACAC-3′

Reverse barcoded library primer:
5′CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGG
AGTTCAGACGTGTGCTCTTCCG-3′

10 BIBLIOGRAPHY

1. Adamson, E. D. & Gardner, R. L. Control of early development. *Br. Med. Bull.* 35, 113–119 (1979).

2. Gardner, R. L. & Rossant, J. Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. *J. Embryol. Exp. Morphol.* 52, 141–152 (1979).

3. Arnold, S. J. & Robertson, E. J. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* 10, 91–103 (2009).

4. Davidson, K. C., Mason, E. A. & Pera, M. F. The pluripotent state in mouse and human. *Dev. Camb. Engl.* 142, 3090–3099 (2015).

5. Tam, P. P. L. & Loebel, D. A. F. Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* 8, 368–381 (2007).

6. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* 16, 516–528 (2014).

7. Brook, F. A. & Gardner, R. L. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc. Natl. Acad. Sci. U. S. A.* 94, 5709–5712 (1997).

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

9. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 78, 7634–7638 (1981).

10. Rowe, R. G. & Daley, G. Q. Induced pluripotent stem cells in disease modelling and drug discovery. *Nat. Rev. Genet.* (2019). doi:10.1038/s41576-019-0100-z

11. Rikhtegar, R. *et al.* Stem cells as therapy for heart disease: iPSCs, ESCs, CSCs, and skeletal myoblasts. *Biomed. Pharmacother. Biomedecine Pharmacother.* 109, 304–313 (2019).

12. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448, 191–195 (2007).

13. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448, 196–199 (2007).

14. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523 (2008).

15. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* 4, 487–492 (2009).

16. Dey, S. K. *et al.* Molecular cues to implantation. *Endocr. Rev.* 25, 341–373 (2004).

17. Bedzhov, I. & Zernicka-Goetz, M. Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell* 156, 1032–1044 (2014).

18. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 90, 8424–8428 (1993).

19. Smith, A. A glossary for stem-cell biology. *Nature* 441, 1060–1060 (2006).

20. Chambers, I. *et al.* Functional expression cloning of Nanog, a

pluripotency sustaining factor in embryonic stem cells. *Cell* 113, 643–655 (2003).

21. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* 9, 625–635 (2007).

22. Okamoto, K. *et al.* A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* 60, 461–472 (1990).

23. Schöler, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J.* 9, 2185–2195 (1990).

24. Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* 10, 440–454 (2012).

25. Teo, A. K. K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* 25, 238–250 (2011).

26. Loh, K. M., Lim, B. & Ang, L. T. Ex uno plures: molecular designs for embryonic pluripotency. *Physiol. Rev.* 95, 245–295 (2015).

27. Itskovitz-Eldor, J. *et al.* Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Mol. Med. Camb. Mass* 6, 88–95 (2000).

28. Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J. Embryol. Exp. Morphol.* 87, 27–45

(1985).

29. Li, X. *et al.* Fibroblast growth factor signaling and basement membrane assembly are connected during epithelial morphogenesis of the embryoid body. *J. Cell Biol.* 153, 811–822 (2001).

30. Yang, M.-J. *et al.* Novel Method of Forming Human Embryoid Bodies in a Polystyrene Dish Surface-Coated with a Temperature-Responsive Methylcellulose Hydrogel. *Biomacromolecules* 8, 2746–2752 (2007).

31. Yoon, B. S. *et al.* Enhanced differentiation of human embryonic stem cells into cardiomyocytes by combining hanging drop culture and 5-azacytidine treatment. *Differ. Res. Biol. Divers.* 74, 149–159 (2006).

32. Martin Gonzalez, J. *et al.* Embryonic Stem Cell Culture Conditions Support Distinct States Associated with Different Developmental Stages and Potency. *Stem Cell Rep.* 7, 177–191 (2016).

33. Zhou, H. *et al.* Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* 4, 381–384 (2009).

34. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113, 631–642 (2003).

35. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95, 379–391 (1998).

36. Martello, G. *et al.* Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* 11, 491–504 (2012).

37. Yeo, J.-C. & Ng, H.-H. The transcriptional regulation of pluripotency. *Cell Res.* 23, 20–32 (2013).

38. Hackett, J. A. & Surani, M. A. Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell* 15, 416–430 (2014).

39. Gonzales, K. A. U. *et al.* Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell* 162, 564–579 (2015).

40. Betschinger, J. *et al.* Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* 153, 335–347 (2013).

41. Licatalosi, D. D. & Darnell, R. B. RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.* 11, 75–87 (2010).

42. Mitchell, S. F. & Parker, R. Principles and properties of eukaryotic mRNPs. *Mol. Cell* 54, 547–558 (2014).

43. Lu, R. *et al.* Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* 462, 358–362 (2009).

44. Han, H. *et al.* MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* 498, 241–245 (2013).

45. Atlasi, Y., Mowla, S. J., Ziaee, S. A. M., Gokhale, P. J. & Andrews, P. W. OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem Cells Dayt. Ohio* 26, 3068–3074 (2008).

46. Gabut, M. *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* 147,

132–146 (2011).

