MASTER EN BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE ET CELLULAIRE

Contribution à la caractérisation de l'expression et des fonctions du microARN miR-767 chez la souris

Svensek, Olivier

Award date:
2015

Awarding institution:
Universite de Namur

Université de Namur
FACULTE DES SCIENCES
Secrétariat du Département de Biologie
Rue de Bruxelles 61 - 5000 NAMUR
Téléphone: + 32(0)81.72.44.18 - Téléfax: + 32(0)81.72.44.20
E-mail: joelle.jonet@unamur.be - http://www.unamur.be

Contribution à la caractérisation de l'expression et des fonctions du microARN miR-767 chez la souris

SVENSEK Olivier

Résumé

Une étude récente a identifié dans des cellules cancéreuses humaines un transcrit alternatif du gène *GABRA3*, produit au départ d'un promoteur alternatif activé par déméthylation de l'ADN. miR-767, un microARN co-transcrit avec *GABRA3* et qui régule l'expression des protéines TET, pourrait jouer un rôle dans le développement cancéreux. L'objectif du présent mémoire est de déterminer si l'expression de *Gabra3* et de miR-767 chez la souris est semblable à ce qui a été observé chez l'homme, et de contribuer à l'étude des fonctions de miR-767. Des expériences de RT-PCR sur tissus de souris ont montré une expression de *Gabra3* dans le cerveau, mais aussi une faible expression dans d'autres tissus, dont le testicule. L'effet de la déméthylation de l'ADN sur l'expression de *Gabra3* a été étudié dans des fibroblastes murins (NIH3T3) traités à la 5'-aza-2'-deoxycytidine, un agent déméthylant. Une faible induction du gène a ainsi pu être observée. En essayant d'identifier un transcrit activé par déméthylation, un nouveau transcrit de *Gabra3* a été identifié dans le cerveau, codant potentiellement pour une protéine allongée du côté N-terminal. Parallèlement, la mise au point d'une détection de la forme mature de miR-767 a été entreprise. Bien que la technique ait été validée par transfection de cellules avec des vecteurs d'expression de miR-767 et par la détection de contrôles synthétiques, elle reste insuffisante pour détecter l'expression endogène du miARN. Enfin, des outils ont été développés afin de modifier de manière ciblée le locus génomique de miR-767 dans des cellules embryonnaires de souris. Des mutations « perte de fonction » ont été obtenues à l'aide de la méthodologie CRISPR/Cas9. De plus, un vecteur est en cours de construction, en vue d'obtenir un allèle knockout conditionnel de miR-767 comprenant un gène rapporteur par recombinaison homologue. Des analyses sur ces cellules et sur des souris transgéniques dérivées de celles-ci devraient permettre d'étudier l'expression et les fonctions de miR-767 in vivo.

Mémoire de master en biochimie et biologie moléculaire et cellulaire

Janvier 2015

Promoteur: O. De Backer

Thanks to Professor Olivier De Backer, for his welcome in his team, his advised guidance and his thrust during all my work in his laboratory.

▫ ◦ □ ◦ ▫

Thanks to G. Van Beersel, D. Hermand, A. Wanet and J. Malaisse for the time and consideration given for the present document.

▫ ◦ □ ◦ ▫

Thanks to Professors X. De Bolle and Benoît Muylkens for their kind technical collaboration.

▫ ◦ □ ◦ ▫

Thanks to François and Elise, for their availability and precious mentoring during the first months of my work.

▫ ◦ □ ◦ ▫

Thanks to Olivia, Dominique, Coco, Typhanie and Cédric, of the laboratory of Genetics, for their precious help and daily support.

▫ ◦ □ ◦ ▫

Thanks again to the members of the laboratory of Genetics, to those of the laboratory of cancer molecular biology (LBMC) and to Max, Lucie and Caroline, my master-student mates, for providing me such a warm and friendly working environment.

▫ ◦ □ ◦ ▫

Thanks to my parents, my grandparents, my brother Sébastien, my sister Laura and my girlfriend Laurine for their priceless love, their moral support and their everlasting, unconditional, and sometimes exasperating confidence in my success.

▫ ◦ □ ◦ ▫

Thanks to Boris, Simon and my closest friends, simply for being there, for lightening my mind in times of doubt and for the rays of sunshine that are our extra-scholar activities.

▫ ◦ □ ◦ ▫

Thanks to Mrs. Blaude, Mrs. Plume, Mrs. Gérard and Mrs. Allard, four of my secondary school teachers, for first transmitting me the taste for Science and Biology.

▫ ◦ □ ◦ ▫

Thanks to Wikipedia.org, which I swore to thank here for its precious readily accessible all-knowingness during the dark hours of past examination sessions in bachelor degree.

TABLE OF CONTENTS

1. INTRODUCTION

Cancer is often regarded as a disease caused by an accumulation of mutations in genes coding for regulators of cell proliferation and survival. This, however, overlooks the epigenetic modifications that occur in cancer cells. It has indeed been shown in recent years that epigenetic mechanisms can be very important contributors to cancer development. This master thesis focuses on miR-767, a microRNA recently discovered to be activated by hypomethylation in cancer cells. In humans, miR-767 might have a function in tumorigenesis, by targeting mRNAs encoding TET enzymes, which are involved in DNA demethylation.

1.1. Cancer and DNA methylation (see [1][2] [3] for reviews)

The main distinctive characteristic of cancer cells, in contrast to normal cells, is their ability to proliferate indefinitely without appropriate control, so that they finally form a tumor. To acquire this capacity, they have to shut off or bypass cell cycle regulatory mechanisms that normally decide whether a cell should divide or not, in response to external and internal signals. This is commonly attributed to mutations in key effectors of those mechanisms, namely the proto-oncogenes and the tumor suppressor genes. Mutations of proto-oncogenes are generally "gain of function", and associated with an increased abundance or exacerbate activity of the protein products (the proto-oncogene becomes an "oncogene" in this context). In contrast, mutations of tumor suppressors are "loss-of function" mutations, causing decreased expression or loss of activity.

A cancer cell, however, differs far more from a normal cell than just by modifications of the DNA sequence. Indeed, the epigenetic state of cancer cells, including their DNA methylation patterns, are often dramatically altered. Both DNA hypomethylation and hypermethylation have been reported. Hypermethylation mainly affects CpG islands of specific promoters, and is associated with a stabilization of transcriptional repression. Genes coding for tumor suppressors or for effectors of DNA repair pathways can be affected in this way, often in a tumor-type-specific manner [4][5][6]. Genes marked by bivalent (both active and passive) histone marks in pluripotent and multipotent stem cells, often involved in development and lineage commitment, seem to be recurrent targets of hypermethylation as well [7][8]. Those genes are expressed at low levels in undifferentiated embryonic stem cells (ESCs), but are poised to activation upon appropriate differentiation signals [9]. Besides CpG islands, alteration of DNA methylation patterns commonly affects CpG shores, which are conserved sequences flanking the CpG islands. These sequences have been shown to bear most of the tissue-specific DNA methylation at the scale of the genome [10]. Cancer-related changes in DNA methylation of CpG shores are also highly correlated with gene expression, and consist of both hypo- and hypermethylation. Of note, those methylation changes correspond to the same regions that are differentially methylated between tissues and involved in differentiation [10][11].

Hypomethylation, on the other hand, is massive and global. Mainly affecting repetitive sequences, it is associated with a loosened chromatin state and promotes DNA rearrangements, as well as reactivation of silenced mobile DNA elements, thereby contributing to genome instability [12][13]. Changes in global gene expression following DNA demethylation could also be mediated by the "partially methylated domains" (PMDs). PMDs are large DNA sequences with a relatively low methylation level. They are found and organized in a tissue-specific fashion in somatic cells, but are nearly completely methylated in pluripotent stem cells [14][15]. Moreover, while they contain few or silenced genes, their locations correlate with nuclear lamina-associated domains, silent regions binding the nuclear membrane [16]. This suggests a role for the PMDs in the establishment of the cell-specific three-dimensional organization of the genome, and an important potential influence on gene expression in a cancerous hypomethylation context. Localized hypomethylation has also been described, but is less understood. It has been hypothesized that demethylation of CpG islands in the promoters of oncogenes could increase their expression, but no evidence of this has been reported so far.

It is important to note that the links between DNA methylation and gene expression are only partially resolved, are dependent of the role of the targeted region and must be regarded in a broader epigenetic context. For example, methylation within gene bodies, by contrast to CpG islands, is generally associated with transcriptional activity rather than with repression [17]. DNA methylation is also found on regulatory sequences, where it can alter binding of transcriptional regulators [18][19]. The downstream effects depend on whether it modulates binding of enhancers or repressors. Furthermore, DNA methylation can indirectly alter regulation of gene expression at the translational level, by modulating expression of microRNAs (or miRNAs). Involvement of miRNAs in malignant transformation is increasingly well-documented (see below). Then, DNA methylation does not play alone. It interacts with numerous epigenetic effectors (enzymes catalyzing histone modifications, for example), whose intricate interplay determines the chromatin architecture and transcription state of DNA [6].

Even though it seems that upregulation of oncogenes is not the primary consequence of hypomethylation, this latter does activate some genes, called "cancer-testis" (CT) or "cancer germline" (CG) [20][21][22]. Expression of these genes is indeed primarily regulated by DNA methylation. They are silenced in most cell types, except in the germline and in cancer cells, where hypomethylation is sufficient to induce their expression. Among the CG category are found the GAGE [23] genes and some members of the MAGE gene family [23]. Interestingly, CG genes are almost exclusively located on the X chromosome. Of great interest from a therapeutic point of view is their potential ability to be used as cancer-specific antigens in immunotherapy [24]. Some of them are suspected to have oncogenic functions.

Thus, reshuffling of the genome methylation patterns is very significant in cancer etiology and offers interesting targets for treatment [2][25]. Four epigenetic drugs are already approved by the FDA, two of them (azacytidine and decitabine) interfering with the DNA methylation process. Many other drugs are undergoing clinical tests and may reach the market in the future. Furthermore, new therapeutic opportunities could arise from the recent discovery of a new family of enzymes involved in active demethylation: the TET proteins.

1.1.1. DNA methylation pathways and the TET proteins (see [26][27][28] for reviews)

1.1.2. DNA (de)methylation mechanisms

Two kinds of epigenetic "writers" regulate DNA methylation: those that add new methyl groups or maintain them, and those that contribute to their removal. DNMT1, DNMT3A and DNMT3B are members of the DNA methyl transferases family and belong to the first category. They catalyze the transfer of a methyl group from S -adenosyl- l -methionine (SAM) to an unmethylated cytosine. DNMT1 has a maintenance activity and acts on hemimethylated DNA substrates, while DNMT3A and DNMT3B have de novo methyltransferase activities. Dysregulation of these proteins are commonly found in cancer [29].

The two FDA-approved drugs aforementioned, 5'-azacytidine and 5'-aza-2'-deoxycytidine (5-azadC, or decitabine), induce global DNA hypomethylation, persistent for days after treatment [25][30]. As nucleotide analogues, both compounds can be incorporated in DNA and form a covalent complex with DNA methyltransferases. Though clinically poorly efficient on solid tumors, they give interesting results on acute myeloid leukemia (AML) and myelodysplasic syndrome (MDS) patients, restoring expression of tumor suppressor genes, sensitivity to apoptosis and proliferation control. Decitabine is also used in culture to induce demethylation and activation of numerous genes, in human as well as mouse cells [31][32].

Recently, the discovery of the TET proteins family, composed of three members, TET1, TET2 and TET3, shed light on the DNA demethylation process. These α-ketoglutarate (α-KG)-dependent dioxygenases catalyze the conversion of 5-methylcytosine (5mC) to

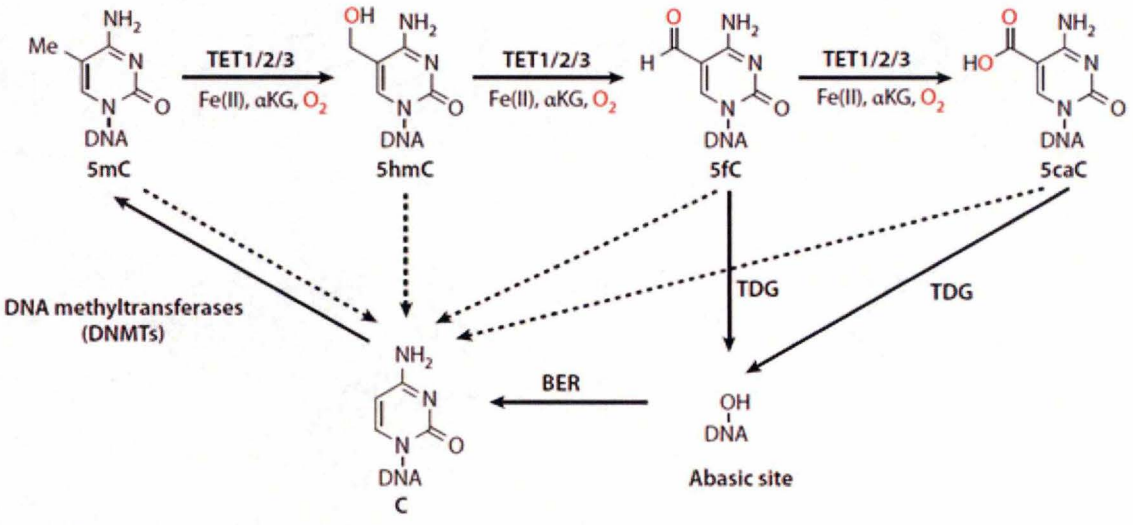

Figure 1. Currently proposed mechanisms for DNA demethylation. The TET dioxygenases catalyse oxidation of 5mC into 5hmC, 5fmC and 5caC. These reactions require the presence of Fe(II) and αKG cofactors. A return to unmodified cytosine is then possible through passive dilution (represented by the dotted arrows), TDG-mediated glycosylation and BER. [26]

5-hydroxymethylcytosine (5hmC) (figure 1) [33]. Their activity relies on the presence of Fe(II) and of an α-ketoglutarate (α-KG) cofactor, produced by isocitrate dehydrogenase enzymes (IDH1 and IDH2). Further oxydation steps catalyzed by the same enzymes can lead to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [34]. Those marks may be specifically recognized by distinct DNA binding proteins and therefore play distinct epigenetics functions on their own. To complete the process towards an unmodified cytosine, two possibilities have been documented. The first one is a replication-dependent passive dilution, as the modified cytosine is no more recognized by the methylation maintenance machinery [35]. The second one is active demethylation, whose exact mechanisms have not yet been fully deciphered. A carboxylase may convert 5caC to an unmodified cytosine, though such enzyme have not been identified so far [36]. DNMT3A and DNMT3B have also been shown to mediate demethylation of 5hmC in absence of a SAM donor, although this may be irrelevant given that SAM is present in all cell types [37]. Thymine-DNA glycosylase (TDG) is currently the best established mediator toward unmodified cytosine, generating an abasic site subsequently repaired through base excision repair (BER). TDG have indeed been shown to have a base excision activity for 5fC and 5caC [38][39]. Glycosylation of 5hmC to 5hmU by AID/APOBEC deaminases followed by BER have also been hypothesized, but is not yet certain. Overexpression of those deaminases together with TET1 has indeed been shown to promote 5hmC demethylation in the adult mouse brain [40].

1.1.3. Biological functions of the TET proteins

Initially, 5hmC was discovered in Purkinje neurons and ESCs, but since then newly developed detection methods have brought considerable information on its genome-wide distribution in various cell types, tissues and conditions, as well as on the distribution of its 5fC and 5caC derivatives [26]. As well, biological importance of the demethylation process in embryonic development, in stem cells and in post-mitotic cells is gradually being highlighted.

1.1.3.1.Embryonic development and epigenetic reprogramming

All three TET proteins are expressed and play major roles at different steps of embryonic development. TET3 is the only one to be upregulated in the one-cell zygote, whereas TET1 and TET2 are most abundant in the inner cell mass (ICM) of the blastocyst [27]. From fertilization to developing embryo, two waves of global DNA demethylation occur, each one involving TET proteins. The first one takes place after fertilization, during preimplantation development, and affects both the paternal and the maternal chromosomes (figure 2A). Gu et al (2011) pointed out the role of TET3 in demethylation occurring in the male pronucleus [41]. While loss of 5mC in the paternal pronucleus is rapid and coincides with gain of 5hmC, suggesting active demethylation, it is replication-dependent in the maternal pronucleus, indicating a passive process. TET3 is abundant in the oocyte cytoplasm and is rapidly localized to the male pronucleus where it appears to mediate 5mC oxidation. The resulting 5hmC are then gradually lost in a replication-dependent manner from the two-cell stage and beyond [42]. TET3-mediated

DNA demethylation seems to be prevented in the maternal pronucleus by PGC7 (and likely by additional factors), recruited by a maternally-enriched histone mark [43].

Gu et al (2011) produced conditional knockout (KO) mice with specific deletion of *Tet3* in the germline. With this model, they demonstrated that while active demethylation occurs normally in zygotes derived from a WT oocyte and a KO sperm cell, it is blocked in *Tet3*-deficient oocytes. Interestingly, demethylation and reactivation of the paternal genes coding for embryonic stem cell factors, such as *Oct4* and *Nanog*, is impaired in these embryos. Those genes are silenced during male gametogenesis, and thus have to be reactivated for embryonic development. [41]

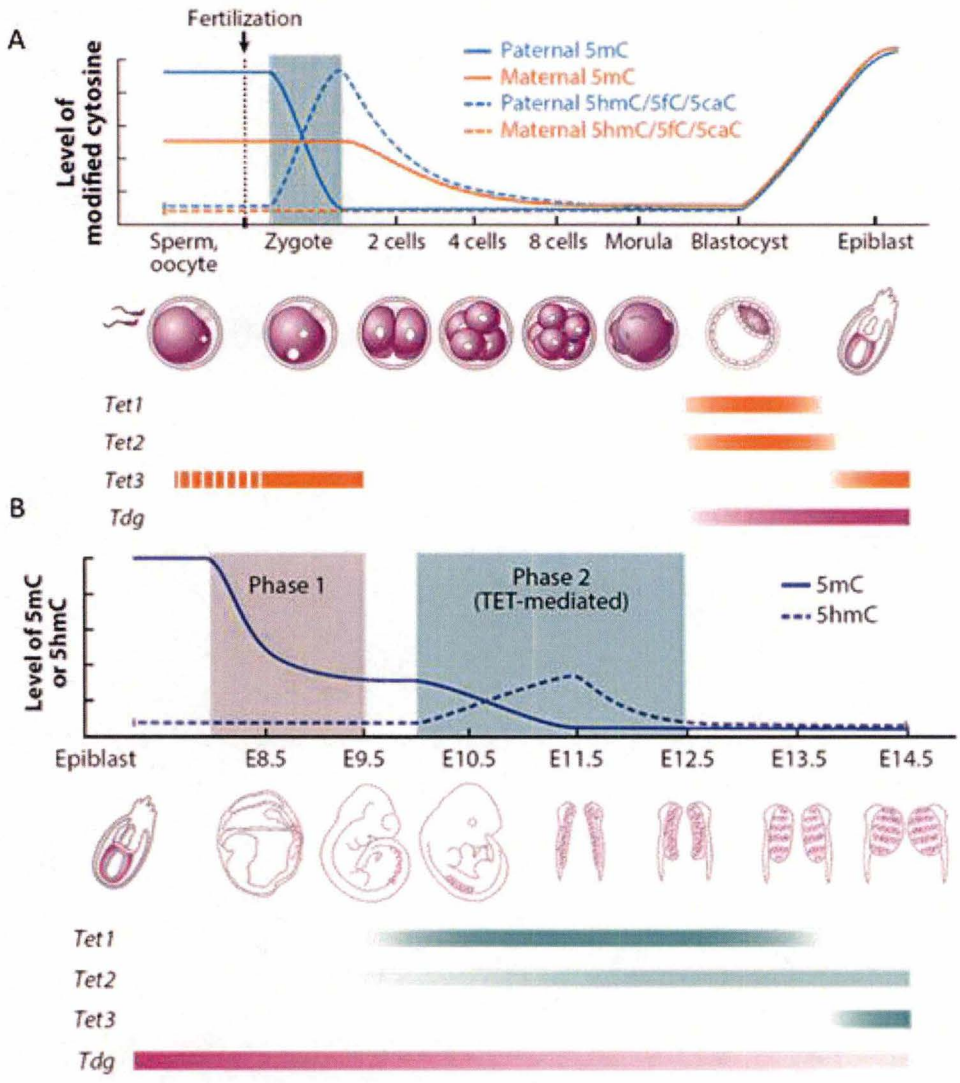

Figure 2. DNA demethylation and TET expression in pre-implantation embryo and in PGC. A) DNA demethylation dynamics in pre-implantation embryo. Modified cytosine levels first drop in the paternal genome after fertilization because of TET3 activity. Both genomes then undergo passive demethylation and remethylation in the ICM. B) Illustration of the 2 phases of demethylation in PGCs, partially mediated by TET proteins. [26]

TET3 does not seem essential for the generation of functional gametes, as specific deletion of *Tet3* in the germline precursors does not affect their epigenetic reprogramming (see below), nor maturation of the gametes and their fertilization capacity. By contrast, females producing *Tet3*-deficient oocytes display reduced fertility due to post-implantation development failure and morphological abnormalities appearing from midgestation. [41]

After this first zygotic DNA demethylation wave, a new methylation pattern is gradually established as cells differentiate in the inner cell mass. While most embryonic cells gradually lose their pluripotency to form somatic tissues, a second wave of demethylation occurs in the primordial germ cells (PGCs) (figure 2B) [44][45]. Those cells are specified from epiblast cells and migrate to the genital ridges to differentiate in the gonads to give rise to gametes, in a sex-specific manner [45]. Epigenetic reprogramming occurs in PGCs, notably to erase methylation imprints inherited from the parents, in order to properly reset new imprints according to the sex of the developing embryo. This second wave occurs in two phases, the first one consisting of a replication-dependent loss of 5mC due to reduced maintenance activity [46]. The second takes place when the cells enter the genital ridges, and coincides with upregulation of TET1 and increased 5hmC marks. It seems that after active oxidation of 5mC by TET1 (and perhaps TET2), 5hmC marks disappear passively through the next rounds of replication [44][47]. It is during this last phase that imprinted genes are demethylated, along with gametogenesis-related genes and CpG islands of the X chromosome [48]. TET1 seems particularly important for imprint erasure, as heterozygous offspring of KO males display embryonic failure related to imprinting abnormalities [49].

1.1.3.2.ESCs and primary differentiation in the 3 germ layers

In ESCs (derived from the ICM), where TET1 and TET2 are expressed, 5mC oxidation products are abundant and enriched at poised regulatory elements and promoters of lineage-specific genes, suggesting a role of TET1 and TET2 in the regulation of these genes in pluripotent stem cells [50][51][52]. By regulating key factors of pluripotency and differentiation such as NANOG, NODAL or LEFTY, they appear to ensure maintenance of an undifferentiated state. Interestingly, both TET1 and TET2 seem to influence differentiation fates [51]. Depletion of TET1 in mESC leads to an overexpression of trophectoderm markers and a decrease of neuroectoderm markers. By contrast, TET2 depletion induces expression of neuroectoderm markers. Injected in animals, ESC can form heterogeneous tumors named teratomas, composed of cells derived from the three embryonic layers. Teratoma formation as well as injection in mouse blastocysts with the same mESC confirmed in vitro results, with a tendency of TET1-depleted cells towards mesoderm, endoderm and even trophectoderm formation, and a tendency towards neuroectoderm differentiation for TET2-depleted cells [51].

Importantly, both proteins are involved in the reprogramming of induced pluripotent stem cells (iPSCs). It is now well-known that differentiated cells can be reprogrammed into pluripotent stem cells by ectopic expression of an association of key transcription factors. The initial Yamanaka cocktail comprised genes encoding 4 factors, namely OCT3/4, SOX2, c-MYC, and KLF4 [53]. Addition of TET1 or TET2 to this cocktail facilitate reprogramming, as these enzymes demethylate *Oct4* and other genes, and physically interact with NANOG. This latter,

another important factor for pluripotency, recruits TET1 and TET2 on DNA binding sites. Enzymatic activity is required for this effect, as overexpression of a catalytically inactive *Tet1* transgene is only able to facilitate reprogramming as long as endogenous TET2 activity is preserved [54]. It has also been shown that *Tet1* can efficiently replace *Oct4* for iPSC induction [55].

1.1.3.3.Toward differentiated somatic tissues

Upon differentiation, TET1 and TET2 levels drop. In somatic tissues, TET2 remains important for the maintenance and differentiation of the hematopoietic lineage [56]. TET3 is the most abundant member of the family in somatic tissues, with a low but ubiquitous expression. It is however expressed at high levels in the brain and in germ cells. In the brain, TET3 seems to contribute to neurogenesis and neural progenitor cells (NPCs) maintenance. In mammals, NPCs are located in the subventricular zone of the brain and contribute to adult neurogenesis [57]. Deletion of *Tet3* in mESCs leads to impeded neuronal differentiation. Furthermore, NPC derived from those cells display increased apoptosis and reduced neuron formation [58].

Active demethylation has a particular role in mature neurons, where stimulation induces expression of numerous neuronal activity-regulated genes. This phenomenon was identified before discovery of the TET proteins and is believed to play a role in memory formation [59]. The first observations have been made in hippocampus, a part of the brain essential for establishment of long-term, short-term and spatial memory. Neuronal activity triggers both DNA methylation and demethylation processes, which represses "memory-suppressor" genes and induces plasticity-promoting genes such as *Bdnf*, coding for a neurotrophin. After learning, long-lasting changes in DNA methylation patterns are also observed in the cortex, which is believed to be the storage site for long-term memories after learning [59]. A key behavioral test to assess memory formation in mice is fear conditioning [60]. Briefly, the animals learn to associate a given environmental context to a displeasing stimulus, and develop a particular fear behavior when they recognize this context. Several behavioral tests have been derived from this basic principle to investigate the mechanisms of memmory.

