

Criblage par ARN interférence pour l'identification de nouveaux gènes impliqués dans la différenciation myogénique

Rayan Alwan

▶ **To cite this version:**

Rayan Alwan. Criblage par ARN interférence pour l'identification de nouveaux gènes impliqués dans la différenciation myogénique. Médecine humaine et pathologie. Université de Limoges, 2017. Français. NNT : 2017LIMO0015 . tel-01566505

HAL Id: tel-01566505
https://theses.hal.science/tel-01566505

Submitted on 21 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de Doctorat

Thèse
pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE LIMOGES

Spécialité : *Génétique et Génomique Moléculaire*

Présentée et soutenue par

Rayan Alwan

Le 23 Juin 2017

Criblage par ARN interférence pour l'identification de nouveaux gènes impliqués dans la différenciation myogénique

Thèse co-dirigée par
le Dr. *Khaled BOUHOUCH et*
le Pr. *Véronique BLANQUET*

JURY

Président du jury

Fabrice LALLOUE, Professeur des Universités, Limoges

Rapporteurs

Fabienne LE PROVOST, Directrice de Recherche INRA, Jouy en Josas
Krzysztof JAGLA, Directeur de Recherche INSERM, Clermont-Ferrand

Examinateurs

Christian COGNARD, Directeur de Recherche CNRS, Poitiers
Laetitia MAGNOL, Maître de conférences, Limoges
Khaled BOUHOUCH, Maître de conférences, Limoges
Véronique BLANQUET, Professeur des Universités, Limoges

Faculté des Sciences et Techniques

Université de Limoges

Ecole Doctorale Bio-Santé 524

Université de Limoges

Thèse de Doctorat

Thèse
pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE LIMOGES

Spécialité : *Génétique et Génomique Moléculaire*

Présentée et soutenue par

Rayan Alwan

Le 23 Juin 2017

Criblage par ARN interférence pour l'identification de nouveaux gènes impliqués dans la différenciation myogénique

Thèse co-dirigée par
le Dr. *Khaled BOUHOUCH et*
le Pr. *Véronique BLANQUET*

JURY

Président du jury

Fabrice LALLOUE, Professeur des Universités, Limoges

Rapporteurs

Fabienne LE PROVOST, Directrice de Recherche INRA, Jouy en Josas
Krzysztof JAGLA, Directeur de Recherche INSERM, Clermont-Ferrand

Examinateurs

Christian COGNARD, Directeur de Recherche CNRS, Poitiers
Laetitia MAGNOL, Maître de conférences, Limoges
Khaled BOUHOUCH, Maître de conférences, Limoges
Véronique BLANQUET, Professeur des Universités, Limoges

Faculté des Sciences et Techniques

Université de Limoges

Ecole Doctorale Bio-Santé 524

Université de Limoges

DEDICACES

*A **mes parents**,*

*A **mes frères et ma sœur***

REMERCIEMENTS

A l'issue de la rédaction de ma thèse, je suis convaincue que la thèse est loin d'être un travail solitaire. En effet, je n'aurais jamais pu réaliser ce travail doctoral sans le soutien d'un grand nombre de personnes dont la générosité, la bonne humeur et l'intérêt manifestés à l'égard de ma recherche m'ont permis de progresser dans cette phase délicate de « l'apprenti-chercheur».

Par ces quelques lignes, je tiens à remercier toutes les personnes qui ont participé de près ou de loin au bon déroulement de cette thèse, en espérant n'avoir oublié personne…

En premier lieu, mes sincères remerciements et ma gratitude vont aux membres du jury, ***Fabrice LALLOUE*** *pour avoir accepté de juger ce travail et d'en présider le jury de soutenance,* ***Fabienne LE PROVOST*** *et* ***Krzysztof JAGLA*** *d'avoir accepté de rapporter cette thèse. Que vous soyez assuré de mon entière reconnaissance et de ma profonde gratitude.*

Merci également à ***Christian COGNARD*** *et* ***Laetitia MAGNOL*** *pour l'intérêt que vous avez porté à mon travail en tant qu'examinateur.*

Egalement, je tiens à remercier mon directeur de thèse, ***Véronique BLANQUET****, pour la confiance qu'il m'a accordée en acceptant d'encadrer ce travail doctoral, pour ses multiples conseils et pour toutes les heures qu'elle a consacrées à diriger cette recherche.*

Je tiens à exprimer ma profonde gratitude à ***Khaled BOUHOUCH****, pour m'avoir encadré, formé, supporté et pour sa patience et son encouragement à finir ce travail tout au long de ces 4 années au laboratoire.*

Il m'est impossible d'oublier ***Anne BLONDEAU DA SILVA*** *pour son aide précieuse pour toutes les analyses statistiques nécessaires. Elle a toujours fait tout son possible pour m'aider.*

Mention spéciale à ***Bilal*** *l'ami de mon cœur qui m'a supporté et m'a permis de me lever motivée, le cœur léger et l'esprit tranquille depuis le début de ma thèse. Très humblement, je voudrais te dire merci pour ton soutien honnête pendant mes périodes de doutes, pour tes multiples encouragements répétés et pour toutes les heures que t'as consacrées pour nos discussions précieuses, nos moments inoubliables ainsi que pour la relecture de cette thèse dans la dernière phase de cette aventure.*

Un très grand merci tout particulier à ***Sébastien LEGARDINIER****, pour nos innombrables et inoubliables conversations, toujours plus intéressantes, avec mes meilleurs souhaits pour la suite de sa carrière.*

Ces remerciements seraient incomplets si je n'en adressais pas à l'ensemble des membres du laboratoire. Les doctorants et les doctorantes dont nous avons partagé des moments inoubliables et même parfois le même bureau. Mélanie, Caroline, Jessica et Alexis, bonne continuation et une pensée bien sûr à Carole qui vient de finir sa thèse. Je vous souhaite tout un très bel avenir.

Et là j'ai failli terminer, on finit toujours par le meilleur... merci à mes parents, à mes frères et à ma sœur. Je n'aurais rien fait de tout cela sans votre amour. Votre présence et vos encouragements sont pour moi les piliers fondateurs de ce que je suis et de ce que je fais. Heureusement que vous êtes là pour me changer les idées. Ils ont tous cru en moi et ouf ! Maintenant j'y suis ! Est-ce un bon endroit pour dire ce genre de choses ? Je n'en connais en tous cas pas de mauvais. Je vous aime ♥.

SOMMAIRE

LISTE DES ABREVIATIONS **3**

LISTE DES FIGURES ET DES TABLEAUX **5**

PUBLICATIONS ET COMMUNICATIONS **11**

INTRODUCTION GENERALE **13**

EXPOSE BIBLIOGRAPHIQUE **19**

Chapitre I : Les grandes stratégies de mutagenèse chez la souris **21**

I. La mutagenèse chimique à l'éthyl-nitroso-urée 21

I.1. Mécanisme d'action 21

I.2. Mise en place d'un crible ENU *in vivo* chez la souris 23

- Crible dominant versus crible récessif 23
- Les cribles phénotypiques 27
- Identification du gène et de la mutation causale 29

I.3. Les cribles ENU à grande échelle et les ressources génétiques 32

I.4. Avantages et limites de la mutagenèse ENU chez la souris 33

II. La mutagenèse insertionnelle chez la souris 35

II.1. Le piégeage de gènes ou «*Gene trap*» 35

II.2. Le piégeage de gènes à grande échelle et les ressources génétiques 41

II.3. Le ciblage de gènes «*Gene targeting*» 45

II.4. Le ciblage de gènes à grande échelle et les ressources génétiques 49

II.5. Avantages et limites du piégeage et du ciblage de gènes 51

Chapitre II : L'ARN interférence **55**

I. Historique 55

II. Les différentes voies de l'ARN interférence 57

II.1. La voie des siRNA 57

II.2. La voie des microARNs 59

II.3. La voie des piRNA 61

III. L'ARNi un outil de la génétique inverse 61
III.1. Design des siRNAs 63
III.2. Les différents modes de délivrance des siRNA 65
III.3. Les contrôles et la quantification de l'effet des siRNAs 71
IV. Les cribles ARNi à grande échelle 71
IV.1 Le nématode C. elegans 73
IV.2. La Drosophile 75
IV.3. Les Mammifères 79
PROJET DE THESE 87
MATERIELS ET METHODES 93
A. Culture cellulaire 94
B. Transfection 97
C. Immunomarquage et coloration des noyaux 99
D. Microscopie à épi-fluorescence 101
E. Analyse et traitement des images 101
1- Quantification de l'aire de la fluorescence 101
2- Indice de fusion 103
3- Détermination de la longueur et de la largeur des Myotubes 103
F. Test statistique 105
G. Vérification de l'extinction par RT-PCR quantitative 105
1- Extraction, dosage et rétrotranscription des ARN totaux 105
2- PCR quantitative (PCRq) 106
H. Gene Ontology 107
RESULTATS 109
I. Stratégie adoptée 111
1- Quel est le modèle cellulaire adéquat ? 111
2- Quels sont les gènes à inactiver ? 113
3- Quel est l'outil d'inactivation à utiliser? 115
4- Quel sont les phénotypes à suivre et comment ? 119
5- Comment analyser les phénotypes pour classer les gènes ? 123

II. Conséquences phénotypiques de l'inactivation des 100 gènes candidats. 124

III. Classification des 100 gènes candidats. 127

IV. Le crible révèle 13 gènes impliqués dans la différenciation et six gènes impliqués dans la prolifération et la différenciation 129

DISCUSSION GENERALE ET PERSPECTIVES 133

PUBLICATION……...149

REFERENCES BIBLIOGRAPHIQUES……..151

LISTE DES ABREVIATIONS

AMMRA: Asian Mouse Mutagenesis and Resource Association
ARNdb: ARN double brin
ARNi : ARN interférence
Best1: Bestrophin 1
CARD: Center for Animal Resources and Development
Cre: Cyclization recombinase
DAPI: 4',6-diamidino-2-phenylindole
DGCR8: DiGeorge Syndrome Critical Region Gene 8
EMMA: Infrafrontier/European Mutant Mouse Archive
EMPReSS: European Mouse Phenotyping Resource for Standardized Screens
ENU: N-ethyl-N-nitrosourea
ERT-2: Estrogen receptor which only binds tamoxifen
ES: Cellules souches embryonnaires
esiRNAs: Endoribonuclease-prepared siRNA
EUCOMM: EUropean Conditional Mouse Mutagenesis Program
EUMODIC: European Mouse Disease Clinic
FLP: Flippase
FRT: Flippase recognition target
GAL4: Galactose-responsive transcription factor GAL4
GAPDH: Glyséraldehyde-3-phosphatase déshydrogénase
GCK: Glucokinase
GFP: Green fluorescent protein
IGTC: International Gene Trap Consortium
IKMC: International Knockout Mouse Consortium
IMPC: the International Mouse Phenotyping Consortium
IRES: Internal ribosome entry site
KOMP: KnockOut Mouse Project
LoxP: Locus of X-over P1
LTR: Long terminal repeat sequence
MGD: Mouse Genome Database
MGI: Mouse Genome Informatics
miRNA: micro RNA
MODY2: Maturity-Onset Diabetes of the Young, Type 2
MRC: Medical Research Council
MRF: Myogenic regulatory factor
MRF4: Myogenic factor 6 (Myf6)
Myf5: Myogenic factor 5
MyoD: Myoblast differentiation protein
Myog: Myogenin
NGS: Next-Generation Sequencing
NIH: National Institutes of Health
NMD: Nonsense-mediated mRNA decay
NorCOMM: North American Conditional Mouse Mutagenesis Project
OMIM: Online Inheritance in Man
ORF: Open reading frame
Pax: Paired box protein
P-body: Processing bodies
piRNA: PIWI interacting RNA
PIWI: P-element Induced WImpy testis
PosMed: Positional Medline
pre-miRNA: precursor-miRNA
pri-miRNA: Primary miRNA
PTGS: Post Transcriptional Gene Silencing
QTL: Quantitative trait locus
RACE: Rapid Amplification of cDNA Ends
RG: Gène rapporteur (reporter gene)
RISC: RNA-induced silencing complex
RNase III: Ribonuclease III
Rom1: Rod Outer Segment Membrane Protein 1
S2: Schneider 2 Drosophila cells
shRNA: Small hairpin RNA
siRNA: Small interfering RNA
Six: Homeobox Protein (Sine Oculis Homeobox Homolog)
SNP: Single-nucleotide polymorphism
SSR: Site-specific recombination
Stx3: Syntaxin 3
TIGM: Texas Institute of Genomic Medicine
UAS: Upstream activating sequence
β2-M: β2- microglobuline
β-gal: β-galactosidase

LISTE DES FIGURES

Figure 1 : Crible de mutagenèse à l'ENU : les différentes étapes.
Figure 2 : Stratégie de croisement pour le crible dominant et récessif.
Figure 3 : Les différents projets internationaux basés sur de la mutagenèse à grande échelle chez la souris.
Figure 4 : Enhancer trap.
Figure 5 : Promoter-trap.
Figure 6 : Poly A trap.
Figure 7 : Les différentes variantes de la technologie « gene targeting ».
Figure 8 : Vecteur de ciblage.
Figure 9 : Principe du système cre/loxP.
Figure 10 : Le système cre/loxP peut aboutir à trois résultats différents selon l'emplacement et l'orientation des sites.
Figure 11 : La variante Knock-in.
Figure 12 : L'ARN interférence.
Figure 13 : Voie de maturation canonique des miRNAs.
Figure 14 : Les conditions de séquence prises en compte pour le design des siRNAs.
Figure 15 : Les trois modes de délivrance de l'ARNdb dans le nématode *C. elegans.*
Figure 16 : Les différents modes de délivrance des siRNAs en utilisant les nanoparticules.
Figure 17 : Le principe de l'électroporation des cellules.
Figure 18 : La transduction des shRNAs via des particules virales.
Figure 19 : Les différentes voies d'injection chez la souris.
Figure 20 : Expression inductible de l'ARNdb par le système UAS-GAL4.
Figure 21 : Hiérarchie des facteurs impliqués dans la myogenèse du muscle squelettique.
Figure 22 : Démarche expérimentale mise en oeuvre.
Figure 23 : Calcul de la taille des myotubes.
Figure 24 : Schéma descriptif différents filtres utilisés pour établir la liste finale des 100 gènes.
Figure 25 : L'analyse des gènes candidats par « *Gene Ontology* » (GO).
Figure 26 : L'efficacité de la délivrance des siRNAs dans les cellules C2C12.
Figure 27 : La quantité relative des transcrits G36, G48 et G18 après inactivation.
Figure 28 : Sites de fixation des siRNAs choisis sur les ARNm des gènes G63 et G32.
Figure 29 : Schéma montrant la reconstruction de la grande image.
Figure 30 : Variation de la densité cellulaire, de la quantité de myotube et de l'indice de fusion dans le contrôle.
Figure 31 : (A) **:** Densité cellulaire, quantité de myotube et indice de fusion des cellules inactivées dans le gène G97. (B) : Immunomarquage des cellules contrôle et des cellules inactivées pour le gène G97.

Figure 32 : Analyse de la quantité de myotubes**.**
Figure 33 : Analyse de la densité cellulaire.
Figure 34 : Analyse de la morphologie des myotubes par immunofluorescence.
Figure 35 : Analyse de l'indice de fusion.

℘ LISTE DES TABLEAUX

Tableau 1 : Les différents projets internationaux à grande échelle réalisés chez la souris.
Tableau 2 : Liste des sondes Taqman utilisées pour quantifier par PCRq les transcrits des gènes analysés.
Tableau 3 : Liste de 100 gènes présélectionnés pour le crible.
Tableau 4 : Les étapes de la classification des gènes selon les critères phénotypiques considérés.

PUBLICATION ET COMMUNICATIONS

PUBLICATION

A siRNA-based screen in C2C12 myoblasts identifies novel genes involved in myogenic differentiation.

Rayan Alwan[a], Angelique Bruel[b], Anne Da Silva[a], Véronique Blanquet[a]*, Khaled Bouhouche[a]*.

Soumis à Experimental Cell Research

COMMUNICATIONS

Communications Orales

Identification of new genes involved in muscle development: using mouse as a model organism.

Rayan ALWAN, Véronique BLANQUET et Khaled BOUHOUCH

17ème séminaire des thésards du Département de Génétique Animale, Jouy-en-Josas, 23-24 Avril 2014.

An efficient RNA interference screening, using C2C12 line, identify new genes involved in myogenic differentiation.

Rayan ALWAN, Véronique BLANQUET et Khaled BOUHOUCH

18ème séminaire des thésards du Département de Génétique Animale, La Rochelle, 21-22 Mai 2015.

Congrès

An efficient RNA interference screening, using C2C12 line, identify new genes involved in myogenic differentiation.

Rayan ALWAN, Véronique BLANQUET and Khaled BOUHOUCH

5ème congrès international de myologie, Palais des congrès de Lyon, 14-18 Mars 2016

.