47. Lu, Y. *et al.* Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. *Cell Stem Cell* 15, 92–101 (2014).

48. Tahmasebi, S. *et al.* Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12360–12367 (2016).

49. Siomi, H. & Dreyfuss, G. RNA-binding proteins as regulators of gene expression. *Curr. Opin. Genet. Dev.* 7, 345–353 (1997).

50. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341 (2018).

51. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* 8, 479–490 (2007).

52. Cléry, A., Blatter, M. & Allain, F. H.-T. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* 18, 290–298 (2008).

53. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* 29, 183–212 (2000).

54. Valverde, R., Edwards, L. & Regan, L. Structure and function of KH domains. *FEBS J.* 275, 2712–2726 (2008).

55. Hennig, J. & Sattler, M. Deciphering the protein-RNA recognition code: Combining large-scale quantitative methods with structural biology. *BioEssays* 37, 899–908 (2015).

56. Xu, Y. *et al.* New Insights into the Interplay between Non-Coding RNAs and RNA-Binding Protein HnRNPK in Regulating Cellular Functions. *Cells* 8, (2019).

57. Wang, Z.-L. *et al.* Comprehensive Genomic Characterization of RNA-Binding Proteins across Human Cancers. *Cell Rep.* 22, 286–298 (2018).

58. Fredericks, A., Cygan, K., Brown, B. & Fairbrother, W. RNA-Binding Proteins: Splicing Factors and Disease. *Biomolecules* 5, 893–909 (2015).

59. Cieply, B. *et al.* Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by Numerous RNA-Binding Proteins. *Cell Rep.* 15, 247–255 (2016).

60. Yeo, G. W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* 16, 130–137 (2009).

61. Lu, X. *et al.* SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat. Cell Biol.* 15, 1141–1152 (2013).

62. Hirsch, C. L. *et al.* Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. *Genes Dev.* 29, 803–816 (2015).

63. Zhang, T. *et al.* Rbm24 Regulates Alternative Splicing Switch in Embryonic Stem Cell Cardiac Lineage Differentiation. *Stem Cells Dayt. Ohio* 34, 1776–1789 (2016).

64. Toh, C.-X. D. *et al.* RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming. *Cell Rep.* 15,

2597–2607 (2016).

65. Elkon, R., Ugalde, A. P. & Agami, R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* 14, 496–506 (2013).

66. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 320, 1643–1647 (2008).

67. Mueller, A. A., Cheung, T. H. & Rando, T. A. All's well that ends well: alternative polyadenylation and its implications for stem cell biology. *Curr. Opin. Cell Biol.* 25, 222–232 (2013).

68. Lackford, B. *et al.* Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* 33, 878–889 (2014).

69. Dai, Q. *et al.* Primordial dwarfism gene maintains Lin28 expression to safeguard embryonic stem cells from premature differentiation. *Cell Rep.* 7, 735–746 (2014).

70. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95 (2014).

71. Batista, P. J. *et al.* m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. *Cell Stem Cell* 15, 707–719 (2014).

72. Wang, L. *et al.* The THO Complex Regulates Pluripotency Gene mRNA Export and Controls Embryonic Stem Cell Self-Renewal and Somatic Cell Reprogramming. *Cell Stem Cell* 13, 676–690 (2013).

73. Chen, Q. & Hu, G. Post-transcriptional regulation of the pluripotent state. *Curr. Opin. Genet. Dev.* 46, 15–23 (2017).

74. Vernet, C. & Artzt, K. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet. TIG* 13, 479–484 (1997).

75. Liu, Z. *et al.* Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 294, 1098–1102 (2001).

76. Chen, T. & Richard, S. Structure-function analysis of Qk1: a lethal point mutation in mouse quaking prevents homodimerization. *Mol. Cell. Biol.* 18, 4863–4871 (1998).

77. Wu, J., Zhou, L., Tonissen, K., Tee, R. & Artzt, K. The quaking I-5 protein (QKI-5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm. *J. Biol. Chem.* 274, 29202–29210 (1999).

78. Ebersole, T. A., Chen, Q., Justice, M. J. & Artzt, K. The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nat. Genet.* 12, 260–265 (1996).

79. Nabel-Rosen, H., Dorevitch, N., Reuveny, A. & Volk, T. The balance between two isoforms of the Drosophila RNA-binding protein how controls tendon cell differentiation. *Mol. Cell* 4, 573–584 (1999).

80. Galarneau, A. & Richard, S. The STAR RNA binding proteins GLD-1, QKI, SAM68 and SLM-2 bind bipartite RNA motifs. *BMC Mol. Biol.* 10, 47 (2009).

81. Justice, M. J. & Bode, V. C. Three ENU-induced alleles of

the murine quaking locus are recessive embryonic lethal mutations. *Genet. Res.* 51, 95–102 (1988).

82. Li, Z. *et al.* Defective smooth muscle development in qkI-deficient mice. *Dev. Growth Differ.* 45, 449–462 (2003).

83. Noveroske, J. K. *et al.* Quaking is essential for blood vessel development. *Genes. N. Y. N 2000* 32, 218–230 (2002).

84. Bohnsack, B. L., Lai, L., Northrop, J. L., Justice, M. J. & Hirschi, K. K. Visceral endoderm function is regulated by quaking and required for vascular development. *Genes. N. Y. N 2000* 44, 93–104 (2006).