Recently, TET1 has been shown to be induced after neuronal stimulation in the hippocampus, regulating target genes and contributing to memory formation and extinction [61]. TET1 overexpression promotes conversion of 5mC to 5hmC in human cells and increases expression of *Bdnf* and other factors of plasticity in mouse brain [40][61]. Mice overexpressing Tet1 in the hippocampus display learning abilities similar to control animals in the fear conditioning paradigm, but display a faster memory extinction. Of note, the catalytic activity of the protein does not seem to be strictly required for this phenotype [61]. In *Tet1* KO mice, 5hmC level is slightly decreased and several genes regulated by neuronal activity are downregulated [62]. Similarly to *Tet1* overexpressing mice, no evident phenotypic changes have been reported in fear memory acquisition. In that case, however, memory extinction is delayed. Interestingly, by contrast to hippocampal neurons, only TET3 levels rise in cortical neurons upon neuronal stimulation in vitro and during fear extinction in vivo [63].

As for TET3, a role in NPC and neuronal differentiation has been shown for TET1. In *Tet1* KO mice, hypermethylation and altered expression of genes regulating neurogenesis leads to impaired proliferation and NPC maintenance, with downstream effects on spatial learning and memory formation. By contrast to *Tet3* KO, however, viability and differentiation potential do not seem affected. [64]

1.1.4. TET proteins and cancer

In the cancer context, the TET proteins generally have tumor suppressor characteristics. As a result, their downregulation in cancer cells comes along with an overall loss of 5hmC [65]. As an exception to the rule, the first evidence of the involvement of these enzymes in cancer was the identification of an oncogenic TET1-MLL fusion protein in rare cases of AML [66]. Of note, the TET acronym (Ten-Eleven-Translocation) arises from this translocation event between *MLL* on chromosome 11 and *TET1* on the chromosome 10. A recent study pointed out the oncogenic function of TET1 in MLL-rearranged leukemia, regardless of the identity of the fusion partner [67]. In those leukemia, *TET1* is significantly upregulated by MLL (particularly by MLL-fusion proteins), which is accordingly enriched in the promoter region of the gene. TET1 and MLL then share several target genes. As many MLL partners are part of chromatin-modifying complexes, it is hypothesized that fusion with one of those partners enhances their synergism, leading to oncogenicity. Among the gene targets shared between MLL and TET1 were found *HOXA9*, *MEIS1*, and *PBX3*, whose activation at least partially mediates the oncogenic effect of TET1. To support the proposed oncogenicity of TET1 in that kind of leukemia, KO of *Tet1* resulted in reduced transformation of MLL-rearranged cells. *Tet1* knockdown reduced viability of those cancerous cells in vitro and delayed development of leukemia in mice [67].

TET2 is to date the member of the TET family which has been the most closely related to cancer. TET2 is indeed very important for the hematopoietic lineage differentiation, and its dysregulation is frequently associated with myeloid malignancies [56][68]. Alteration of TET2-mediated demethylation also commonly results from mutations in IDH1 (isocitrate dehydrogenase), IDH2, FH (fumarate hydratase) or SDH (succinate dehydrogenase). These mutations cause accumulation of α-KG analogs (2-hydroxyglutarate, fumarate and succinate, respectively) that block the α-KG cofactor binding site and therefore inhibit TET activity. In solid tumors, the impact of *TET* mutations is less obvious, but levels of all three proteins are often reduced. In various types of tumors, including melanoma, breast, prostate, lung and liver tumors, TET levels are significantly reduced, with a correlated decrease in 5hmC and with a negative impact on prognosis [69][70]. In line with this, TET1 has been shown to prevent metastasis in prostate and breast cancers by inducing expression of the tissue inhibitors of metalloproteinase (TIMP) family proteins [71]. Those proteins are indeed known to modulate matrix metalloproteinase (MMP) proteins, which are important drivers of metastasis.

Importantly, numerous miRNAs have been shown to regulate TET expression and to be deregulated in cancer (see below).

1.2. MicroRNAs and cancer

microRNAs dysregulation is another source of plasticity exploited by cancer cells to reshuffle gene expression and to improve their competitive edge. Most human protein-coding genes can indeed be targeted by one or more miRNAs, and a given miRNA may regulate numerous target genes. The biogenesis of those small non-coding RNAs is much conserved and has been studied in a wide range of species (figure 3). miRNAs transcription is carried out by RNA Pol II. Some miRNAs are embedded in mRNAs, whereas others are "intergenic" and are part of non-coding transcripts. In the first case, they generally share the promoter of their host gene, although it has been shown that they often have more than one transcription start site (TSS) and can be transcribed on their own [72]. miRNAs belonging to this category are usually intronic and do not prevent host gene processing, but others that are located in exons can destabilize their host RNA. In all cases, they are frequently organized in co-transcribed clusters. Transcription by RNA Pol II leads to a long pri-miRNA, containing one or several stem-loop structures. These structures are recognized by the microprocessor complex (composed of RNAse III Drosha and its cofactor DGCR8), which processes them into individual ~70 bp-long pre-miRNAs. Pre-miRNAs are then translocated through the nuclear pore complex by the protein exportin 5. [73]

In the cytoplasm, pre-miRNAs are cleaved by the RNAse II endonuclease Dicer (which interacts with the protein TRBP, bearing a dsRNA binding domain) at the level of their terminal loop to form an RNA duplex. This latter can subsequently be loaded on argonaute (AGO) proteins to form the RNA-induced silencing complex (RISC). For the complex to become effective, one of the RNA strands, the passenger strand (also named miRNA*), is removed. Classically, the strand displaying the lowest binding stability with its complementary strand at its 5' end is conserved. Regulation of miRNA biogenesis is possible at every step from transcription initiation to assembly of the effective RISC complex, on every protein intervening in the process and on the miRNAs themselves. Once fully processed, they regulate expression of their target mRNAs, typically through binding complementary sequences located in the 3' UTR, inducing translational repression and decay of the target mRNA. The miRNA-mRNA pairing is most of times only partially complementary, except for the miRNA "seed", a domain spreading from nucleotide position 2 to 7 from the 5' end of the mature miRNA, which is generally a perfect match for the target. This seed region is the most reliable indicator for target prediction, and is sufficient to predict conserved targets of miRNAs [74]. [73]

During cancer development, regulation of miRNA levels can be modulated at the transcription level or in later steps [75]. Both genetic and epigenetic events influence miRNA transcription. Deletions and chromosomal rearrangements lead to loss or amplification of many miRNAs, which are often located in fragile regions [76]. DNA methylation also affects expression of CpG-associated miRNAs in various types of cancer, promoting expression of oncogenic

miRNAs and repressing tumor-suppressor miRNAs [77]. For example, miR-148a, miR-34b/c, and miR-9, belonging to the tumor-suppressor category, are repressed by hypermethylation in human metastatic cancer cells. Restoration of their expression using expression vectors reduces metastatic progression in vivo [78]. On the contrary, oncogenic miRNAs tend to be activated by hypomethylation, as has been shown in prostate cancer, where the tumor-suppressor PTEN is downregulated by several miRNAs induced by hypomethylation [79].

Epigenetic effectors themselves can be regulated by miRNAs. For example, miR-29 downregulates *DNMT3A* and *DNMT3B*, and its ectopic expression leads to reactivation of tumor-suppressor miRNAs silenced by DNA hypermethylation in cancer cells [80]. The TET

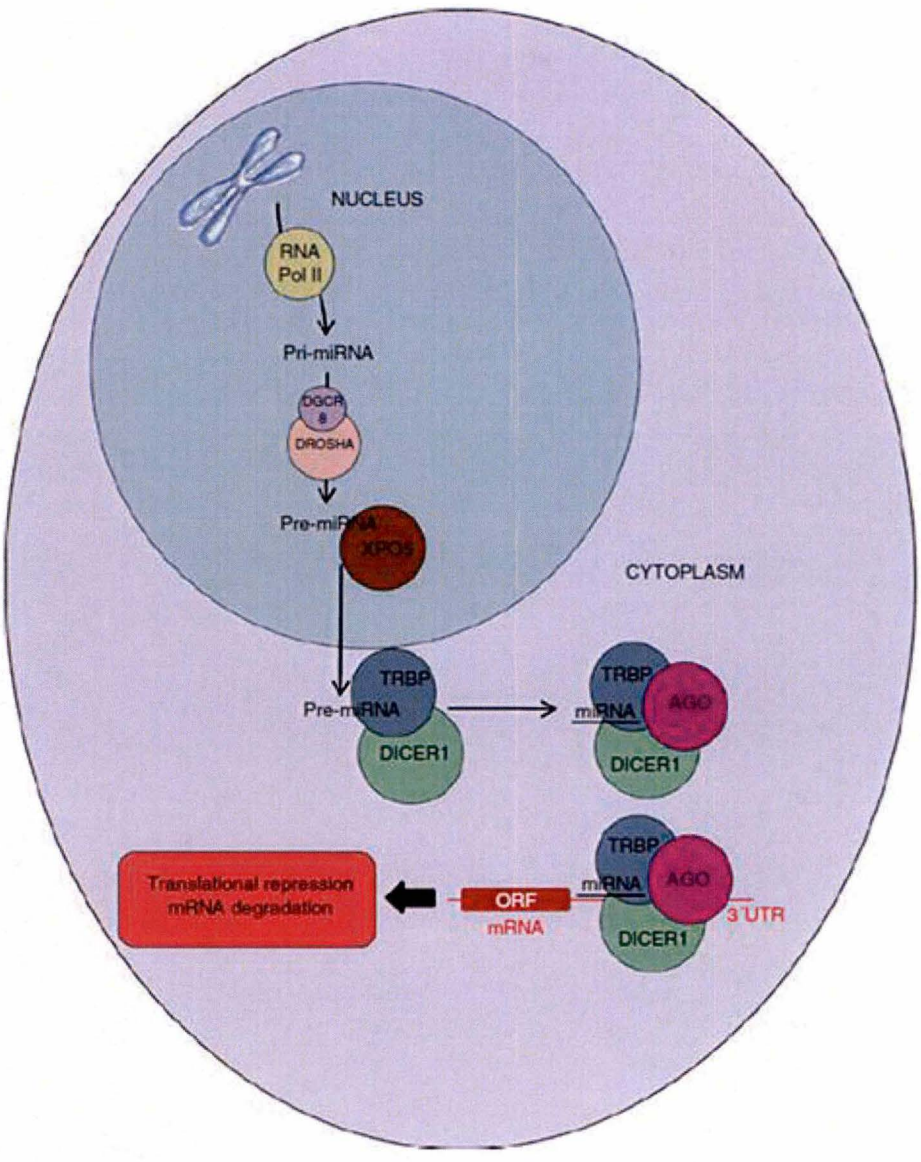

Figure 3. microRNA biosynthesis. After transcription by RNA polymerase II, the pri-miRNA is processed by DROSHA into a pre-miRNA, which is exported in the cytoplasm by exportin 5 (XPO5). There, it is cleaved by DICER and loaded on AGO proteins to form the RISC complex, which degrades or prevents translation of target mRNAs. ORF = open reading frame. [77]

proteins are no exception to the rule, being the targets of at least 30 miRNAs. The most oncogenic of them appears to be miR-22, whose high expression level in breast cancer cells is associated with upregulation of genes involved in metastasis and with decreased survival rates [81]. The mechanism by which miR-22 promotes metastasis is a striking example of interplays between TET proteins and miRNAs (figure 4). Overexpressed, it reduces translation of all three TET enzymes, with a corresponding increase in the levels of 5hmC. This in turn causes repression of miR-200, a tumor-suppressor miRNA, by impaired demethylation of CpG islands located in its promoter. miR-200 exerts its tumor suppressive functions by targeting transcriptional repressors, thereby limiting epithelial-mesenchymal transition (EMT) and stemness in cancer cells. Silencing of the TET-miR-200 axis therefore promotes stemness acquisition, EMT, invasion and metastasis [81]. Overexpression of miR-22 also promotes hematopoietic malignancies through repression of *TET2* [82][83]. mir-29 is also able to modulate TET levels and DNA demethylation process, but seems to have a protective function regarding tumorigenesis, though its overexpression has been observed in AML, along with other *TET2*-targeting miRNAs [84][85].

Evidences of miRNA deregulation in various diseases have initiated research for therapeutic strategies [86][87]. Several approaches are being developed, including the use of inhibiting antisense oligonucleotides (anti-miRs). However, these RNA drugs are facing multiple difficulties in terms of administration, toxicity, delivery and stability. Nevertheless, some therapeutics have entered clinical development, including for cancer. MRX34, the first to enter clinical test, is currently in Phase I testing on patients with liver cancer or hepatic metastasis [88]. It mimics the miR34 family, frequently lost in cancer, by targeting numerous oncogenes. Similarly, Let-7, another potential candidate for miRNA mimic development, is under study.

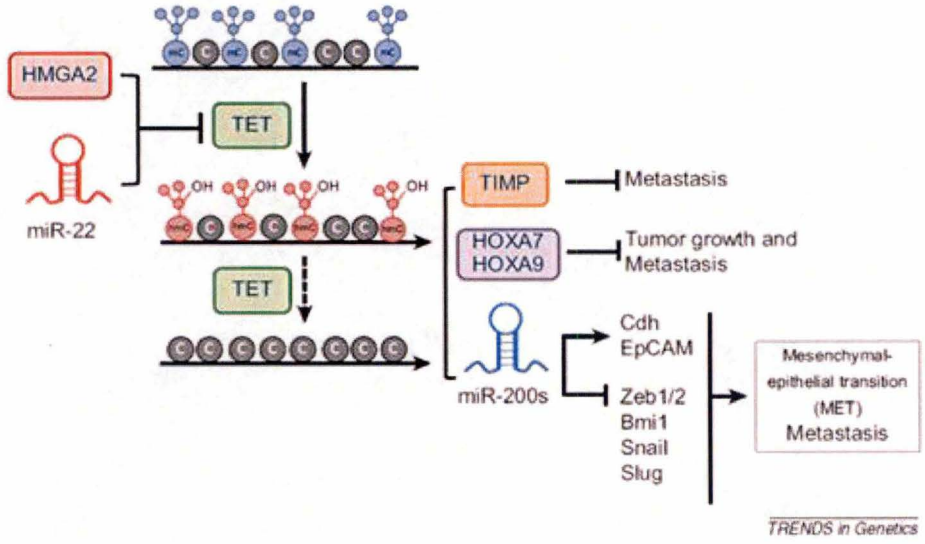

Figure 4. Interplays between TET proteins and miRNA regulation. In breast cancers, overexpression of the oncogenic miRNA miR-22 results in downregulation of the TET proteins and of TET-regulated tumor-suppressor genes. miR-200, a tumor-suppressor miRNA, is also downregulated. This promotes metastasis, tumor growth and MET. (HMGA2 is another gene upregulated in breast cancers with similar effects) [65]

By contrast, an inhibitor is being developed for miR-21, an oncomiR that target *PTEN*, among other tumor suppressors.

1.3. A cancer-germline transcript for *GABRA3*

As described in the previous section, molecules targeting miRNAs have interesting therapeutic potential. A limitation of this approach, however, is the possible off-target effects, as a given targeting treatment is very likely to affect several miRNAs of the same family in several organs. Therefore, the more a miRNA has an expression restricted to cancer cells, the more it is an interesting target for a potential treatment. Given that miRNA expression can be altered along with modifications of DNA methylation patterns, C. De Smet and his collaborators hypothesized that hypomethylation could induce miRNA expression in a cancer-germline manner, as it does for CG protein coding genes mentioned above [89]. They screened databases for human miRNAs specifically expressed in testis (and in no more than one other tissue) and located on the X-chromosome. Among these, they focused on miR-105 and miR-767, both located in the first intron of the gene *GABRA3*, which encodes a subunit of a $GABA_A$ ionotropic

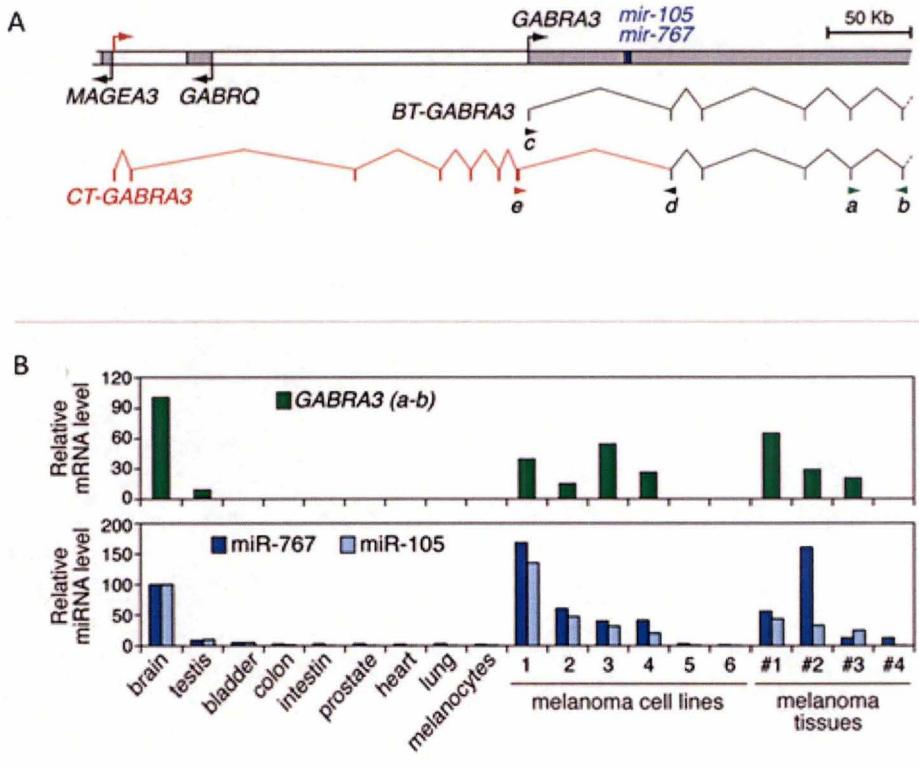

Figure 5. Expression of the alternative CG *GABRA* transcript hosting miR-105 and miR-767 in humans. A) Transcription of CG-*GABRA3* (referred here as CT, for 'cancer-testis') from an alternative TSS, under control of a bidirectional promoter activated by demethylation. B) RT-qPCR detection of *GABRA3* mRNA and of its host miRNAs in a panel of human tissues and in tumor cell lines. The primers used for *GABRA3* are shown in A. [89]

receptor (figure 5A). Interestingly, whereas *GABRA3* expression appeared restricted to brain and testis in healthy conditions, it is also ectopically expressed in several types of human cancers (figure 5B). In non-small cell lung cancer (NSLC), ectopic expression of *GABRA3* is correlated with bad prognosis [90][91][92]. An impact on cancer development could result from miRNA expression, but it could also be mediated by the GABRA3 protein itself, as has been suggested in hepatocellular carcinoma [93].

According to De Smet et al (2014), expression of human miR-767 and miR-105 mirrors expression of their host gene (figure 5B). However, while the team showed that *GABRA3* expression is restricted to brain and testis, they detected low expression of miR-105 and/or miR-767 in other tissues (bladder, colon, intestine, prostate, heart and lung). De Smet et al identified an alternative transcript for *GABRA3* that displays a cancer-germline expression profile [89]. The transcription start site (TSS) of this transcript is localized 247kb on the 5' side of canonical *GABRA3* TSS, near a bidirectional promoter activated upon demethylation and that also drive expression of the CG gene *MAGEA3* (figure 5A). The *GABRA3* CG-transcript only differs from the canonical one by the presence of 7 additional 5' exons and the absence of the canonical first exon. Importantly, the *CG-GABRA3* RNA is specifically expressed in testis (it is not detected in the brain) and is induced in cancer cells and in cells treated with the demethylating drug 5-azadC, together with miR-105 and miR767.

Based on in silico predictions, further investigation revealed that miR-767 is able to target *TET1* and *TET3* mRNAs and to reduce abundance of their encoded proteins. Transfection with luciferase reporter genes showed that miR-767 efficiently binds to the 3'UTR region of both *TET1* and *TET3* mRNAs. Regulation of endogenous TET1 and TET3 levels was confirmed by transfection with synthetic miR-767 molecules and antisense oligonucleotides. Changes of TET proteins and of 5hmC levels were assessed by western blot and slot blot assays, respectively. Finally, an inverse correlation between expressions of *GABRA3* and *TET1/3* was highlighted in lung carcinoma based on microarray expression data.

Very recently, another study pointed out and characterized a functional involvement of miR-105 in breast cancer metastasis [94]. Transferred through exosomes secretion to endothelial cells, miR-105 downregulates ZO-1, a key component of tight junctions, thereby weakening the endothelial barrier and promoting escape of cancer cells in the bloodstream. Given that miR-105 and miR-767 are co-transcribed, it would be worth investigating whether they both play a role in cancer, or if expression of one is simply driven by positive selection of the other. As evidence have been provided about the relationship between TET expression and metastasis, miR-767 could work in synergy with miR-105 to promote cancer invasiveness, migration and metastasis formation.

As a regulator of *TET1* and *TET3*, miR-767 may thus play various functions in epigenetic reprogramming, in embryonic development, in pluripotency maintenance, in activity-dependent gene regulation in neurons and in gametogenesis. For obvious practical and ethical reasons, these potential functions cannot be analyzed in humans, where their study would be limited to cultured cells. Conversely, a mouse model would enable deep and accurate analyses in a whole organism, including behavioral phenotyping. Moreover, the possibility to associate

a reporter gene with the gene of interest in transgenic mice should facilitate characterization of the tissue expression pattern of miR-767.

To date, little is known about the expression and biological functions of *Gabra3* in other tissues than the brain. Only one knockout model exists for *Gabra3*. Such a model would be inappropriate given the difficulty to discriminate phenotypes caused by the lack of protein or the lack of its associated miRNAs. No obvious developmental or morphological defect has been reported for these mice, but they display subtle behavioral phenotypes. Some of these phenotypes have been attributed to hyperactivity of dopaminergic neurons due to the absence of α3 GABA receptor, as administration of a dopamine receptor antagonist (haloperidol) restored the normal phenotype [95]. Another phenotype observed in those mice was a frustrative response in the sucrose preference test. In this test, when sucrose concentration was reduced in drinking water, consumption of the KO mice decreased, whereas consumption of the WT mice remained stable. This was hypothesized to betray frustration, as the reward was not as important as expected, and could be attributed to the lack of the GABA receptor and dopaminergic dysregulation [96]. In the same study, *Gabra3* KO mice displayed a faster memory extinction after fear conditioning. This observation could be linked to the fact that mice overexpressing TET1 also display such a phenotype. One could wonder whether it could also be attributed to *TET1* overexpression caused by a reduced expression of miR-767. However, it is likely that expression of miR-105 and miR-767 remains unchanged in the *Gabra3* KO mice, as the mutation is a duplication of exon 4, downstream of the miRNA cluster.

Thus, it would be first interesting to verify whether expression of miR-105 and miR-767 is affected in the *GABRA3* KO mice. If it is, absence of both miRNA might play a part in the phenotypes observed in these mice. If it is not, questions regarding the viability and the effects of a specific inactivation of miR-767 should be reconsidered.

1.4. The CRISPR-Cas9 system

The standard method of gene targeting by homologous recombination (HR) in mice is quite a heavy procedure [97]. It requires to construct a targeting vector with long homology arms surrounding the modified region. This vector is then transfected in mESCs where it can be used as a repair template, resulting in the replacement of the region flanked by the homology arms However, such recombination events occur at a low frequency. The CRISPR-cas9 system is an alternative method recently developed to achieve genome engineering with much higher efficiency [98].

CRISPR (for Clustered Regularly Interspaced Short Palindromic Repeats), refers to bacterial genomic loci involved in adaptative immune defense (figure 6). Those loci, constituted of a succession of repeated elements interspaced by variable sequences called spacers (CRISPR arrays), also include CRISPR-associated genes (Cas). CRISPR repeats are partially palindromic and tend to form hairpin structures, while spacers are variable sequences acquired during

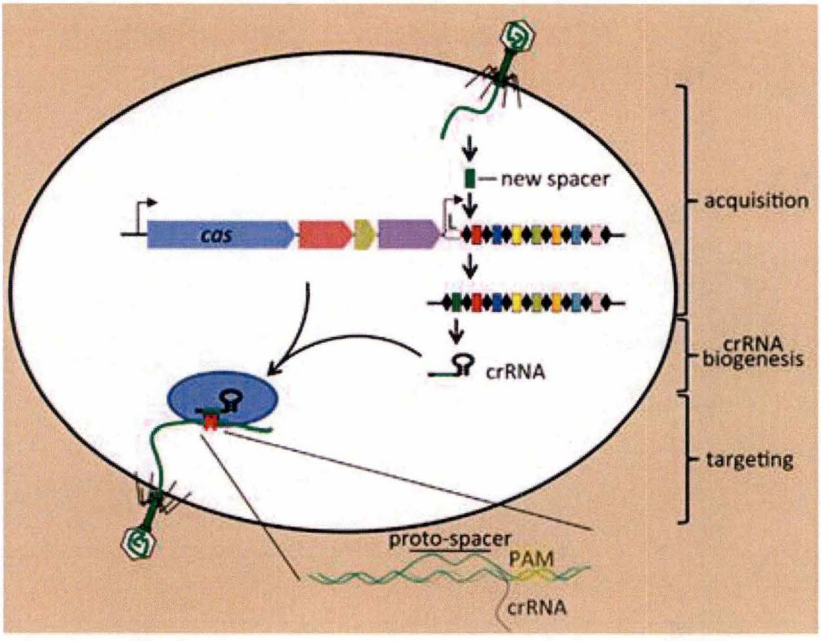

Figure 6. CRISPR system in bacterial immune defence. Clustered Regularly Interspaced Short Palindromic Repeats contain foreign sequences acquired from infecting viruses during previous infection events. Integrated in the bacterial genome, those sequences can be processed in crRNA, able to guide Cas-encoded nucleases to corresponding protospacers (in immediate proximity of a PAM) on exogenous nucleic acid molecules, which are thus degraded. [99]

previous phage infectious events. They are thus homologous to foreign genetic sequences, called protospacers. CRISPR arrays are generally transcribed as a single RNA, further processed into shorter CRISPR RNAs (crRNAs), whose function is to guide the catalytic activity of a Cas-encoded nuclease. A Cas-targeted protospacer is always associated with a protospacer adjacent motif (PAM), which varies depending on the type of CRISPR system. To summarize, crRNAs form a complex with Cas-coded nucleases, and guide them to allow a specific cleavage of a protospacer in the immediate proximity of a PAM. CRISPR-Cas systems can be sorted into three classes, namely type I, type II and type III [99].