INTRODUCTION GENERALE

La biologie s'est clairement engagée ces deux dernières décennies dans l'analyse des processus biologiques à grande échelle et à haut débit. Un paradigme nouveau s'est développé du fait des approches utilisées. Ainsi, au lieu de séquencer le locus d'un seul gène, on décrypte le génome entier ; au lieu de regarder l'expression d'un seul gène, spécifique durant le développement ou durant un stade physiologique, des stratégies ont été développées pour déterminer l'expression des dizaines de milliers de gènes en même temps. Un outil indispensable dans la compréhension du fonctionnement complexe de nos organismes est l'utilisation de modèles animaux. L'étude des phénotypes et des schémas de croisement d'expression des gènes dans des organismes modèles sont essentiels dans l'analyse du rôle des gènes dans les maladies humaines. La souris est utilisée depuis plus d'une centaine d'années pour l'étude des bases génétiques des procédés biologiques et des maladies chez l'homme. C'est un excellent modèle pour définir la fonction des gènes des mammifères, elle présente en effet de grandes similitudes génétiques, immunologiques, reproductives, physiologiques et pathologiques avec l'homme.

L'étude fonctionnelle des gènes est abordée par la mise en œuvre de deux stratégies expérimentales efficaces et complémentaires permettant de perturber la fonction du gène : (i) la génétique classique qui correspond à l'observation et la description d'un phénotype particulier engendré par une mutation qui va servir de base à l'identification fonctionnelle du gène et de son produit, (ii) la génétique inverse qui part d'un gène défini et va vers la mutagenèse ciblée et l'analyse du phénotype consécutif à la modification dirigée introduite dans ce gène ; elle recouvre un large éventail de méthodes d'inactivation génique en utilisant surtout les connaissances issues du séquençage des génomes. Les deux stratégies reposent sur des techniques de mutagénèse de plus en plus efficaces, accompagnées de méthodes complémentaires pour détecter les phénotypes et analyser les fonctions des gènes. Ainsi, dans le but d'identifier et d'analyser la fonction des gènes à l'échelle du génome entier chez la souris, deux approches complémentaires dites «*phenotypic driven*» et «*gene driven*» ont été mises en place et appliquées. Parmi les mutagenèses réalisées selon la première approche, celle utilisant l'agent mutagène *N*-ethyl-N-nitrosourea (ENU) a été particulièrement utilisée avec succès pour générer un grand nombre de nouveaux mutants

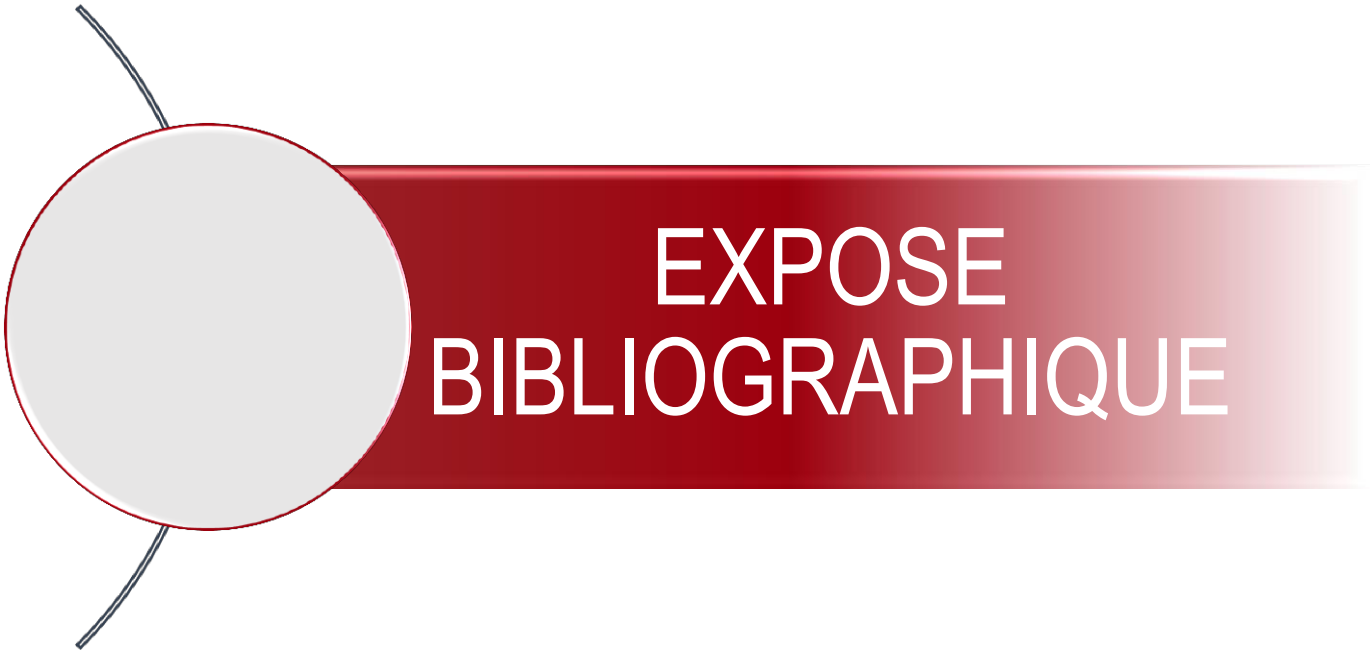
pour des études systématiques de gènes murins et/ou pour l'identification et la caractérisation fonctionnelle de certains gènes liés à des maladies humaines. L'approche «*gene driven*» quant à elle consiste en l'introduction des mutations insertionnelles dirigées ou aléatoires dans le génome de la souris. Pour muter tous les gènes codants, plusieurs projets à grande envergure dits « *knockout programs* » ont été développés dans les cellules souches embryonnaires de souris (ES) en utilisant l'une des deux techniques « *gene trap* » (piégeage de gène) ou « *gene targeting* » (ciblage de gènes), pour générer des souris hébergeant les modifications génétiques correspondantes.

Ces études fonctionnelles à haut débit ont aussi pris de l'élan avec la découverte de l'ARN interférence (ARNi), ce qui a permis de mettre en place des cribles fonctionnels à l'échelle génomique. L'ARNi est un mécanisme de régulation génique utilisant des petits ARNs non codants reconnaissant des séquences cibles sur l'ARNm induisant sa destruction ou la répression de sa traduction. En quelques années, l'ARNi est devenu un outil d'inactivation génique très performant pour éteindre d'une façon systématique et ciblée n'importe quel gène avec rapidité, efficacité et faible coût.

Au sein du laboratoire de Génétique Moléculaire Animale, l'équipe dans laquelle j'ai réalisé mes travaux de thèse s'intéresse au déterminisme génétique du développement musculaire et porte une attention particulière aux gènes susceptibles d'intervenir dans le contrôle de la prolifération et la différenciation des cellules musculaires, les caractéristiques du muscle et aux gènes qui pourraient contrôler la qualité du produit « muscle » (viande). Pour atteindre ses objectifs, l'équipe combine des stratégies de criblage de gènes/QTL, soit chez la souris soit chez le bovin, avec des analyses fonctionnelles de gènes d'intérêt. Ainsi, après avoir développé chez la souris un crible de mutagenèse chimique (éthyl-nitroso-urée) sensibilisé pour identifier des mutations conduisant à des anomalies du système musculaire, il a été choisi de mettre en place un crible fonctionnel en utilisant l'ARNi comme outil d'inactivation génique dans la lignée myoblastique C2C12. Le but est l'identification de nouveaux gènes impliqués dans la différenciation myogénique. L'ARNi a été utilisé avec succès dans des cribles à grande échelle chez des eucaryotes inférieurs tels que la drosophile et le nématode *C. elegans* pour isoler de nouveaux gènes muscle spécifiques. Ce genre de cribles à haut débit semble être difficile à réaliser chez les mammifères et en particulier dans un tissu hétérogène et complexe comme le muscle squelettique. C'est pourquoi le laboratoire a

choisi de réaliser un crible fonctionnel grâce au knock-down de gènes par transfection de *Small interfering RNA* (siRNA) dans des cellules C2C12. Cette lignée est largement utilisée pour étudier les mécanismes moléculaires fondamentaux des processus biologiques liés au développement musculaire.

Les résultats de ce crible qui font l'objet d'un article soumis à *Experimental Cell Research* seront présentés dans la deuxième partie de ce manuscrit après avoir décrit dans une première partie les différentes approches/stratégies utilisées pour identifier et analyser de nouveaux gènes, la technologie ARNi et son utilisation dans les cribles à haut débit ainsi que le développement musculaire. Enfin, la troisième partie de mon manuscrit sera consacrée à une discussion/perspectives sur les résultats obtenus ainsi que la pertinence de la méthode utilisée.


EXPOSE
BIBLIOGRAPHIQUE

Chapitre I : Les grandes stratégies de mutagenèse chez la souris

Trois méthodes ont été largement utilisées pour introduire des mutations à grande échelle dans la lignée germinale de la souris, à savoir la mutagenèse (i) physique **(Rinchik, 1991)**, (ii) chimique **(Russell et al., 1979; Munroe et al., 2000)** et, (iii) insertionnelle **(Schnieke et al., 1983; Wagner et al., 1983; Jenkins and Copeland, 1985; Robertson et al., 1986; Doetschman et al., 1987; Thomas and Capecchi, 1987; Gossler et al., 1989; Spence et al., 1989)**. Seules, la mutagenèse chimique médiée par l'éthyl-nitroso-urée (ENU) et celle basée sur les mutations insertionnelles par « *gene trap* » et « *gene targeting* » seront décrites. Je traiterai en plus, dans cette partie du manuscrit, l'utilisation de l'ARNi comme outil d'inactivation génique en génomique fonctionnelle.

I. La mutagenèse chimique à l'éthyl-nitroso-urée

I.1. Mécanisme d'action

L'ENU est un agent mutagène synthétique qui provoque principalement des mutations ponctuelles aléatoires de l'ordre d'une mutation toutes les 700-1000 paires de bases. Il agit par alkylation des acides nucléiques et aucun procédé métabolique n'est nécessaire pour son activation. L'ENU peut transférer son groupement éthyle sur les bases azotées de l'ADN, ce qui engendre un mésappariement et la substitution d'une paire de bases s'il n'y a pas de réparation de l'ADN **(Justice et al., 1999)**. Ces mécanismes de réparation jouent un rôle très important dans la mutagenèse, et il y a une relation directe entre la dose d'ENU et l'efficacité de ces mécanismes. Cet agent chimique possède le taux le plus élevé de mutation dans les cellules souches spermatogoniales pré-méiotiques, avec une fréquence de 1 :700/locus/gamète *in vivo* chez la souris, pour une dose et une souche murine donnée **(Favor, 1998)**. Une dose efficace et tolérée pour la plupart des lignées murines correspond à l'injection intrapéritonéale de 270mg ENU/kg administrés en trois injections sur 3 semaines.

Les mutations ponctuelles introduites par l'ENU sont majoritairement des transversions de type A-T vers T-A et des transitions de type A-T vers G-C. Les changements

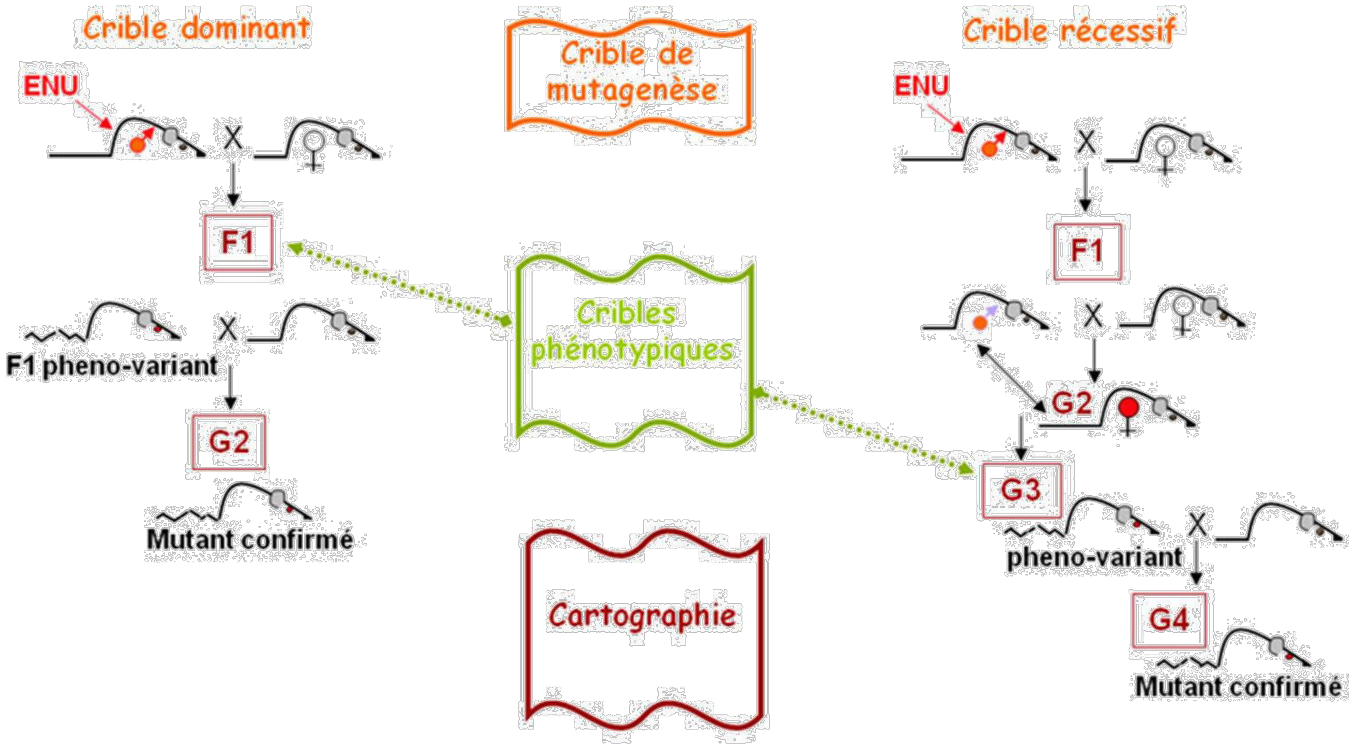


Figure 1 : Crible de mutagenèse à l'ENU : les différentes étapes.
L'ENU est injecté par voie intrapéritonéale à des souris mâles. Ces mâles mutagenéisés sont croisés à des souris femelles sauvages pour produire des descendants porteurs de mutations potentielles dominantes. Les pédigrées récessifs peuvent être produits en croisant les mâles fondateurs (père) (G1) avec les souris femelles de deuxième génération (Backcross) ou par croisement des femelles G2 avec des mâles G2 (intercross) «croisement frère-sœur » **(Hardisty-Hughes et al., 2010)**. Dans tous les cas, les descendants sont ensuite soigneusement phénotypés au travers de différents cribles phénotypiques comprenant chacun plusieurs paramètres d'analyse. Les valeurs aberrantes sont identifiées de façon qualitative ou en utilisant des méthodes statistiques et, une fois la transmission et la pénétrance d'un phénotype spécifique confirmées, la localisation et l'identification du gène candidat est entreprise par cartographie en se basant sur les SNP ou en utilisant la technique NGS. http://www.umr1061.unilim.fr/spip.php?article98.

introduits par l'ENU conduisent dans 64% des cas à des mutations faux-sens, 26% à des erreurs d'épissage, et 10% à des mutations non-sens **(Justice et al., 1999)**.

Contrairement à la mutagenèse dite insertionnelle, l'ENU peut générer en plus des mutants nuls (allèle amorphe), tous les autres types d'allèles : la perte partielle de fonction (allèle hypomorphe), le gain de fonction (allèle hypermorphe), un dominant négatif (allèle antimorphe) et le gain de nouvelle fonction (allèle néomorphe) **(Godinho and Nolan, 2005)**. Plusieurs formes alléliques générées pour le même gène, appelées « série allélique » **(Quwailid et al., 2004)**, fournissent beaucoup plus d'informations sur la fonction du gène qu'une simple délétion du gène (Knockout) **(Davis and Justice, 1998; Horner and Caspary, 2011)**.

I.2. Mise en place d'un crible ENU *in vivo* chez la souris

Un crible ENU comprend trois étapes principales: la mutagenèse proprement dite, les cribles phénotypiques et l'identification de la mutation causale par cartographie. Les trois étapes sont illustrées dans la **Figure1**.

- *Crible dominant versus crible récessif*

On distingue deux types de cribles ENU fonctionnels à l'échelle génomique : dominant et récessif pour lesquels les schémas de croisement sont bien établis. Ainsi, une fois que les mâles retrouvent leur fertilité suite au traitement à l'ENU **(Figure 2)**, ils sont croisés à des femelles sauvages pour produire la descendance de la première génération (G1).