85. Sidman, R. L., Dickie, M. M. & Appel, S. H. MUTANT MICE (QUAKING AND JIMPY) WITH DEFICIENT MYELINATION IN THE CENTRAL NERVOUS SYSTEM. *Science* 144, 309–311 (1964).

86. Paronetto, M. P. *et al.* Sam68 regulates translation of target mRNAs in male germ cells, necessary for mouse spermatogenesis. *J. Cell Biol.* 185, 235–249 (2009).

87. Lukong, K. E. & Richard, S. Motor coordination defects in mice deficient for the Sam68 RNA-binding protein. *Behav. Brain Res.* 189, 357–363 (2008).

88. Fumagalli, S., Totty, N. F., Hsuan, J. J. & Courtneidge, S. A. A target for Src in mitosis. *Nature* 368, 871–874 (1994).

89. Taylor, S. J. & Shalloway, D. An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature* 368, 867–871 (1994).

90. Lin, Q., Taylor, S. J. & Shalloway, D. Specificity and determinants of Sam68 RNA binding. Implications for the biological

function of K homology domains. *J. Biol. Chem.* 272, 27274–27280 (1997).

91. Paronetto, M. P., Achsel, T., Massiello, A., Chalfant, C. E. & Sette, C. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J. Cell Biol.* 176, 929–939 (2007).

92. Sánchez-Jiménez, F. & Sánchez-Margalet, V. Role of Sam68 in post-transcriptional gene regulation. *Int. J. Mol. Sci.* 14, 23402–23419 (2013).

93. Rajan, P. *et al.* Regulation of gene expression by the RNA-binding protein Sam68 in cancer. *Biochem. Soc. Trans.* 36, 505–507 (2008).

94. Lukong, K. E. & Richard, S. Sam68, the KH domain-containing superSTAR. *Biochim. Biophys. Acta* 1653, 73–86 (2003).

95. Najib, S., Martín-Romero, C., González-Yanes, C. & Sánchez-Margalet, V. Role of Sam68 as an adaptor protein in signal transduction. *Cell. Mol. Life Sci. CMLS* 62, 36–43 (2005).

96. Matter, N., Herrlich, P. & König, H. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature* 420, 691–695 (2002).

97. Babic, I., Cherry, E. & Fujita, D. J. SUMO modification of Sam68 enhances its ability to repress cyclin D1 expression and inhibits its ability to induce apoptosis. *Oncogene* 25, 4955–4964 (2006).

98. Chawla, G. *et al.* Sam68 regulates a set of alternatively spliced exons during neurogenesis. *Mol. Cell. Biol.* 29, 201–213 (2009).

99. Paronetto, M. P. & Sette, C. Role of RNA-binding proteins in

mammalian spermatogenesis. *Int. J. Androl.* 33, 2–12 (2010).

100. Naro, C. *et al.* Functional Interaction between U1snRNP and Sam68 Insures Proper 3′ End Pre-mRNA Processing during Germ Cell Differentiation. *Cell Rep.* 26, 2929-2941.e5 (2019).

101. Richard, S. *et al.* Ablation of the Sam68 RNA binding protein protects mice from age-related bone loss. *PLoS Genet.* 1, e74 (2005).

102. Richard, S. *et al.* Sam68 haploinsufficiency delays onset of mammary tumorigenesis and metastasis. *Oncogene* 27, 548–556 (2008).

103. La Rosa, P. *et al.* Sam68 promotes self-renewal and glycolytic metabolism in mouse neural progenitor cells by modulating Aldh1a3 pre-mRNA 3'-end processing. *eLife* 5, e20750 (2016).

104. Li, N., Hébert, S., Song, J., Kleinman, C. L. & Richard, S. Transcriptome profiling in preadipocytes identifies long noncoding RNAs as Sam68 targets. *Oncotarget* 8, (2017).

105. Song, J. & Richard, S. Sam68 Regulates S6K1 Alternative Splicing during Adipogenesis. *Mol. Cell. Biol.* 35, 1926–1939 (2015).

106. Galarneau, A. & Richard, S. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat. Struct. Mol. Biol.* 12, 691–698 (2005).

107. Ryder, S. P., Frater, L. A., Abramovitz, D. L., Goodwin, E. B. & Williamson, J. R. RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. *Nat. Struct. Mol. Biol.* 11, 20–28 (2004).

108. Darbelli, L. & Richard, S. Emerging functions of the Quaking RNA-binding proteins and link to human diseases: Emerging

functions of the Quaking proteins. *Wiley Interdiscip. Rev. RNA* 7, 399–412 (2016).

109. Lu, Z. *et al.* The quakingviable mutation affects qkI mRNA expression specifically in myelin-producing cells of the nervous system. *Nucleic Acids Res.* 31, 4616–4624 (2003).

110. Hardy, R. J. *et al.* Neural cell type-specific expression of QKI proteins is altered in quakingviable mutant mice. *J. Neurosci. Off. J. Soc. Neurosci.* 16, 7941–7949 (1996).

111. Larocque, D. *et al.* The QKI-6 and QKI-7 RNA binding proteins block proliferation and promote Schwann cell myelination. *PloS One* 4, e5867 (2009).