The *Streptococcus pyogenes* Cas9 nuclease and associated CRISPR RNAs, (type II), have been manipulated to develop a new genome engineering technology [98][100][101]. This system involves two crRNAs, namely a precursor crRNA (precrRNA) which contains the spacers interspaced by direct repeats, and a trans-activating crRNA (tracrRNA). These two RNAs guide the Cas9 nuclease to protospacers immediately preceding a 5'-NGG PAM sequence (figure 7B). A codon-optimized S. pyogenes Cas9 (SpCas9) has been engineered for use of this system in mammalian cells. It has been shown that this nuclease, combined with a tracrRNA and a precrRNA, is sufficient to generate targeted double-strand breaks (DSB) in the genome of mammalian cells [100]. The precrRNA and tracrRNA can even be fused together into a single guide RNA chimera (sgRNA) (figure 7A) [100][102]. The DSB generated are preferentially repaired by the error-prone non-homologous end joining (NHEJ), frequently leading to formation of short deletions or insertions, although homology-directed repair (HDR) with a recombination template (endogenous or introduced in the cells) is also possible

[103][104][105]. Key mutations in one of the catalytic domains of SpCas9 allowed to convert the nuclease into a nickase (SpCas9n) which can still be used to induce HDR rather than NHEJ and indel formation, although recombination frequency is much reduced compared to that of the WT Cas9 [100]. The wild type Cas9 has two nuclease domains, namely RuvC and HNH. Two kinds of nickase mutants have been developed by inactivating one or the other domain. The nickase used in this work has an aspartate-to-alanine substitution (D10A) in the RuvC I domain and only cleaves the strand which is complementary to the RNA guide.

A considerable drawback of the CRISPR-Cas9 technology is its off-target activity. Although the target specificity is ensured by a more or less 20 bp sequence, single mismatches between crRNA and target DNA are deleterious for cleavage efficiency only in the 8-14 bp region

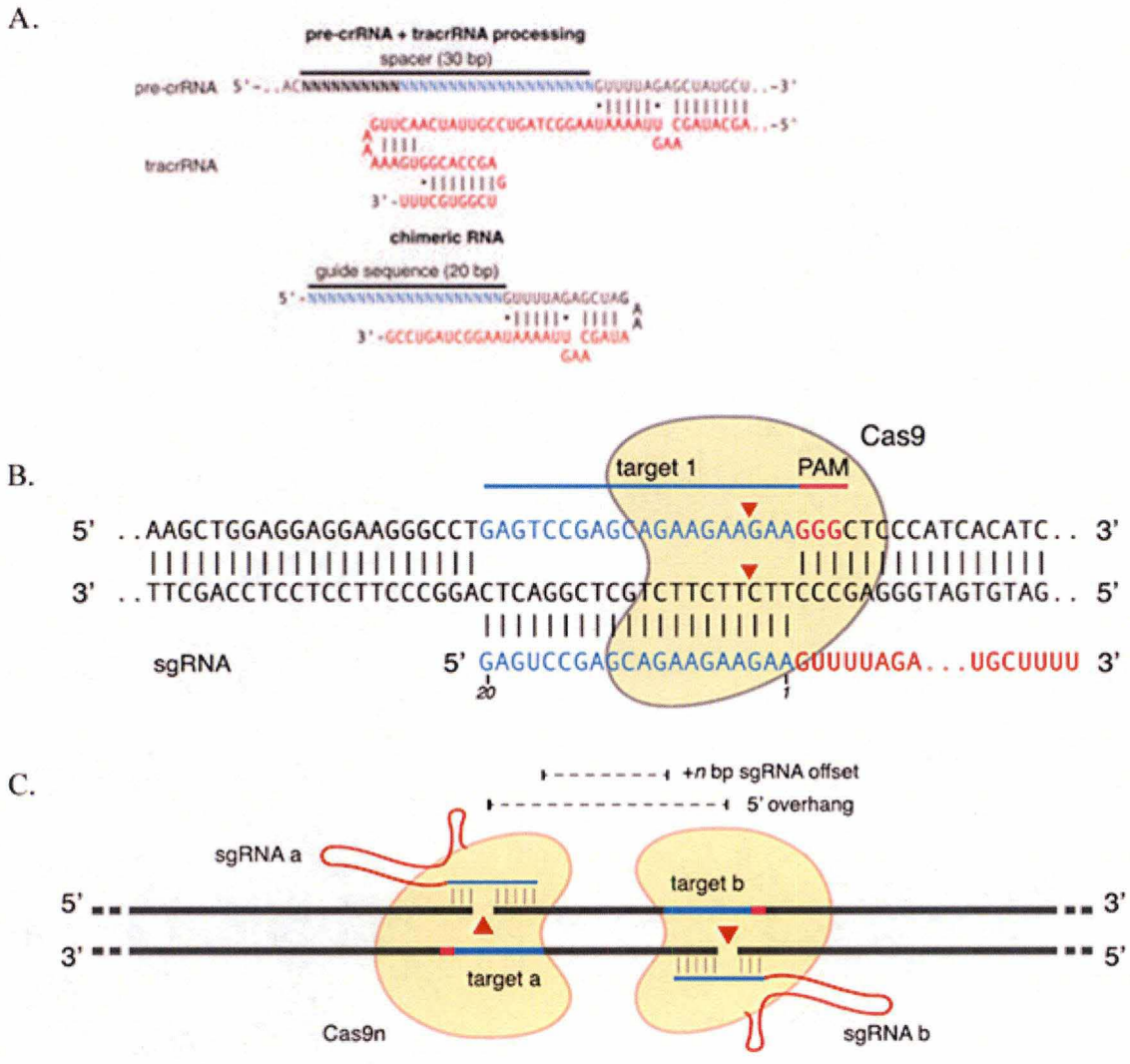

Figure 7. CRISPR/Cas9 system for genome editing. A) Representation of the 2 possible kinds of guide processing: 1 precrRNA + 1 tracrRNA, or a single chimeric sgRNA. [100] B) Targeting of a protospacers with the WT Cas9. The red arrows represent cleavage of DNA strands to generate a DSB. [107] C) Targeting with the nCas9, with 2 guides specific for 2 protospacers. The offset and overhang that characterise such a pair of RNA guides are represented by dotted lines. [107]

immediately upstream of the PAM, while they are relatively well-tolerated elsewhere in the guide sequence [106]. A solution to reduce off-targets is the use of a pair of RNAs, each guiding Cas9n at different but neighboring sites (figure 7C). This results in simultaneous nicks on both DNA strands of the targeted locus. A given pair of guiding RNAs is characterized by its offset and the kind of overhangs it generates, both affecting DSB generation efficiency [107]. The offset is defined as the distance between the 5' ends (distal to the PAM) of the two protospacers. Of note, offset and overhang type are interdependent, a positive offset always creating a 5'overhang with the D10A Cas9 nickase. For this nickase, optimum efficiency for NHEJ and indel formation is reached with 5' overhangs and offsets ranging from -4 to 20 bp.

Double nicks seem to be perceived by the cells as a DSB, and are therefore also repaired either by NHEJ or by HDR. This strategy highly increases specificity without reducing cleavage efficiency [100][107]. Importantly, HDR-mediated recombination frequency induced by the D10A nickase is similar to that of the WT Cas9, and thus higher than that induced by single nicks. [107][108]. Single-strand oligonucleotide (ssODN) are the most broadly used templates for HDR, but double-strand DNA donor vectors have also been used efficiently [109][110][111].

The Cas9 system may be applied for genome engineering in many cultured cell types, including in ES cells, and can therefore be used in transgenesis. Interestingly, Cas9n as well as WT Cas9 can be used to mediate efficient genome engineering following injection into one-cell embryos, with direct generation of transgenic animals [112][108][109][113].

1.5. Objectives

The final objective of this work is to obtain cells and mice with a loss of function mutation of miR-767. These model will allow investigation of the miRNA biological functions. In addition, a reporter gene inserted within the host gene *Gabra3* should allow a reliable and accurate detection of *Gabra3*, miR-105 and miR-767 transcription during development and in adult tissues.

Another objective of the work is to determine the expression pattern of miR767 and *Gabra3* in mouse tissues and to compare it with expression in humans. Importantly, induction of miR-767 by DNA demethylation and the cancer-germline expression pattern are to be confirmed in this species.

2. RESULTS

2.1. miR-767 and *Gabra3*: in silico data

In the mouse gene *Gabra3*, the second intron contains sequences similar to those of human miR-105 and miR-767. Whereas miR-767 is surrounded by two copies of miR-105 (miR-105-1 on the 5' side and miR-105-2 on the 3' side) in humans, there is only one copy of miR-105 in mice, on the 5' side of miR-767 (figure 8A). Alignment of miR-767 and miR-105 sequences between humans and mice revealed a perfect conservation of the seed sequence of miR-767 between the two species, but not of the seed sequence of miR-105 (figure 8B). Predictions based on seed pairing also indicate the *Tet* mRNAs as potential targets of miR-767, with several putative binding sites in the 3'UTR of *Tet1*, *Tet2* and *Tet3* mRNAs (Diana Micro-T, Miranda). However, such predictions are not sufficient to substantiate a regulatory impact of miR-767 on those mRNAs.

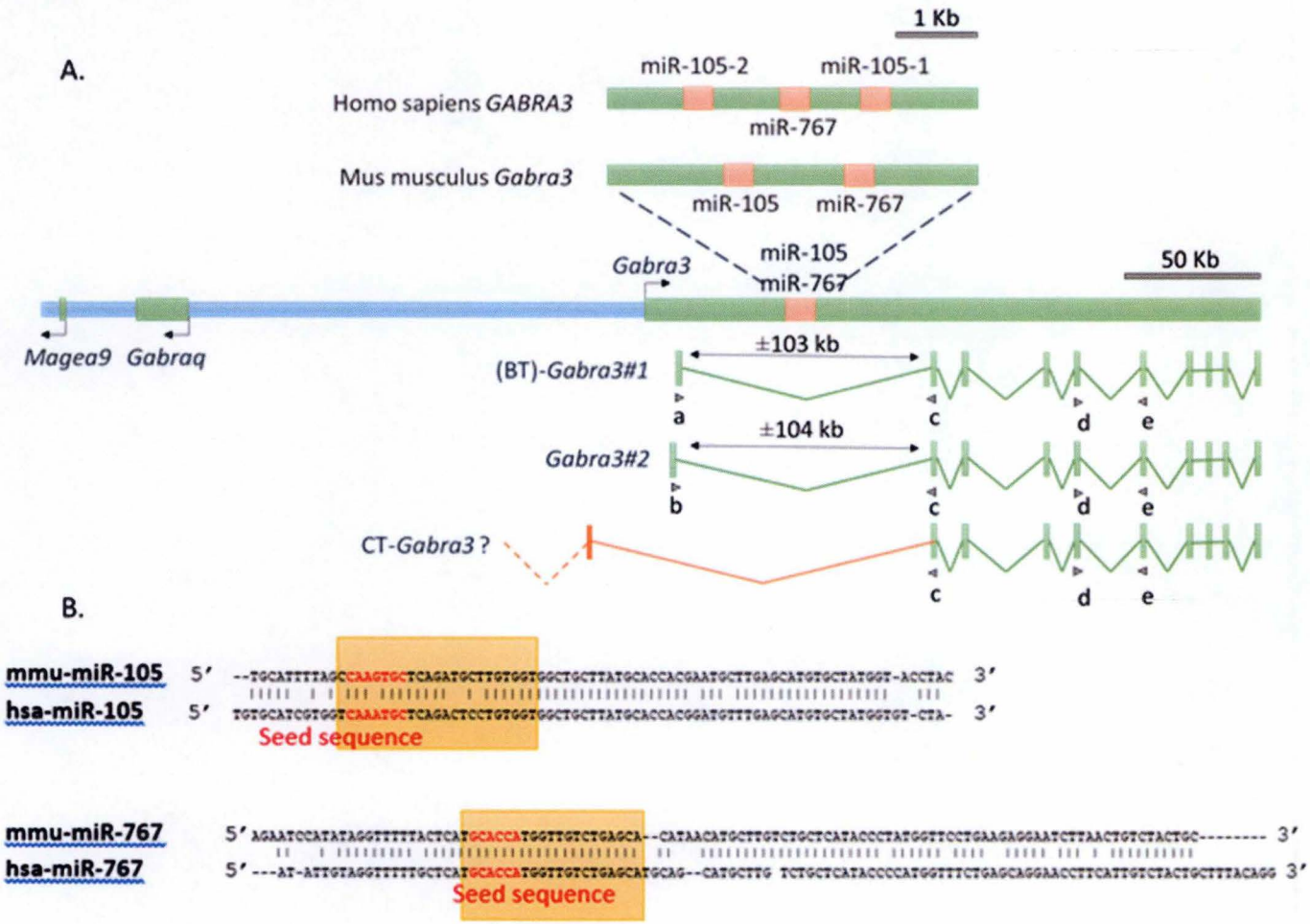

Figure 8. Genomic loci of mouse *Gabra3*, miR-105 and miR-767. A) Schematic representation of *Gabra3* gene and surroundings and of its different transcripts. The primers used for RT-qPCR are represented by the arrows. B) Alignment of mouse (mmu) miR-105 and miR-767 genomic sequences against their human (hsa) equivalents. In boxes are the sequences of the mature miRNAs. Their seed sequences are colored in red.

The canonical protein-coding transcript for *Gabra3* (NM_008067.4, labelled #1 here, see figure 8A) in mice is similar to the human one, comprising 10 exons and a conserved ORF spreading from exon 2 to exon 10. This ORF encodes a protein of 492 amino acids (aa). The first exon contains a part of the 5'UTR in both species. In humans, transcription of the CG mRNA is driven by a bidirectional promoter shared with the gene *MAGEA3*. This genomic region is not conserved in mice, as *Magea9* (annotated as a pseudogene) is found in the region corresponding to *MAGEA3*. A cluster of other MAGE genes is also found in the neighborhood of *MAGEA3* in humans, but not in mice. Sequence alignment of the third exon of *MAGEA3* with *Magea9* shows a 381 bp region with 67% identity. Despite being annotated as a pseudogene, 2 transcripts have been reported for *Magea9* (Ensembl Genome Browser), indicating that it is transcribed. These transcripts both at least partially contain the 381 bp region just mentioned.

Interestingly, a second protein-coding transcript of *Gabra3* has been predicted based on cDNA evidence in rats (Rattus norvegicus). The sequence of this transcript perfectly matches the mouse genome (ENSMUST00000114554.2 on www.ensembl.org, labelled #2 here, see figure 8A). This putative transcript is similar to the canonical mRNA #1, except for the first exon, which is different, and for the 10th exon, which is shorter. The first exon of mRNA #2 is located on the 5' side of that of the transcript #1 in the genome, thus in closer proximity of the pseudogene *Magea9*. If this mRNA exists in mice, it would result in a longer protein (533 aa instead of 492).

It is worth to note that contradictions exist regarding *GABRA3*/*Gabra3* expression. RNAseq data from human tissues report the highest *GABRA3* expression level in the brain, and lower levels in colon, breast, prostate and testis (Illumina Body Map, www.genecards.org). Microarrays in humans even suggest a nearly ubiquitous expression of the gene (biogps.org). These last data are however unlikely reliable, given the observations of Loriot et al (2014) [89]. Similar microarray data in mice indicate a high expression in the brain and low expression in a wide range of other tissues (biogps.org). In summary, it seems that human *GABRA3* and mouse *Gabra3* are indeed predominantly expressed in brain and lowly expressed in testis, but that it is not strictly restricted to these tissues.

2.2. Targeting miR-767 in ESCs: production of a conditional allele

As aforementioned, a mouse miR-767 with possible specific inactivation of the miRNA and a reporter gene allowing to track its expression would be a valuable tool. To obtain such a model, a strategy was designed inspired by Park et al (2012), who used standard gene targeting in ESCs to generate conditional alleles of miRNAs containing a LacZ reporter gene (figure 9) [114]. It was initially planned to flank the genomic locus of miR-767 by a 2 kb 5' homology arm (containing miR-105) and a 10 kb 3' homology arm in the targeting vector. The shorter 5' homology arm should allow a PCR-screening of the transfected cells to identify clones with the recombined allele. After HR, the endogenous region between those arms will be replaced by a

modified locus containing a LacZ-neo cassette. In addition, recombination sites for the Cre and Flp recombinases (respectively LoxP and FRT sites) within this replacement locus should allow the generation of useful variants of the recombined allele. The LacZ-neo cassette, composed of a LacZ reporter (with a 5' splicing acceptor site and a 3' poly-A site) and a neomycin selection marker (NeoR), will be located downstream from the 5' homology arm and upstream from the floxed pre-miRNA.

Monitoring miR-767 using such an allele presents several advantages. The first one is the possibility to easily detect expression in situ. Although Northern blot, microarrays and sequencing have provided huge data about miRNA expression, they are performed from isolated cells or homogenized tissues and lack spatiotemporal resolution. In situ hybridization with LNA (locked nucleic acid)-modified oligonucleotide probes is the most widely used technique for in situ detection of miRNAs [115]. Nevertheless, given the necessary small size of the probes, specific hybridization is difficult to obtain, and the signal to background ratio is generally low. Another technique have been developed to visualize expression in vivo, with transgenic mice expressing a miRNA "sensor". This sensor consist of a constitutively expressed LacZ reporter with a 3'UTR bearing complementary sequences for a given miRNA. In presence of this miRNA, the LacZ reporter is downregulated and the cells appear paler after LacZ staining [116]. Here again, the weakness of this system is specificity, as other factors could regulate expression of the LacZ sensor, including other miRNAs with a similar binding sequence. Specificity is the second advantage of our approach, as LacZ staining will here

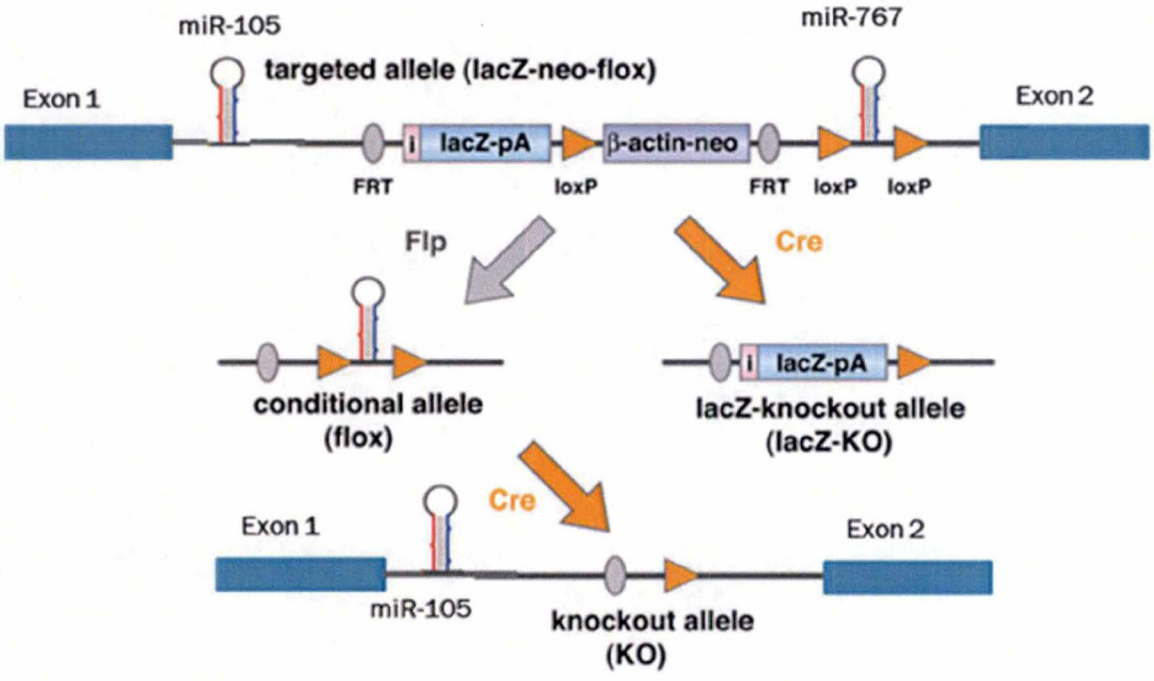

Figure 9. Strategy for conditional KO of miR-767 and flexibility of the targeted allele. In the 'lacZ-neo-flox' allele, thanks to a splicing acceptor site (i), lacZ mRNA will be fused to the beginning of *Gabra3* mRNA. This allele will thus be KO for *Gabra3*, because of the polyadenylation signal (pA) downstream of the lacZ reporter. Gabra3 expression will be restored in the 'flox' and 'KO' allele, but not in the 'lacZ-KO' allele. (Figure derived from ref [114])

systematically indicate *Gabra3* expression. A drawback, however, is the fact that the reporter will reflect transcription of miR-767 and its host gene, rather than the presence of the mature miRNA. Expression of miR-767 could indeed be regulated post-transcriptionally.

As already mentioned, the targeted allele will offer great flexibility, as it can be rearranged for different purposes by the Cre and Flp recombinase (figure 9). Crossing with mice expressing Cre will remove the floxed miRNA and NeoR, resulting in a LacZ reporter allele that will be KO for miR-767. This allele should provide reliable spatial and temporal information on *Gabra3* expression, without any possible interference of the ß-actin promoter that drives NeoR expression. Crossing with Flp-expressing mice will remove the LaZ and NeoR genes, leading to a clean functional conditional KO allele, with only two LoxP sites flanking miR-767. Animals bearing this allele can be used for further crossings with transgenic mice expressing Cre or CreERT2 (in the germline or in specific tissues), intended to delete the miRNA in a precise spatiotemporal way.

The starting plasmid used to construct the miR-767 targeting vector contained an ampicillin resistance gene (AmpR) for bacterial selection and a ccdB cassette for gateway recombination. The latter was composed of the ccdB gene and a chloramphenicol resistance gene (CmR), surrounded by AttR sites for LR gateway recombination. Figure 10A shows the sequential steps of construction. The first step was the insertion of the floxed miR-767 sequence on the 3' site of the ccdB cassette. This miR-767 insert (645 bp) was obtained by gene synthesis, with LoxP sites designed at a distance of 225 bp on both sides of miR-767 genomic sequence. HindIII restrictions sites were included at both extremities of the synthetic fragment to facilitate its insertion into the plasmid. A FseI site and a PacI site were also included, to allow respectively the later insertion of the 3' homology arm and the linearization of the targeting vector before transfection in mESCs.

The next step was the insertion of the short 5' homology arm. This 2 kb fragment was obtained by PCR amplification using a BAC DNA (bMQ120-c13) as a template, and inserted on the 5' side of the ccdB gene. Then, an LR gateway recombination was performed to replace the ccdB-containing gateway cassette by the LacZ-neo cassette. This gateway reaction requires recognition by an LR clonase of AttL and AttR sites, respectively on the donor plasmid and on the destination plasmid. Those sites are then recombined to form AttB and AttP sites, with exchange of the gateway cassettes between the two plasmids. Removal of the ccdB gene allowed positive selection of the recombined vector, using non-ccdB competent bacteria that cannot survive if they contain a non-recombined plasmid bearing the ccdB gene.

The last and most challenging step was the addition of the long 3' homology arm. By contrast to the 5' arm, the region covered by this arm (38,9% C-G) contains numerous T-stretches and other highly repetitive sequences, rendering PCR amplification difficult. These repetitive sequences can be highlighted by dot plotting the sequence of that fragment against itself (figure 11C). This reveals a large duplicated region of over 1 kb in the central region, represented by 2 parallel lines on both sides of the central alignment. Several palindromes of over 200 bp are also present, represented by smaller secant lines.