- **Les cribles dominants**

Pour les cribles dominants, qui sont relativement rapides, les souris G1 sont soumises à la batterie de tests phénotypiques et les phénovariants hétérozygotes pour les mutations induites par ENU, sont considérées comme mutants confirmés si 50% de leurs descendances (G2) issues de croisement avec la souche sauvage montrent le phénotype d'intérêt **(Figure 2a)**. Un certain nombre de phénotypes et de gènes ont été identifiés par ce type de crible. Citons l'exemple du gène codant pour l'enzyme glucokinase (GCK) qui a été identifié dans le cadre de trois projets ENU : *The Munich ENU Project* **(van Bürck et al., 2010)**, *The MRC Harwell Mutagenesis Project* **(Toye et al., 2004)** et *The RIKEN ENU Mutagenesis Project*

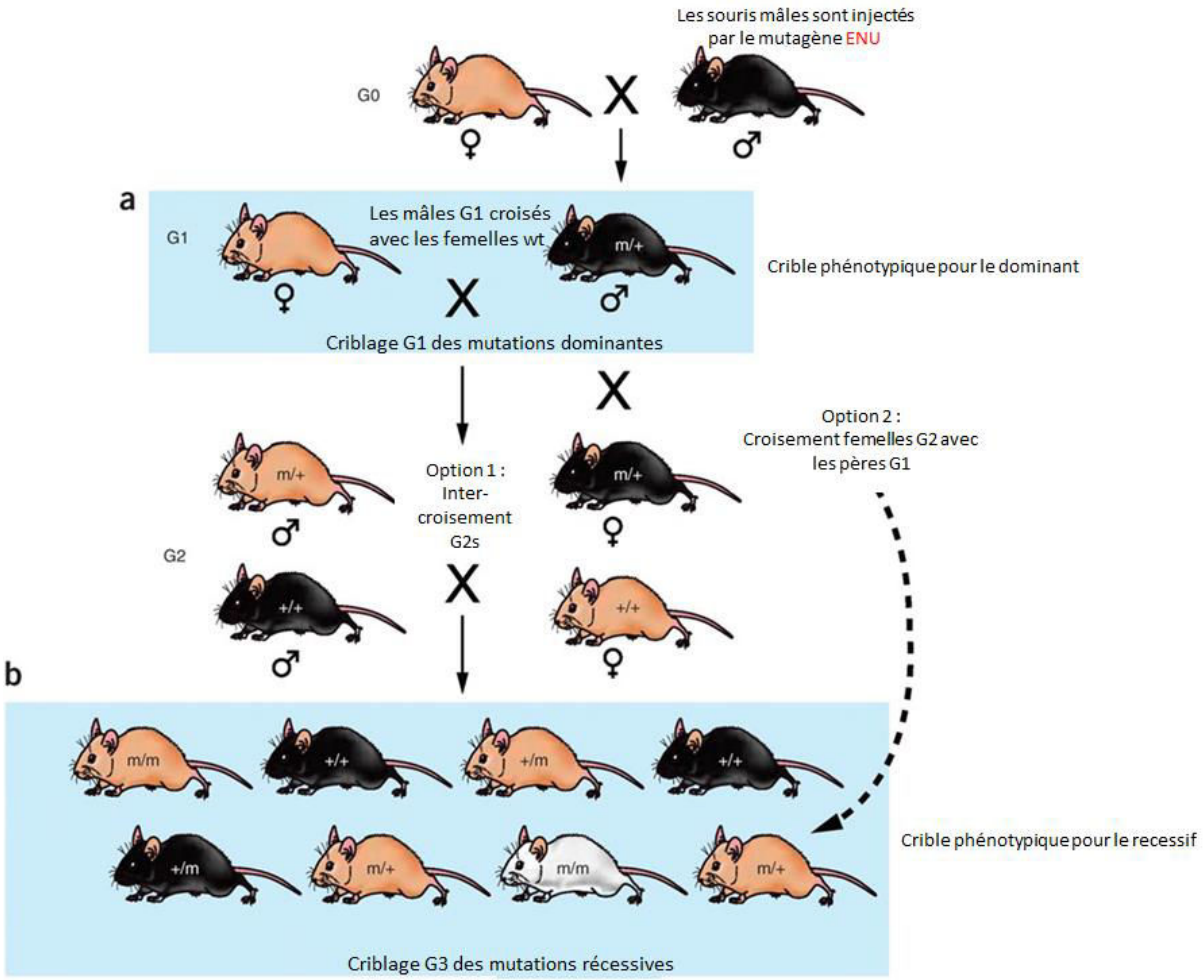


Figure 2 : Stratégie de croisement pour le crible dominant et récessif.

(a) Dans un crible dominant, une souris mâle est mutée avec l'ENU et est croisée à une souris femelle sauvage (la génération G0). Pour chaque souris, les génotypes sont indiqués. Une mutation hétérozygote est présente dans la lignée germinale chez le mâle. Si cette mutation conduit à un phénotype dominant, elle peut être détectée, suite à un crible phénotypique spécifique dans la génération G1. Un crible récessif nécessite plus de croisements. Les souris G1 seront hétérozygotes pour les mutations induites par l'ENU et ne montreront pas de phéno-variation. Ainsi, des souris G1 seront croisées avec des souris sauvages pour produire des souris G2. La moitié des souris G2 aura hérité de la mutation présente dans la souris G1. Les souris G3 (b) peuvent être produites de plusieurs façons, soit par backcross G1xG2, soit par intercross G2xG2. Les souris G3 résultantes seront phénotypes **(Hardisty-Hughes et al., 2010)**.

(Inoue et al., 2004). L'activité GCK est nécessaire pour la glycolyse, les souris homozygotes pour quelques mutations identifiées meurent quelques jours après la naissance, alors que les souris hétérozygotes ont montré une réduction de 50% de l'activité de l'enzyme. Ces hétérozygotes présentent une hyperglycémie sévère et finissent par développer un diabète similaire à une forme particulière de diabète chez l'homme : le diabète de type MODY2 (*Maturity-Onset Diabetes of the Young* ou diabète de type adulte chez le jeune). Les différentes mutations ponctuelles identifiées par les trois cribles fournissent une série allélique pour le gène GCK (des allèles nuls suite à des mutations non-sens et d'autres hypomorphes dont les mutations ont perturbé des sites d'épissage), ceci va aider à mieux comprendre le rôle de cette enzyme dans la maladie humaine MODY2.

- **Les cribles récessifs**

Bien que le criblage des mutations dominantes soit relativement rapide, les données publiées à partir de ces cribles à grande échelle révèlent que uniquement 2% des mutations fonctionnelles sont dominantes **(de Angelis et al., 2000; Nolan et al., 2000)**. De plus, il faut noter que la majorité des maladies génétiques humaines sont dues à des mutations récessives. Le schéma de croisement d'un crible récessif est détaillé sur la **Figure 2b** **(Cordes, 2005; Kile and Hilton, 2005)**. Les cribles récessifs ont pu identifier de nombreux phénotypes liés à des gènes impliqués dans plusieurs processus cellulaires comme par exemple les voies de signalisation **(Herron et al., 2002; Fernandez et al., 2009)**. Des mutations récessives liées à des maladies génétiques humaines ont pu être aussi récupérées dans les grands projets ENU. A titre d'exemple, le projet de mutagenèse *RIKEN* a généré grâce à un crible récessif un modèle murin de dystrophie neuroaxonale infantile **(Wada et al., 2009)**. La mutation a été identifiée, c'est une transition G vers A dans le gène codant pour l'enzyme phospholipase A2 (iPLA2ß). La mutation supprimant l'activité enzymatique conduit à une perturbation de l'homéostasie de la membrane cellulaire, ce qui entraîne une dégénérescence axonale. De même, le projet de mutagenèse ENU de « *the Genomics Institute of the Novartis Research Foundation* » a pu identifier une mutation dans un gène qui n'était pas préalablement considéré comme étant associé à un diabète **(Lloyd et al., 2010)**. C'est une mutation ponctuelle touchant le gène codant pour la protéine Sec61α1. Les souris mutantes présentent un diabète et une stéatose hépatique provoquée par une augmentation de l'apoptose des cellules pancréatiques. Les auteurs suggèrent que l'apoptose serait probablement due à une perturbation de la fonction

du réticulum endoplasmique (RE) et la maturation des protéines qui entraînerait un stress dans le RE conduisant à l'apoptose **(Lloyd et al., 2010)**.

- **Les cribles sensibilisés**

Malgré le fait que les cribles à grande échelle (récessif ou dominant) sont capables de récupérer un grand nombre de mutations fonctionnelles, ils sont très laborieux et prennent beaucoup de temps, nécessitent de grandes structures d'élevage d'animaux et un grand nombre de personnel qualifié. Pour réduire les coûts et accroître la faisabilité, deux alternatives ont été proposées :

(i) développer des projets de mutagenèse ENU à petite échelle focalisés sur le criblage des phénotypes associés à des processus physiologiques plus spécifiques comme : les phénotypes musculaires **(Miller et al., 2010; Magnol et al., 2011)**, la chute des cheveux **(Kim et al., 2010)**, l'audition et la fonction vestibulaire **(Hardisty-Hughes et al., 2010; El Hakam Kamareddin et al., 2015)**, les maladies cardiovasculaires **(Fernandez et al., 2009)**, le développement de lymphocytes T **(Kakugawa et al., 2009)** et l'anémie **(Lambe et al., 2009)**.

(ii) réaliser des cribles dits « sensibilisés » qui utilisent un fond génétique modifié ou des conditions physiologiques ou environnementales altérées, dans le but de récupérer des mutations quasi irrécupérables dans le contexte sauvage **(Cordes, 2005; Kile and Hilton, 2005)**. Ce type de mutations indétectables concerne en particulier les gènes à allèles haplosuffisants ou les gènes paralogues à effet redondant. Le crible se réalise dans un contexte génétique sensibilisé pour pouvoir accentuer les phénotypes d'intérêt recherchés **(Horner and Caspary, 2011; Buchovecky et al., 2013)**. Cette catégorie de cribles a pour objectif l'identification de nouveaux acteurs impliqués dans le même processus ou la même voie de signalisation **(Curtis et al., 2004; Blewitt et al., 2005)**, la mise en évidence de nouvelles interactions génétiques **(Kauppi et al., 2008)** ou encore la caractérisation, par génomique comparative, des orthologues des gènes identifiés dans une autre espèce **(Magnol et al., 2011)**.

- *Les cribles phénotypiques*

Le succès d'un crible ENU repose sur la conception des tests phénotypiques spécifiques, fiables et discriminants. Pour identifier des mutations fonctionnelles introduites par l'ENU dans un gène donné, il faut cribler approximativement 1000 génomes (entre 300 et 5000 voire plus) **(Nolan et al., 2002)**. Ces cribles phénotypiques correspondent à une batterie

très large de tests qui sont utilisés afin d'analyser un maximum de paramètres impliqués dans l'expression d'une mutation. Ces cribles sont des cribles primaires qui doivent être faciles à mettre en œuvre, non invasifs et rapides. Les analyses menées grâce à ces tests permettent d'établir une « carte de santé » des souris **(Brown SDM, 1998; de Angelis and Balling, 1998; de Angelis et al., 2000; Nolan et al., 2000)**. SHIRPA *(Smith Kline Beecham, Harwell, MRC Mouse Genome Centre, Imperial College, Royal London Hospital, Phenotype Assessment)* est l'un des premiers protocoles de criblage standardisé développé, qui met en place près de 40 tests pour identifier des phénotypes de types comportementaux et neurologiques **(Rogers et al., 1997)**. D'autres projets plus récents tels que EMPReSS *(European Mouse Phenotyping Resource for Standardized Screens)* ont été aussi conçus pour réaliser une plus grande variété de tests comprenant des analyses sanguines capables d'identifier les troubles d'autres systèmes physiologiques comme les systèmes hormonaux, métaboliques et immunologiques (http://empress.har.mrc.ac.uk). EMPReSS utilise également des techniques d'imagerie par résonance magnétique et les rayons X pour examiner les fonctions cardiovasculaires et respiratoires ainsi que la composition corporelle de la souris. Des protocoles standardisés pour la nécropsie et l'histopathologie ont également été développés pour détecter et référencer des anomalies tissulaires chez la souris **(Sundberg et al., 2008)**. La plupart des mutants identifiés par les cribles phénotypiques ne sont pas maintenus sous forme de lignées vivantes mais seront stockés sous forme de sperme ou d'embryons congelés **(Justice et al., 1999; Marschall et al., 2009)**.

- *Identification du gène et de la mutation causale*

Pour lancer la cartographie d'une mutation, les souris phénovariantes et/ou leurs descendants sont d'abord croisés avec des souris de fond génétique différent. Les recombinaisons méiotiques *via* ces croisements vont ensuite permettre grâce à la liaison génétique qui peut avoir lieu entre la mutation et des marqueurs polymorphes (microsatellites et/ou SNP), de balayer le génome, repérer le chromosome portant la mutation et limiter l'intervalle à une région chromosomique de 20 à 60 Mb **(Wiltshire et al., 2003)**. La recherche de gènes candidats dans l'intervalle peut être effectuée grâce à :

(i) l'utilisation des « *genome browser* » comme celui de la base de données *Ensembl* (http://www.ensembl.org/index.html) ou celui de *RIKEN bioresource center* (http://omicspace.riken.jp/db/genome.html). *RIKEN* a également mis au point un outil

appelé *Positional Medline* (PosMed) qui suggère et classe les gènes candidats dans la région cartographiée à partir de recherches des phénotypes dans différentes bases de données **(Yoshida et al., 2009)**. Les exons et les régions flanquantes des gènes candidats de l'intervalle sont ensuite séquencés pour identifier la mutation. Un modèle murin pour la dégénérescence rétinienne humaine a été analysé avec PosMed **(Sato et al., 2010)** et la mutation causale a été cartographiée dans la région proximale du chromosome 19. Malgré le fait qu'il y a plus de 400 gènes dans la région cartographiée, l'outil PosMed a désigné trois gènes candidats qui pourraient être responsables de la dégénération rétinienne: *Rom1*, *Stx3* et *Best1*. Les exons des trois gènes ont été séquencés et une seule mutation faux sens dans le gène *Rom1* a été détectée.

(ii) La révolution technologique des méthodes de séquençage à très haut débit, couramment appelés NGS, sont d'une aide inestimable pour raccourcir le temps de travail laborieux pour affiner l'intervalle contenant la mutation causale, l'analyse de liaison et la détection de la mutation. En conséquence, de plus en plus des parties de génomes ou même des génomes entiers de souris phénovariantes sont séquencés ou en cours de séquençage dans le but d'identifier le gène responsable et détecter la mutation causale **(Simon et al., 2012; Bull et al., 2013; Wang et al., 2014)**. Ces nouvelles technologies sont efficaces et rapides (150-200 nouvelles identifications par an, www.irdirc.org) mais génèrent généralement des milliers de variations potentielles des gènes candidats qui doivent être analysées, interprétées pour identifier le gène responsable et la mutation causale. Plusieurs tentatives d'automatiser les comparaisons entre génotypes et phénotypes (l'étape critique nécessaire à la compréhension de ces maladies héréditaires) par une approche systématique, commencent à se mettre en place. En effet, le consortium *the Monarch Initiative* (www.monarchinitiative.org) a intégré dans une même base de donnée toutes les données concernant les maladies humaines et les organismes modèles à partir des bases de données déjà existant comme *Online Inheritance in Man* (OMIM) **(Amberger et al., 2015)**, *the Mouse Genome Database* (MGD) **(Eppig et al., 2015b)**, *the International Mouse Phenotyping Consortium (IMPC)* **(Koscielny et al., 2014)**, dans le but d'accélérer l'identification de nouveaux gènes responsables **(Haendel et al., 2015)**.

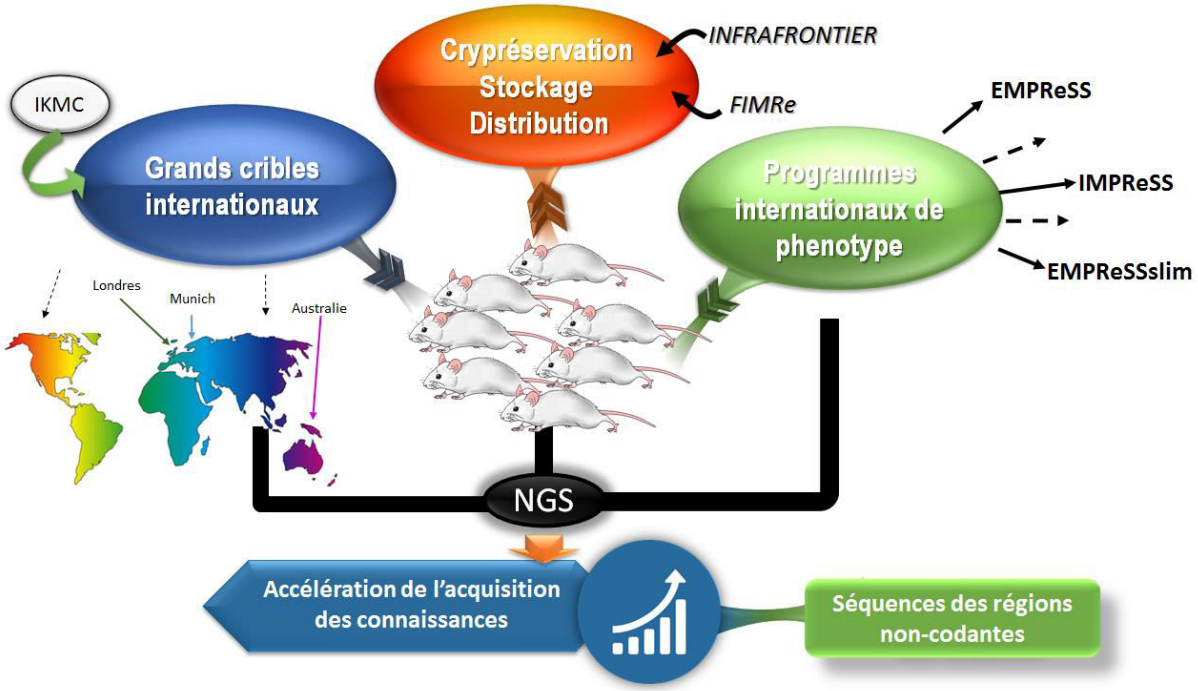


Figure 3 : Les différents projets internationaux basés sur de la mutagenèse à grande échelle chez la souris.