112. Kondo, T. *et al.* Genomic organization and expression analysis of the mouse qkI locus. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 10, 662–669 (1999).

113. Lorenzetti, D. *et al.* The neurological mutant quaking(viable) is Parkin deficient. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 15, 210–217 (2004).

114. Hardy, R. J. Molecular defects in the dysmyelinating mutant quaking. *J. Neurosci. Res.* 51, 417–422 (1998).

115. Chénard, C. A. & Richard, S. New implications for the QUAKING RNA binding protein in human disease. *J. Neurosci. Res.* 86, 233–242 (2008).

116. Hogan, E. L. & Greenfield, S. Animal Models of Genetic Disorders of Myelin. in *Myelin* (ed. Morell, P.) 489–534 (Springer US, 1984). doi:10.1007/978-1-4757-1830-0_14

117. Hardy, R. J. QKI expression is regulated during neuron-glial cell fate decisions. *J. Neurosci. Res.* 54, 46–57 (1998).

118. Wu, J. I., Reed, R. B., Grabowski, P. J. & Artzt, K. Function of quaking in myelination: regulation of alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4233–4238 (2002).

119. Zhang, Y. & Feng, Y. Distinct molecular mechanisms lead to diminished myelin basic protein and 2',3'-cyclic nucleotide 3'-phosphodiesterase in qk(v) dysmyelination. *J. Neurochem.* 77, 165–172 (2001).

120. Li, Z., Zhang, Y., Li, D. & Feng, Y. Destabilization and mislocalization of myelin basic protein mRNAs in quaking dysmyelination lacking the QKI RNA-binding proteins. *J. Neurosci. Off. J. Soc. Neurosci.* 20, 4944–4953 (2000).

121. Larocque, D. *et al.* Nuclear retention of MBP mRNAs in the quaking viable mice. *Neuron* 36, 815–829 (2002).

122. Hayakawa-Yano, Y. & Yano, M. An RNA Switch of a Large Exon of Ninein Is Regulated by the Neural Stem Cell Specific-RNA Binding Protein, Qki5. *Int. J. Mol. Sci.* 20, 1010 (2019).

123. de Bruin, R. G. *et al.* The RNA-binding protein quaking maintains endothelial barrier function and affects VE-cadherin and β-catenin protein expression. *Sci. Rep.* 6, 21643 (2016).

124. Conn, S. J. *et al.* The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125–1134 (2015).

125. Gupta, S. K. *et al.* Quaking Inhibits Doxorubicin-Mediated Cardiotoxicity Through Regulation of Cardiac Circular RNA Expression. *Circ. Res.* 122, 246–254 (2018).

126. Bonnet, A. *et al.* Quaking RNA-Binding Proteins Control Early Myofibril Formation by Modulating Tropomyosin. *Dev. Cell* 42, 527-541.e4 (2017).

127. Brand, T. Heart development: molecular insights into cardiac specification and early morphogenesis. *Dev. Biol.* 258, 1–19 (2003).
128. Costello, I. *et al.* The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nat. Cell Biol.* 13, 1084–1091 (2011).
129. Kattman, S. J., Huber, T. L. & Keller, G. M. Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. *Dev. Cell* 11, 723–732 (2006).
130. 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).
131. Kelly, R. G., Brown, N. A. & Buckingham, M. E. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. *Dev. Cell* 1, 435–440 (2001).
132. Cai, C.-L. *et al.* Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev. Cell* 5, 877–889 (2003).
133. Zaffran, S., Kelly, R. G., Meilhac, S. M., Buckingham, M. E. & Brown, N. A. Right ventricular myocardium derives from the anterior heart field. *Circ. Res.* 95, 261–268 (2004).
134. Hoffmann, A. D., Peterson, M. A., Friedland-Little, J. M., Anderson, S. A. & Moskowitz, I. P. sonic hedgehog is required in pulmonary endoderm for atrial septation. *Dev. Camb. Engl.* 136, 1761–1770 (2009).
135. Christoffels, V. M. *et al.* Formation of the venous pole of the heart from an Nkx2-5-negative precursor population requires Tbx18.

Circ. Res. 98, 1555–1563 (2006).

136. Mommersteeg, M. T. M. *et al.* The sinus venosus progenitors separate and diversify from the first and second heart fields early in development. *Cardiovasc. Res.* 87, 92–101 (2010).

137. Brand, T. Heart development: molecular insights into cardiac specification and early morphogenesis. *Dev. Biol.* 258, 1–19 (2003).

138. Kirby, M. L. Molecular embryogenesis of the heart. *Pediatr. Dev. Pathol. Off. J. Soc. Pediatr. Pathol. Paediatr. Pathol. Soc.* 5, 516–543 (2002).

139. Lohr, J. L. & Yost, H. J. Vertebrate model systems in the study of early heart development: Xenopus and zebrafish. *Am. J. Med. Genet.* 97, 248–257 (2000).

140. Kirby, M. L., Gale, T. F. & Stewart, D. E. Neural crest cells contribute to normal aorticopulmonary septation. *Science* 220, 1059–1061 (1983).

141. Kwon, C. *et al.* Canonical Wnt signaling is a positive regulator of mammalian cardiac progenitors. *Proc. Natl. Acad. Sci. U. S. A.* 104, 10894–10899 (2007).