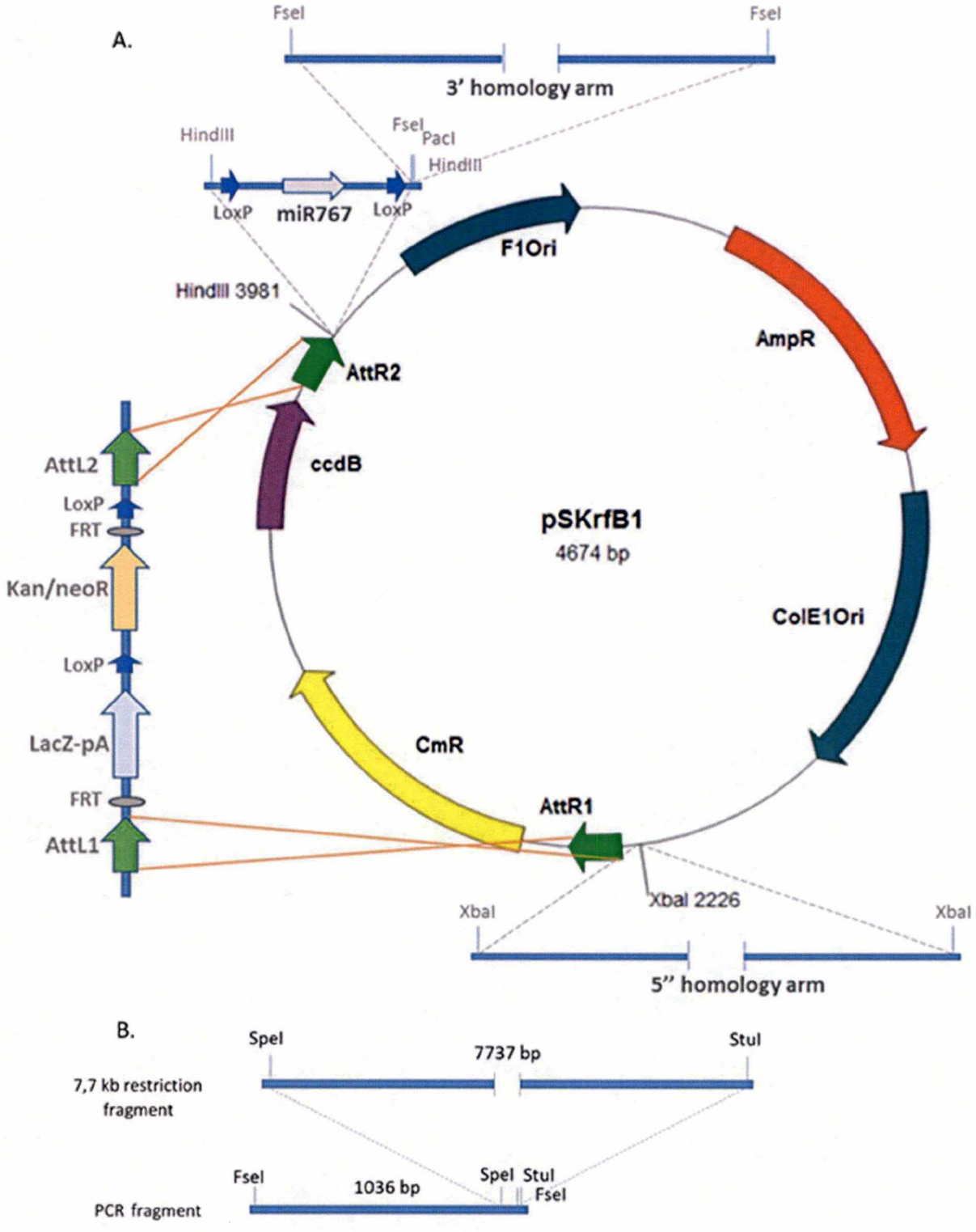

Figure 10. Construction of a targeting vector for standard homologous recombination. A) Schematic representation of the construction steps to obtain the targeting vector. B) Alternative strategy for the cloning of the long 3' arm, with inclusion of a 7,7 kb, SpeI-StuI restriction fragment within a 1 kb PCR product, to be cloned in the final vector after FseI restriction.

Therefore, an alternative strategy was developed (figure 10B). A 1 kb fragment, covering the 5' part of the arm before any repetitive pattern, would be PCR-amplified and cloned into an intermediate plasmid. SpeI and StuI restriction sites within this fragment would then allow insertion of a restriction fragment containing the 3' part of the arm (7,7 Kb).

Whereas the 1 Kb PCR fragment was easily cloned into the intermediate plasmid, addition of the SpeI-StuI restriction fragment proved to be difficult. The StuI restriction site within the PCR fragment overlapped a dcm methylation site. As StuI is blocked by such methylation patterns, the construct was amplified in a dam-/dcm- bacterial strain.

Twenty-five fragments were supposed to be generated after double restriction of the BAC clone bMQ120-c13 with SpeI and StuI. As the sticky ends generated by SpeI are incompatible with the blunt end of StuI cutting, only 8 BAC fragments could potentially be cloned into the vector. Southern blot was performed on bMQ120-c13 digested with SpeI and StuI to confirm the presence of the expected 7,7 Kb band. The hybridization probe used for Southern analysis was

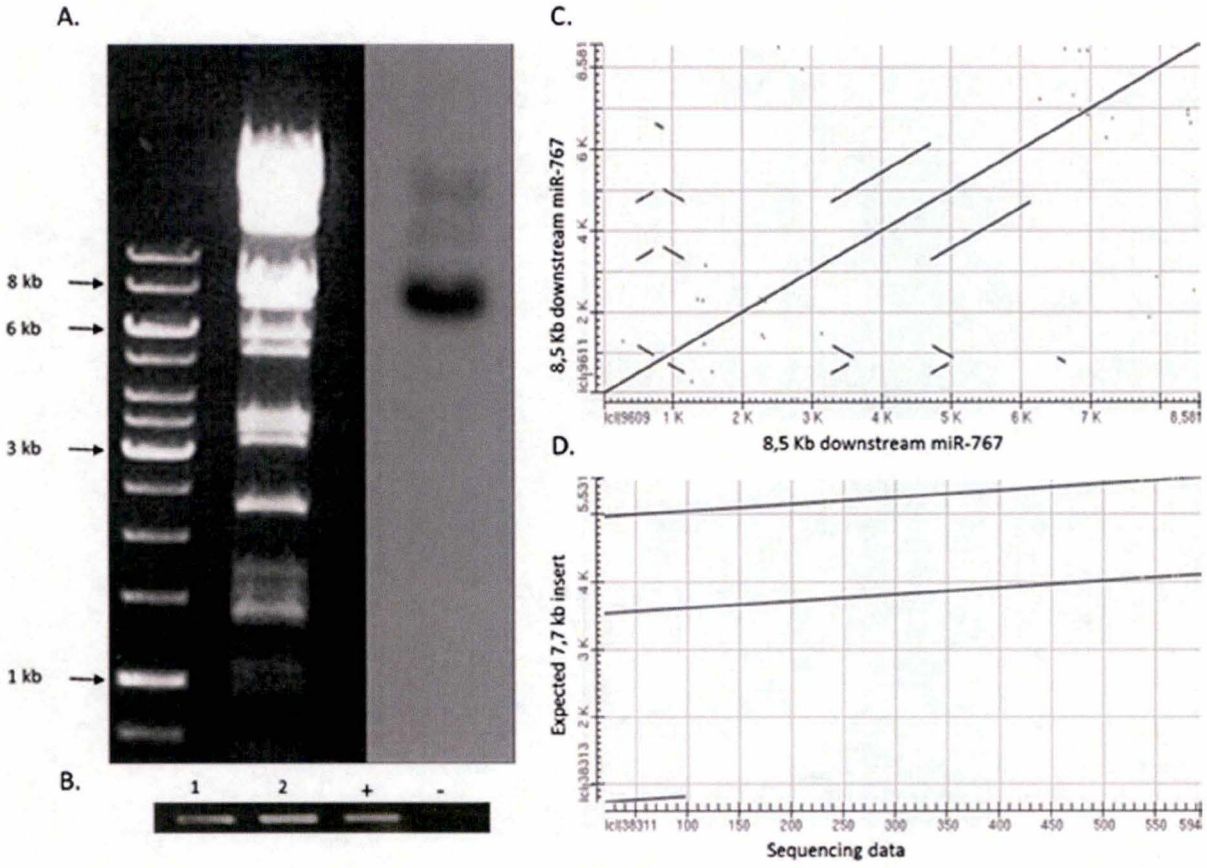

Figure 11. Assembly of the 3' homology arm. A) Southern blot on the restriction product of BAC DNA with SpeI and StuI. Electrophoresis of the restriction product is shown on agarose gel (on the left) and after Southern blot (on the right). B) PCR screen of the 2 positive clones picked on agar plates, with positive (BAC DNA) and negative (water) controls (oligos 5 and 6 in appendix B). C) Dot plot of the 8,5 kb region immediately downstream of miR-767 against itself, comprising the 1 kb PCR fragment and the 7,7 kb restriction fragment. D) Dot plot of the sequence of one of the fragments cloned into the intermediate construct against the expected 7,7 kb restriction fragment.

a 20 bp-long oligonucleotide, designed to specifically bind the fragment (available in appendix B). As expected, a unique band of about 8 kb was observed (figure 11A). The whole restriction product was ligated with the double-digested vector, and the ligation product was introduced in bacteria by electroporation. As a PCR screening of colonies randomly picked on agar plates gave no positive result, screening was performed by hybridization, using the same oligonucleotide probe used for the Southern analysis. By this way, positive colonies were detected, and 2 were isolated. Both were also positive in PCR screening with primers located inside the insert (figure 11B).

Restriction profile and sequencing of the plasmid DNA extracted from these 2 clones, however, revealed that neither contained the expected insert. Instead of the 7,7 Kb fragment, these inserts were shorter and matched the sequence of the expected 3' homology arm in a discontinuous way, suggesting that rearrangements occurred (figure 11D).

2.3. CRISPR/Cas9-mediated KO of miR-767 in mouse ES cells

Given the difficulties to obtain the desired targeting vector, an alternative strategy was elaborated using the recently developed CRISPR-Cas9 technology. This strategy was meant to generate miR-767 KO cells by causing short deletions in the sequence of pre-miR-767 that would make proper processing into a mature miRNA impossible, either by profoundly altering the hairpin structure or by a deletion affecting the mature miR-767 sequence. The Cas9 nickase was selected over the WT nuclease for its higher specificity and reduced off-target activity. Two distinct pairs of guides were designed using the online "CRISPR Design Tool" (http://crispr.mit.edu/, Zhang Laboratory). Each pair was composed of two RNAs, each targeting a protospacer on one DNA strand within the pre-miR sequence (figure 12A). The offsets were 14 bp (pair 1) and 25 bp (pair 2), creating 5' overhangs of respectively 54 bp and 65 bp. Targeting two different loci presents two advantages. First, some RNA guiding pairs may be more efficient than others. Testing several pairs thus increases the odds of success. Second and most important, other targets also means other off-targets. Therefore, the mutants obtained with distinct pairs are expected to contain different (if any) off-target mutations, which are unlikely to cause similar artefactual phenotypes. Two different plasmids were also tested to compare their efficiency, namely pX334 and pX335. While pX334 separately produces a precrRNA and a tracrRNA, pX335 codes for a chimeric sgRNA. Another difference between the 2 plasmids is the absence of selection marker in pX335 (which thus has to be co-transfected with another plasmid to allow selection), while pX334 contains a puromycin selection marker.

CRISPR/Cas9-mediated targeting of miR-767 was performed in mESCs, so that KO clones could potentially be used to generate KO mice. Furthermore, as mESCs express *Tet1* and *Tet2* mRNAs, 2 potential targets of mouse miR-767, the effect of miR-767 KO on TET expression levels can be assessed in those cells, provided that expression of *Gabra3* can be induced. After

transfection, short selection and cloning by dilution, 39 clones were isolated. DNA was extracted from those clones, and a DNA fragment covering the region targeted by the RNA guides was amplified by PCR, in order to sequence the miR-767 locus (data not shown). PCR

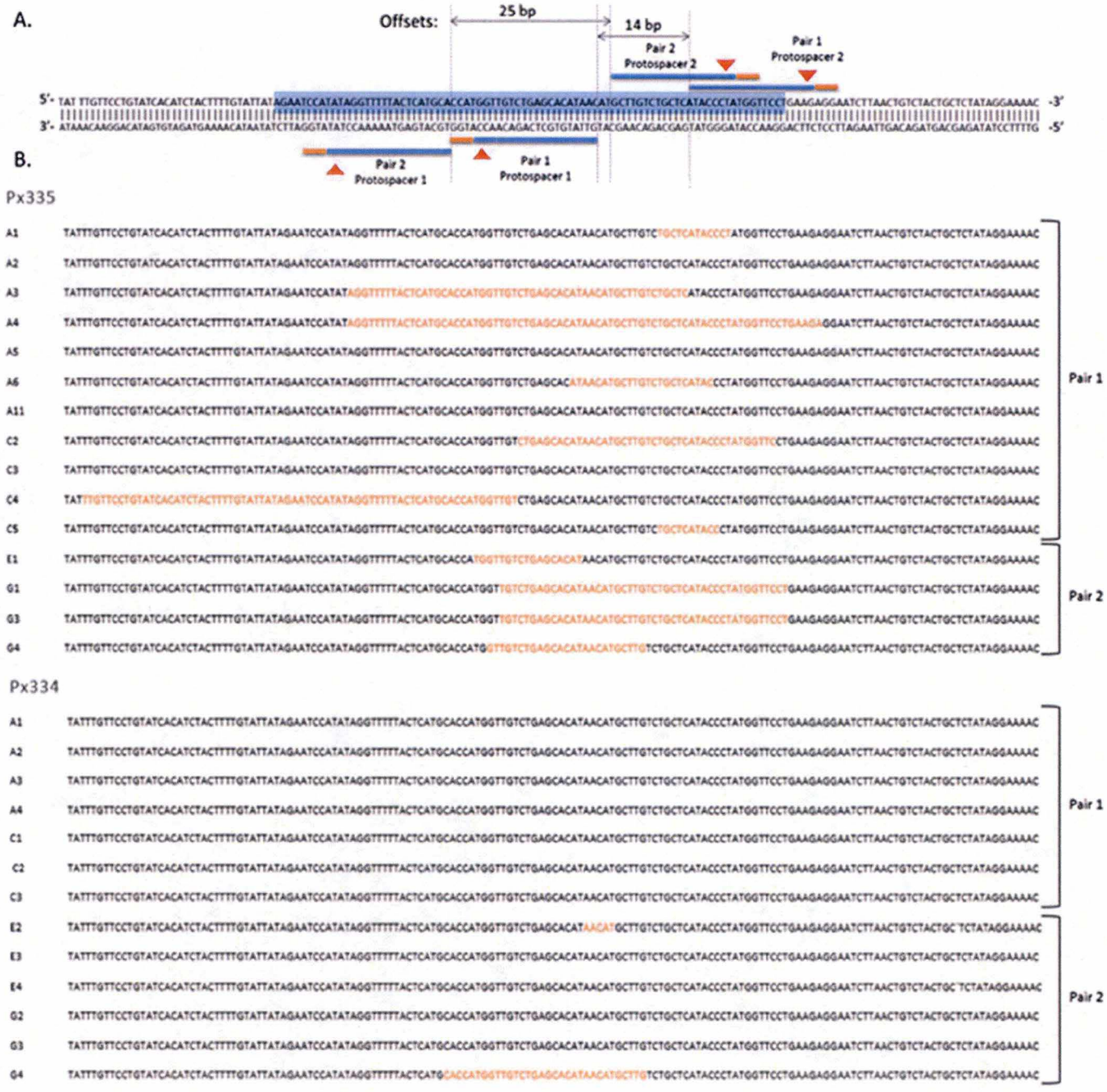

Figure 12. Simple KO of miR-767 using the Cas9n. A) Schematic representation of the targeting strategy, with the 2 sgRNA pairs used, their offsets and the overhangs produced. The PAMs are represented in orange color, next to the protospacers (in blue). In blue box is the genomic sequence of miR-767 (corresponding to the pre-miRNA). Sites of cleavage by Cas9n are represented by the red arrows. B) Genotyping of the isolated ESC clones. Deleted sequences are marked in orange.

amplification failed for 10 clones transfected with pX335 (producing chimeric sgRNAs). Sequencing of the 29 remaining clones revealed that 11 out of 16 clones generated by the pX335-derived plasmids, and 2 out of 13 with the pX334-derived plasmids, had small deletions in the pre-miR sequence (figure 12B). Interestingly, efficiency does not seem reduced when using the RNA guides with the 24 bp offset, although the recommended optimal range is 4 to 20 bp. Indeed, all the clones obtained with these guides were mutated with pX335, and 2 out of 6 with pX334.

Encouraged by this successful targeting through NHEJ, we decided to use the CRISPR/Cas9 technology to facilitate HR using our "incomplete" targeting vector, containing 2 short recombination arms (the 2 kb 5' arm and the 1 kb PCR fragment as 3' arm), as a double-strand DNA donor molecule (figure 13A). For "spontaneous" homologous recombination in ES cells, long homology arms are required. By contrast, the CRISPR technology has been shown to promote HR using DNA donors containing smaller homology arms. Targeted DSB to promote HDR events will be mediated by the Cas9n guided by two sgRNA pairs on both sides of the endogenous mir-767 locus.

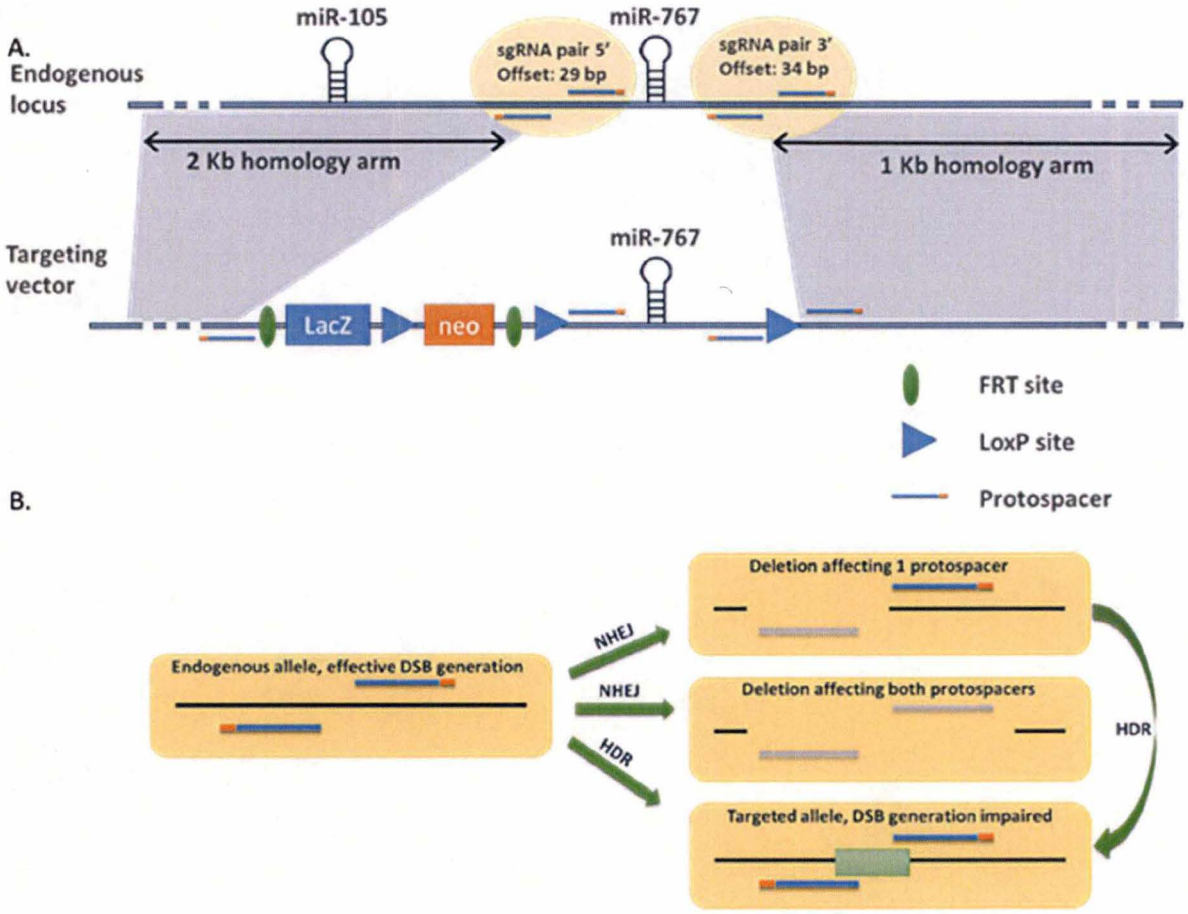

Figure 13. Targeting strategy for homologous recombination using the Cas9n. A) Schematic representation of the genomic locus targeted and the dsDNA donor. Two sgRNA pairs will be used on both sides of miR-767, that will be specific for the unmodified, endogenous allele. B) Schema of the possibilities of modification of the allele.

A crucial point when designing sgRNAs for HR is to ensure that the guides are specific for the WT allele, but will not be able to target the locus after HR has occurred, thereby damaging the targeted allele by indel formation. For this reason, the sgRNA pairs previously used for direct KO were not suitable here. Two new sgRNA pairs were designed, targeting both sides of the miR-767 locus (figure 13A). The first one is located near the 5' homology arm and overlaps the insertion site of the LacZ-neo cassette, so that after HR, the 2 sgRNAs will be separated by around 7 kb, limiting their action to single nicks that will be too distant to generate DSB. This pair has an offset of 29 bp and will generate a 5' overhang of 69 bp. The second sgRNA pair is located near the 3' arm, has an offset of 34 bp and will generate a 74 bp 5' overhang. In the recombined allele, addition of a LoxP site will lengthen the offset, up to 68 bp, which is expected to importantly reduce the possibility of additional indel formation. Because these requirements limit the choice of potential protospacers, the offsets of both sgRNA pairs are slightly out of the recommended range. Nevertheless, this range has been established for indel formation through NHEJ, and although efficiency is then decreased with increasing offsets, indels are still observed with offsets up to 100 bp [107]. This means that indels might still occur with the 3' sgRNA pair, even after HR.

CRISPR/Cas9-mediated gene targeting will be attempted with a single or two DSB. One cut with the 5' sgRNA pair may be sufficient to induce recombination. However, the neomycin selection will only apply as far as the NeoR gene. miR-767 locus and its surrounding sequences, included in the dsDNA donor, are also homologous to the endogenous locus. These sequences could therefore mediate recombination events that would not comprise the terminal 3' LoxP site. Two DSB generated with both sgRNA pairs could enhance occurrence of recombination events including the entire modified miR-767 allele. Moreover, a second sgRNA pair could serve as a "second chance" in case of a first undesired indel event (figure 13B).

2.4. Expression of *Gabra3* in mouse tissues and in demethylated cells

Another focus of the project was to characterize *GABRA3* expression profile in mouse and to determine whether DNA hypomethylation could induce expression of a CG-transcript as was shown in humans. To this purpose, pairs of primers were designed in several exons of Gabra3 to perform RT-PCR analyses. As no alternative (CG)-transcript has been reported in the mouse so far, we designed the primers (see figure 8A, materials and methods and appendix B) in the known protein-coding transcript, which should correspond to the BG mRNA. A first pair of primers (a-c), should be specific for the putative BG transcript, with the sense primer (a) located in the first exon and the antisense primer (c) in the second exon. A second pair of primers (d-e) located in downstream exons should amplify cDNA corresponding to both BG and CG transcripts.

2.4.1. *Gabra3* expression in mouse tissues

Gabra3 expression was first assessed in a small panel of tissues by RT-PCR. *Gabra3* mRNA was only detected in the brain after a single row of amplification, with both sets of primers (data not shown). To increase sensitivity, nested PCR was performed, using 2 additional primers. This allowed detection of *Gabra3* in brain and testis as expected, but also in liver and intestine (figure 14A). However, those results are mainly qualitative, and even extremely low expression levels could result in a signal after so many PCR cycles (60 cycles). In order to obtain quantitative data, qRT-PCR was carried out (figure 14B). Amplification with primers a-c, supposed to be specific for the BG transcript, repeatedly provided aberrant CT amplification signals. PCR amplifications with the other pair of primers (d-e) were consistent with our previous observations. The highest expression was observed in the brain, as expected, and lower levels were observed in other organs, namely testis (about 0,06% of brain expression), spleen (about 0,93% of brain expression) and intestines (about 0,12% of brain expression). Expression appeared more important in spleen and intestine than in testis. This is not in favor of the proposed BG expression profile of *Gabra3*, but is rather consistent with expression data reported in databases for both mouse and human *GABRA3*/*Gabra3* genes.