Consortium	Website
The European Conditional Mouse Mutagenesis Programme (EUCOMM)	www.eucomm.org
The EUCOMMTools: Tools for Functional Annotation of the Mouse Genome (EUCOMMTools)	www.eucommtools.org
The European Mouse Phenotyping Resource of Standardised Screens (EMPRESS)	www.empress.har.mrc.ac.uk
The European Union Mouse Research for Public Health and Industrial Applications (EUMORPHIA)	www.eumorphia.org
The European Mouse Disease Clinic (EUMODIC)	www.EUMODIC.org
The EuroPhenome Mouse Phenotyping Resource	www.europhenome.org
The European Mouse Mutant Archive (EMMA)	www.emmanet.org
The European Infrastructure for Phenotyping and Archiving of model mammalian genomes (INFRAFRONTIER)	www.infrafrontier.eu
The International Knockout Mouse Consortium (IKMC)	www.knockoutmouse.org
The International Mouse Phenotyping Consortium (IMPC)	www.mousephenotype.org
The Knockout Mouse Project (KOMP)	www.komp.org
The Knockout Mouse Phenotyping Programme (KOMP2)	http://commonfund.nih.gov/KOMP2/
The Wellcome Trust Sanger Institute Mouse Genetics Project (SANGER-MGP)	http://www.sanger.ac.uk/mouseportal/

Tableau 1 : Les différents projets internationaux à grande échelle réalisés chez la souris (Ayadi et al., 2012).

I.3. Les cribles ENU à grande échelle et les ressources génétiques

L'idée de lancer un projet de mutagenèse chez la souris à grande échelle a été soulevée dans les années 1990 quand le projet de séquençage du génome humain était en cours et devait révéler de nombreux gènes candidats pour des maladies héréditaires. Les projets de mutagenèse à grande échelle ont commencé chez la souris en 1997 autour de deux programmes lancés au *German research Center for environment and Health* en Allemagne **(de Angelis et al., 2000)** et au *Medical research Council* au Royaume Unie **(Nolan et al., 2000)**. En 1999, deux autres projets ont suivi à *RIKEN* au Japon et à *l'Australian national university* en Australie. Au cours des années 2000, de nombreux autres projets de mutagenèse ENU chez la souris ont été planifiés et réalisés pour produire et analyser, chacun, les phénotypes d'environ 10000 souris G1 **(Figure 3)** **(Gondo, 2008)**. Le **Tableau 1** montre plus d'exemples de projets de mutagenèse ENU à grande échelle.

Toutes les ressources génétiques ainsi générées sont maintenant répertoriées dans des bases de données comme *Infrafrontier/European Mutant Mouse Archive* (EMMA) **(Wilkinson et al., 2010; Infrafrontier Consortium, 2015; Raess et al., 2016)**, *the Jackson Laboratory Repository* (jax) **(Ostermeier et al., 2008)**, *The Center for Animal Resources and Development* (CARD) **(Nakagata and Yamamura, 2009)**, the *Riken Bio Resource Center* **(Yoshiki et al., 2009)** ou *Mouse Genome Informatics* (MGI) **(Eppig et al., 2015a)**.

I.4. Avantages et limites de la mutagenèse ENU chez la souris

L'utilisation de la mutagenèse ENU pour identifier les gènes impliqués dans les processus biologiques et/ou les maladies génétiques humaines présente un certain nombre d'avantages par rapport à d'autres techniques. La nature aléatoire de la mutagenèse, l'analyse systématique et simultanée d'un grand nombre de souris et le fait qu'elle ne nécessite aucune connaissance préalable de la fonction ou de la séquence des gènes étudiés, ne font de la mutagenèse ENU une technique particulièrement plus intéressante que les techniques conventionnelles de mutagénèses insertionnelles comme le ciblage de gènes. Cette dernière nécessite au moins la connaissance de la séquence du gène étudié et ne peut étudier qu'un seul ou un petit nombre de gènes à la fois. De plus, les mutagenèses insertionnelles génèrent uniquement des allèles nuls (Knockout), alors que la mutagenèse ENU peut produire les différents types d'allèles modélisant mieux les variations génétiques

naturelles trouvées chez l'homme. En induisant des substitutions d'acides aminés, des changements de conformation de domaines protéiques ou encore des modifications des produits de l'épissage. Les allèles produits par mutagenèse ENU révèlent des fonctions géniques distinctes qui ne sont pas identifiables par la production d'allèles nuls par mutagenèse insertionnelle.

Malgré le fait que les cribles à grande échelle basés sur la mutagenèse ENU sont potentiellement capables de récupérer un grand nombre de mutations fonctionnelles, ils restent des projets très laborieux, nécessitent beaucoup de temps, et demandent un effort organisationnel et logistique assez conséquent pour l'élevage d'animaux avec un grand nombre de personnel qualifié. Les différentes étapes de la cartographie pour localiser les mutations induites par l'ENU restent le processus le plus long du crible malgré la réduction considérable de ce temps observée suite à l'utilisation rapide et peu coûteuse des technologies de séquençage NGS.

II. La mutagenèse insertionnelle chez la souris

Au cours des années 2000, de nombreux projets visant à déléter et étudier tous les gènes de la souris ont été lancés dans le but d'une annotation fonctionnelle aussi bien du génome de la souris que celui de l'homme qui venaient d'être publiés **(Austin CP et al., 2004; Johan Auwerx1, 2004)**. Dans ce chapitre, je me limiterai à la description de deux technologies de mutagenèse insertionnelle : le piégeage de gènes et le ciblage de gènes. Le ciblage génique est une mutagénèse dirigée qui constitue en l'introduction d'une construction d'ADN dans les cellules ES (*Embryonic stem cells*) pour générer la mutation souhaitée par recombinaison homologue entre les séquences de la construction et celles du locus ciblé **(Mansour et al., 1988; Capecchi, 1989; Bradley A, 1992)**. Contrairement au ciblage de gènes, le piégeage de gènes est une méthode de mutagénèse aléatoire dans laquelle l'insertion du transgène dans un gène endogène actif transcriptionnellement, conduit à la pertubation de sa transcription et à son identification.

II.1. Le piégeage des gènes ou «*Gene trap*»

La technique de piégeage de gènes a été développée chez la souris d'abord par Gossler et al en 1989 **(Gossler et al., 1989)** puis par plusieurs autres laboratoires **(Gossler et al., 1989; Friedrich and Soriano, 1991; Skarnes et al., 1992)**. Le principe de la technique, qu'il s'agisse de

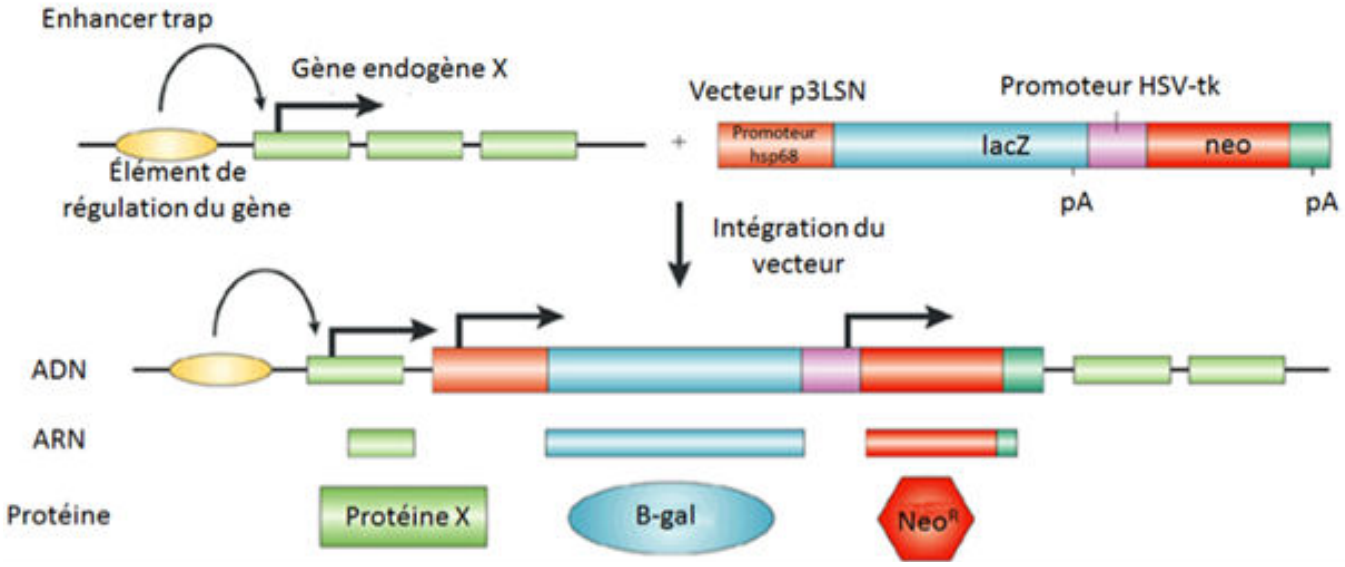


Figure 4 : Enhancer trap.

La construction plasmidique contient un gène rapporteur lacZ et un gène de résistance qui est sous le contrôle d'un promoteur fort, pour piéger un gène endogène X. Dans cet exemple, le vecteur piégeage «enhancer-trap » p3LSN contient un promoteur minimal inductible (hsp68) en amont de lacZ. L'insertion du vecteur à proximité de l'enhancer du gène X conduira à la transcription et à la traduction du gène rapporteur lacZ. Ce vecteur génère généralement des mutations hypomorphes plutôt que nulles **(Stanford et al., 2001).**

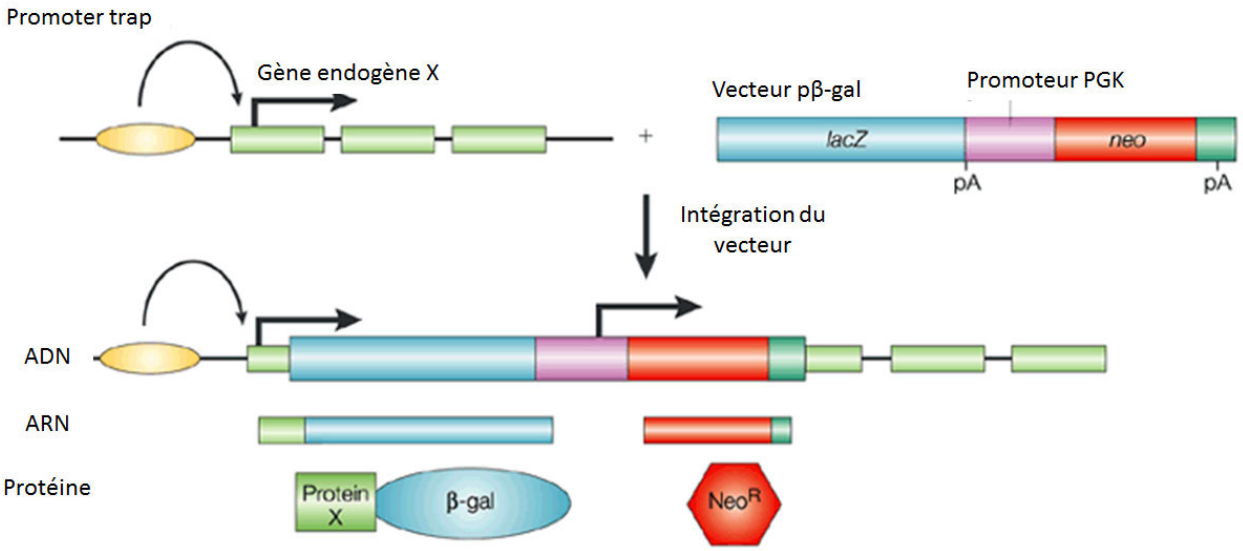


Figure 5 : Promoter-trap.

Si le vecteur βgal, dépourvu de promoteur, s'insère dans la séquence codante du gène X, cela conduit à la transcription du gène rapporteur lacZ. Lors de l'activation du gène X, un transcrit et une protéine de fusion entre la séquence en amont du gène X et lacZ sont générés **(Stanford et al., 2001)**.

piéger des séquences régulatrices (*enhancer-trap*) **(O'Kane and Gehring, 1987)**, des promoteurs (*promoter-trap*) **(Hicks et al., 1997)** ou des séquences de polyadénylation (*polyA-trap*) **(Niwa et al., 1993)**, repose sur l'introduction aléatoire dans le génome de la souris d'un gène rapporteur (souvent le gène LacZ) dépourvu partiellement ou totalement de ses éléments régulateurs de transcription **(Figure 4, 5, 6)**. L'expression du gène rapporteur dépendra donc des éléments de régulation agissant en *cis* de l'unité de transcription à l'intérieur ou à proximité de laquelle l'insertion a eu lieu et sera donc le reflet fidèle de l'expression de cette dernière. Les vecteurs de piégeage, contenant en plus du gène rapporteur un marqueur de sélection, sont introduits dans les cellules d'intérêt, les cellules ES, par électroporation ou par infection rétrovirale. La construction « piège » va interrompre un gène endogène transcrit et donc se comporter comme un agent mutagène. L'insertion de cette séquence étrangère représente un hameçon moléculaire facilitant le clonage et l'identification du gène muté **(Camus, 1998)**.

Les vecteurs les plus utilisés sont ceux des « *enhancer trap* » qui contiennent un site accepteur d'épissage et un signal de polyadénylation flanquant le gène rapporteur *βgeo* (fusion entre les phases codantes des gènes Néomycine et β galactosidase) qui sera intégré dans les introns des gènes **(Figure 4)** **(Gossler et al., 1989; Friedrich and Soriano, 1991)**. Une fois intégré, le gène rapporteur, exprimé à partir du promoteur du gène interrompu, va nous renseigner sur le profil de transcription de ce dernier **(Skarnes et al., 1992)**. L'interruption de la transcription dans ce cas peut être très sévère et conduira souvent à des allèles hypomorphes voire nuls quand l'intégration du vecteur s'effectue dans la partie 5' du gène endogène. Le deuxième type de piégeage appelé « *promoter trap* » est basé sur l'utilisation d'un vecteur sans site accepteur d'épissage conduisant à une intégration dans un des exons du gène endogène **(Figure 5)** **(von Melchner et al., 1992; Hicks et al., 1997)**. Ce type de piégeage semble être efficace quand l'intégration s'effectue dans un gène à un seul exon mais reste moins efficace que le premier type en particulier quand il s'agit d'une mutagenèse à grande échelle. Enfin, le troisième type de piégeage dit « *polyA trap* » est conçu pour les gènes qui ne sont pas transcrits dans les cellules d'intérêt ; les vecteurs de ce type contiennent un marqueur de sélection contrôlé par un promoteur ubiquitaire avec un site donneur d'épissage du coté 3' qui remplace la séquence de polyadénylation **(Figure 6)** **(Niwa et al., 1993; Zambrowicz and Friedrich, 1998)** . Le vecteur de « *polyA trap* » doit s'intégrer dans le

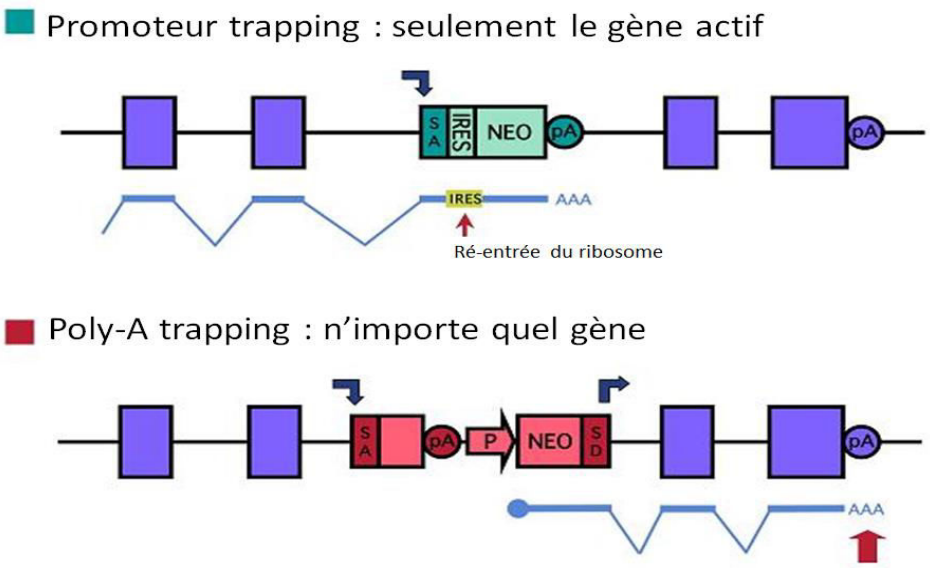


Figure 6 : Poly A trap.