142. Kwon, C. *et al.* A regulatory pathway involving Notch1/beta-catenin/Isl1 determines cardiac progenitor cell fate. *Nat. Cell Biol.* 11, 951–957 (2009).

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

144. Bruneau, B. G. Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb. Perspect.*

Biol. 5, a008292 (2013).

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

146. Dunwoodie, S. L. Combinatorial signaling in the heart orchestrates cardiac induction, lineage specification and chamber formation. *Semin. Cell Dev. Biol.* 18, 54–66 (2007).

147. Miquerol, L. & Kelly, R. G. Organogenesis of the vertebrate heart. *Wiley Interdiscip. Rev. Dev. Biol.* 2, 17–29 (2013).

148. Schleich, J.-M., Abdulla, T., Summers, R. & Houyel, L. An overview of cardiac morphogenesis. *Arch. Cardiovasc. Dis.* 106, 612–623 (2013).

149. Lyons, I. *et al.* Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev.* 9, 1654–1666 (1995).

150. Yamagishi, H. *et al.* The combinatorial activities of Nkx2.5 and dHAND are essential for cardiac ventricle formation. *Dev. Biol.* 239, 190–203 (2001).

151. Redkar, A., Montgomery, M. & Litvin, J. Fate map of early avian cardiac progenitor cells. *Dev. Camb. Engl.* 128, 2269–2279 (2001).

152. He, A., Kong, S. W., Ma, Q. & Pu, W. T. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5632–5637 (2011).

153. Zhou, P., He, A. & Pu, W. T. Regulation of GATA4 transcriptional activity in cardiovascular development and disease. *Curr. Top. Dev. Biol.* 100, 143–169 (2012).

154. Perrino, C. & Rockman, H. A. GATA4 and the two sides of gene expression reprogramming. *Circ. Res.* 98, 715–716 (2006).

155. Rivera-Feliciano, J. Development of heart valves requires Gata4 expression in endothelial-derived cells. *Development* 133, 3607–3618 (2006).

156. Sutherland, L. C., Rintala-Maki, N. D., White, R. D. & Morin, C. D. RNA binding motif (RBM) proteins: a novel family of apoptosis modulators? *J. Cell. Biochem.* 94, 5–24 (2005).

157. Yang, J. *et al.* RBM24 is a major regulator of muscle-specific alternative splicing. *Dev. Cell* 31, 87–99 (2014).

158. Fetka, I., Radeghieri, A. & Bouwmeester, T. Expression of the RNA recognition motif-containing protein SEB-4 during Xenopus embryonic development. *Mech. Dev.* 94, 283–286 (2000).

159. Grifone, R. *et al.* The RNA-binding protein Rbm24 is transiently expressed in myoblasts and is required for myogenic differentiation during vertebrate development. *Mech. Dev.* 134, 1–15 (2014).

160. Maragh, S. *et al.* Identification of RNA binding motif proteins essential for cardiovascular development. *BMC Dev. Biol.* 11, 62 (2011).

161. Miller, R. A., Christoforou, N., Pevsner, J., McCallion, A. S. & Gearhart, J. D. Efficient array-based identification of novel cardiac genes through differentiation of mouse ESCs. *PloS One* 3, e2176 (2008).

162. Xu, X. Q., Soo, S. Y., Sun, W. & Zweigerdt, R. Global expression profile of highly enriched cardiomyocytes derived from human embryonic stem cells. *Stem Cells Dayt. Ohio* 27, 2163–2174

(2009).

163. Liu, J., Kong, X., Zhang, M., Yang, X. & Xu, X. RNA binding protein 24 deletion disrupts global alternative splicing and causes dilated cardiomyopathy. *Protein Cell* 10, 405–416 (2019).

164. Beraldi, R. *et al.* Rbm20-deficient cardiogenesis reveals early disruption of RNA processing and sarcomere remodeling establishing a developmental etiology for dilated cardiomyopathy. *Hum. Mol. Genet.* 23, 3779–3791 (2014).

165. Arrell, D. K. Delineating RBM20 regulation of alternative splicing in dilated cardiomyopathy. *Circ. Cardiovasc. Genet.* 7, 732–733 (2014).

166. Guo, W. *et al.* RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat. Med.* 18, 766–773 (2012).

167. Maatz, H. *et al.* RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *J. Clin. Invest.* 124, 3419–3430 (2014).

168. Methawasin, M. *et al.* Experimentally Increasing the Compliance of Titin Through RNA Binding Motif-20 (RBM20) Inhibition Improves Diastolic Function In a Mouse Model of Heart Failure With Preserved Ejection Fraction. *Circulation* 134, 1085–1099 (2016).

169. Weeland, C. J., van den Hoogenhof, M. M., Beqqali, A. & Creemers, E. E. Insights into alternative splicing of sarcomeric genes in the heart. *J. Mol. Cell. Cardiol.* 81, 107–113 (2015).

170. Zhou, A. *et al.* RNA Binding Protein, HuR, Regulates *SCN5A* Expression Through Stabilizing MEF2C transcription factor mRNA. *J. Am. Heart Assoc.* 7, (2018).