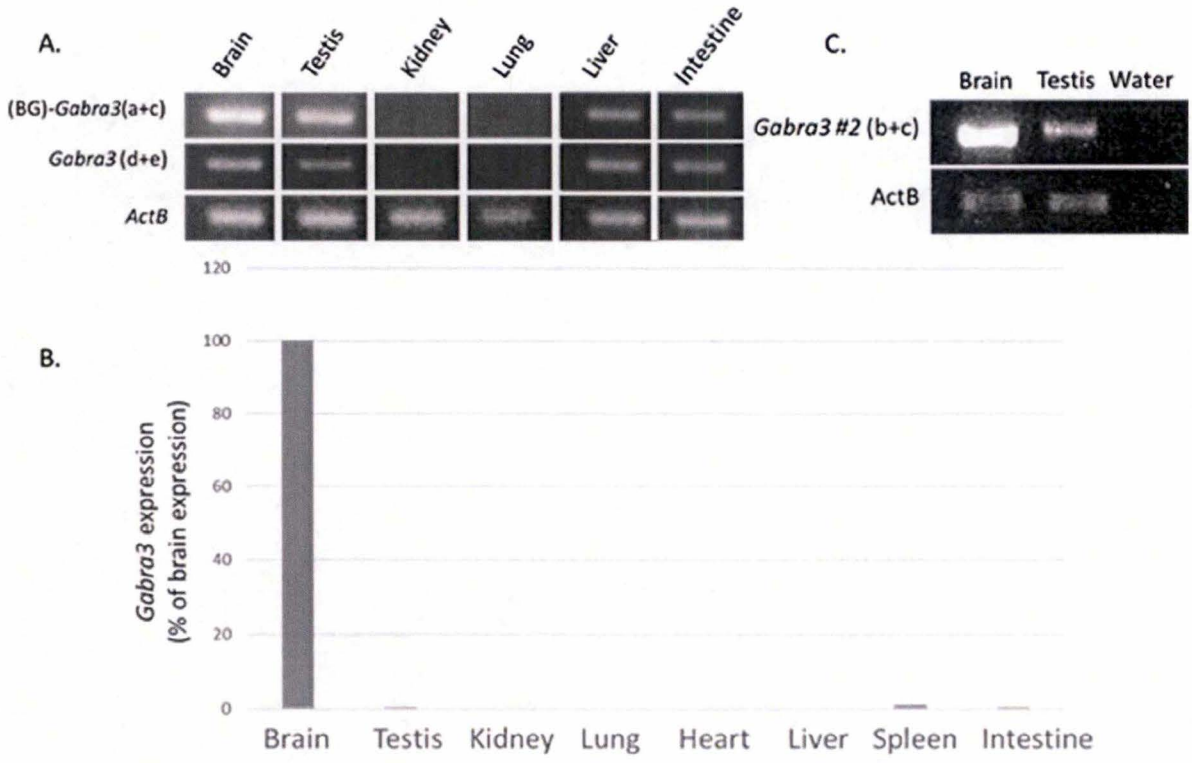

Figure 14. *Gabra3* expression in tissues. A) Nested RT-PCR in several tissues, with primers specific for the putative BG transcript (a+c), or for total *Gabra3* mRNAs (d+e). B) RT-qPCR on (total) Gabra3 (d+e) in some tissues, normalized by GAPDH expression, and expressed as a percentage of brain expression. C) Nested RT-PCR on the putative transcript #2 in brain and testis.

Nested RT-PCR was also performed to specifically detect the putative transcript #2 of *Gabra3* in brain and testis, with primers located in the first and second exons (b-c). Expression was detected in both tissues, with a stronger signal in brain (figure 14C). Therefore, this transcript does not correspond to a CG variant.

2.4.2. Induction of a putative CG-*Gabra3* by DNA demethylation

Although the canonical Gabra3 mRNA (#1) did not seem to be restricted to brain and testis as previously suggested, It was still to be determined whether an alternative CG transcript (specific of cancer cells and of the germline), could be induced by DNA demethylation in mouse cells. For that purpose, NIH3T3 mouse fibroblasts were treated with the DNMT inhibitor 5-azadC. This drug is known to reduce proliferation and to display an important cytotoxicity. Those effects should be avoided as much as possible, as 5-azadC-induced demethylation is passive and therefore replication-dependent. To determine a concentration of the drug to use, cells were first cultured for 3 days in medium containing 1µM, 2µM and 4µM of 5-azadC, and viable cells were counted. Cytotoxicity was clearly observed, but no evident difference was observed between the different concentrations (figure 15A). Of note, important cell mortality was noticed from the second day of treatment, suggesting that the duration of treatment may be more important for toxicity than the dose, at least in the range of concentrations used here.

We tried to determine a duration of treatment suitable to induce DNA demethylation. Mouse 3T3 fibroblasts were cultured with 2 µM 5-azadC for 24h, 48h and 72h before RNA extraction. Cells were also counted to evaluate the toxic effect of the drug with time. An important cytotoxicity was observed (figure 15B), the total number of cells decreasing after 3 days of treatment. DNA demethylation was evaluated indirectly, by following the expected induction of *Magea5*, a CG gene that has been shown to be induced in L1210 lymphocytic leukemia cells treated with 5-azadC [117]. Expression of *Magea5* was not detected by RT-qPCR in untreated cells, but was indeed induced by 5-azadC treatment. Induction was already detectable after 24h, and increased with time (figure 15C). After 72h, Magea5 was expressed up to about 0,3 % of its expression level in testis. By contrast, no significant induction of *Gabra3* could be observed (data not shown). Actually, given high cell mortality, RNA amounts extracted from the surviving cells and engaged in RT-qPCR were small for some samples. *Gabra3* mRNA was only detected in the last PCR cycles, in both untreated and treated cells.

To try to improve induction of *Magea5* and to induce *Gabra3*, another protocol of DNA demethylation was implemented. Instead of a continuous treatment, 5-azadC was administered to the cells twice: at time "zero" and after 48h. Each time, the treatment was maintained for 24h. This turned up to be less aggressive for the cells, which could be harvested 120h after the first administration, allowing more time for the demethylation process to take place. Two concentrations of 5-azadC (2 µM and 4 µM) were tested. RNA levels of *Gabra3* and *Magea5* were estimated by RT-qPCR. *MageA5* RNA was clearly induced (figure 15D). As no evident difference was observed between the 2 concentrations of 5-azadC regarding Magea5 induction,

the data from cells treated with 2 µM and 4 µM were pooled and compared with those previously obtained after only 72h of continuous treatment (figure 15E). Here, *MageA5* reached

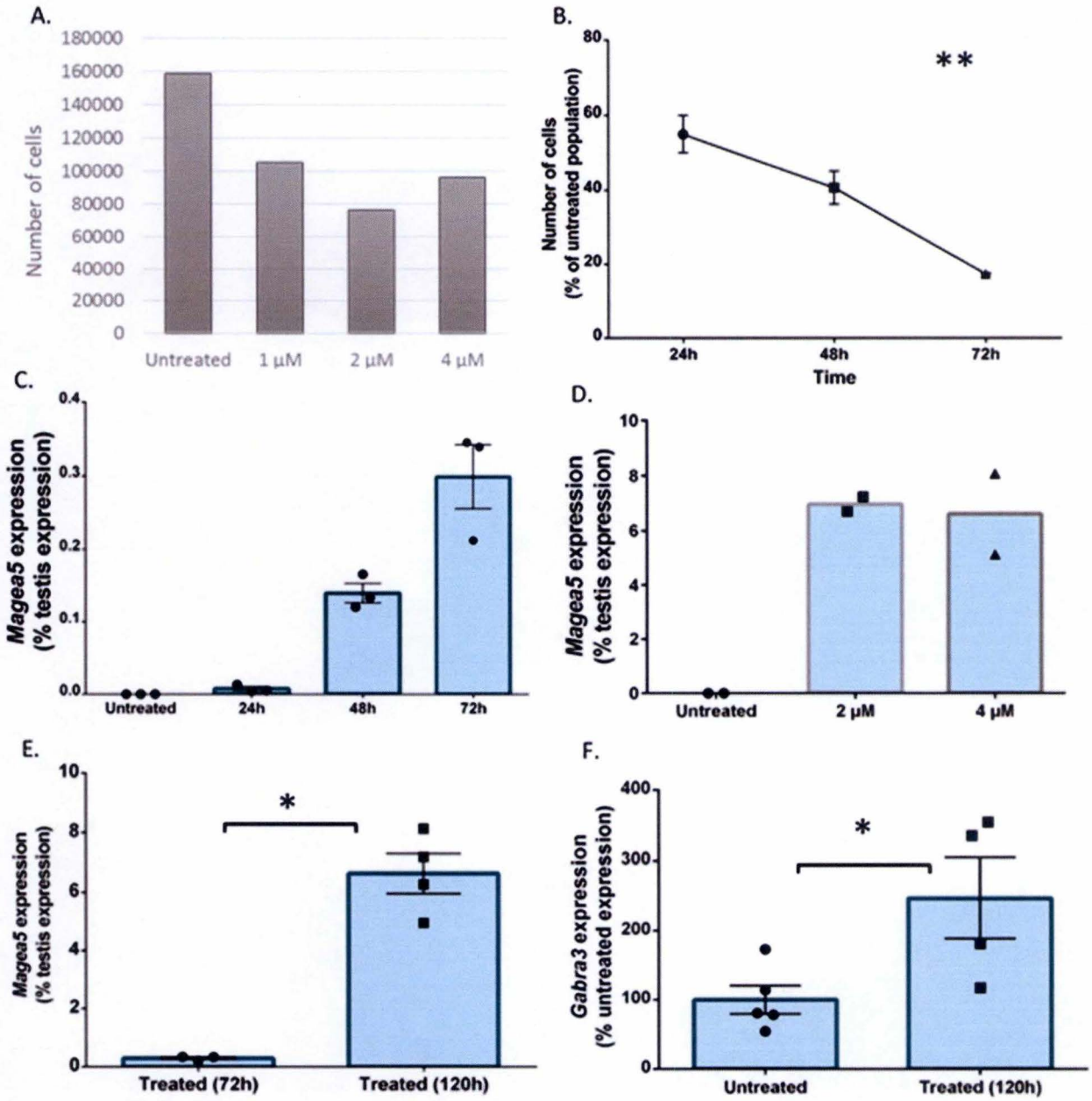

Figure 15. DNA demethylation in 3T3 fibroblasts with 5-azadC. A) Cell counts (average of 2 replicates) after 3 days of treatment with different concentrations of the drug. B) Percentage of surviving cells (compared to untreated cells) treated with 2 µM of 5-azadC after 24h, 48h and 72h (mean ±SEM). Experiment was performed on 3 replicates, and statistical analysis was made with ANNOVA1. C) *Magea5* induction with time, expressed as a percentage of expression in testis (mean ±SEM). Cells were treated at 2 µM in triplicates. D) *Magea5* expression in untreated cells, and after 120h of treatment. Cells were treated in duplicates for the 2 concentrations. E) Comparison of *Magea5* expression after a continuous treatment of 72h (N=3) and after a discontinuous treatment of 120h (N=4) (mean ±SEM). Statistical analysis was performed using non parametric Mann-Whitney U test. F) *Gabra3* induction after 120h of treatment (N=4) compared to untreated cells (N=5), expressed as a percentage of average expression in untreated cells (mean ±SEM). Statistical analysis was made using non-parametric Mann-Whitney U test.

expression levels representing about 7% of its expression in testis, which was significantly more than after 72h of treatment (around 0,3%). A weak (2,5 fold) but significant induction of *Gabra3* was also observed in cells treated for 120h compared to untreated cells (figure 15F). Importantly, expression of *Gabra3* was already detected in untreated cells.

In all qRT-PCR experiments described, the BG specific primers still systematically generated aberrant results, rendering impossible any comparison between the levels of BG transcripts and total Gabra3 transcripts (BG + putative CG).

2.4.2. 5'RACE on *Gabra3* mRNAs

To our knowledge, no previous study has been carried out to characterize the 5' end of mouse *Gabra3* mRNA(s). In order to determine the 5' end of the canonical *Gabra3* transcript (#1), and with the hope of identifying a CG- transcript induced by DNA hypomethylation, a 5'RACE (for 5' Rapid Amplification of cDNA Ends) experiment was carried out on RNA from mouse brain and testis, and from NIH3T3 cells treated with 5-azadC.

In summary, the 5'RACE method consists to generate a circular cDNA containing the 5' region of the mRNA to be characterized, surrounded by known sequences (figure 16A). This cDNA is produced by targeted reverse transcription (RT), and is self-ligated. The reverse transcription was specifically achieved on *Gabra3* transcripts, using an antisense primer located in exon 5. After degradation of the RNA template, the ssDNA was circularized, and nested PCR were performed to amplify the 5' region (see material and methods and appendix B for sequences of the RT primer and of the PCR primers). On the linear cDNA, the sense and antisense (S and R) PCR primers are oriented respectively toward the 3' end and the 5' end of the transcript. On the circularized cDNA, they amplify the 5' region of interest.

No PCR products could be obtained using RNA from testis or demethylated 3T3 cells (figure 16B). By contrast, PCR products of several sizes were produced from the brain (figure 16B). These fragments were cloned by "TA cloning" and sequenced. Alignment (BLAST) of the sequences with the mouse genome lead to the identification of 2 distinct transcripts of *Gabra3* (figure 16C). The first one corresponded to the known canonical transcript (#1), whose complete sequence was already recorded in databases. The second one, by contrast, had not yet been reported. In this novel transcript (#3), the first exon of the canonical mRNA is replaced by two alternative exons of respectively 86 bp (exon 1) and 198 bp (exon 2). These exons are located approximatively 92 kb and 84 kb upstream from exon 3. The third exon of the alternative transcript #3 corresponds to exon 2 of the canonical transcript #1. Translation of this novel mRNA should result in a protein containing 51 additional aa, with an ORF starting in exon 2 (figure 17). The putative aa sequence corresponding to transcript #3 was analyzed using the InterPro database (http://www.ebi.ac.uk/interpro/), but no additional protein domain could be identified.

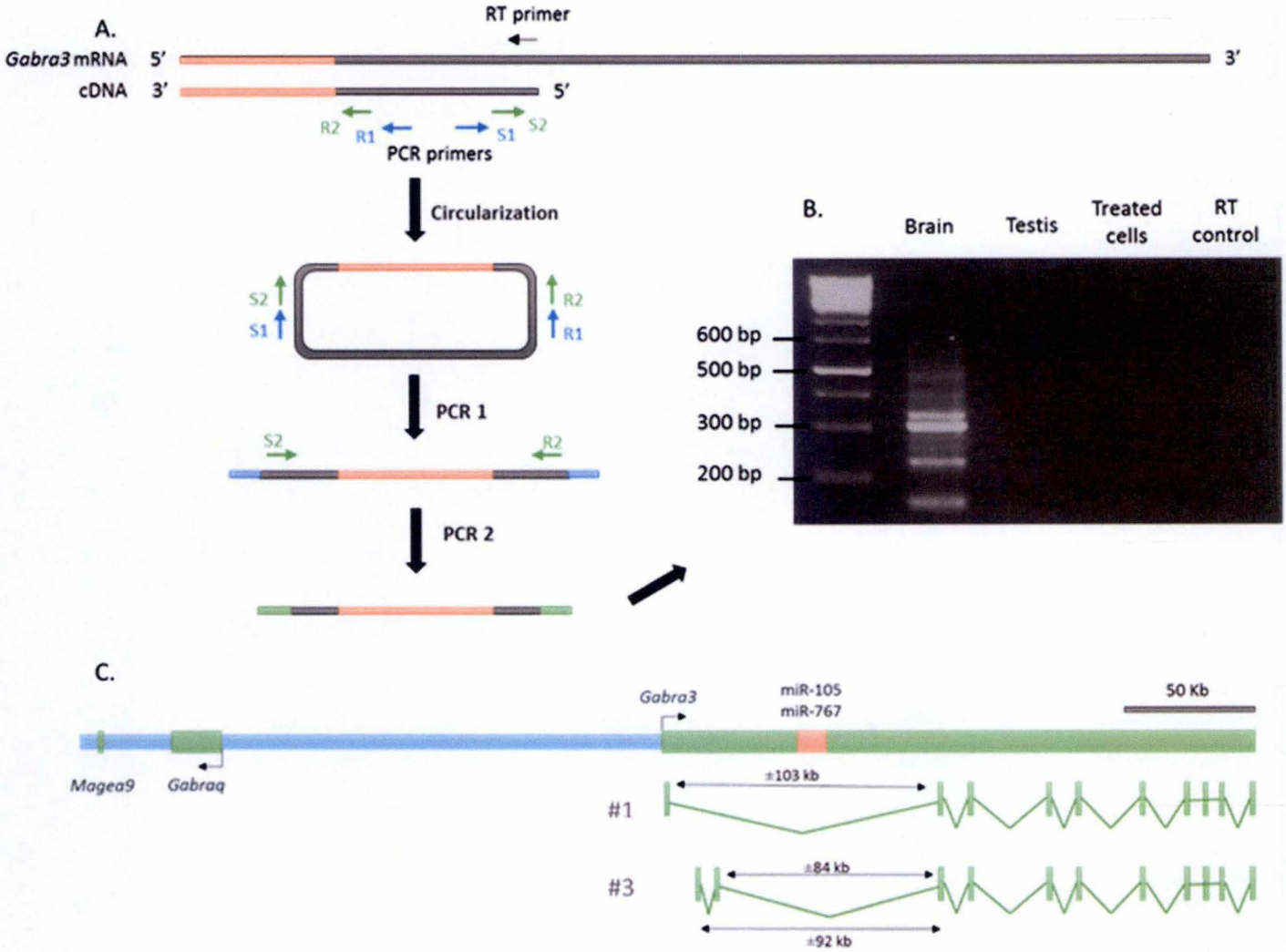

Figure 16. 5'RACE on mouse *Gabra3*. A) Schematic overview of the 5'RACE technique. The RT primer and the sense (S) and antisense (R) PCR primers are represented by the arrows. B) Electrophoresis of the nested PCR poducts. The 'RT control' corresponds to brain RNA that did not undergo reverse transcription. C) Comparison of the 2 transcripts of *Gabra3* identified by 5'RACE: the canonical transcript #1 and the novel transcript #3.

Other 5'RACE products of different sizes were also observed (figure 16B). They corresponded to chimeric cDNAs or to truncated transcripts. The reason why no RACE products were obtained from testis and demethylated cells is not known, but this may be due to the low expression level *Gabra3* in these samples. Of note, the putative transcript #2 was not found among the brain RACE products.

By analyzing the CG-transcripts obtained by 5'RACE in humans, our collaborators from UCL noticed that some of the alternative exons they characterized matched exon predictions established from RNAseq data. It was thus asked whether similarly predicted mouse exons could also be part of a CG-*Gabra3* mRNAs. Two such predictions were tested by RT-PCR. For each prediction, two sense primers (sequences available in appendix B) were designed and used on cDNA samples from brain and testis together with an antisense primer (c) located in

exon 2 of *Gabra3*. After 2 rows of amplification, a PCR product was obtained from testis cDNA (data not shown). However, its sequence did not match with chromosome X, indicating that it resulted from an unspecific amplification.

GA	CAC	TCG	TGA	TGA	CTT	AAG	TTA	CTT	CGA	GAC	GGT	TCT	TAT	GTC	AGC	ACA	CCT	TCT
	H	S	*	*	L	K	L	L	R	D	G	S	Y	V	S	T	P	S
CAC	CTT	CTC	TTA	GTA	CGG	GAT	TAT	GAC	AAG	CAT	CCT	CTA	AAT	GGA	CGG	AGA	ATG	TTA
H	L	L	L	V	R	D	Y	D	K	H	P	L	N	G	R	R	M	L
AGG	GTC	AAG	AAG	TTA	AGA	AAA	GTT	ATC	TGA	CTC	AGG	AAG	GTC	AGA	AAC	ATT	ATG	AAA
R	V	K	K	L	R	K	V	I	*	L	R	K	V	R	N	I	M	K
GCA	ATA	AAC	AAT	GTC	TGG	GGA	ACA	CAT	ACT	CCC	TCC	CTG	ATT	CTC	TGT	GGA	AAT	GGC
A	I	N	N	V	W	G	T	H	T	P	S	L	I	L	C	G	N	G
TGC	CAG	ACA	CTG	AGC	TCA	AGA	GAT	TCA	GCT	TTT	ATG	GCC	CCT	CAT	TCA	TAT	GAA	GAT
C	Q	T	L	S	S	R	D	S	A	F	M	A	P	H	S	Y	E	D
CTC	ACA	GGT	CTC	TTC	AAG	TTG	CTG	TCT	AAG	AAG	ATG	ATA	ATC	ACA	CAA	ATG	TGG	CAC
L	T	G	L	F	K	L	L	S	K	K	M	I	I	T	Q	M	W	H
TTT	TAT	GTG	ACC	AGA	GTT	GTA	CTT	CTT	CTC	CTG	ATT	AGT	ATT	CTC	CCT	GGA	ACC	ACT
F	Y	V	T	R	V	V	L	L	L	L	I	S	I	L	P	G	T	T
AGC	CAA	GGG	GAA															
S	Q	G	E															

Figure 17. 5' sequence of the novel transcript #3. The sequence is noted from 5' to 3', with its translation into aa sequence in the same reading frame as the canonical protein. The additional aa sequence is indicated in red, and is followed by the canonical sequence (in black).

2.5. Characterization of miR-767 expression

Because miR-767 is hosted within the second intron of Gabra3, miR-767 and Gabra3 expressions are expected to be coordinated. However, mouse miR-767 expression has not yet been documented. It was thus decided to perform experiments aimed to detect the mature miRNA. A qRT-PCR detection method was chosen, in order to obtain quantitative data. Mature miRNAs are too short to allow hybridization of 2 PCR primers, but strategies have been developed to circumvent this problem. In the RT-qPCR platform chosen here (Exiqon), a 3' poly-A tail is added before reverse transcription, which allows hybridization of an anchored RT primer bearing a 5' extension (a "universal" tag) (figure 18A). qPCR is then performed using a miRNA-specific primer and a universal primer complementary to the extension. These PCR primers are LNA primers, which improve specificity of the system [118]. This method was applied on RNA from mouse tissue and NIH3T3 cells treated with 5-azadC. No signal was obtained, possibly because of a too low expression level of miR-767.

As low expression of *Gabra3* and of its associated miRNAs could account for absence of detection, optimization of the RNA extraction method was undertaken. RNA samples tested so far had indeed been extracted using TRIzol reagent and ethanol precipitation, which could result in poor yield of miRNAs. A specialized purification kit (miRNEASY Mini kit, QUIAGEN) was used to extract RNA from several parts of the brain (olfactory bulb, cortex and cerebellum)

and from testis. Once more, no detection was observed by qRT-PCR. miR-105 expression was also analyzed, and was detected in cortex and testis, with about 2-fold higher expression in testis (figure 18B). However, as qPCR signals were only observed in the last amplification cycles (32-35), these results may not be fully reliable. U6 and Snord68 RNAs (the 2 endogenous controls) and a spike-in control (synthetic RNA molecules provided with the Exiqon amplification kit) were detected in all those RT-qPCR experiments, indicating that the quality of RNA was appropriate, and that the RT and PCR reactions worked (data not shown).

To validate the qRT-PCR detection method for miR-767 and to ensure that absence of detection did not result of a failure of the miR-767-specific primers, it was decided to include positive controls. A first control was to force expression of miR-767 and miR-105 in murine cells, by transfecting expression vectors driving their expression. Genomic fragments containing miR-767 or miR-105 were cloned into pCA-ß-EGFPm5-mU6, downstream of the strong RNA pol III mouse U6 promoter. In this vector, expression of the cloned miRNA is coupled with expression of the fluorescent protein EGFP [119].

The miR-767 and miR-105 vectors were independently transfected in NIH3T3 fibroblasts, and transfected EGFP-positive cells were observed at the microscope to confirm expression of the transgene (figure 18C). The fraction of EGFP-positive cells appeared low, indicating a low efficiency of transfection. RNA was extracted from the transfected cells using the miRNEASY Mini kit (QUIAGEN), and RT-qPCR was carried out to detect both miR-105 and miR-767. A synthetic miR-767 RNA oligonucleotide was used as a spike-in positive control. Based on a previously established relation between CT values and spike-in synthetic oligonucleotide concentrations, 2 amounts of the synthetic miR-767 were engaged in 2 independent RT-qPCR reactions (4 µl of dilution at 400 amol/µl or 200 fmol/µl in 10 µl of PCR reaction) [120].

As expected, miR-105 and miR-767 were detected only in the cells transfected with the corresponding expression vector (figure 18D). This demonstrates that murine cells are able to process both miR-105 and miR-767 sequences into mature miRNAs, and should thus also be able to process them from their endogenous host: the pre-mRNA of *Gabra3*. Detected levels of miR-767 were much lower than those of miR-105 (around 5% of miR-105). However, it is impossible to determine which factor(s) account for this difference: transfection efficiency (although no evident difference was observed), efficiency of miRNA processing, miRNA stability or RT-qPCR detection efficiency. The synthetic miR-767 spike-in was also detected, with CT values approximately corresponding to expectations according to the starting amounts (figure 18E).