Le plasmide contient un promoteur suivi d'un gène rapporteur et d'un site donneur d'épissage sans séquence signal poly (A). Le gène rapporteur est exprimé si le vecteur s'insère dans la partie codante d'un gène dans la bonne orientation. L'efficacité d'un piégeage de gènes avec un vecteur « poly (A) trap » est très faible (adaptée de http://bsw3.naist.jp/kawaichi/naistrap-e.html).

dernier intron du gène endogène dans la bonne orientation avec l'exon en aval ; en utilisant la séquence de polyadénylation du gène muté, ce type de piégeage conduit soit à la production d'une protéine tronquée soit à la dégradation de l'ARNm chimère par le mécanisme de surveillance des ARNs le *nonsense-mediated mRNA decay* (NMD) **(Shigeoka et al., 2005b)**. Pour éviter la dégradation du transcrit chimère codant pour le marqueur de sélection, une séquence permettant le recrutement direct du ribosome au niveau du codon de démarrage de la traduction appelée « *internal ribosome entry site* » (IRES) est insérée derrière le codon de terminaison du marqueur de sélection ; ceci aboutit à la non reconnaissance du transcrit chimère par le NMD **(Shigeoka et al., 2005b)**. Comme mentionné précédemment, les constructions « piège » sont introduites dans les cellules ES par électroporation de plasmides ou par infections rétrovirales. Contrairement aux vecteurs rétroviraux qui s'intègrent de préférence dans le premier intron du gène endogène, les vecteurs plasmidiques s'intègrent d'une façon aléatoire tout le long du gène **(Chen et al., 2004; De-Zolt et al., 2006)**. C'est pour cette raison que les vecteurs rétroviraux sont beaucoup plus adaptés pour générer des allèles nuls ou très hypomorphes tandis que les plasmidiques produisent uniquement des mutations hypomorphes voire non mutagènes dans certains cas. La présence de point chaud d'intégration ainsi que les insertions complexes, compliquant les analyses ultérieures des clones, représentent les inconvénients majeurs des deux méthodes **(Hansen et al., 2003; Chen et al., 2004)**. La forte mortalité observée chez les embryons issus de lignées de souris hébergeant deux allèles nuls produit par piégeage du gène a conduit plusieurs équipes de recherche à développer des variantes conditionnelles de la technique en insérant les systèmes de recombinaison *cre-lox* et/ou *flp-FRT* en différentes configurations autour des éléments essentiels du vecteur de piégeage. Ces constructions sont conçues pour faciliter des modifications post insertionnelles (élimination ou inversion) du locus du gène endogène interrompu. Ceci permet d'étudier l'effet de l'allèle létal à un stade de développement plus avancé ou même à l'âge adulte **(Schnütgen et al., 2005; Xin et al., 2005; Singla et al., 2010)**.

Contrairement aux mutagenèses aléatoires chimiques ou physiques, les séquences des vecteurs de piégeage sont utilisées pour identifier facilement le gène muté ; cela est généralement réalisé soit par 5' ou 3' RACE des transcrits de fusion générés par l'épissage de la séquence endogène interrompue à la séquence de piégeage, soit par PCR inverse après

digestion de l'ADN génomique par une enzyme de restriction et circularisation des fragments générées **(Silver and Keerikatte, 1989)**.

II.2. Le piégeage de gènes à grande échelle et les ressources génétiques

Les premiers cribles « piégeage de gènes » ont été réalisés pour identifier des gènes nécessaires au développement. Friedrich et Soriano ont pu avoir de nombreuses lignées de souris chimères issues de pools de cellules ES transfectées par un vecteur « *promoter trap* ». Sur les 24 lignes obtenues, 9 n'étaient pas capables de générer une descendance homozygote viable, indiquant que les gènes invalidés sont indispensables au développement de la souris **(Friedrich and Soriano, 1991**). Dans un deuxième crible, un grand nombre de gènes, parmi les 279 gènes caractérisés par leurs profils d'expression, ont été identifiés comme étant régulés spatio-temporellement durant l'étape de l'embryogenèse **(Wurst et al., 1995)**. D'autres cribles « *gene trap* » ont suivi dans le but d'identifier des gènes régulés ou exprimés différemment en fonction des cellules, des tissus ou encore de l'état physiologique ou de différenciation des cellules. Ce type de crible a été utilisé avec succès pour identifier par exemple un gène codant pour « *lysosomal cysteine protease cathepsin B* » régulé durant la différenciation des myoblastes *in vitro* **(Gogos et al., 1996)**, des gènes surexprimés ou sous exprimés suite à un cascade de signalisation dans les cellules de Sertoli **(Vidal et al., 2001)** ou des gènes spécifiques d'un tissu comme l'hippocampe **(Steel et al., 1998)** ou l'oreille interne **(Yang et al., 1997)**. De plus, des variantes de la technique adaptées au type cellulaire ou à la nature de la protéine (cytoplasmique, membranaire, secrétée etc.) ont pu être réalisées pour identifier de nouveaux gènes qui contrôlent les connexions neuronales **(Leighton et al., 2001)**, qui s'expriment différemment dans cellules souches cardiovasculaires **(Kuhnert and Stuhlmann, 2004)**, qui codent pour des protéines membranaires ou secrétées **(Skarnes et al., 1995)** ou qui régulent le développement **(Gossler et al., 1989; Friedrich and Soriano, 1991)**.

Le piégeage de gène a pris de l'importance au cours des années 1990 et a été utilisé à grande échelle pour la production des mutations insertionnelles aléatoires dans les cellules souches **(Zambrowicz and Friedrich, 1998)**, conduisant à la construction des grandes banques de cellules ES hébergeant des gènes piégés **(Figure 3)** :

(i) la banque « *Omnibank* » développée par « *Lexicon Genetics* » qui contient la plus grande collection de mutants ES, avec 270000 clones, couvre près de 60% du génome de la souris **(Zambrowicz et al., 2003)**. Les données concernant cette banque sont répertoriées aussi bien dans le site web de « *lexico genetics* » (http://www.lexicon-genetics.com/index.php) que dans la base de donnée « *NCBI Genome Survey Sequences Database* ».

(ii) L'objectif du projet « *the International Gene Trap Consortium* » (IGTC), initié et réalisé par les grandes équipes de recherche travaillant sur le modèle murin, était de construire des banque de cellules ES mutantes par « *gene trap* » et les mettre à la disposition de toute la communauté scientifique **(International Gene Trap et al., 2004)**. L'IGTC a généré une collection de 50000 lignées de cellules ES hébergeant des mutations dans 9000 gènes soit 40% du génome de la souris. Il faut noter que contrairement à *Omnibank*, l'IGTC a utilisé une panoplie de vecteurs de piégeage pour interrompre un grand nombre de loci dans un petit nombre de clones **(International Gene Trap et al., 2004)**. Les séquences interrompues sont archivées dans la base de données du site web de l'IGTC (http://www.genetrap.org/) **(Nord et al., 2006)** et visualisées par le « *genome browser* » de la souris sur la base de données *Ensembl* (http://www.ensembl.org/Mus_musculus/Info/Index).

D'autres cribles, de moindre importance par rapport au nombre de gènes interrompus, ont été développés en utilisant : (i) la technique conventionnelle comme celui de *Baygenomics* **(Stryke et al., 2003)**, (ii) des variantes conditionnelles comme celui de *German Gene trap Consortium* (http://genetrap.gsf.de/) **(Schnütgen et al., 2005)** qui a produit 4000 clones de cellules ES en insérant le vecteur *flip-excision* dans prés de 1000 gènes, ou (iii) celui développé par De-zolt et al. pour identifier 716 gènes codant pour des protéines transmembranaires ou secrétées, dont 61% parmi eux n'étaient pas identifiables par l'approche conventionnelle de la technique **(De-Zolt et al., 2006)**. De nombreuses mutations insertionnelles de type « *gene trap* » ont été cherchées dans les banques ES parce qu'elles conduisent chez la souris à des phénotypes similaires à des syndromes de maladies humaines. Les modèles murins « *gene trap* » ainsi générés sont utilisés pour comprendre les mécanismes moléculaires des pathologies, pour identifier des marqueurs pour un diagnostic précoce ou encore pour mettre en place la thérapie nécessaire.

Les maladies comme le cancer **(Nobukuni et al., 2005; Ishiguro et al., 2010; Robles-Oteiza et al., 2015)**, les anémies **(Khoriaty et al., 2016; Sun et al., 2016b)**, les maladies cardio-vasculaires, les

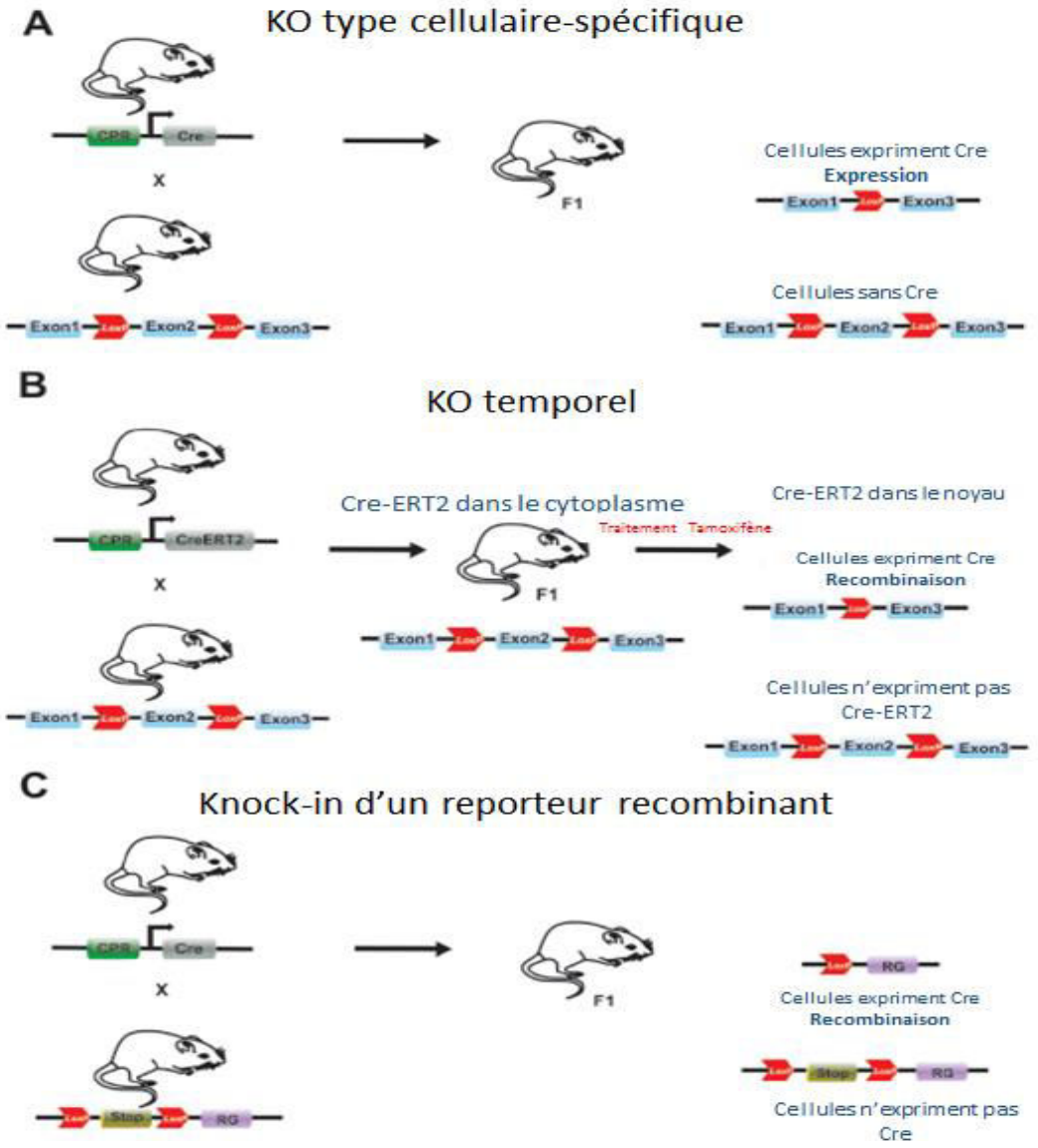


Figure 7 : les différentes variantes de la technologie « *gene targeting* ».

Knockout conditionnel (A) : Des souris peuvent être générées en associant deux lignées différentes de souris transgéniques: l'une portant une enzyme SSR (par exemple Cre) sous le contrôle d'un promoteur spécifique de type cellulaire (CPR) et l'autre contenant des régions codantes du gène d'intérêt flanqué par LoxP. La progéniture de cet accouplement portant les deux transgènes subirait une inactivation génétique uniquement dans les cellules exprimant Cre qui va recombiner la région LoxP flanqué. Dans les cellules sans expression de Cre, aucune recombinaison ou inactivation du gène ne se produira.

Knockout conditionnel spatio-temporel (B) : Le contrôle temporel de la recombinaison à médiation Cre peut être réalisé par l'utilisation d'une protéine Cre chimère (Cre-ERT2). Le Cre-ERT2 ne se translate que vers le noyau et catalyse la recombinaison après le traitement au tamoxifène.

Knock-in conditionnel (C) : Les souris rapporteuses de recombinaison knockin portent habituellement une construction avec un codon d'arrêt flanqué par LoxPsequences suivi d'un gène rapporteur (RG). Chez les souris portant ce transgène, RG n'est activé que dans les cellules exprimant Cre, où le codon d'arrêt a été excisé. Dans les cellules Cre négatives, le RG n'est pas exprimé **(ROCHA-MARTINS et al., 2015)** .

cardiomyopathies **(Becker et al., 2014; Quintero-Rivera et al., 2015)**, les neuropathologies **(Trimborn et al., 2010; Gow et al., 2013; Sul et al., 2013; Ylikallio et al., 2013)**, les maladies génétiques liées au chromosome X **(Cox et al., 2010)**, les maladies métaboliques **(Michailidou et al., 2008)** ou les maladies hépatorénales **(Wright et al., 2009)** ne sont que quelques exemples.

II.3. Le ciblage de gènes «*Gene targeting*»

La technique « *gene targeting* » ou ciblage de gènes qui permet d'inactiver un gène chez la souris et de vérifier ainsi son rôle, est réalisée par recombinaison homologue dans les cellules ES **(Evans and Kaufman, 1981; Martin, 1981; Smithies et al., 1985)**. En 2007, le prix Nobel de médecine et de physiologie a récompensé deux Américains, Mario Capecchi et Oliver Smithies, et un Britannique Sir Martin Evans, dont les travaux pionniers ont été les bases de cette technique qui est aujourd'hui très largement utilisée dans la recherche biomédicale **(Mak, 2007)**. Les mutations introduites par la technique « *gene targeting* » sont essentiellement des délétions des gènes endogènes par la méthode de remplacement du gène appelés couramment « Knockout » ; toutefois, d'autres types d'altérations peuvent être introduites comme par exemple les mutations ponctuelles afin d'étudier des maladies humaines, les délétions conditionnelles par recombinases ou des fragments d'ADN endogènes ou exogènes insérés dans des sites spécifiques du génome de la souris nommées « Knock-in » **(Figure 7)**.

Dans la méthode conventionnelle du Knockout, les vecteurs plasmidiques utilisés contiennent à la fois un marqueur de sélection positif (souvent le gène codant la résistance à la Néomycine) et un marqueur de sélection négatif **(Figure 8)** **(Mansour et al., 1988; Valenzuela et al., 2003)**; le marqueur de sélection positif est souvent flanqué par des séquences homologues à celles qui encadrent le gène d'intérêt. L'introduction d'un tel vecteur dans les cellules ES par électroporation permet, par recombinaison homologue, le remplacement du gène mis à l'étude par le marqueur positif et conduit à la génération des cellules ES résistantes à la néomycines et délétées totalement ou partiellement dans le gène en question. Le marqueur de sélection négatif quant à lui est utilisé pour éliminer les cellules ES hébergeant des intégrations ectopiques du vecteur **(Capecchi, 1989)**. Après la double sélection, les cellules ES sont injectées dans des blastocystes pour produire des souris chimères qui seront croisées avec la lignée sauvage pour avoir une descendance hétérozygote pour la délétion. Le

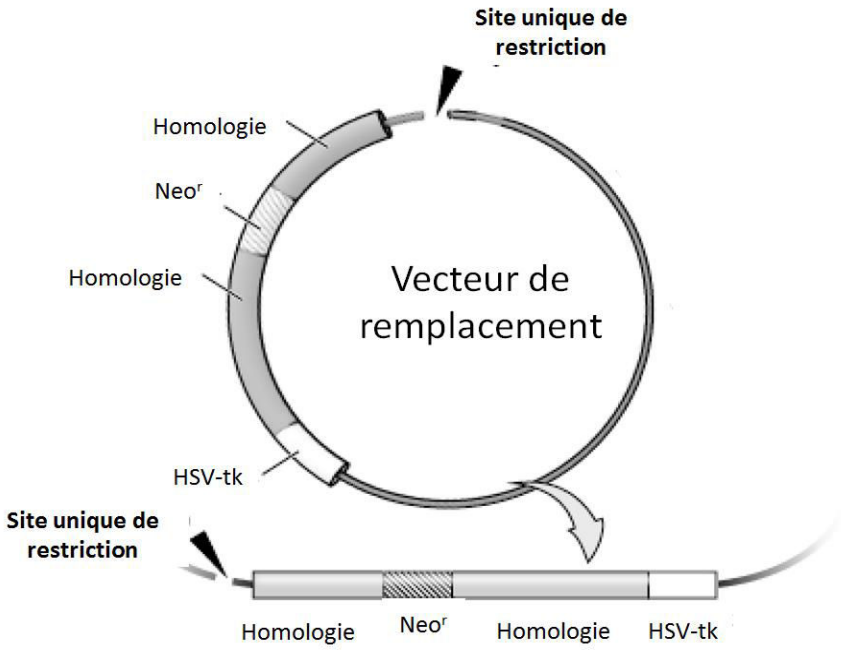


Figure 8 : vecteur de ciblage.