171. Zaffran, S., Astier, M., Gratecos, D. & Sémériva, M. The held out wings (how) Drosophila gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. *Dev. Camb. Engl.* 124, 2087–2098 (1997).

172. Nir, R., Grossman, R., Paroush, Z. & Volk, T. Phosphorylation of the Drosophila melanogaster RNA-binding protein HOW by MAPK/ERK enhances its dimerization and activity. *PLoS Genet.* 8, e1002632 (2012).

173. Justice, M. J. & Hirschi, K. K. The role of quaking in mammalian embryonic development. *Adv. Exp. Med. Biol.* 693, 82–92 (2010).

174. Wray, J. *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat. Cell Biol.* 13, 838–845 (2011).

175. Wesselschmidt, R. L. The Teratoma Assay: An In Vivo Assessment of Pluripotency. in *Human Pluripotent Stem Cells* (eds. Schwartz, P. H. & Wesselschmidt, R. L.) 767, 231–241 (Humana Press, 2011).

176. Nelakanti, R. V., Kooreman, N. G. & Wu, J. C. Teratoma formation: a tool for monitoring pluripotency in stem cell research. *Curr. Protoc. Stem Cell Biol.* 32, 4A.8.1-17 (2015).

177. Li, Q.-H. *et al.* Retardation of the G2-M phase progression on gene disruption of RNA binding protein Sam68 in the DT40 cell line. *FEBS Lett.* 525, 145–150 (2002).

178. Biedermann, B., Hotz, H.-R. & Ciosk, R. The Quaking family of RNA-binding proteins: coordinators of the cell cycle and

differentiation. *Cell Cycle Georget. Tex* 9, 1929–1933 (2010).

179. Hayakawa-Yano, Y. *et al.* An RNA-binding protein, Qki5, regulates embryonic neural stem cells through pre-mRNA processing in cell adhesion signaling. *Genes Dev.* 31, 1910–1925 (2017).

180. Grépin, C., Nemer, G. & Nemer, M. Enhanced cardiogenesis in embryonic stem cells overexpressing the GATA-4 transcription factor. *Dev. Camb. Engl.* 124, 2387–2395 (1997).

181. Sjöblom, B., Salmazo, A. & Djinović-Carugo, K. Alpha-actinin structure and regulation. *Cell. Mol. Life Sci. CMLS* 65, 2688–2701 (2008).

182. Luther, P. K. *et al.* Understanding the Organisation and Role of Myosin Binding Protein C in Normal Striated Muscle by Comparison with MyBP-C Knockout Cardiac Muscle. *J. Mol. Biol.* 384, 60–72 (2008).

183. Chu, M., Gregorio, C. C. & Pappas, C. T. Nebulin, a multi-functional giant. *J. Exp. Biol.* 219, 146–152 (2016).

184. Warkman, A. S. *et al.* Developmental expression and cardiac transcriptional regulation of Myh7b, a third myosin heavy chain in the vertebrate heart. *Cytoskeleton* 69, 324–335 (2012).

185. Matsson, H. *et al.* Alpha-cardiac actin mutations produce atrial septal defects. *Hum. Mol. Genet.* 17, 256–265 (2008).

186. Fritz-Six, K. L. *et al.* Aberrant myofibril assembly in tropomodulin1 null mice leads to aborted heart development and embryonic lethality. *J. Cell Biol.* 163, 1033–1044 (2003).

187. Jin, J.-P., Zhang, Z. & Bautista, J. A. Isoform diversity, regulation, and functional adaptation of troponin and calponin. *Crit. Rev. Eukaryot. Gene Expr.* 18, 93–124 (2008).

188. Li, M. X. & Hwang, P. M. Structure and function of cardiac troponin C (TNNC1): Implications for heart failure, cardiomyopathies, and troponin modulating drugs. *Gene* 571, 153–166 (2015).

189. Dilg, D. *et al.* HIRA Is Required for Heart Development and Directly Regulates Tnni2 and Tnnt3. *PloS One* 11, e0161096 (2016).

190. Remme, C. A. Cardiac sodium channelopathy associated with *SCN5A* mutations: electrophysiological, molecular and genetic aspects: Cardiac sodium channelopathy associated with *SCN5A* mutations. *J. Physiol.* 591, 4099–4116 (2013).

191. Takeshima, H. *et al.* Embryonic lethality and abnormal cardiac myocytes in mice lacking ryanodine receptor type 2. *EMBO J.* 17, 3309–3316 (1998).

192. Yang, H.-T. *et al.* The ryanodine receptor modulates the spontaneous beating rate of cardiomyocytes during development. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9225–9230 (2002).

193. Wiese, C. *et al.* Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by Tbx18 and Tbx3. *Circ. Res.* 104, 388–397 (2009).

194. Bartulos, O. *et al.* ISL1 cardiovascular progenitor cells for cardiac repair after myocardial infarction. *JCI Insight* 1, (2016).

195. Peng, W. *et al.* Dysfunction of Myosin Light-Chain 4 (MYL4) Leads to Heritable Atrial Cardiomyopathy With Electrical, Contractile, and Structural Components: Evidence From Genetically-Engineered Rats. *J. Am. Heart Assoc.* 6, (2017).

196. Sheikh, F., Lyon, R. C. & Chen, J. Functions of myosin light chain-2 (MYL2) in cardiac muscle and disease. *Gene* 569, 14–20

(2015).