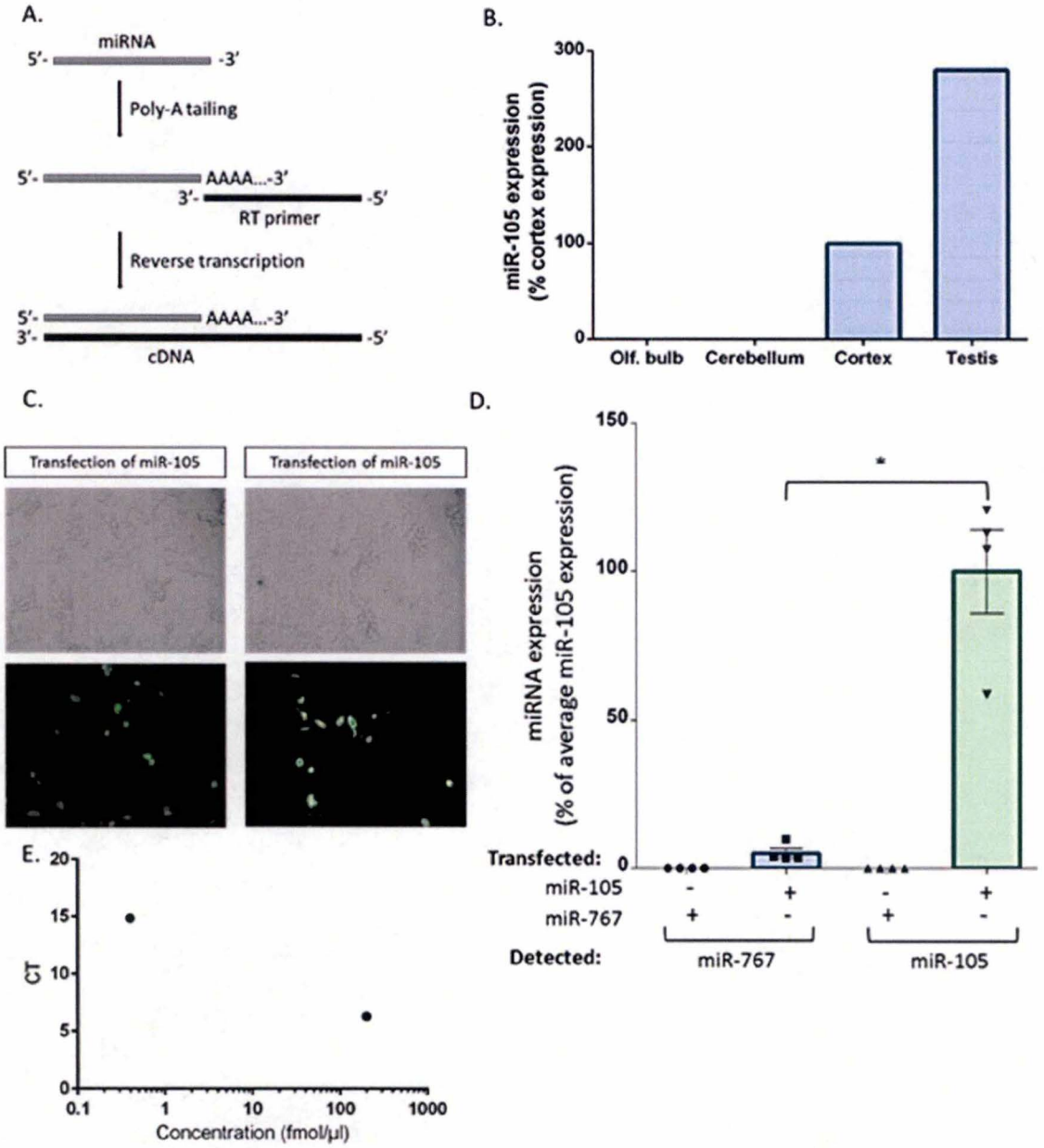

Figure 18. RT-qPCR on miR-105 and miR-767. A) Universal tailing strategy during the reverse transcription step. B) Detection of miR-105 in brain and testis, expressed as percentage of expression in the cortex. C) Verification of the fluorescence of the cells transfected with miR-105 and miR-767. Two pictures of the same area were taken, 1 in normal light and 1 in fluorescence. D) Detection of miR-105 and miR-767 after transfection, expressed as a percentage of miR-105 average expression in cells transfected with miR-105 (mean ±SEM). A Mann-Whitney U test was performed for statistical analysis. E) CT values obtained by miR-767 detection at the 2 concentrations tested for synthetic miR-767 spike-in.

3. DISCUSION AND PERSPECTIVES

In conclusion, much remains to be investigated regarding *Gabra3* and miR-767 expression in mice. The canonical transcript of *Gabra3* (#1)was detected in brain and testis, but also in spleen and intestine. These results do not fit entirely with the initially proposed BG expression profile, but are consistent with microarray and RNAseq data from human and mouse tissues indicating that *Gabra3* is mainly expressed in the brain, but also at low levels in other tissues, including testis. This, however, does not exclude the possibility of a CG-specific transcript that would be induced by DNA hypomethylation.

To assess this, mouse NIH3T3 fibroblasts were treated with 5-azadC to induce DNA demethylation. Although the treatment induced the CG gene *Magea5*, it only resulted in a low induction of *Grabra3*, which was already detected in untreated cells. This could be due to shallow demethylation, and the procedure to demethylate cells with 5-azadC may be optimized. Indeed, Lim et al (2010) reported that while continuous exposure to 5-azadC at concentrations from 1 to 5 µM was efficient to promote demethylation of repetitive elements in NIH3T3 cells, a dose of 0,1 µM was much more efficient to demethylate genic regions [121]. Induction of *Magea5* and *Gabra3* might thus be improved by optimizing the dose range, and maybe by increasing the duration of 5-azadC exposure. Other factors could also impact on *Gabra3* expression, some of them depending on the cell type. Demethylation of other cell types should be attempted. It would also be interesting to investigate the methylation and expression status of *Gabra3* in a panel of mouse malignancies, compared with healthy tissues.

Our qPCR data revealed that *Gabra3* displays low expression levels, excepted in the brain. In such qPCR analysis, there is a restricted range where the CT values are a linear function of mRNA concentration. As the CT values were frequently over 30 (meaning that more than 30 PCR cycles were required for detection), one could wonder whether expression levels inferred from those results are accurate in terms of quantification. Establishing the range of linear relation between *Gabra3* expression and CT values would be necessary to determine the relevance of those results.

Regarding *Gabra3* alternative transcripts, 5'RACE was performed on brain, testis and DNA-demethylated cells, to characterize the 5' part of *Gabra3* mRNA and to try to identify a CG transcript. A novel transcript was identified in the brain, potentially coding for a protein of 543 aa, with a putative N-terminal extension of 51 aa with respect to the canonical protein encoded by *Gabra3* #1. No additional domain was identified in this extension. Still, the N-terminal signal peptide (aa 1 to 28) of the canonical protein would be shifted 51 aa downstream. If such a protein is produced, the presence of the extension should have repercussions on the protein addressing, trafficking or properties. Although potentially interesting regarding *Gabra3* functions, this mRNA is of limited interest for our project. Indeed, it was identified in the brain and does not correspond to a possible CG transcript. The 5'RACE experiment failed to identify *Gabra3* transcripts in testis and in NIH3T3 cells treated with 5-azadC, where the putative CG-specific alternative transcription was expected to occur. The reason for this is unknown, but other 5'RACE should be attempted. Ideally, this experiment

should be performed using RNA containing only *Gabra3* transcripts induced by DNA hypomethylation. Such template could originate either from tumor cells derived from tissues which do not normally express *Gabra3*, or from cells in which *Gabra3* would have been sufficiently induced by 5-azadC treatment.

Given that the human genomic locus of *MAGEA3*, on the 5' side of *GABRA3*, is not conserved in mice, it is possible that a DNA demethylation-induced transcription from an alternative TSS does not exist in mice. Although studying the biological function of miR-767 in mice would still be relevant, its possible expression (and thus function) in mouse cancer is still an open question. As already mentioned, a CG-transcript could possibly be identified by reiterating cell demethylation and 5'RACE, or by characterizing expression of *Gabra3* in mouse cancer cell lines. It would also be interesting to determine whether a transcript originating from *Magea9* could be induced by DNA demethylation. To date, *Magea9* is annotated as a pseudogene, but non-coding transcripts have been observed and reported. RT-PCR could be performed on cDNA from demethylated cells or testis to detect *Magea9* mRNA. If the promoter of *Magea9* can be activated by DNA methylation as the human *MAGEA3*, it may also drive transcription of a CG transcript of *Gabra3*.

Regarding the expression of the mature forms of miR-767 and miR-105, the detection method must be improved to further analyze mouse tissues or cells. In our hands, detection of endogenous expression appeared barely possible in brain and testis for miR-105, and was not possible for miR-767. Transfection of NIH3T3 cells with expression vectors bearing the sequences of premiR-105 or pre-miR-767 allowed detection of both miRNAs, but a lower signal was observed for miR-767. This may be due to a less efficient processing of miR-767, to a differential post-transcriptional regulation, or to a differential efficiency of the RT-qPCR assay. A synthetic miR-767 spike-in, included in the RT-qPCR procedure, was successfully amplified at 2 different concentrations. Such spike-in control could also be included for miR-105, allowing to establish a relation curve between spike-ins concentration and corresponding CT values for both miR-105 and miR-767. Any differential technical efficiency of detection would thus be highlighted, and could be taken into account. Anyway, those results indicate that murine cells are able to produce mature forms of miR-105 and miR-767 from their pre-miRNAs. Therefore, it would be reasonable to consider that *Gabra3* expression is a good indicator of miR-105 and miR-767 expression.

Little can be done to further improve the sensitivity of the RT-qPCR assay. The RNA extraction method used is supposed to be efficient[1]. It is still possible to enrich extractions for short (<200 bp) RNAs, but this procedure also reduces the global extraction yield and has been shown to have a variable effect on the miRNAs, inducing changes in relative levels [122]. Other qPCR platforms exist that could be tested, some of them using TaqMan probes and a different reverse transcription strategy. However, sensitivity could be lower than that of the SYBR Green method. Some post-transcriptional modifications in the 3' part of miRNAs can affect detection levels of those platforms more than those using the universal tagging strategy [123]. Indeed,

[1] « Evaluation of miRNA Extraction Methods », D. J. Hollingshead et al, GPCL, University of Pittsburgh, 2007

mature miRNAs can exist in several forms called isomiRNAs, the most abundant of which is not necessarily the mature form referred in miRBase [124].

Frozen brains from the *Gabra3* KO mice have kindly been provided by Dr. Uwe Rudolph. Although no functional protein is produced in those mice, expression of miR-105 and miR-767 are probably unaffected. Indeed, the mutation in *Gabra3* consists of a duplication of exon 4, downstream from the miRNAs. The mutant *Gabra3* mRNA could be degraded by nonsense-mediated mRNA decay [95]. Nevertheless, as miRNA cleavage by Drosha has been shown to occur transcriptionally, miR-105 and miR767 should be excise from their pri-miRNA before degradation [125]. It would thus be worth investigating expression of miR-105 and miR-767 in those mice. However, as expression is barely detectable in WT mice brain, detection in *Gabra3* KO brains may be hazardous. Improving the detection method of miR-105 and miR-767 would be necessary before risking to waste those brain in chancy RT-qPCR experiments.

Although base-pairing predictions indicate that the *Tet* mRNAs are potential targets of miR-767 in mice as in humans, they are not sufficient to conclude in a conserved regulatory function of miR-767 on those proteins. Therefore, luciferase assays should be performed on transfected cells, using a miR-767 mimic molecule and plasmids encoding a luciferase gene grafted with the 3'UTRs of the *Tet* mRNAs. Conservation of this regulation is indeed important for the relevance of the characterization of miR-767 functions in mice.

Particular emphasis has been placed in the generation of miR-767 KO. Using the CRISPR/Cas9 technology, several clones of mESCs with a simple loss-of-function mutation of the miRNA have been obtained. According to databases (GEO Profiles and biogps.org), *Gabra3* expression may be detected at very low levels in these cells. Treatment of those cells with 5-azadC could be attempted in the aim to induce the gene. It will first be verified that the miR-767 mutations do not affect expression of *Gabra3*. If DNA demethylation induces *Gabra3* mRNA and miR-767 in those cells, then the effect of miR-767 deficiency on TET1 and TET2 protein expression levels could be assessed. If the protein levels are upregulated, it will not only confirm the inhibitory function of miR-767 on *Tet1* and *Tet2* mRNAs, but also indirectly provide evidence of mature miR-767 expression.

As ESCs are pluripotent, they can potentially be differentiated in all the cell types originating from the inner cell mass, including neurons expressing *Gabra3*. Protocols have been established to differentiate mESCs into neural precursors able to generate neurons [126]. Provided that *Gabra3* is expressed in neurons derived from mESCs, and as TET1 and TET3 are thought to intervene in this differentiation process, a comparison could be made between differentiations of WT and miR-767-deficient mESCs.

The conditional KO allele with an embedded LacZ reporter (LacZ-neo-flox) would be precious to obtain reliable information about *Gabra3* expression, complementary to those obtained by qPCR analyses. When attempting to clone the long 3' homology arm of the targeting vector, rearrangements occurred, likely due to plasmid instability caused by the repetitive sequences contained in this region. Palindromic repeats, which are found at several places in the sequence, are indeed known to confer instability or "unviability" of the host plasmid by forming secondary structures [127]. Consequently, the initially desired targeting vector could not be achieved.

Another strategy has been undertaken, using the CRISPR/Cas9 technology to facilitate HR events, allowing to reduce the size of the homology arms of replacement targeting vectors. A shorter targeting vector will thus be used as a dsDNA donor template for HR in mESCs.

Plasmids driving expression of the Cre or the Flp recombinases will be transfected in the targeted cells to ensure that these recombinase are able to modify the allele into the LacZ-KO allele, the conditional (flox) allele or the KO allele. Finally, we will verify that the KO of miR-767 does not prevent expression of its host gene, by assessing Gabra3 expression in cells with the KO allele. If *Gabra3* cannot be induced by demethylation in mESCs, these cells could be differentiated in *Gabra3*-expressing neurons. After validation of the targeted allele in mESCs, transgenic mice will be generated by blastocyst injection.

The LacZ-KO allele will allow detection of *Gabra3*/miR-767/miR-105 in tissues, with a cellular resolution. This reporting may not be entirely reliable for the presence of the mature miR-767, but it is reliable for its transcription. It has been shown in mice that *Gabra3* mRNA is ubiquitous and abundant in the brain during embryonic development, but displays a post-natal decline, as a switch occurs to the α1 subunit of the $GABA_A$ receptor [128]. In the adult brain, *Gabra3* expression is restricted to cortical neurons [129]. Co-transcripted with *Gabra3* in early stages of life, miR-767 may thus have an important function in brain development and neuron differentiation, possibly by regulating the TET proteins.

Post-transcriptional modification of *Gabra3* mRNA through RNA editing has been shown to play a role in the switch towards the α1 subunit after birth. Adenosine-to-inosine (A-to-I) RNA editing is mediated by the ADAR deaminases, which recognize dsRNA that can result from mRNA folding. Inosine in exonic sequences is then recognized as a guanosine during translation. RNA editing predominantly occurs in the brain [130]. Adenosine to inosine conversion by ADAR1 and ADAR2 on a specific site of exon 9 of *Gabra3* mRNA results in an aa change (isoleucine to methionine), with altered pharmacological properties of GABA response [131][132]. This process is conserved between mice and humans, and the protein resulting from edited mRNA displays an altered trafficking, with reduced surface level due to increased internalization and lysosomal degradation [133]. Interestingly, RNA editing of *Gabra3* is low during embryonic development and increases after birth, with around 100% of *Gabra3* transcripts edited in adult brain [131][132][133].

RNA editing is also known to edit pri-miRNAs and regulate miRNA biogenesis and functions. Editing can result in enhanced or reduced cleavage by Drosha, influence mature strand selection, or even modify the seed sequence, diversifying target recognition [134]. As *Gabra3* is a known substrate for A-to-I RNA editing, one could wonder whether its 2 host miRNAs could also be concerned. This might explain why miR-105 was detected at higher level than miR-767 in brain by RT-qPCR, or account for a differential processing or detection of miR-105 and miR-767. Lower editing during development may be significant for miR-767 functions at this stage. Besides, as editing predominantly occurs in the brain, post-transcriptional regulation may confer to miR-105 and miR767 an expression pattern more restricted than that of *Gabra3*, so that a CG profile cannot be excluded.

Mice bearing the flox allele should express *Gabra3*, miR-105 and miR-767 and display a phenotype similar to WT animals. RT-qPCR experiments will be performed to validate those expectations. Crossing of those mice with Cre-expressing mice will result in mice KO for miR-767. Again, RT-qPCR will be performed to verify that expression of *Gabra3* is not impaired. KO mice will be compared to those with the functional conditional allele. Morphological analysis could be made during development, as well as a comparison of the TET levels between KO and conditional mice. Behavioral phenotyping of adult animals should also be made, focusing on memory formation and extinction. A comparison of miR-767 KO mice and *Gabra3* KO mice in the tests where these latter displayed phenotypic particularities should allow to distinguish the respective effects of the miRNA and *Gabra3* deficiencies.

In conclusion, it is still to be determined whether a CG alternative transcription of the *Gabra3* is induced by DNA hypomethylation in mouse cells. Generation of a CG transcript of *Gabra3* hosting miR-105 and miR-767, as well as the potential role of miR-767 in cancer, might be specific to humans. However, the characterization of miR-767 functions in a mouse model remains an interesting perspective, provided that its regulatory impact on the expression of the TET proteins can be established in this species. The tools developed in the present work should contribute to this characterization, by the generation of a KO allele of miR-767 in mice.

4. MATERIALS AND METHODS

Agarose Gel Electrophoresis

DNA fragments were separated by electrophoresis on agar gels (0,7%, 1% or 2%, according to the length of the fragments) with ethidium bromide staining. Agarose was dissolved in TAE 1X buffer. DNA was mixed with 6X Orange DNA loading dye (Fermentas) prior to electrophoresis. GeneRuler 100 bp or 1 kb (Fermentas) DNA ladders were used for sizing of the fragments.

Vector constructions

Construction generalities

All PCR products intended for construction were purified with the Wizard® SV Gel and PCR Clean-Up System (Promega, REFA9282). Vectors were dephosphorylated prior to ligation by Shrimp alkaline phosphatase (rSAP, NEB). Vectors and inserts were purified by phenol-chloroform and precipitated with ethanol. Ligations were performed by the T4 DNA ligase (NEB). After each ligation step, the constructs were introduced in bacteria by electroporation and submitted to antibiotic selection. DB3.1 competent cells were used for the constructs bearing the CcdB gene. Otherwise, DH5α or DH10B E.coli strains were used. Colonies were isolated on agar plates and PCR screened. DNA was extracted from the cultured PCR-positive clones with GeneJET Plasmid Miniprep kit (ThermoScientific, #K0502) for cultures up to 5 ml or Nucleobond® PC500 (Macherey-Nagel, REF740574.25) for larger cultures (from 100ml). Proper insertions were checked by Sanger sequencing (3130 Genetic Analyzer, Applied Biosystems). Both PCR-amplified homology arms (2 kb and 1 kb) were completely sequenced. All PCR screens were made with the DreamTaq Green PCR Master Mix (Fermentas).

Construction of the targeting vector for HR

pSKrfB1, which served as a base for construction of the targeting vector for homologous recombination, was kindly provided by Prof. X. De Bolle. The 645 bp fragment containing the floxed miR-767 (see the appendix for complete sequence) was ordered as a synthetic DNA fragment from IDT® (Integrated DNA Technologies), integrated within the plasmid pIDTSMART-KAN. It was inserted in the vector following digestion with HindIII (NEB). The donor plasmid for LR Gateway recombination, pL1L2-GTIRES, was kindly provided by Dr. Barry Rosen. Nourseothricin (CloNAT, WERNER BioAgents) was used for selection of this plasmid in bacteria. LR gateway reaction was performed by Gateway® LR Clonase® II Enzyme Mix (Invitrogen™). The 2 kb 5' homology arm was PCR amplified (1 and 2 in appendix B) by LongAmpTM Taq DNA polymerase (NEB). The BAC DNA template was bMQ120-c13, a 129s7/AB2.2 BAC clone from the bMQ library (Source Bioscience Lifesciences™). The arm was cloned in the vector after digestion with XbaI (NEB). A restriction site for AgeI, borne by the sense primer (oligo 1) was added during PCR amplification, for the final linearization of the vector. LongAmpTM Taq DNA polymerase was also used for amplification of the 1 Kb PCR fragment (oligos 3 and 4 in appendix B), bearing SpeI and StuI restriction sites to insert the 7,7 kb restriction fragment. This amplicon was then

cloned in pIDTSMART-KAN (the intermediate plasmid) following FseI (NEB) digestion. The resulting construct was introduced and amplified in a dam-/dcm- strain (GM2163, NEB) of E. coli, kindly provided by Prof. X. De Bolle. Restriction of the vector and of bMQ120-c13 to produce the 7,7 kb restriction fragment was performed by SpeI (NEB) and StuI (NEB).

CRISPR vectors

Assembly of the CRISPR vector was achieved following the tips of the ZhangLab website (http://crispr.genome-engineering.org/) and a protocol of Ran et al, 2013 [101]. The starting vectors, pX334-U6-DR-BB-DR-Cbh-NLS-hSpCas9n(D10A)-NLS-H1-shorttracr-PGK-puro (Plasmid #42333) and pX335-U6-Chimeric_BB-CBh-hSpCas9n(D10A) (Plasmid #42335), were ordered from Addgene. For pX335, oligonucleotides (oligos 15-22 in appendix B) were designed on a 20 bp basis, corresponding to the chosen protospacers, immediately 5' of a 5'-NGG PAM (which was not included in the oligos). A 5'G was added on oligos when they did not already begin by this base. Additional bases were added on 5' to create overhangs compatible with the ends generated by restriction with BbsI. Oligos were designed in partially complementary pairs for each protospacer targeting. The annealed oligos corresponded to the following pattern:

```
5'-CACC(G)NNNNNNNNNNNNNNNNNNN
          |||||||||||||||||||||
         (C)NNNNNNNNNNNNNNNNNNNAAAC-5'
```

For cloning into pX334, oligos (oligos 8-14 in appendix B) were designed on a 30 bp basis, namely 20 bp corresponding to the sequence cloned in pX335 and an extension of 10 bp on the 5' side (distal to the PAM). Here, a 5' G was not needed. The final duplexes differed in their 5' overhangs with respect to the duplexes cloned into pX335:

```
5'-AAACNNNNNNNNNNNNNNNNNNNNNNNNNNNG T
          |||||||||||||||||||||||||||||||
         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAAAT-5'
```

Before annealing and ligation, all oligos were phosphorylated with the T4 Polynucleotide kinase (NEB). The CRISPR vectors were restricted by BbsI (NEB) prior to the cloning of the guide duplexes.

miRNA expression vectors

The miRNA expression vector, pCA-ß-EGFPm5-mU6, was kindly provided by Prof. B. Muylkens. The genomic sequences of miR-105 and miR-767, surrounded on both sides by 20-30 bp, were PCR-amplified (oligos 50-53 in appendix B) with the DreamTaq Green PCR Master Mix (Fermentas). Restriction sites were added on the primers for subsequent insertion in the vector, namely for ApaI (NEB) at the 5'end and EcoRIV (Promega) at the 3' end.

Southern blot and Screening by hybridization

Following SpeI-StuI restriction of bMQ120-c13, 2µg of DNA were engaged in electrophoresis on a 0,7% agarose gel and blotted on a nylon membrane (hybond). The oligonucleotide probe (oligo 5 in appendix B) was marked with Adenosine 5'-triphosphate, [γ-^{32}P]- (EasyTides,

PerkinElmer) by the T4 polynucleotide kinase (NEB). It was then purified through a CHROMA SPIN-10 Column (Clontech). Prehybridization was carried out during 1h at 42°C in PerfectHyb™ Plus Hybridization Buffer (Sigma-Aldrich). Proper hybridization was carried out overnight at 58°C. The membrane was washed in SSC 6X at room temperature and then in SSC 2X at 55°C, to remove non-specific binding of the probe. It was then left for radioactive impression on a photosensitive film (Amersham Hyperfilm™ MP, GE Healthcare) for 4 h before development.

For the screening of bacteria containing the 7,7 kb SpeI-StuI restriction fragment, the radioactive probe used for Southern blot was re-used. Electroporated bacteria were plated on a nylon membrane (hybond) placed on agar, with antibiotic selection. After overnight growth at 37°C, the colonies on the nylon membrane were transferred on a second one (the copy). This copy was placed at 37°C to allow bacterial growth. Hybridization was then performed similarly to hybridization on the Southern blot membrane, after bacterial lysis and DNA fixation, according to a protocol described elsewhere[2].

Cell culture

E14.Tg2a mESCs were maintained on 0,1% gelatin-coated culture dishes, in G-MEM, (Lonza) supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin (Lonza), 1 mM sodium pyruvate, 1 mM nonessential amino acids, 100 µM ß-mercaptoethanol and 1500 U/ml LIF (Millipore, ESG1106, ESGRO® Leukemia Inhibitory Factor).

Mouse fibroblastic NIH 3T3 cells were maintained in DMEM (GIBCO) supplemented with 10% (v/v) foetal bovine serum (Lonza) and 1% penicillin/streptomycin (Lonza).

5-azadC treatment

Demethylation agent 5-azadC (Sigma-Aldrich, REFA3656 Sigma) was received as a powder, dissolved in DMSO at a concentration of 50mM, aliquoted and stored at -80°C. When treating the cells, appropriate dilution of the stock solution was made in culture medium before pouring in culture dishes.

During the preliminary experiment to determine a dose to use, cells were maintained in 5-azadC at indicated concentrations, without interruption for 72 h. Each concentration was tested in duplicates. As the product is highly unstable in culture conditions, the medium was changed daily, including in the untreated condition. Cells from all conditions were harvested and counted with trypan blue coloration after the 72h period.

In the second experiment to assess toxicity and demethylation with time, cells were all plated at the same density, and treated in triplicates at a concentration of 2µM. Again, treatment was uninterrupted with daily medium replacement. Treated cells were harvested after 24h, 48h and

[2] Sambrook and Russel, Molecular Cloning, A Laboratory Manual, 3th edition, Cold Spring Harbor, New York, 2001, (1.129-1.142)

72h. Untreated cells were harvested after 48h. Cell counts with trypan blue were made for each condition. The starting cell number was identical for all wells.

In the experiment with discontinuous treatment, two doses were tested, namely 2µM and 4µM. Conditions were made in duplicates. The first dose was administered at 60-70% confluency, and cells remained in 5-azadC during 24h. After this period, the medium was replaced by a drug-free medium. A second similar administration was made 48h after the first one, and cells were collected for RNA extraction 72h later, after a total of 120h of treatment.

Transfections

Transfection in mESCs was carried out using the FuGENE® HD Transfection Reagent (Promega) according the manufacturer recommendations. One day before transfection, the cells were counted up and plated in 6-well plates at a density of $3x10^5$ cells per well in order to reach optimum density on the next day. A total of 3,3 µg of plasmid DNA was first mixed in Opti-MEM® Reduced Serum Medium (Gibco®) in a total volume of 158 µl. An equal proportion of DNA was calculated for all the plasmid to be transfected. For conditions with pX334, 2 plasmids were mixed, corresponding to the 2 RNA guides of a targeting pair. For conditions with pX335, an additional plasmid was needed to provide puromycin resistance. Cells were passaged in a new plate the day after transfection. Puromycin selection was started after a few hours the same day and was maintained for 48h, at a concentration of 1,5 µg/ml.