Deux bras d'homologie flanquent un marqueur de sélection positif (neo^r). Un marqueur de sélection négatif (HSV-tk) est placé adjacent à l'un des bras de ciblage. Un site d'enzyme de restriction unique est situé entre le squelette du vecteur et le bras d'homologie. Lorsqu'il est linéarisé, le squelette du vecteur protégera alors le HSV-tk contre les nucléases **(Hall et al., 2009)**.

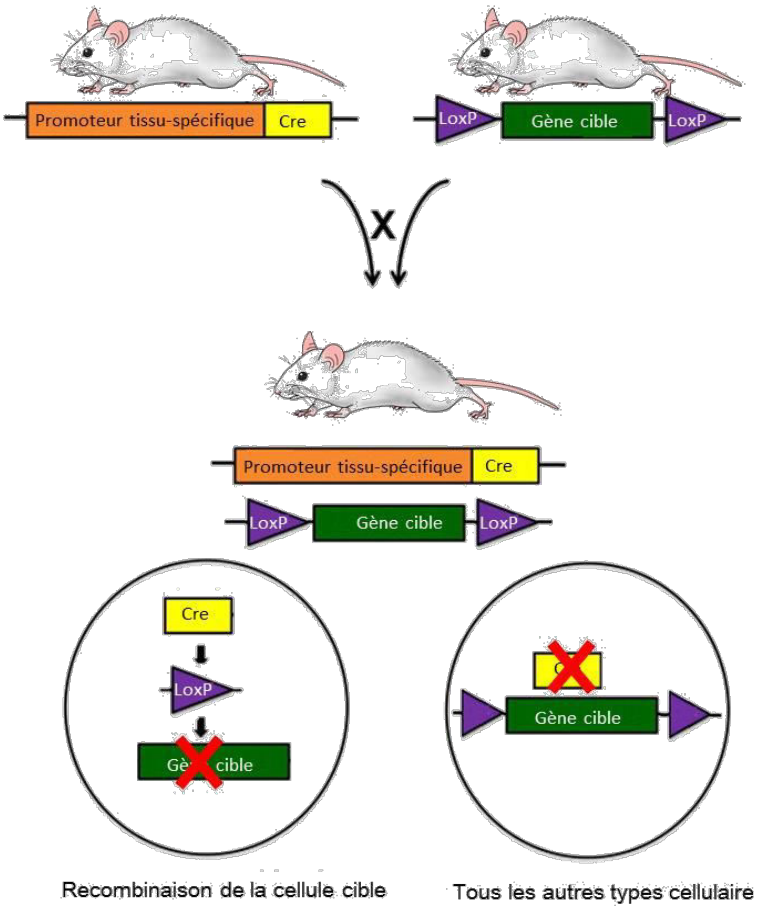


Figure 9 : Principe du système *Cre-loxP*.

Ce système permet de générer des knockouts tissu-spécifiques. Les souris exprimant la recombinase Cre sous le contrôle d'un promoteur tissu-spécifique sont accouplées à des souris dans lesquelles le gène d'intérêt est flanqué de sites loxP (souris "floxées"). Dans la progéniture, la recombinaison du gène floxé ne se produit que dans les cellules dans lesquelles Cre est exprimée, conduisant à un ciblage «tissu-spécifique» **(Harno et al., 2013)**.

Knockout conventionnel peut conduire dans certains cas à une mortalité embryonnaire si le gène délété joue un rôle essentiel au cours du développement.

Pour éviter ce problème de mortalité, des variantes conditionnelles de la technique basées sur l'utilisation de nouveaux vecteurs fondés sur le système de recombinase *Cre-loxP* du bactériophage Pl ou le système équivalent de la levure *Flp/Frt* ont été mises en place **(Figure 9)**. Ces vecteurs permettent d'une part, de créer des mutations « propres », où les gènes de sélection sont éliminés du locus étudié et ne peuvent donc pas avoir d'influence sur les gènes voisins ; d'autre part, ils permettent une expression du gène muté contrôlée dans le temps, ce qui rend possible l'étude, à l'âge adulte, de gènes dont l'absence ou la mutation sont létales au stade embryonnaire **(Kühn and Schwenk, 1997)**. La recombinaison par l'enzyme Cre (ou *Flp*) entre deux sites *loxP* (ou *Frt*) ayant des orientations opposées conduit à l'inversion du fragment d'ADN compris entre les deux sites *loxP* (ou *Frt*). En revanche, la recombinaison entre deux sites de même orientation conduit à la délétion du fragment d'ADN. Cette réaction est réversible et une molécule portant un site *loxP* (ou *Frt*) peut s'intégrer dans une autre molécule d'ADN arborant un site *loxP* (ou *Frt*) **(Figure 10)** **(Michael et al., 1999)**. Pour générer des souris délétées dans un gène avec cette variante conditionnelle de la technique, deux lignées de souris sont nécessaires : la première, appelée « *floxée* », contient l'exon du gène d'intérêt flanqué par les sites *loxP* (ou *Frt*) et la deuxième, nommée « transgénique », héberge un transgène contenant le gène codant pour la recombinasse Cre (ou *Flp*) ; l'expression du gène *Cre* pouvant être sous le contrôle d'un promoteur spécifique de tissu ou inductible ou les deux à la fois **(Michael et al., 1999)**. Dans la descendance issue du croisement des deux lignées, la recombinase va permettre la délétion de l'exon flanqué par les sites *loxP* (ou *Frt*) dans le tissu spécifique et/ou durant le stade de développement étudié. Il existe actuellement plus de 500 lignées exprimant la recombinase Cre de façon tissu spécifique dont les données sont répertoriées dans la base de données *CreXMice* du laboratoire de Andras Nagy (http://nagy.mshri.on.ca/cre/) **(Nagy and Mar, 2001; Nagy et al., 2009)**.

La technique knock-in, un raffinement du technique conventionnel knockout, consiste en l'introduction, en phase avec la séquence codante du gène cible, de la phase codante d'un gène marqueur **(Figure 11)**. Une fois la recombinaison homologue effectuée dans le gène cible choisi, l'allèle modifié exprime une protéine de fusion contenant la

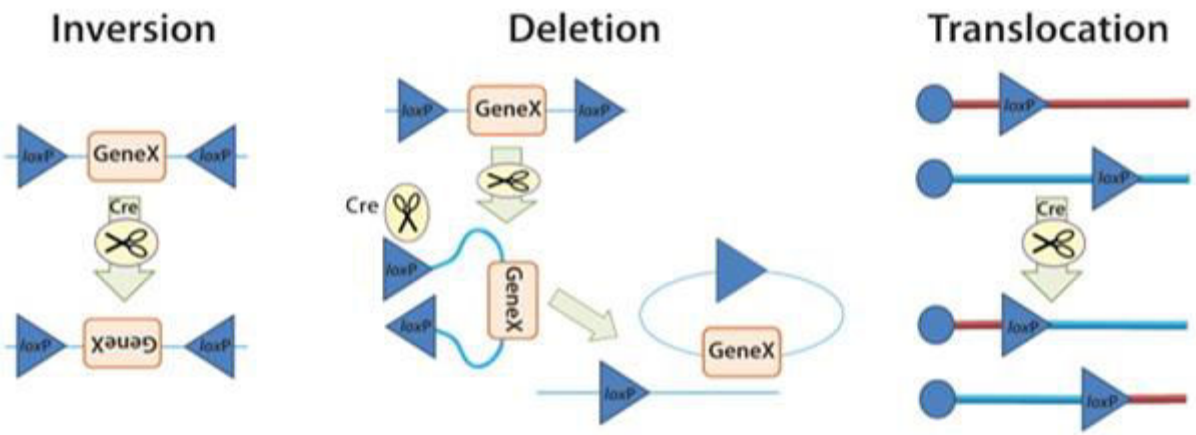


Figure 10 : Le système *Cre-loxP* peut aboutir à trois résultats différents selon l'emplacement et l'orientation des sites.

Inversion: Si les sites loxP sont sur le même brin d'ADN et sont dans des orientations opposées, la recombinaison entraîne une inversion et la région de l'ADN entre les sites loxP est inversée. **Délétion**: Si les sites sont orientés dans le même sens, la séquence entre les sites loxP est excisée comme un fragment circulaire d'ADN et n'est pas maintenue. **Translocation**: Si les sites loxP sont sur des molécules d'ADN distinctes, un événement de translocation est généré entre les sites loxP loxP (www.addgene.org/cre-lox/).

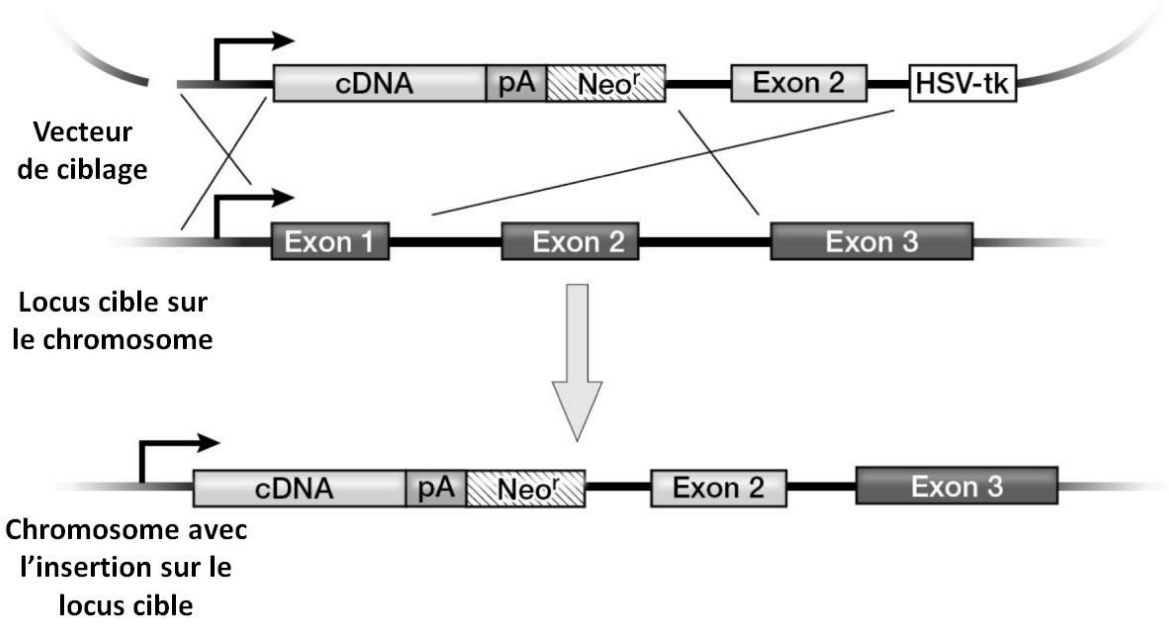


Figure 11 : La variante Knock-in.

Le knock-in est une variante du knock-out.

Dans cette variante, mis à part une invalidation du gène cible, un gène d'intérêt est introduit dans le locus. Après recombinaison homologue, ce gène d'intérêt est placé sous le contrôle du promoteur et des séquences régulatrices du gène cible. Il est donc synthétisé à la place du gène cible. Le plus souvent, on choisit d'insérer un gène rapporteur qui une fois exprimé peut être observé, et quantifié **(Hall et al., 2009)**.

protéine du gène cible et la protéine du marqueur. Le choix du gène marqueur dépend évidemment du but recherché ; par exemple, il peut s'agir de la séquence codante de la β-gal dont l'activité est aisément repérable, y compris à l'échelle cellulaire. Ainsi l'expression de la β-gal mime-t-elle l'expression du gène cible, ce qui permet la détermination de son profil d'expression et/ou sa localisation **(Li et al., 1997)** mais peut également être très utile pour suivre le destin des cellules exprimant normalement le gène cible dans le contexte d'un animal porteur de la mutation à l'état homozygote **(Schneider-Maunoury et al., 1993; Tajbakhsh, 1996)**. Une autre variante de la technique knock-in concerne le cas de gènes appartenant à une famille multigénique ayant un patron d'expression différent et dont l'inactivation respective entraîne des phénotypes différents ou contrastés. Les protéines codées par ces gènes apparentés ont souvent une fonction équivalente et c'est le changement de profil d'expression qui est à l'origine des phénotypes observés et non une fonction différente. En effet, d'après quelques études **(Hanks et al., 1995; Hanks et al., 1998; Saga, 1998; Acampora et al., 1999; Geng et al., 1999; Suda et al., 1999)**, c'est généralement ce type de situation qui est observé, ce qui souligne l'importance des régions régulatrices d'un gène comme cibles d'évolution et de diversification fonctionnelles.

II.4. Le ciblage de gènes à grande échelle et les ressources génétiques

La technique « *gene targeting* » a été utilisée à grande échelle dans le but de muter d'une façon systématique tous les gènes codant de la souris, créant ainsi des ressources génétiques précieuses pour la recherche biomédicale **(Austin CP et al., 2004; Johan Auwerx, 2004)**. Trois grands projets de délétion (*Knockout programs*) ont été élaborés : le KOMP (*KnockOut Mouse Project*) réalisé par le *National Institutes of Health (NIH, USA)*; l'EUCOMM (*EUropean Conditional Mouse Mutagenesis Program*) **(Friedel et al., 2007)** mis en place par la commission européenne (EC) et le NorCOMM (*North American Conditional Mouse Mutagenesis Project*) developpé par *Genome Canada*. Avec deux autres projets réalisés par l'institut TIGM (*Texas Institute of Genomic Medicine)* **(Collins et al., 2007)** et par l'association AMMRA (*Asian Mouse Mutagenesis and Resource Association*), tous les projets ont été finalement regroupés sous le consortium international IKMC (*International Knockout Mouse Consortium*) **(collins, 2007)**. Toutes les ressources génétiques générées par tous les projets sont stockées et accessibles sur le site web officiel d'IKMC (www.knockoutmouse.org). Les objectifs principaux du site web de l'IKMC et de sa base de données sont : mettre à disposition une infrastructure pour

sélectionner par priorité les gènes à déléter dans les centres de production de l'IKMC, faciliter la coordination des travaux et suivre leurs progrès au sein de l'IKMC et rendre toutes les données facilement accessibles à la communauté scientifique **(Figure 3)** **(Skarnes et al., 2011; Ayadi et al., 2012)**. Dans le cadre de l'IKMC, 26000 knockouts dans des cellules ES ont été produits, 15000 parmi eux sont générés par la technique « *gene targeting* » dont 13000 sont conditionnelles. Ces mutations sont introduites dans plus de 18500 gènes dans des cellules ES et plus de 4900 lignées de souris mutantes ont été générées. Ces lignées de souris sont maintenant caractérisées par des programmes de phénotypage à grande échelle organisés au sein de l'IMPC et d'EUMODIC **(White et al., 2013; de Angelis et al., 2015)**. En outre, plus de 19000 ES clones pour environ 9000 différents gènes ont été distribués à la communauté scientifique pour des études fonctionnelles **(Rosen et al., 2015)**.

Récemment, dans le cadre de l'effort fourni par l'IMPC pour caractériser phénotypiquement les 5000 souris knockout générées par l'IKMC, Dickson et al ont identifié parmi les 1751 lignées knockout étudiées près de 410 gènes, dont le Knockout est létal au stade embryonnaire. En utilisant un pipeline d'analyse phénotypique et de l'imagerie à haute résolution, des phénotypes sont identifiés et analysés essentiellement pour le Knockout des gènes non connus ; parmi eux plusieurs gènes impliqués dans des maladies humaines ont été révélés. Ces lignées Knockout ont permis d'une part, l'étude fonctionnelle des gènes chez les mammifères et elles représentent d'autre part, des modèles précieux pour comprendre et décortiquer les mécanismes moléculaires sous-jacents des maladies héréditaires humaines **(Dickinson et al., 2016)**.

II.5. Avantages et limites du piégeage et du ciblage de gènes

La technique « *gene trap* » est une mutagénèse aléatoire à grande échelle mais semble être incapable de créer des mutants dans tous les gènes murins. De plus, le nombre de lignées de cellules ES générées par « *gene trap* » est beaucoup plus grand que le nombre de gènes de souris connus. Uniquement 50 à 70% des gènes de la souris ont été piégés ; Ceci veut dire qu'il y a un certain nombre de gènes qui a été piégé plusieurs fois et que certains vecteurs sont insérés dans des régions non codantes **(Abuin et al., 2007; Roma et al., 2007; Roma et al., 2008)**. En effet, de nombreuses expériences ont montré un certain biais dans l'intégration des vecteurs de piégeage et que certaines régions génomiques sur les chromosomes dites

« points froids », contrairement aux points chauds, ne peuvent pas être invalidées par cette technique **(Austin CP et al., 2004; Shigeoka et al., 2005a)**. Contrairement au « *gene trap* », la technique « *gene targeting* » peut générer des Knockout dans n'importe quel gène de la souris et il a été montré qu'il était efficace pour déléter presque tous les gènes de la souris, en particulier avec sa variante conditionnelle *cre/Lox* qui induit d'une façon précise la délétion du gène d'intérêt à la fois dans le temps et dans l'espace **(Austin CP et al., 2004; Barbaric et al., 2007)**. Toutefois, cette technique reste quand même une tache expérimentale assez lourde et fastidieuse à mettre en œuvre qui prend énormément de temps et qui nécessite des compétences professionnelles élevées.