197. Chernyavskaya, Y., Ebert, A. M., Milligan, E. & Garrity, D. M. Voltage-gated calcium channel CACNB2 (β2.1) protein is required in the heart for control of cell proliferation and heart tube integrity. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 241, 648–662 (2012).

198. Zhang, J. *et al.* Functional cardiac fibroblasts derived from human pluripotent stem cells via second heart field progenitors. *Nat. Commun.* 10, 2238 (2019).

199. Büttner, P. *et al.* Identification of Central Regulators of Calcium Signaling and ECM–Receptor Interaction Genetically Associated With the Progression and Recurrence of Atrial Fibrillation. *Front. Genet.* 9, 162 (2018).

200. Moncman, C. L. & Wang, K. Targeted disruption of nebulette protein expression alters cardiac myofibril assembly and function. *Exp. Cell Res.* 273, 204–218 (2002).

201. Horsthemke, M. *et al.* A novel isoform of myosin 18A (Myo18Aγ) is an essential sarcomeric protein in mouse heart. *J. Biol. Chem.* 294, 7202–7218 (2019).

202. Fenix, A. M. *et al.* Muscle-specific stress fibers give rise to sarcomeres in cardiomyocytes. *eLife* 7, (2018).

203. Ushijima, T. *et al.* The actin-organizing formin protein Fhod3 is required for postnatal development and functional maintenance of the adult heart in mice. *J. Biol. Chem.* 293, 148–162 (2018).

204. Pinto, Y. M. & Reckman, Y. J. Formins Emerge as a Cause of Hypertrophic Cardiomyopathy: New Genes for Thick Hearts. *J. Am. Coll. Cardiol.* 72, 2468–2470 (2018).

205. Ochoa, J. P. *et al.* Formin Homology 2 Domain Containing 3 (FHOD3) Is a Genetic Basis for Hypertrophic Cardiomyopathy. *J. Am. Coll. Cardiol.* 72, 2457–2467 (2018).

206. Olesen, M. S. *et al.* A novel KCND3 gain-of-function mutation associated with early-onset of persistent lone atrial fibrillation. *Cardiovasc. Res.* 98, 488–495 (2013).

207. Nader, M. *et al.* Tail-anchored membrane protein SLMAP is a novel regulator of cardiac function at the sarcoplasmic reticulum. *Am. J. Physiol. Heart Circ. Physiol.* 302, H1138-1145 (2012).

208. Ishikawa, T. *et al.* A novel disease gene for Brugada syndrome: sarcolemmal membrane-associated protein gene mutations impair intracellular trafficking of hNav1.5. *Circ. Arrhythm. Electrophysiol.* 5, 1098–1107 (2012).

209. Nader, M. The SLMAP/Striatin complex: An emerging regulator of normal and abnormal cardiac excitation-contraction coupling. *Eur. J. Pharmacol.* 858, 172491 (2019).

210. Correll, R. N. *et al.* STIM1 elevation in the heart results in aberrant Ca^{2+} handling and cardiomyopathy. *J. Mol. Cell. Cardiol.* 87, 38–47 (2015).

211. Parks, C., Alam, M. A., Sullivan, R. & Mancarella, S. STIM1-dependent Ca(2+) microdomains are required for myofilament remodeling and signaling in the heart. *Sci. Rep.* 6, 25372 (2016).

212. Rosenberg, P., Katz, D. & Bryson, V. SOCE and STIM1 signaling in the heart: Timing and location matter. *Cell Calcium* 77, 20–28 (2019).

213. Liebscher, I. *et al.* New functions and signaling mechanisms for the class of adhesion G protein-coupled receptors. *Ann. N. Y.*

Acad. Sci. 1333, 43–64 (2014).

214. Patra, C. *et al.* Organ-specific function of adhesion G protein-coupled receptor GPR126 is domain-dependent. *Proc. Natl. Acad. Sci. U. S. A.* 110, 16898–16903 (2013).

215. Hu, Z., Liang, M. & Soong, T. Alternative Splicing of L-type CaV1.2 Calcium Channels: Implications in Cardiovascular Diseases. *Genes* 8, 344 (2017).

216. Hedley, P. L. *et al.* The genetic basis of Brugada syndrome: a mutation update. *Hum. Mutat.* 30, 1256–1266 (2009).

217. Napolitano, C., Splawski, I., Timothy, K. W., Bloise, R. & Priori, S. G. Timothy Syndrome. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).

218. Fukuyama, M. *et al.* Long QT syndrome type 8: novel CACNA1C mutations causing QT prolongation and variant phenotypes. *Eur. Eur. Pacing Arrhythm. Card. Electrophysiol. J. Work. Groups Card. Pacing Arrhythm. Card. Cell. Electrophysiol. Eur. Soc. Cardiol.* 16, 1828–1837 (2014).

219. Zhu, Y., Luo, J., Jiang, F. & Liu, G. Genetic analysis of sick sinus syndrome in a family harboring compound CACNA1C and TTN mutations. *Mol. Med. Rep.* (2018). doi:10.3892/mmr.2018.8773

220. Previs, M. J., Beck Previs, S., Gulick, J., Robbins, J. & Warshaw, D. M. Molecular mechanics of cardiac myosin-binding protein C in native thick filaments. *Science* 337, 1215–1218 (2012).