Transfection in 3T3 cells with miRNA expression vectors was carried out in 12-well plates, using Lipofectamine® 2000 Transfection Reagent (Invitrogen™) and following the manufacturer recommendations (5µg of DNA and 2µl of Lipofectamine). Cells were transfected in 4 replicates for each miRNA expression vector. DNA-lipid complexes were made in Opti-MEM® Reduced Serum Medium prior to transfection. Cells were collected for RNA extraction 1 day after transfection.

Serial dilution

To isolate clonal populations of targeted mESCs, cells were counted, appropriately diluted and plated in 96-well plates at densities of 15 and 45 cells per plate. After a period of 7 days, individual clones were identified and gathered on a new plate for amplification.

DNA/RNA isolation

For DNA isolation from cultured cells, cells were lysed overnight at 37°C in a lysis solution (100 mM TrisHCl, 5 mM EDTA, 0,2% SDS and 200 mM NaCl) containing 100 µg/ml of proteinase K. DNA was then isolated by ethanol precipitation.

For RT-PCR experiments on *Gabra3*, total RNA was isolated following purification with RIBOzol® Reagent (AMRESCO®) and chloroform, by isopropanol and ethanol precipitation. For analysis of miR-767 and miR-105 expressions, total RNA was purified through spin columns using the miRNeasy Mini Kit (QIAGEN) following the manufacturer instruction.

RT-(q)PCR

For mRNA reverse transcription, the GoScript™ Reverse Transcription System (Promega, #A5000) was used following the manufacturer recommendations, with random RT primers. For reverse transcription of miRNAs, the Universal cDNA Synthesis Kit II (Exiqon, #203301) was used, following the manufacturer recommendations. A spike in control provided by the kit was included in the samples.

For qPCR amplifications aimed to detect mRNAs, the Maxima SYBR Green/ROX qPCR Master Mix (Fermentas) was used. Expression data were normalized by *Gapdh* expression for expression in tissues. The gene *ActB* was then chosen for normalizations in cells (and normalization of the testis sample used to express the levels of *Magea5*). The primers used are available in appendix B: oligos 31 and 33 for BG-*Gabra3* (a and c), 34 and 35 for total *Gabra3* (d and e), 38 and 39 for *Magea5*, 53 and 54 for *Gapdh* and 55 and 56 for *ActB*.

For qPCR amplifications aimed to detect small RNAs, the ExiLENT SYBR® Green master mix (Exiqon, #203403) was used, using specific primers (miRCURY LNA™ primer set) for miR-767 (n° 205243) and miR-105 (n°205090). Snord68 (primer set n°203911) and U6 snRNA (primer set n° 203907) were tested as endogenous controls, and Snord68 was selected for all miRNA normalizations. The miR-767 synthetic spike in was ordered from Eurogentec (see appendix B for sequence).

For nested PCR, the DreamTaq Green PCR Master Mix (Fermentas) was used. Products of the first row of amplification (30 cycles) were diluted 100 fold before the second amplification (30 cycles). The primers used for (semi)nested-PCR on *Gabra3* were those used in qPCR, nested with an additional primer. For the 2 RNAseq exon predictions, 2 primers were designed in each predicted sequence, and sequentially used with 2 primers located in *Gabra3* mRNA #1.

	BG-*Gabra3*	Total *Gabra3*	RNAseq pred. 1	RNAseq pred.2
N° oligos 1st PCR	31+33	34+37	33+46	33+47
N° oligos 2nd PCR	31+36	34+35	36+45	36+48

5'RACE

5'RACE was performed on RNA from brain, testis and NIH3T3 cells treated with 5-azadC (continuous treatment, 120h) already isolated, with the 5' RACE Core Set (Takara, Clontech, #6122). Reverse transcription was performed with an RT primer specific for *Gabra3* (n°40 in appendix B). Nested PCR was performed with the DreamTaq Green PCR Master Mix with 2 pairs of primers available in appendix B (sequentially: n°41+43 and n°42+45).

Statistical analysis

Statistical analysis on RT-qPCR data to compare 2 means were performed on the ΔCT values (CT values normalized with the endogenous control) with a non-parametric Mann-Whitney U test (one tailed to compare Magea5 induction between 72h and 120h of 5-azadC treatment and Gabra3 induction, and two-tailed to compare expression of miR-105 and miR-767 in transfected cells).

TA cloning

The PCR products obtained in testis for RNAseq prediction 1 and the PCR products obtained by 5'RACE in the brain were cloned in the PCR®2.1 Vector using The Original TA Cloning kit (Invitrogen, #45-0046) and sequenced by Sanger sequencing (3130 Genetic Analyzer, Applied Biosystems).

5. REFERENCES

[1] C. Plass, S. M. Pfister, A. M. Lindroth, O. Bogatyrova, R. Claus, and P. Lichter, "Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer.," *Nat. Rev. Genet.*, vol. 14, no. 11, pp. 765–80, Nov. 2013.

[2] M. a Dawson and T. Kouzarides, "Cancer epigenetics: from mechanism to therapy.," *Cell*, vol. 150, no. 1, pp. 12–27, Jul. 2012.

[3] A. Portela and M. Esteller, "Epigenetic modifications and human disease.," *Nat. Biotechnol.*, vol. 28, no. 10, pp. 1057–68, Oct. 2010.

[4] J. F. Costello, M. C. Frühwald, D. J. Smiraglia, L. J. Rush, G. P. Robertson, X. Gao, F. A. Wright, J. D. Feramisco, P. Peltomäki, J. C. Lang, D. E. Schuller, L. Yu, C. D. Bloomfield, M. A. Caligiuri, A. Yates, R. Nishikawa, H. Su Huang, N. J. Petrelli, X. Zhang, M. S. O'Dorisio, W. A. Held, W. K. Cavenee, and C. Plass, "Aberrant CpG-island methylation has non-random and tumour-type-specific patterns.," *Nat. Genet.*, vol. 24, no. 2, pp. 132–8, Feb. 2000.

[5] D. J. Smiraglia, L. J. Rush, M. C. Frühwald, Z. Dai, W. A. Held, J. F. Costello, J. C. Lang, C. Eng, B. Li, F. A. Wright, M. A. Caligiuri, and C. Plass, "Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies.," *Hum. Mol. Genet.*, vol. 10, no. 13, pp. 1413–9, Jun. 2001.

[6] S. B. Baylin and P. A. Jones, "A decade of exploring the cancer epigenome — biological and translational implications," *Nat. Rev. Cancer*, vol. 11, no. 10, pp. 726–734, 2011.

[7] J. E. Ohm, K. M. McGarvey, X. Yu, L. Cheng, K. E. Schuebel, L. Cope, H. P. Mohammad, W. Chen, V. C. Daniel, W. Yu, D. M. Berman, T. Jenuwein, K. Pruitt, S. J. Sharkis, D. N. Watkins, J. G. Herman, and S. B. Baylin, "A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing.," *Nat. Genet.*, vol. 39, no. 2, pp. 237–42, Feb. 2007.

[8] H. Easwaran, S. E. Johnstone, L. Van Neste, J. Ohm, T. Mosbruger, Q. Wang, M. J. Aryee, P. Joyce, N. Ahuja, D. Weisenberger, E. Collisson, J. Zhu, S. Yegnasubramanian, W. Matsui, and S. B. Baylin, "A DNA hypermethylation module for the stem/progenitor cell signature of cancer.," *Genome Res.*, vol. 22, no. 5, pp. 837–49, May 2012.

[9] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander, "A bivalent chromatin structure marks key developmental genes in embryonic stem cells.," *Cell*, vol. 125, no. 2, pp. 315–26, Apr. 2006.

[10] A. Doi, I.-H. Park, B. Wen, P. Murakami, M. J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G. Q. Daley, and A. P. Feinberg, "Differential methylation of tissue- and cancer-specific CpG island shores

distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.," *Nat. Genet.*, vol. 41, no. 12, pp. 1350–3, Dec. 2009.

[11] R. A. Irizarry, C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J. B. Potash, S. Sabunciyan, and A. P. Feinberg, "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.," *Nat. Genet.*, vol. 41, no. 2, pp. 178–86, Feb. 2009.

[12] F. Gaudet, J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt, and R. Jaenisch, "Induction of tumors in mice by genomic hypomethylation.," *Science*, vol. 300, no. 5618, pp. 489–92, Apr. 2003.

[13] E. Lee, R. Iskow, L. Yang, O. Gokcumen, P. Haseley, L. J. Luquette, J. G. Lohr, C. C. Harris, L. Ding, R. K. Wilson, D. A. Wheeler, R. A. Gibbs, R. Kucherlapati, C. Lee, P. V Kharchenko, and P. J. Park, "Landscape of somatic retrotransposition in human cancers.," *Science*, vol. 337, no. 6097, pp. 967–71, Aug. 2012.

[14] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker, "Human DNA methylomes at base resolution show widespread epigenomic differences.," *Nature*, vol. 462, no. 7271, pp. 315–22, Nov. 2009.

[15] D. I. Schroeder, P. Lott, I. Korf, and J. M. LaSalle, "Large-scale methylation domains mark a functional subset of neuronally expressed genes.," *Genome Res.*, vol. 21, no. 10, pp. 1583–91, Oct. 2011.

[16] B. P. Berman, D. J. Weisenberger, J. F. Aman, T. Hinoue, Z. Ramjan, Y. Liu, H. Noushmehr, C. P. E. Lange, C. M. van Dijk, R. A. E. M. Tollenaar, D. Van Den Berg, and P. W. Laird, "Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains.," *Nat. Genet.*, vol. 44, no. 1, pp. 40–6, Jan. 2012.

[17] A. K. Maunakea, R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. M. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang, and J. F. Costello, "Conserved role of intragenic DNA methylation in regulating alternative promoters.," *Nature*, vol. 466, no. 7303, pp. 253–7, Jul. 2010.

[18] M. Bobinet, V. Vignard, L. Florenceau, F. Lang, N. Labarriere, and A. Moreau-Aubry, "Overexpression of meloe gene in melanomas is controlled both by specific transcription factors and hypomethylation.," *PLoS One*, vol. 8, no. 9, p. e75421, Jan. 2013.

[19] T. Schoofs, C. Rohde, K. Hebestreit, H.-U. Klein, S. Göllner, I. Schulze, M. Lerdrup, N. Dietrich, S. Agrawal-Singh, A. Witten, M. Stoll, E. Lengfelder, W.-K. Hofmann, P. Schlenke, T. Büchner, K. Hansen, W. E. Berdel, F. Rosenbauer, M. Dugas, and C. Müller-Tidow, "DNA methylation changes are a late event in acute promyelocytic

leukemia and coincide with loss of transcription factor binding.," *Blood*, vol. 121, no. 1, pp. 178–87, Jan. 2013.

[20] C. De Smet, C. Lurquin, B. Lethé, C. D. E. Smet, B. Lethe, and T. Boon, "DNA Methylation Is the Primary Silencing Mechanism for a Set of Germ Line- and Tumor-Specific Genes with a CpG-Rich Promoter DNA Methylation Is the Primary Silencing Mechanism for a Set of Germ Line- and Tumor-Specific Genes with a CpG-Rich Promoter," 1999.

[21] C. De Smet and A. Loriot, "DNA hypomethylation and activation of germline-specific genes in cancer.," *Adv. Exp. Med. Biol.*, vol. 754, pp. 149–66, Jan. 2013.

[22] J. Cannuyer, A. Loriot, G. K. Parvizi, and C. De Smet, "Epigenetic hierarchy within the MAGEA1 cancer-germline gene: promoter DNA methylation dictates local histone modifications.," *PLoS One*, vol. 8, no. 3, p. e58743, Jan. 2013.

[23] M. Sang, L. Wang, C. Ding, X. Zhou, B. Wang, L. Wang, Y. Lian, and B. Shan, "Melanoma-associated antigen genes - an update.," *Cancer Lett.*, vol. 302, no. 2, pp. 85–90, Mar. 2011.

[24] O. L. Caballero and Y.-T. Chen, "Cancer/testis (CT) antigens: potential targets for immunotherapy.," *Cancer Sci.*, vol. 100, no. 11, pp. 2014–21, Nov. 2009.

[25] D. Dhanak and P. Jackson, "Development and classes of epigenetic drugs for cancer.," *Biochem. Biophys. Res. Commun.*, Jul. 2014.

[26] L. Shen, C.-X. Song, C. He, and Y. Zhang, "Mechanism and function of oxidative reversal of DNA and RNA methylation.," *Annu. Rev. Biochem.*, vol. 83, pp. 585–614, Jan. 2014.

[27] R. M. Kohli and Y. Zhang, "TET enzymes, TDG and the dynamics of DNA demethylation.," *Nature*, vol. 502, no. 7472, pp. 472–9, Oct. 2013.

[28] H. Wu and Y. Zhang, "Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation.," *Genes Dev.*, vol. 25, no. 23, pp. 2436–52, Dec. 2011.

[29] D. Subramaniam, R. Thombre, A. Dhar, and S. Anant, "DNA methyltransferases: a novel target for prevention and therapy.," *Front. Oncol.*, vol. 4, p. 80, Jan. 2014.

[30] B. Hackanson and M. Daskalakis, "Decitabine.," *Recent Results Cancer Res.*, vol. 201, pp. 269–97, Jan. 2014.

[31] D. Mossman, K.-T. Kim, and R. J. Scott, "Demethylation by 5-aza-2'-deoxycytidine in colorectal cancer cells targets genomic DNA whilst promoter CpG island methylation persists.," *BMC Cancer*, vol. 10, p. 366, Jan. 2010.

[32] a Shore, a Karamitri, P. Kemp, J. R. Speakman, and M. a Lomax, "Role of Ucp1 enhancer methylation and chromatin remodelling in the control of Ucp1 expression in murine adipose tissue.," *Diabetologia*, vol. 53, no. 6, pp. 1164–73, Jun. 2010.

[33] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao, "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.," *Science*, vol. 324, no. 5929, pp. 930–5, May 2009.

[34] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang, "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.," *Science*, vol. 333, no. 6047, pp. 1300–3, Sep. 2011.

[35] V. Valinluck and L. C. Sowers, "Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1.," *Cancer Res.*, vol. 67, no. 3, pp. 946–50, Feb. 2007.

[36] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, and T. Carell, "Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing.," *Angew. Chem. Int. Ed. Engl.*, vol. 51, no. 26, pp. 6516–20, Jun. 2012.

[37] C.-C. Chen, K.-Y. Wang, and C.-K. J. Shen, "The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases.," *J. Biol. Chem.*, vol. 287, no. 40, pp. 33116–21, Sep. 2012.

[38] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, and G.-L. Xu, "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.," *Science*, vol. 333, no. 6047, pp. 1303–7, Sep. 2011.

[39] A. Maiti and A. C. Drohat, "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites.," *J. Biol. Chem.*, vol. 286, no. 41, pp. 35334–8, Oct. 2011.

[40] J. U. Guo, Y. Su, C. Zhong, G. Ming, and H. Song, "Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain.," *Cell*, vol. 145, no. 3, pp. 423–34, Apr. 2011.

[41] T.-P. Gu, F. Guo, H. Yang, H.-P. Wu, G.-F. Xu, W. Liu, Z.-G. Xie, L. Shi, X. He, S. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabó, G. P. Pfeifer, J. Li, and G.-L. Xu, "The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes.," *Nature*, vol. 477, no. 7366, pp. 606–10, Sep. 2011.

[42] A. Inoue and Y. Zhang, "Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos.," *Science*, vol. 334, no. 6053, p. 194, Oct. 2011.

[43] T. Nakamura, Y.-J. Liu, H. Nakashima, H. Umehara, K. Inoue, S. Matoba, M. Tachibana, A. Ogura, Y. Shinkai, and T. Nakano, "PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos.," *Nature*, vol. 486, no. 7403, pp. 415–9, Jun. 2012.

[44] J. A. Hackett, R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down, and M. A. Surani, "Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine.," *Science*, vol. 339, no. 6118, pp. 448–52, Jan. 2013.

[45] J. A. Hackett, J. J. Zylicz, and M. A. Surani, "Parallel mechanisms of epigenetic reprogramming in the germline.," *Trends Genet.*, vol. 28, no. 4, pp. 164–74, Apr. 2012.

[46] K. Kurimoto, Y. Yabuta, Y. Ohinata, M. Shigeta, K. Yamanaka, and M. Saitou, "Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice.," *Genes Dev.*, vol. 22, no. 12, pp. 1617–35, Jun. 2008.

[47] S. Yamaguchi, K. Hong, R. Liu, A. Inoue, L. Shen, K. Zhang, and Y. Zhang, "Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming.," *Cell Res.*, vol. 23, no. 3, pp. 329–39, Mar. 2013.

[48] M. M. Dawlaty, A. Breiling, T. Le, G. Raddatz, M. I. Barrasa, A. W. Cheng, Q. Gao, B. E. Powell, Z. Li, M. Xu, K. F. Faull, F. Lyko, and R. Jaenisch, "Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development.," *Dev. Cell*, vol. 24, no. 3, pp. 310–23, Mar. 2013.

[49] S. Yamaguchi, L. Shen, Y. Liu, D. Sendler, and Y. Zhang, "Role of Tet1 in erasure of genomic imprinting.," *Nature*, vol. 504, no. 7480, pp. 460–4, Dec. 2013.

[50] S. Ito, A. C. D'Alessio, O. V Taranova, K. Hong, L. C. Sowers, and Y. Zhang, "Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.," *Nature*, vol. 466, no. 7310, pp. 1129–33, Aug. 2010.

[51] K. P. Koh, A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley, and A. Rao, "Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells.," *Cell Stem Cell*, vol. 8, no. 2, pp. 200–13, Mar. 2011.

[52] Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, and Y. G. Shi, "Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells.," *Mol. Cell*, vol. 42, no. 4, pp. 451–64, May 2011.

[53] K. Takahashi and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.," *Cell*, vol. 126, no. 4, pp. 663–76, Aug. 2006.

[54] Y. Costa, J. Ding, T. W. Theunissen, F. Faiola, T. A. Hore, P. V Shliaha, M. Fidalgo, A. Saunders, M. Lawrence, S. Dietmann, S. Das, D. N. Levasseur, Z. Li, M. Xu, W. Reik, J. C. R. Silva, and J. Wang, "NANOG-dependent function of TET1 and TET2 in establishment of pluripotency.," *Nature*, vol. 495, no. 7441, pp. 370–4, Mar. 2013.

[55] Y. Gao, J. Chen, K. Li, T. Wu, B. Huang, W. Liu, X. Kou, Y. Zhang, H. Huang, Y. Jiang, C. Yao, X. Liu, Z. Lu, Z. Xu, L. Kang, J. Chen, H. Wang, T. Cai, and S. Gao,

"Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming.," *Cell Stem Cell*, vol. 12, no. 4, pp. 453–69, Apr. 2013.

[56] E. Solary, O. a Bernard, a Tefferi, F. Fuks, and W. Vainchenker, "The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases.," *Leukemia*, vol. 28, no. 3, pp. 485–96, Mar. 2014.

[57] S. M. G. Braun and S. Jessberger, "Adult neurogenesis: mechanisms and functional significance.," *Development*, vol. 141, no. 10, pp. 1983–6, May 2014.

[58] T. Li, D. Yang, J. Li, Y. Tang, J. Yang, and W. Le, "Critical Role of Tet3 in Neural Progenitor Cell Maintenance and Terminal Differentiation.," *Mol. Neurobiol.*, May 2014.

[59] J. J. Day and J. D. Sweatt, "DNA methylation and memory formation.," *Nat. Neurosci.*, vol. 13, no. 11, pp. 1319–23, Nov. 2010.

[60] R. E. Brown, L. Stanford, and H. M. Schellinck, "Developing Standardized Behavioral Tests for Knockout and Mutant Mice," *ILAR J.*, vol. 41, no. 3, pp. 163–174, Jan. 2000.

[61] G. a Kaas, C. Zhong, D. E. Eason, D. L. Ross, R. V Vachhani, G.-L. Ming, J. R. King, H. Song, and J. D. Sweatt, "TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation.," *Neuron*, vol. 79, no. 6, pp. 1086–93, Sep. 2013.

[62] A. Rudenko, M. M. Dawlaty, J. Seo, A. W. Cheng, J. Meng, T. Le, K. F. Faull, R. Jaenisch, and L.-H. Tsai, "Tet1 is critical for neuronal activity-regulated gene expression and memory extinction.," *Neuron*, vol. 79, no. 6, pp. 1109–22, Sep. 2013.

[63] X. Li, W. Wei, Q.-Y. Zhao, J. Widagdo, D. Baker-Andresen, C. R. Flavell, A. D'Alessio, Y. Zhang, and T. W. Bredy, "Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 19, pp. 7120–5, May 2014.

[64] R. Zhang, Q. Cui, K. Murai, Y. C. Lim, Z. D. Smith, S. Jin, P. Ye, L. Rosa, Y. K. Lee, H. Wu, W. Liu, Z. Xu, L. Yang, Y. Ding, F. Tang, A. Meissner, C. Ding, Y. Shi, and G. Xu, "Tet1 Regulates Adult Hippocampal Neurogenesis and Cognition," *Stem Cell*, vol. 13, no. 2, pp. 237–245, 2013.

[65] Y. Huang and A. Rao, "Connections between TET proteins and aberrant DNA modification in cancer.," *Trends Genet.*, pp. 1–11, Aug. 2014.

[66] R. B. Lorsbach, J. Moore, S. Mathew, S. C. Raimondi, S. T. Mukatira, and J. R. Downing, "TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23).," *Leukemia*, vol. 17, no. 3, pp. 637–41, Mar. 2003.

[67] H. Huang, X. Jiang, Z. Li, Y. Li, C.-X. Song, C. He, M. Sun, P. Chen, S. Gurbuxani, J. Wang, G.-M. Hong, A. G. Elkahloun, S. Arnovitz, J. Wang, K. Szulwach, L. Lin, C.

Street, M. Wunderlich, M. Dawlaty, M. B. Neilly, R. Jaenisch, F.-C. Yang, J. C. Mulloy, P. Jin, P. P. Liu, J. D. Rowley, M. Xu, C. He, and J. Chen, "TET1 plays an essential oncogenic role in MLL-rearranged leukemia.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 29, pp. 11994–9, Jul. 2013.

[68] M. Ko, Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, and A. Rao, "Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2.," *Nature*, vol. 468, no. 7325, pp. 839–43, Dec. 2010.

[69] C. G. Lian, Y. Xu, C. Ceol, F. Wu, A. Larson, K. Dresser, W. Xu, L. Tan, Y. Hu, Q. Zhan, C.-W. Lee, D. Hu, B. Q. Lian, S. Kleffel, Y. Yang, J. Neiswender, A. J. Khorasani, R. Fang, C. Lezcano, L. M. Duncan, R. a Scolyer, J. F. Thompson, H. Kakavand, Y. Houvras, L. I. Zon, M. C. Mihm, U. B. Kaiser, T. Schatton, B. a Woda, G. F. Murphy, and Y. G. Shi, "Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma.," *Cell*, vol. 150, no. 6, pp. 1135–46, Sep. 2012.

[70] H. Yang, Y. Liu, F. Bai, J.-Y. Zhang, S.-H. Ma, J. Liu, Z.-D. Xu, H.-G. Zhu, Z.-Q. Ling, D. Ye, K.-L. Guan, and Y. Xiong, "Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation.," *Oncogene*, vol. 32, no. 5, pp. 663–9, Jan. 2013.

[71] C.-H. Hsu, K.-L. Peng, M.-L. Kang, Y.-R. Chen, Y.-C. Yang, C.-H. Tsai, C.-S. Chu, Y.-M. Jeng, Y.-T. Chen, F.-M. Lin, H.-D. Huang, Y.-Y. Lu, Y.-C. Teng, S.-T. Lin, R.-K. Lin, F.-M. Tang, S.-B. Lee, H. M. Hsu, J.-C. Yu, P.-W. Hsiao, and L.-J. Juan, "TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases.," *Cell Rep.*, vol. 2, no. 3, pp. 568–79, Sep. 2012.

[72] A. M. A. S. Monteys, R. M. Spengler, J. I. Wan, L. Tecedor, K. A. Lennox, Y. I. Xing, and B. L. Davidson, "Structure and activity of putative intronic miRNA promoters," pp. 495–505, 2010.

[73] M. Ha and V. N. Kim, "Regulation of microRNA biogenesis.," *Nat. Rev. Mol. Cell Biol.*, vol. 17, no. July, Jul. 2014.

[74] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.," *Cell*, vol. 120, no. 1, pp. 15–20, Jan. 2005.

[75] S. a Melo and M. Esteller, "Dysregulation of microRNAs in cancer: playing with fire.," *FEBS Lett.*, vol. 585, no. 13, pp. 2087–99, Jul. 2011.

[76] G. A. Calin, C. Sevignani, C. D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini, and C. M. Croce, "Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 9, pp. 2999–3004, Mar. 2004.

[77] P. Lopez-Serra and M. Esteller, "DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer.," *Oncogene*, vol. 31, no. 13, pp. 1609–22, Mar. 2012.

[78] A. Lujambio, G. A. Calin, A. Villanueva, S. Ropero, M. Sánchez-Céspedes, D. Blanco, L. M. Montuenga, S. Rossi, M. S. Nicoloso, W. J. Faller, W. M. Gallagher, S. A. Eccles, C. M. Croce, and M. Esteller, "A microRNA DNA methylation signature for human cancer metastasis.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 36, pp. 13556–61, Sep. 2008.