Chapitre II : L'ARN interférence

I. Historique

L'ARNi est un mécanisme général de régulation de l'expression génique fondé sur la dégradation séquence spécifique des ARNs transcrits. La première mise en évidence de ce phénomène, bien que celle-ci ne fût identifiée comme telle que plusieurs années plus tard, était dans les années 1990 par l'équipe de R. Jorgensen lors de ses recherches sur les mécanismes de coloration de Pétunia **(Napoli et al., 1990)**. Les auteurs, souhaitant obtenir un renforcement de la coloration des fleurs, ont introduit une deuxième copie du gène de la pigmentation dans le génome de la plante, et ont obtenus contre toute attente, une extinction de la coloration **(Napoli et al., 1990)**. Ces expériences appelées, co-suppression, ont mis en évidence un mécanisme très répandu chez les plantes, impliqué dans la protection du génome contre les éléments transposables et les virus à ARN **(Vaucheret and Fagard, 2001)**. Agissant sur les ARNs déjà transcrits, ce mécanisme a été baptisé « *Post Transcriptional Gene Silencing* » ou PTGS. Au cours des années 90, de nombreuses analyses moléculaires réalisées sur *C. elegans* **(Guo and Kemphues, 1995)** et *N. crassa* **(Cogoni et al., 1996)**) hébergeant des transgènes inactifs ont montré que la transcription aussi bien du gène résident que le transgène est toujours active, mais les transcrits sont rapidement dégradés dans le cytoplasme. C'est le 19 février 1998, que paraît l'article considéré aujourd'hui comme fondateur de l'ARNi et dans lequel le rôle suppresseur des petites molécules d'ARN non codants est clairement établi **(Fire et al., 1998)**. Ce travail sera récompensé huit ans plus tard par le prix Nobel de physiologie et de médecine. Cette découverte a marqué le monde de la génétique expérimentale, la biologie cellulaire et, plus récemment, de la médecine.

A. Fire et al. ont découvert que l'introduction d'un ARN double brin (ARNdb) dans les cellules de *C. elegans* pouvait réduire spécifiquement l'expression des protéines en se liant à leurs ARNm et ceci pendant plusieurs générations. Ce phénomène d'extinction génique baptisé ARN interférence, a permis d'analyser et de comprendre la fonction de nombreuses protéines **(Fire et al., 1998)**. Au cours des années suivantes, de nombreuses équipes parviennent à mettre en évidence le phénomène d'ARNi chez d'autre espèces comme la

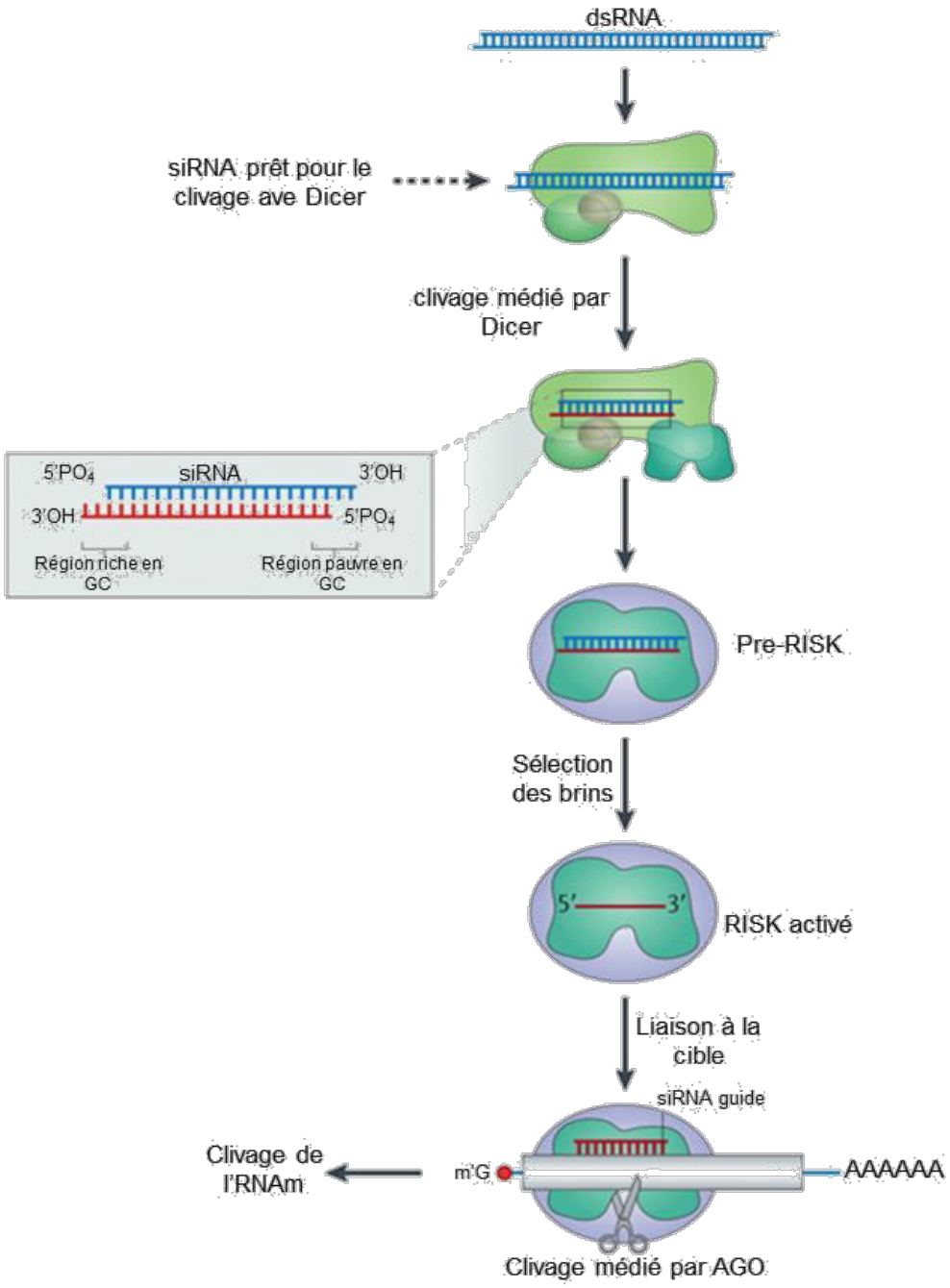


Figure 12 : L'ARN interference.

Les précurseurs de siARN d'origine endogène ou exogène, comprenant une longue région double brin, sont pris en charge par l'enzyme Dicer qui va les maturer dans le cytoplasme. Les siARN matures sont ensuite pris en charge par le complexe RISC (RNA induced silencing complex) qui va d'abord éliminer l'un des deux brins, le brin passager et garder le brin guide puis avec les autres protéines du complexe RISC, chercher l'ARNm cible pour le dégrader avec l'activité Slicer de la protéine Ago d'une manière séquence spécifique **(Davidson and McCray, 2011)**.

drosophile **(Kennerdell and Carthew, 1998)**, le xénope **(Oelgeschlager et al., 2000)**, la souris **(Svoboda et al., 2000)**, le trypanosome **(Ngô et al., 1998)**, la planaire, le poisson zèbre **(Zamore et al., 2000)** les protistes **(Ruiz et al., 1998)**, etc.

L'ARNi désigne actuellement l'ensemble des régulations spécifiques de l'expression des gènes qui sont dirigées par une molécule d'ARN de petite taille qui sert de séquence guide. Ces régulations peuvent induire l'arrêt de la transcription du gène ciblé, c'est la voie nucléaire de l'ARNi **(Castel and Martienssen, 2013)**, ou agir au niveau de l'ARNm soit par un mécanisme de dégradation soit par une inhibition de la traduction ; il s'agit des voies de l'extinction post-transcriptionnelle des gènes : la voie des siRNA (*small interfering RNA*), la voie des miRNA (*micro RNA*) **(Naqvi et al., 2009)** et la voie des piRNA (*PIWI interacting RNA*) **(Siomi et al., 2011)**.

II. Les différentes voies de l'ARN interférence

Les ARN interférents sont de petits ARNs non codants, double brin de 20 à 30 nucléotides (nts). Ils sont issus de la maturation d'ARNs précurseurs plus longs dans le noyau et/ou dans le cytoplasme. Ces ARNs agissent sous forme simple brin en s'hybridant à la séquence d'ARN cible et ainsi empêcher l'expression du gène en question soit en clivant cet ARN soit en inhibant sa traduction en protéine. Cette association avec l'ARN cible se fait à travers des complexes protéiques associés au petit ARN et impliqués aussi bien dans sa maturation que dans sa fonction **(Wilson et al., 2015)**.

II.1. La voie des siRNA

Les siARNs sont des petits ARN double brin d'environ 21 nts, impliqués dans la dégradation séquence spécifique des ARNm cibles. Ils sont formés de deux brins de 19 nts complètement complémentaires et de deux nts sortants sur chacune des extrémités 3' **(Figure 12)**. Ces petit ARNs sont issus d'ARNdb qui peut être d'origine endogène ou exogène ; l'ARNdb exogène peut être virale ou introduit par injection dans l'organisme ou encore transcrit à partir d'un transgène dupliqué avec répétition inverse. Ils sont capables d'inhiber l'expression d'un gène par complémentarité parfaite avec les séquences d'ARNm. Un mécanisme à deux phases a été proposé pour la dégradation des transcrits induite par l'ARNdb. Dans la première phase appelée phase initiatrice, l'ARN double brin est clivé en

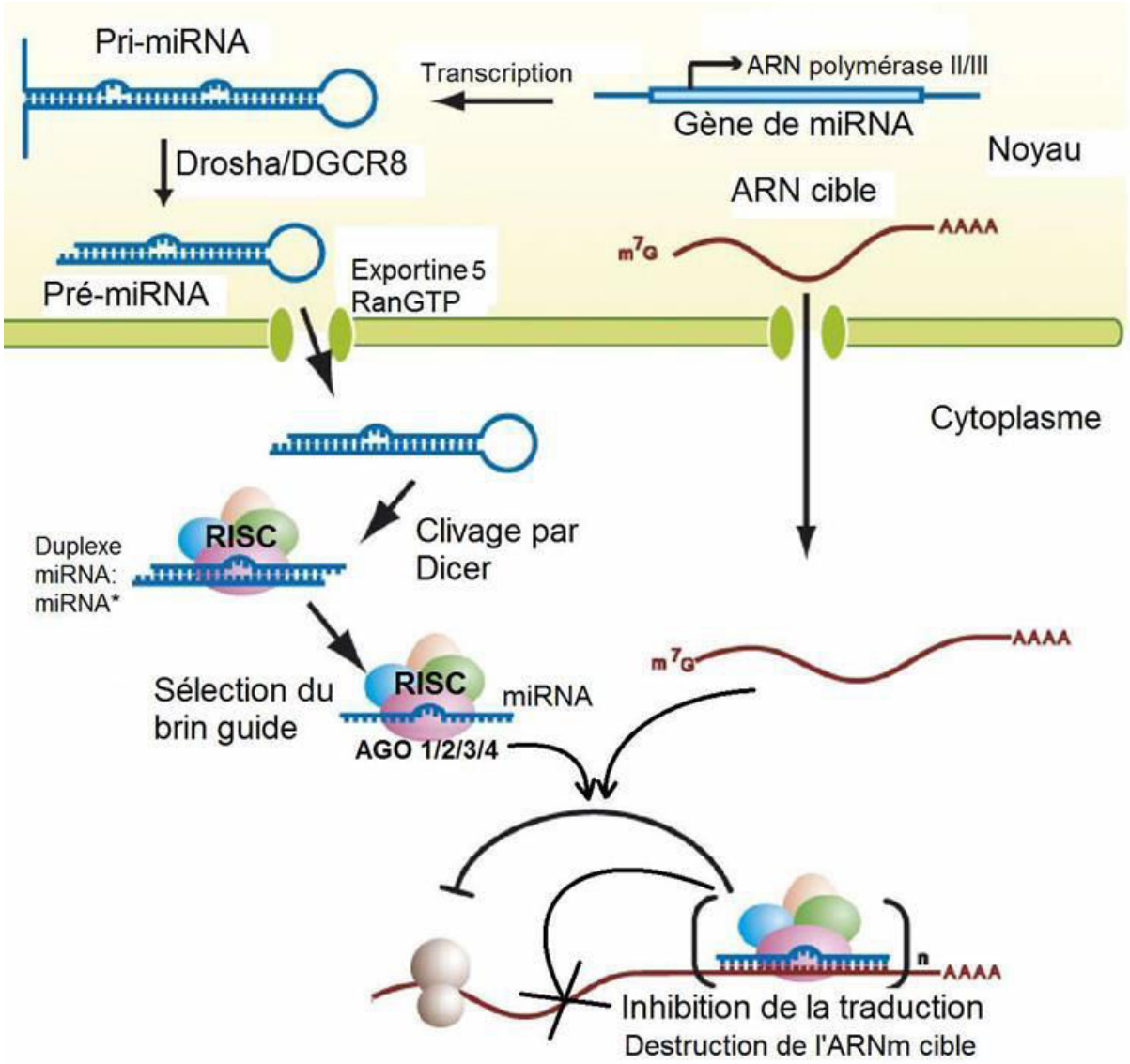


Figure 13 : Voie de maturation canonique des miRNAs.

Les gènes de miRNAs sont transcrits sous la forme de précurseurs appelés « pri-miRNA ». Ces précurseurs sont clivés dans le noyau par un complexe nommé microprocesseur (Drosha et DGCR) pour avoir un produit intermédiaire appelé « pre-miRNA » en forme tige boucle. Le pre-miRNA est ensuite transporté dans le cytosol par l'exportine 5 pour être clivé par Dicer. Le miRNA s'incorpore en forme de petit ARN double brin « miRNA/miRNA* » dans le complexe RISC qui va garder que le brin miRNA. Selon le degré de la complémentarité, l'association entre le miRNA et l'ARNm va conduire soit à l'inhibition de la traduction soit à la dégradation de l'ARNm cible d'après **(Kurreck, 2009)**.

siRNAs par une ribonucléase de type III appelée Dicer **(Paroo, 2010)**. Au cours de la deuxième phase dite phase effectrice, un brin du siRNA (le brin guide) avec sa région essentielle appelée « *seed region* » (une série de 5 ou 6 nucléotides proche du début de l'extrémité 3') vont guider une machinerie de dégradation appelée RISC (*RNA-induced silencing complex*) qui va cibler l'ARNm pour le cliver grâce à l'activité endonucléolytique (slicer) de la protéine Argonaute **(Tolia and Joshua-Tor, 2007; Meister, 2013)**. Une fois clivé, l'ARNm cible sera dégradé dans les sens 5'-3'et 3'-5' dans des structures cellulaires impliquées dans le stockage et la dégradation des ARN appelées *Processing bodies* (P-body) **(Eulalio et al., 2007)**. De nombreuses études ont montré que cette voie de l'ARNi est spécifique de la lignée somatique et impliquée dans l'intégrité du génome en le protégeant contre les ADN parasites comme les transposons et les virus **(Ghildiyal and Zamore, 2009)**.

II.2. La voie des microARNs

Les miRNAs sont une catégorie d'ARN simple brin non codant d'environ 22 nts. Leur maturation requiert l'intervention de deux enzymes de type RNaseIII : Drosha et Dicer **(Figure 13)**. La biosynthèse des miRNAs débute dans le noyau par la transcription par l'ARN polymérase II de gènes des miRNAs en long précurseurs : les pri-miRNAs. Ces précurseurs en tige boucle seront excisés par un complexe protéique « le microprocesseur » (Drosha/DGCR8), pour donner naissance aux pré-miRNAs (ces précurseurs sont parfois directement produits par la machinerie du splicing dans le cas de plusieurs miRNAs appelés mirtrons parce qu'ils sont localisés dans les introns des gènes codants) **(Babiarz et al., 2008)**. Les pré-miRNAs sont ensuite transportés dans le cytoplasme par la protéine exportine 5 où ils sont maturés par Dicer qui hydrolyse la structure boucle. Le petit ARN double brin libéré (miRNA/miRNA*) sera pris en charge par le complexe RISC, un seul brin (le miRNA) est incorporé dans la protéine Argonaute **(Bushati and Cohen, 2007; Tolia and Joshua-Tor, 2007)**. Enfin le complexe RISC chargé est guidé vers l'extrémité 3' UTR de l'ARNm cible par une complémentarité partielle de séquence ; cette association conduit à l'inhibition de la traduction de l'ARNm cible par la perte de sa coiffe en 5' et de sa polyadénylation en 3' ainsi qu'une dégradation par des exonucléases **(Bushati and Cohen, 2007) (Brodersen and Voinnet, 2009)**. Les gènes régulés par cette voie sont impliqués dans un grand nombre de fonctions physiologiques essentielles telles que le métabolisme **(Moore et al., 2010)**, la

croissance **(Shenoy and Blelloch, 2014)**, la différenciation cellulaire **(Ivey and Srivastava)**, le développement des plantes **(Chen, 2008)** etc.