221. Birket, M. J. *et al.* Contractile Defect Caused by Mutation in MYBPC3 Revealed under Conditions Optimized for Human PSC-Cardiomyocyte Function. *Cell Rep.* 13, 733–745 (2015).

222. Delaney, J. T. *et al.* A KCNJ8 mutation associated with early

repolarization and atrial fibrillation. *Eur. Eur. Pacing Arrhythm. Card. Electrophysiol. J. Work. Groups Card. Pacing Arrhythm. Card. Cell. Electrophysiol. Eur. Soc. Cardiol.* 14, 1428–1432 (2012).

223. Zhou, L. *et al.* Tbx5 and Osr1 interact to regulate posterior second heart field cell cycle progression for cardiac septation. *J. Mol. Cell. Cardiol.* 85, 1–12 (2015).

224. Liu, Y. *et al.* Sox17 is essential for the specification of cardiac mesoderm in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3859–3864 (2007).

225. Varjosalo, M. & Taipale, J. Hedgehog: functions and mechanisms. *Genes Dev.* 22, 2454–2472 (2008).

226. Livi, C. M., Klus, P., Delli Ponti, R. & Tartaglia, G. G. catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinforma. Oxf. Engl.* 32, 773–775 (2016).

227. Zhao, H. *et al.* The chromatin remodeler Chd4 maintains embryonic stem cell identity by controlling pluripotency- and differentiation-associated genes. *J. Biol. Chem.* 292, 8507–8519 (2017).

228. Tian, T. V. *et al.* Whsc1 links pluripotency exit with mesendoderm specification. *Nat. Cell Biol.* 21, 824–834 (2019).

229. Ye, J. & Blelloch, R. Regulation of pluripotency by RNA binding proteins. *Cell Stem Cell* 15, 271–280 (2014).

230. Guallar, D. & Wang, J. RNA-binding proteins in pluripotency, differentiation, and reprogramming. *Front. Biol.* 9, 389–409 (2014).

231. Boy, S. *et al.* XSEB4R, a novel RNA-binding protein involved in retinal cell differentiation downstream of bHLH

proneural genes. *Dev. Camb. Engl.* 131, 851–862 (2004).

232. He, B. *et al.* MicroRNA-155 promotes the proliferation and invasion abilities of colon cancer cells by targeting quaking. *Mol. Med. Rep.* 11, 2355–2359 (2015).

233. Sargent, C. Y., Berguig, G. Y. & McDevitt, T. C. Cardiomyogenic Differentiation of Embryoid Bodies Is Promoted by Rotary Orbital Suspension Culture. *Tissue Eng. Part A* 15, 331–342 (2009).

234. Subramania, S. *et al.* SAM68 interaction with U1A modulates U1 snRNP recruitment and regulates mTor pre-mRNA splicing. *Nucleic Acids Res.* 47, 4181–4197 (2019).

235. Ingolia, N. T. Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol.* 470, 119–142 (2010).

236. Narita, N., Bielinska, M. & Wilson, D. B. Cardiomyocyte differentiation by GATA-4-deficient embryonic stem cells. *Dev. Camb. Engl.* 124, 3755–3764 (1997).

237. Chen, Y. *et al.* Novel trigenic CACNA1C/DES/MYPN mutations in a family of hypertrophic cardiomyopathy with early repolarization and short QT syndrome. *J. Transl. Med.* 15, 78 (2017).

238. Boczek, N. J. *et al.* Identification and Functional Characterization of a Novel CACNA1C-Mediated Cardiac Disorder Characterized by Prolonged QT Intervals With Hypertrophic Cardiomyopathy, Congenital Heart Defects, and Sudden Cardiac Death. *Circ. Arrhythm. Electrophysiol.* 8, 1122–1132 (2015).

239. Arshad Rafiq, M., Koopmann, T. T., Zahavich, L. A., Fatah, M. & Hamilton, R. M. An Inherited Arrhythmia Syndrome with Long QT, Sudden Death and Depolarization Disorder Due to an In-Frame

Deletion in Exon 16 of the CACNA1C Gene. *J. Gener. Seq. Appl.* 04, (2017).

240. Splawski, I. *et al.* Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell* 119, 19–31 (2004).

241. Splawski, I. *et al.* Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8089–8096; discussion 8086-8088 (2005).

242. Antzelevitch, C. *et al.* Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death. *Circulation* 115, 442–449 (2007).

243. He, J. J., Henao-Mejia, J. & Liu, Y. Sam68 functions in nuclear export and translation of HIV-1 RNA. *RNA Biol.* 6, 384–386 (2009).

244. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods San Diego Calif* 37, 376–386 (2005).

245. Li, X. *et al.* Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol. Cell* 67, 214-227.e7 (2017).

246. Wilusz, J. E. A 360° view of circular RNAs: From biogenesis to functions. *Wiley Interdiscip. Rev. RNA* 9, e1478 (2018).

247. Tan, W. L. W. *et al.* A landscape of circular RNA expression in the human heart. *Cardiovasc. Res.* cvw250 (2017). doi:10.1093/cvr/cvw250