[79] J. Weischenfeldt, R. Simon, L. Feuerbach, K. Schlangen, D. Weichenhan, S. Minner, D. Wuttig, H.-J. Warnatz, H. Stehr, T. Rausch, N. Jäger, L. Gu, O. Bogatyrova, A. M. Stütz, R. Claus, J. Eils, R. Eils, C. Gerhäuser, P.-H. Huang, B. Hutter, R. Kabbe, C. Lawerenz, S. Radomski, C. C. Bartholomae, M. Fälth, S. Gade, M. Schmidt, N. Amschler, T. Haß, R. Galal, J. Gjoni, R. Kuner, C. Baer, S. Masser, C. von Kalle, T. Zichner, V. Benes, B. Raeder, M. Mader, V. Amstislavskiy, M. Avci, H. Lehrach, D. Parkhomchuk, M. Sultan, L. Burkhardt, M. Graefen, H. Huland, M. Kluth, A. Krohn, H. Sirma, L. Stumm, S. Steurer, K. Grupp, H. Sültmann, G. Sauter, C. Plass, B. Brors, M.-L. Yaspo, J. O. Korbel, and T. Schlomm, "Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer.," *Cancer Cell*, vol. 23, no. 2, pp. 159–70, Feb. 2013.

[80] M. Fabbri, R. Garzon, A. Cimmino, Z. Liu, N. Zanesi, E. Callegari, S. Liu, H. Alder, S. Costinean, C. Fernandez-Cymering, S. Volinia, G. Guler, C. D. Morrison, K. K. Chan, G. Marcucci, G. A. Calin, K. Huebner, and C. M. Croce, "MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 40, pp. 15805–10, Oct. 2007.

[81] S. J. Song, L. Poliseno, M. S. Song, U. Ala, K. Webster, C. Ng, G. Beringer, N. J. Brikbak, X. Yuan, L. C. Cantley, A. L. Richardson, and P. P. Pandolfi, "MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling.," *Cell*, vol. 154, no. 2, pp. 311–24, Jul. 2013.

[82] S. J. Song, K. Ito, U. Ala, L. Kats, K. Webster, S. M. Sun, M. Jongen-Lavrencic, K. Manova-Todorova, J. Teruya-Feldstein, D. E. Avigan, R. Delwel, and P. P. Pandolfi, "The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation.," *Cell Stem Cell*, vol. 13, no. 1, pp. 87–101, Jul. 2013.

[83] P. Dalerba and M. F. Clarke, "Oncogenic miRNAs and the perils of losing control of a stem cell's epigenetic identity.," *Cell Stem Cell*, vol. 13, no. 1, pp. 5–6, Jul. 2013.

[84] S. Morita, T. Horii, M. Kimura, T. Ochiya, S. Tajima, and I. Hatada, "miR-29 represses the activities of DNA methyltransferases and DNA demethylases.," *Int. J. Mol. Sci.*, vol. 14, no. 7, pp. 14647–58, Jan. 2013.

[85] J. Cheng, S. Guo, S. Chen, S. J. Mastriano, C. Liu, A. C. D'Alessio, E. Hysolli, Y. Guo, H. Yao, C. M. Megyola, D. Li, J. Liu, W. Pan, C. A. Roden, X.-L. Zhou, K. Heydari, J. Chen, I.-H. Park, Y. Ding, Y. Zhang, and J. Lu, "An extensive network of TET2-targeting MicroRNAs regulates malignant hematopoiesis.," *Cell Rep.*, vol. 5, no. 2, pp. 471–81, Oct. 2013.

[86] a G. Bader, D. Brown, J. Stoudemire, and P. Lammers, "Developing therapeutic microRNAs for cancer.," *Gene Ther.*, vol. 18, no. 12, pp. 1121–6, Dec. 2011.

[87] Z. Li and T. M. Rana, "Therapeutic targeting of microRNAs: current status and future challenges," *Nat. Rev. Drug Discov.*, vol. 13, no. 8, pp. 622–38, Jul. 2014.

[88] A. Bouchie, "First microRNA mimic enters clinic.," *Nat. Biotechnol.*, vol. 31, no. 7, p. 577, Jul. 2013.

[89] A. Loriot, A. Van Tongelen, J. Blanco, S. Klaessens, J. Cannuyer, and N. Van Baren, "A novel cancer-germline transcript carrying pro-metastatic miR-105 and TET - targeting miR-767 induced by DNA hypomethylation in tumors," *Epigenetics*, vol. 9, no. 8, pp. 1163–1171, 2014.

[90] Y. Liu, F. Guo, M. Dai, D. Wang, Y. Tong, J. Huang, J. Hu, and G. Li, "Gammaaminobutyric acid A receptor alpha 3 subunit is overexpressed in lung cancer.," *Pathol. Oncol. Res.*, vol. 15, no. 3, pp. 351–8, Sep. 2009.

[91] Y. Lu, W. Lemon, P.-Y. Liu, Y. Yi, C. Morrison, P. Yang, Z. Sun, J. Szoke, W. L. Gerald, M. Watson, R. Govindan, and M. You, "A gene expression signature predicts survival of patients with stage I non-small cell lung cancer.," *PLoS Med.*, vol. 3, no. 12, p. e467, Dec. 2006.

[92] X. Zhang, R. Zhang, Y. Zheng, J. Shen, D. Xiao, J. Li, X. Shi, L. Huang, H. Tang, J. Liu, J. He, and H. Zhang, "Expression of gamma-aminobutyric acid receptors on neoplastic growth and prediction of prognosis in non-small cell lung cancer.," *J. Transl. Med.*, vol. 11, p. 102, Jan. 2013.

[93] Y. Liu, "Gamma-aminobutyric acid promotes human hepatocellular carcinoma growth through overexpressed gamma-aminobutyric acid A receptor α3 subunit," *World J. Gastroenterol.*, vol. 14, no. 47, p. 7175, 2008.

[94] W. Zhou, M. Y. Fong, Y. Min, G. Somlo, L. Liu, M. R. Palomares, Y. Yu, A. Chow, S. T. F. O'Connor, A. R. Chin, Y. Yen, Y. Wang, E. G. Marcusson, P. Chu, J. Wu, X. Wu, A. X. Li, Z. Li, H. Gao, X. Ren, M. P. Boldin, P. C. Lin, and S. E. Wang, "Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis.," *Cancer Cell*, vol. 25, no. 4, pp. 501–15, Apr. 2014.

[95] B. K. Yee, R. Keist, L. von Boehmer, R. Studer, D. Benke, N. Hagenbuch, Y. Dong, R. C. Malenka, J.-M. Fritschy, H. Bluethmann, J. Feldon, H. Möhler, and U. Rudolph, "A schizophrenia-related sensorimotor deficit links alpha 3-containing GABAA receptors to a dopamine hyperfunction.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 47, pp. 17154–9, Nov. 2005.

[96] R. Fiorelli, U. Rudolph, C. J. Straub, J. Feldon, and B. K. Yee, "Affective and cognitive effects of global deletion of alpha3-containing gamma-aminobutyric acid-A receptors.," *Behav. Pharmacol.*, vol. 19, no. 5–6, pp. 582–96, Sep. 2008.

[97] M. R. Capecchi, "Altering the Genome Homologous Recombination by From ES Cells to Germ Line Chimera," *Science (80-.).*, vol. 236, no. 1987, pp. 1288–1292, 1989.

[98] P. D. Hsu, E. S. Lander, and F. Zhang, "Development and Applications of CRISPR-Cas9 for Genome Engineering.," *Cell*, vol. 157, no. 6, pp. 1262–1278, Jun. 2014.

[99] R. Barrangou and L. a Marraffini, "CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity.," *Mol. Cell*, vol. 54, no. 2, pp. 234–44, Apr. 2014.

[100] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, P. D. Hsu, X. Wu, W. Jiang, and L. A. Marraffini, "Multiplex Genome Engineering Using CRISPR/Cas Systems," *Science (80-.).*, vol. 339, no. 6121, pp. 819–823, 2013.

[101] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. a Scott, and F. Zhang, "Genome engineering using the CRISPR-Cas9 system.," *Nat. Protoc.*, vol. 8, no. 11, pp. 2281–308, Nov. 2013.

[102] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.," *Science*, vol. 337, no. 6096, pp. 816–21, Aug. 2012.

[103] S. W. Cho, S. Kim, J. M. Kim, and J.-S. Kim, "Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease.," *Nat. Biotechnol.*, vol. 31, no. 3, pp. 230–2, Mar. 2013.

[104] M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna, "RNA-programmed genome editing in human cells.," *Elife*, vol. 2, p. e00471, Jan. 2013.

[105] N. K. Pyzocha, F. A. Ran, P. D. Hsu, and F. Zhang, "RNA-guided genome editing of mammalian cells.," *Methods Mol. Biol.*, vol. 1114, pp. 269–77, Jan. 2014.

[106] P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang, "DNA targeting specificity of RNA-guided Cas9 nucleases.," *Nat. Biotechnol.*, vol. 31, no. 9, pp. 827–32, Sep. 2013.

[107] F. A. Ran, P. D. Hsu, C. Lin, and J. S. Gootenberg, "Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity," *Cell*, vol. 154, no. 6, pp. 1380–1389, 2013.

[108] B. Shen, W. Zhang, J. Zhang, J. Zhou, J. Wang, L. Chen, L. Wang, A. Hodgkins, V. Iyer, X. Huang, and W. C. Skarnes, "Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects.," *Nat. Methods*, vol. 11, no. 4, pp. 399–402, Apr. 2014.

[109] S. J. Gratz, F. P. Ukken, C. D. Rubinstein, G. Thiede, L. K. Donohue, A. M. Cummings, and K. M. O'Connor-Giles, "Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila.," *Genetics*, vol. 196, no. 4, pp. 961–71, Apr. 2014.

[110] K. Li, G. Wang, T. Andersen, P. Zhou, and W. T. Pu, "Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System.," *PLoS One*, vol. 9, no. 8, p. e105779, Jan. 2014.

[111] D. J. R. and B. G. Daniel J. Dickinson, Jordan D. Ward, "Engineering the Caenorhabditis elegans Genome Using Triggered Homologous Recombination," *Nat. Methods*, vol. 10, no. 10, pp. 1028–1034, 2013.

[112] H. Wang, H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang, and R. Jaenisch, "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering.," *Cell*, vol. 153, no. 4, pp. 910–8, May 2013.

[113] A. Y.-F. Lee and K. C. K. Lloyd, "Conditional targeting of Ispd using paired Cas9 nickase and a single DNA template in mice.," *FEBS Open Bio*, vol. 4, pp. 637–42, Jan. 2014.

[114] C. Y. Park, L. T. Jeker, K. Carver-Moore, A. Oh, H. J. Liu, R. Cameron, H. Richards, Z. Li, D. Adler, Y. Yoshinaga, M. Martinez, M. Nefadov, A. K. Abbas, A. Weiss, L. L. Lanier, P. J. de Jong, J. A. Bluestone, D. Srivastava, and M. T. McManus, "A resource for the conditional ablation of microRNAs in the mouse.," *Cell Rep.*, vol. 1, no. 4, pp. 385–91, Apr. 2012.

[115] E. Wienholds, W. P. Kloosterman, E. Miska, E. Alvarez-Saavedra, E. Berezikov, E. de Bruijn, H. R. Horvitz, S. Kauppinen, and R. H. A. Plasterk, "MicroRNA expression in zebrafish embryonic development.," *Science*, vol. 309, no. 5732, pp. 310–1, Jul. 2005.

[116] J. H. Mansfield, B. D. Harfe, R. Nissen, J. Obenauer, J. Srineel, A. Chaudhuri, R. Farzan-Kashani, M. Zuker, A. E. Pasquinelli, G. Ruvkun, P. a Sharp, C. J. Tabin, and M. T. McManus, "MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression.," *Nat. Genet.*, vol. 36, no. 10, pp. 1079–83, Oct. 2004.

[117] E. De Plaen, O. De Backer, D. Arnaud, B. Bonjean, P. Chomez, V. Martelange, P. Avner, P. Baldacci, C. Babinet, S. Y. Hwang, B. Knowles, and T. Boon, "A new family of mouse genes homologous to the human MAGE genes.," *Genomics*, vol. 55, no. 2, pp. 176–84, Jan. 1999.

[118] A. R. Kore, M. Hodeib, and Z. Hu, "Chemical Synthesis of LNA-mCTP and its application for MicroRNA detection.," *Nucleosides. Nucleotides Nucleic Acids*, vol. 27, no. 1, pp. 1–17, Jan. 2008.

[119] R. Bron, B. J. Eickholt, M. Vermeren, N. Fragale, and J. Cohen, "Functional knockdown of neuropilin-1 in the developing chick nervous system by siRNA hairpins phenocopies genetic ablation in the mouse.," *Dev. Dyn.*, vol. 230, no. 2, pp. 299–308, Jun. 2004.

[120] Y. Li and K. V Kowdley, "Method for microRNA isolation from clinical serum samples.," *Anal. Biochem.*, vol. 431, no. 1, pp. 69–75, Dec. 2012.

[121] H. W. Lim, M. Iwatani, N. Hattori, S. Tanaka, S. Yagi, and K. Shiota, "Resistance to 5-aza-2'-deoxycytidine in genic regions compared to non-genic repetitive sequences.," *J. Reprod. Dev.*, vol. 56, no. 1, pp. 86–93, Mar. 2010.

[122] N. Redshaw, T. Wilkes, A. Whale, S. Cowen, J. Huggett, and C. A. Foy, "A comparison of miRNA isolation and RT-qPCR technologies and their effects on quantification accuracy and repeatability.," *Biotechniques*, vol. 54, no. 3, pp. 155–64, Mar. 2013.

[123] P. Chugh and D. P. Dittmer, "Potential pitfalls in microRNA profiling.," *Wiley Interdiscip. Rev. RNA*, vol. 3, no. 5, pp. 601–16.

[124] F. Kuchenbauer, R. D. Morin, B. Argiropoulos, O. I. Petriv, M. Griffith, M. Heuser, E. Yung, J. Piper, A. Delaney, A.-L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C. L. Hansen, M. A. Marra, and R. K. Humphries, "In-depth characterization of the microRNA transcriptome in a leukemia progression model.," *Genome Res.*, vol. 18, no. 11, pp. 1787–97, Nov. 2008.

[125] M. Morlando, M. Ballarino, N. Gromak, F. Pagano, I. Bozzoni, and N. J. Proudfoot, "Primary microRNA transcripts are processed co-transcriptionally.," *Nat. Struct. Mol. Biol.*, vol. 15, no. 9, pp. 902–9, Sep. 2008.

[126] M. Bibel, J. Richter, K. Schrenk, K. L. Tucker, V. Staiger, M. Korte, M. Goetz, and Y.-A. Barde, "Differentiation of mouse embryonic stem cells into a defined neuronal lineage.," *Nat. Neurosci.*, vol. 7, no. 9, pp. 1003–9, Sep. 2004.

[127] T. Kimura, Y. Kaga, H. Ohta, M. Odamoto, Y. Sekita, K. Li, N. Yamano, K. Fujikawa, A. Isotani, N. Sasaki, M. Toyoda, K. Hayashi, M. Okabe, T. Shinohara, M. Saitou, and T. Nakano, "Induction of primordial germ cell-like cells from mouse embryonic stem cells by ERK signal inhibition.," *Stem Cells*, vol. 32, no. 10, pp. 2668–78, Oct. 2014.

[128] D. R. Leach, "Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair.," *Bioessays*, vol. 16, no. 12, pp. 893–900, Dec. 1994.

[129] D. J. Laurie, W. Wisden, and P. H. Seeburg, "The distribution of thirteen GABAA receptor subunit mRNAs in the rat brain. III. Embryonic and postnatal development.," *J. Neurosci.*, vol. 12, no. 11, pp. 4151–72, Nov. 1992.

[130] W. Wisden, D. J. Laurie, H. Monyer, and P. H. Seeburg, "The distribution of 13 GABAA receptor subunit mRNAs in the rat brain. I. Telencephalon, diencephalon, mesencephalon.," *J. Neurosci.*, vol. 12, no. 3, pp. 1040–62, Mar. 1992.

[131] J. B. Li and G. M. Church, "Deciphering the functions and regulation of brain-enriched A-to-I RNA editing.," *Nat. Neurosci.*, vol. 16, no. 11, pp. 1518–22, Nov. 2013.

[132] J. Ohlson, J. S. Pedersen, D. Haussler, and M. Ohman, "Editing modifies the GABA(A) receptor subunit alpha3.," *RNA*, vol. 13, no. 5, pp. 698–703, May 2007.

[133] E. Y. Rula, A. H. Lagrange, M. M. Jacobs, N. Hu, R. L. Macdonald, and R. B. Emeson, "Developmental modulation of GABA(A) receptor function by RNA editing.," *J. Neurosci.*, vol. 28, no. 24, pp. 6196–201, Jun. 2008.

[134] C. Daniel, H. Wahlstedt, J. Ohlson, P. Björk, and M. Ohman, "Adenosine-to-inosine RNA editing affects trafficking of the gamma-aminobutyric acid type A (GABA(A)) receptor.," *J. Biol. Chem.*, vol. 286, no. 3, pp. 2031–40, Jan. 2011.

[135] J. Krol, I. Loedige, and W. Filipowicz, "The widespread regulation of microRNA biogenesis, function and decay.," *Nat. Rev. Genet.*, vol. 11, no. 9, pp. 597–610, Sep. 2010.

APPENDIX

Appendix A: Abbreviations

Abbr.	Meaning
5-azadC	5'-aza-2'-deoxycytidine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
α-KG	α-ketoglutarate
AGO	Argonaute
AML	Acute myeloid leukemia
AmpR	Ampicillin resistance
BER	Base excision repair
Cas9n	Cas9 nickase
CG	Cancer-germline
crRNA	CRISPR RNA
CT	Cancer-testis
CmR	Chloramphenicol resistance
ds	Double-strand
DSB	Double-strand break
EMT	Epithelial-mesenchymal transition
(m)ESC	(Mouse) embryonic stem cell
HDR	Homology directed repair
HR	Homologous recombination
ICM	Inner cell mass
IDH	Isocitrate dehydrogenase
KO	Knockout
LNA	Locked nucleic acid

MDS	Myelodysplasic syndrome
MMP	Matrix metalloproteinase
NeoR	Neomycin selection marker
NHEJ	Non-homologous end joining
NPC	Neural progenitor cells
NSLC	Non-small cell lung cancer
PAM	Protospacer adjacent motif
PGC	Primordial germ cell
PMD	Partially methylated regions
precrRNA	Precursor crRNA
RISC	RNA-induced silencing complex
sgRNA	Single guide RNA
SAM	S -adenosyl- l -methionine
(Sp)Cas9	(S. pyogenes) Cas9
(Sp)Cas9n	(S. pyogenes) Cas9 nickase
ssODN	Single-stranded oligonucleotide
tracrRNA	Trans-activating crRNA
ss	Single-strand
TDG	Thymine-DNA glycosylase
TIMP	Tissue inhibitors of metalloproteinase
TSS	Transcription start site
WT	Wild-type

Appendix B: Primers and synthetic sequences

Synthetic floxed miR-767

HindIII >LoxP

5'-CGACACTGCTCGATCCGCTCGCACCCCAAC|AAGCTT|ATAACTTCGTATAGCATACATTATACGAAGTTATCTACTCTTAAGCTATAG

TCAAACTATTTGCTTGTAATGTCTGGAAATACACGTTGGTGCAATGCTTCATCATCAGAGAATTGTATGAATATATGTTGTGTTATTGCCTT

TAGCATCATTGCCATGTCTTGGTGGCCTTATGAATGCTACCTTTATTCTCATCTCTTGTTTATCTTGGCATATTCCCGTGTTATTTGTTCCTGT

>miR-767

ATCACATCTACTTTTGTATTAT|AGAATCCATATAGGTTTTTACTCATGCACCATGGTTGTCTGAGCACATAACATGCTTGTCTGCTCATA

CCCTATGGTTCCTGAAGAGGAATCTTAACTGTCTACTGCTCTATAGGAAAACAGTGTTTTATGTATCAGTTTATATTTTTGGTTGCTTTTTC

CTAGTACATGGACCTCTGCTCCCTTAACATACTGGTGTTTTTCCTGTACCATCTCTATCTGTTCCAATTCTATGGATGTTAACTTGCCACTAG

>LoxP

ACTTCCTGTGCATCATAGTCCAGTGCTCAAATAAATCCCATGCACTATGACTGCTTAAGCATGTAACAATAGTTTCTCA|ATAACTTCGTAT

FseI PacI HindIII

AGCATACATTATACGAAGTTAT|GGCCGGCC|TTAATT|AAGCTTCTTACGGATCGACGAGAGCAGCGCGACTGGAT-3'

Synthetic miR-767 RNA spike in

5'- UGCACCAUGGUUGUCUGAGCA-3'

Primers

n°	Name	Sequence (5' to 3')
1	5'arm_Fwd	AGTTCTAGACCGGTCTCAGTTTATGCTGACTGCTCA
2	5'arm_Rev	AGTTCTAGATAGATGGTCAAGTATATTAGGGAAAA
3	PCR_1kb_Fwd	ATAGGCCGGCCTGTGTTTTATCAGTATGATCTTC
4	PCR_1kb_Rev	ATAGGCCGGCCAGGCCTTCTCCTTCCACTGACAGGC
5	Probe_SB	GGTTGTGAGCCTAGCCTTTA
6	seq_BL5	CATTTTGGATAGTATGGGT
7	CRISPRpX334_1A_UP	aaacAGACAAGCATGTTATGTGCTCAGACAACCAgt
8	CRISPRpX334_1A_DOWN	taaaacTGGTTGTCTGAGCACATAACATGCTTGTCT
9	CRISPRpX334_1B_UP	aaacTTGTCTGCTCATACCCTATGGTTCCTGAAGgt
10	CRISPRpX334_1B_DOWN	taaaacCTTCAGGAACCATAGGGTATGAGCAGACAA
11	CRISPRpX334_2A_UP	aaacACAACCATGGTGCATGAGTAAAAACCTATAgt
12	CRISPRpX334_2A_DOWN	taaaacTATAGGTTTTTACTCATGCACCATGGTTGT
13	CRISPRpX334_2B_UP	aaacCACATAACATGCTTGTCTGCTCATACCCTAgt
14	CRISPRpX334_2B_DOWN	taaaacTAGGGTATGAGCAGACAAGCATGGTTATGTG
15	CRISPRpX335_1A_UP	caccGTTATGTGCTCAGACAACCA
16	CRISPRpX335_1A_DOWN	aaacTGGTTGTCTGAGCACATAAC

17	CRISPRpX335_1B_UP	caccgATACCCTATGGTTCCTGAAG
18	CRISPRpX335_1B_DOWN	aaacCTTCAGGAACCATAGGGTATc
19	CRISPRpX335_2A_UP	caccgTGCATGAGTAAAAACCTATA
20	CRISPRpX335_2A_DOWN	aaacTATAGGTTTTTACTCATGCAc
21	CRISPRpX335_2B_UP	caccGCTTGTCTGCTCATACCCTA
22	CRISPRpX335_2B_DOWN	aaacTAGGGTATGAGCAGACAAGC
23	HR_5'A_UP	caccGTAGATGGTCAAGTATATTA
24	HR_5'A_DOWN	aaacTAATATACTTGACCATCTAC
25	HR_5'B_UP	caccgTAATGTCTGGAAATACACGT
26	HR_5'B_DOWN	aaacACGTGTATTTCCAGACATTAc
27	HR_3'A_UP	caccgCTTAAGCAGTCATAGTGCAT
28	HR_3'A_DOWN	aaacATGCACTATGACTGCTTAAGc
29	HR_3'B_UP	caccgATGATCTTCTCCTAGTCTGA
30	HR_3'B_DOWN	aaacTCAGACTAGGAGAAGATCATc
31	RTgabra3Fa	AGAAAGAAAAGAAGCCAGGC
32	RTgabra3Fb	CACCCATAGAGATGGAAGGG
33	RTgabra3Rc	GCTTGGGAGAGAGTCCTCC
34	RTgabra3Fd	CTCACAATATGACCACACCC
35	RTgabra3Re	CAGCTTTGGTATAGGCATAGC
36	RTgabra3Rc2	GTCGTCTTGATTCCCCTTGG
37	RTgabra3Rg	TTCTCTTCAGATGAAAGTGGG
38	RTmagea5F	GAAAGGAGTTCGCCTTGCC
39	RTmagea5R	TCCTCCTCGGATGTCTCCA
40	G3RACE_RT	GTC-ATA-TTG-TGA-GCC
41	G3RACE_S1	AGACAGACATGGCATGATGAA
42	G3RACE_S2	TGAAGATCCTTCCACTGAACAA
43	G3RACE_R1	ATATCTGGGGCATGCTTGGG
44	G3RACE_R2	CCCAGGTTCTTGTCGTCTTG
45	RTrnaseq1n	GGTCCTAGACCCACTTAGAG
46	RNAseq1	AACAGCGCAACCACCCAGA
47	RTrnaseq2	CCATCTGGGGAGTCAATACG
48	RNAseq2n	GTCAATACGCCAAATCTCCG
49	FapaI_premir767	TTTGGGCCCATATTCCCGTGTTATTTGTTCC
50	RecoRV_premir767	TTTGATATCGGAGCAGAGGTCCATGTACTAGGA
51	FApaI_premiR105	TTTGGGCCCTCGCCTGTAACATGGCATTAAC
52	REcoRV_premiR105	TTTGATATCGAGTAATGAATGGCTTTGTTCC
53	mGapdh S	CGTGCCGCCTGGAGAA
54	mGapdh AS	GATGCCTGCTTCACCACCTT
55	ActB_S	GGCTGTATTCCCCTCCATCG
56	ActB_AS	CCAGTTGGTAACAATGCCATGT