II.3. La voie des piRNA

Les piRNAs (*PIWI Interacting RNA*) sont des petits ARN non codants de 26 à 31 nucléotides. Ils s'associent avec les protéines de type PIWI pour former des complexes ribonucléoprotéiques impliqués notamment dans la répression des retrotransposons au sein de la lignée germinale. La biogenèse des piRNAs est complètement différente de celle des siRNAs et des miRNAs, indépendante de l'action de l'enzyme Dicer et les protéines PIWI elles-mêmes sont directement y impliquées **(Bushati and Cohen, 2007; Luteijn et al., 2012)**.

III. L'ARNi un outil de la génétique inverse

L'ARNi, en particulier la voie des siRNAs, est devenu une technique couramment utilisée pour réaliser l'analyse fonctionnelle des gènes. Il s'agit d'introduire de façon exogène dans un organisme (*in vivo*), dans des cellules primaires (*ex vivo*) ou encore dans des lignées cellulaires (*in vitro*), de l'ARN ou de l'ADN capable de produire des siRNAs dirigés contre un ou plusieurs gènes à inactiver. Les progrès techniques réalisés en matière d'introduction des acides nucléiques dans des cellules ou des organismes entiers permettent aujourd'hui de diminuer l'expression de n'importe quel gène d'une façon à la fois rapide, spécifique et efficace. Cette méthode appelée Knockdown permet donc d'inactiver un gène mais de manière plus subtile que les méthodes Knockout et Knock-in utilisées dans les approches « *gene trap* » et « *gene targeting* » décrites plus haut. Actuellement, l'outil ARNi est largement utilisé dans les laboratoires académiques dont les retombées dans le domaine de la santé sont en pleine expansion **(Gavrilov and Saltzman, 2012)**.

L'efficacité du knockdown dépend en grande partie d'une part de la séquence du siRNA et d'autre part des conditions d'introduction des acides nucléiques ainsi que de leurs stabilités dans la cellule ou dans l'organisme. Aussi bien des chercheurs que des industriels ont réussi à mettre au point différentes stratégies veillant à optimiser ces étapes selon le type cellulaire ou le degré de complexité de l'organisme.

III.1. Design des siRNAs

Les scientifiques ont commencé l'analyse fonctionnelle des gènes par inactivation en introduisant en premier lieu de l'ARNdb de grande taille dans les cellules ou les organismes entiers. Cette méthode d'introduction a montré un grand succès pour investiguer la fonction des gènes chez plusieurs organismes eucaryotes **(Fire et al., 1998; Kennerdell and Carthew, 1998; Ngô et al., 1998; Ruiz et al., 1998; Oelgeschlager et al., 2000; Svoboda et al., 2000; Zamore et al., 2000)**. Cependant jusqu'en 2001, il était impossible d'utiliser cette méthode dans les cellules de mammifères. Les mammifères ont développé une réponse antivirale particulière. En effet, la présence d'ARNdb de grande taille (supérieur à 30 nts) induit l'activation de la voie interféron qui aboutit à la dégradation de tous les ARNs cellulaires, quelle que soit leur séquence, conduisant à la mort de la cellule infectée. Ce n'est qu'en 2001 que l'équipe de Tuschl a eu l'idée d'introduire directement des siRNAs, nettement plus petits que l'ARNdb, dans les cellules de mammifères. Cette manipulation a provoqué l'inactivation du gène étudié sans déclencher la réponse interféron non spécifique **(Elbashir et al., 2001; Stark et al., 2003)**. Pour cette raison, des tentatives de production des siRNAs par digestion enzymatique de l'ARNdb transcrit *in vitro* ont été réalisées; ces siRNAs, appelés esiRNAs pour « *endoribonuclease-prepared siRNA* », sont générés par l'action enzymatique de l'endoribonucléase RNase III d'*E. coli* **(Yang et al., 2002)** ou du Dicer recombinant humain **(Kawasaki et al., 2003)**. Actuellement, les siRNAs sont plutôt désignés et produits par des industriels avec beaucoup plus d'efficacité, de stabilité et sans toxicité.

Le design des siRNAs synthétiques est un facteur crucial pour déterminer leur efficacité à cibler l'ARNm. Chaque brin du siRNA est capable de se lier au complexe RISC, uniquement le brin complémentaire de l'ARNm est capable de diriger ce complexe jusqu'à sa cible pour entraîner sa dégradation **(Stein, 2001)**. Le brin sens quant à lui peut provoquer une diminution de l'efficacité de l'inactivation si c'est lui qui est majoritairement pris en charge par le complexe RISC. De plus, tous les siRNAs potentiels dirigés contre un ARNm cible donné n'ont pas la même efficacité d'inactivation **(Holen et al., 2002)**. L'inactivation non spécifique, appelé « *off target effect* » se produit lorsque la région essentielle du siRNA « *seed region* » est identique à celle d'un autre ARNm qui n'est pas destiné à être ciblé **(Jackson et al., 2006)**. Pour cette raison, les fabricants des siRNA ont développé

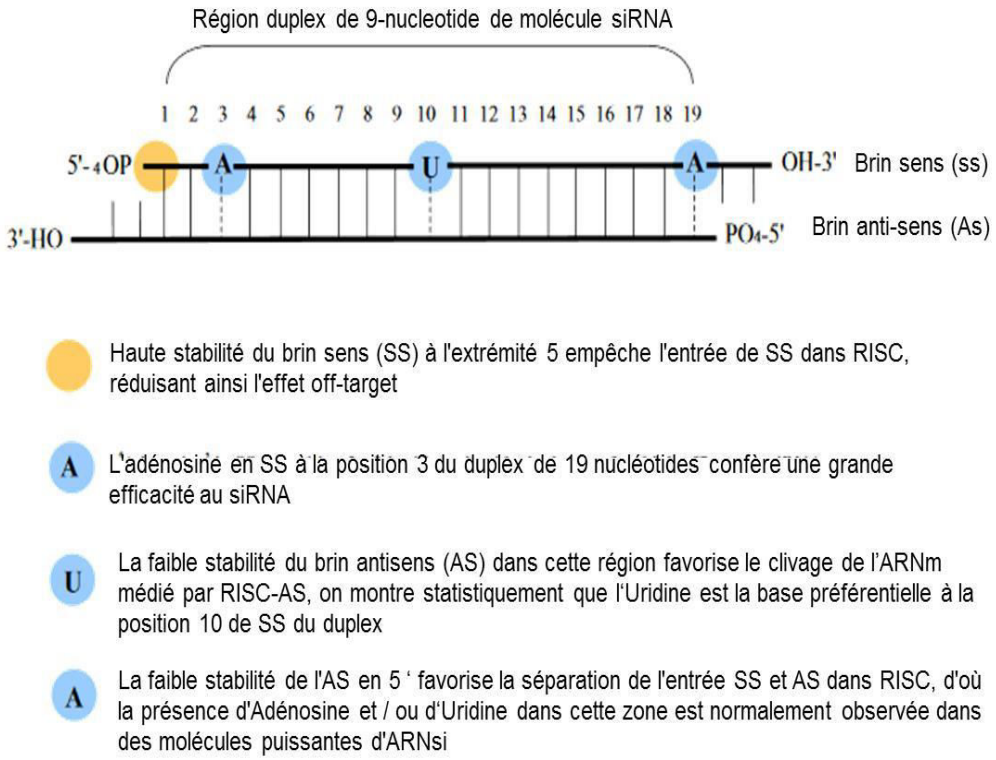


Figure 14 : Les conditions de séquence prises en compte pour le design des siRNA (Leung and Whittaker, 2005).

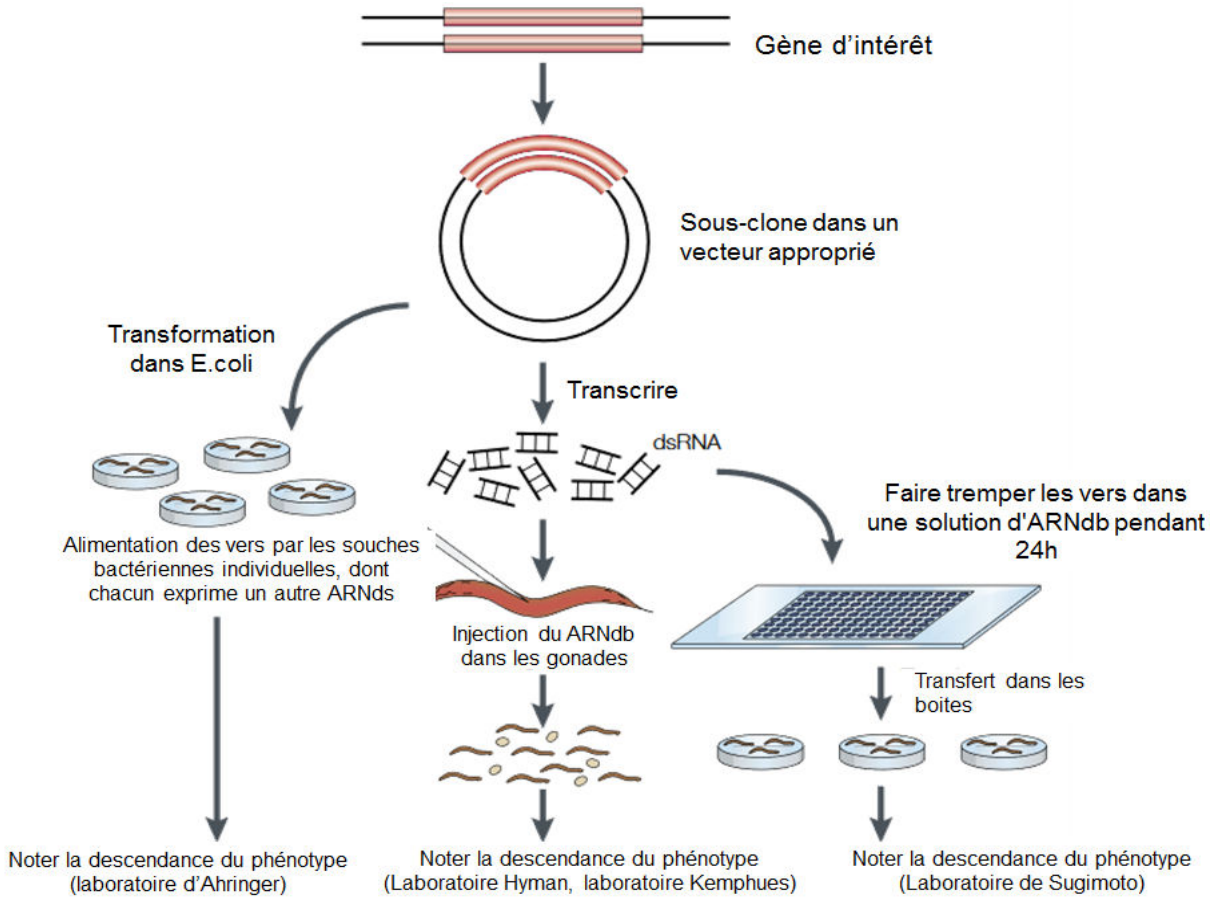


Figure 15 : Les trois modes de délivrance de l'ARNdb dans le nématode *C. elegans* : Feeding, injection et soaking (Kim, 2001).

plusieurs algorithmes permettant de choisir les séquences des siRNAs en fonction de plusieurs critères connus pour améliorer leurs potentiels d'inactivation :

(i) l'utilisation de duplex de 21 nts contenant un domaine central double brin de 19 nts et 2 nts sortants à l'extrémité 3' de chaque brin **(Elbashir et al., 2001)** l'utilisation de 2'-deoxythymidines pour ces deux nts sortants semble avoir une action protectrice contre les exonucléases.

(ii) la stabilité intrinsèque des duplexes joue un rôle essentiel dans le mécanisme d'inactivation ; en effet, pour empêcher l'incorporation du brin sens dans le complexe RISC et éviter l'effet « *off target* », il est nécessaire que cette stabilité soit plus faible du coté 5' que du côté 3' du brin antisens **(Hutvagner, 2001)** **(Figure 14)**. D'autre part, une faible stabilité autour de la région centrale permet de faciliter le clivage de l'ARNm par l'activité slicer du complexe RISC activé **(Mittal, 2004)**.

(iii) une base adénosine en position 3 et 19 du brin sens semble être requise pour obtenir une grande efficacité d'extinction et l'absence de base C ou G en position 19 du brin sens favorise la sélection du brin antisens par le complexe RISC **(Mittal, 2004)**.

Autant de critères pour pallier aux contraintes structurales, thermodynamiques et/ou d'hybridation croisée qui diminuent l'efficacité des siRNAs. Ainsi, après l'étape de design, les industriels choisissent 3 à 4 siRNA les plus intéressants pour les synthétiser afin de les valider et les sélectionner en fonction de leurs effets.

III.2. Les différents modes de délivrance des siRNAs

Plusieurs types de modes de délivrance sont apparus. Quel que soit le mode, l'ARN doit se trouver sous forme de molécules d'ARNdb (de petite ou de grande taille) dans la cellule pour permettre l'initiation du mécanisme, la production des siRNAs et la dégradation de l'ARNm cible.

(a) Soaking, feeding et bathing : le « *soaking* » est une méthode d'immersion concerne uniquement *C. elegans* ; les larves sont immergés dans une solution concentrée d'ARNdb synthétisés *in vitro* et identique au gène cible. L'effet de l'inactivation est ensuite analysé sur la descendance **(Tabara et al., 1998; Maeda et al., 2001)** **(Figure 15)**. Le « *feeding* », originellement développée chez *C. elegans* **(Timmons and Fire, 1998; Hammell and Hannon, 2012)**, est utilisée également chez les ciliés comme la Paramécie **(Galvani and Sperling, 2002)**. Elle consiste à alimenter *C. elegans* ou la paramécie par des

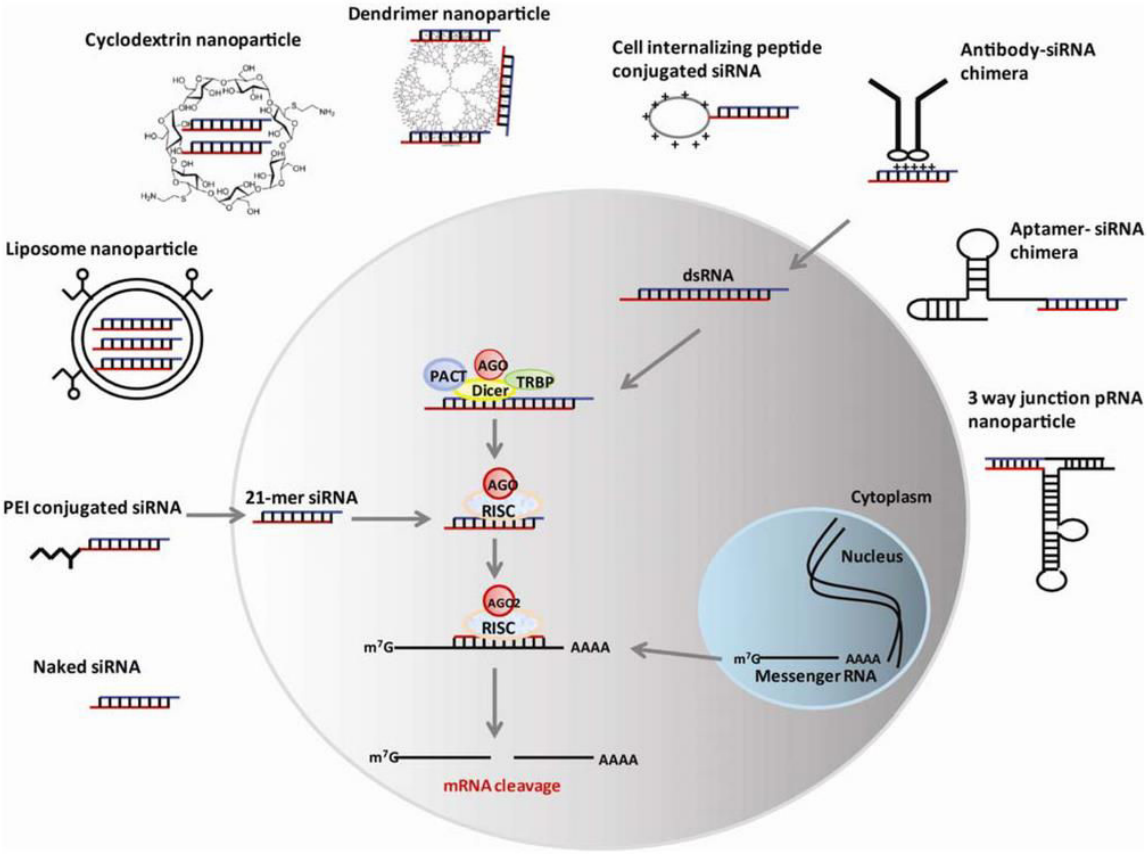


Figure 16 : Les différents modes de délivrance des siRNAs en utilisant les nanoparticules.
(Adapté de http://igtrcn.org/rnai-based-therapeutics-delivery-is-key/).

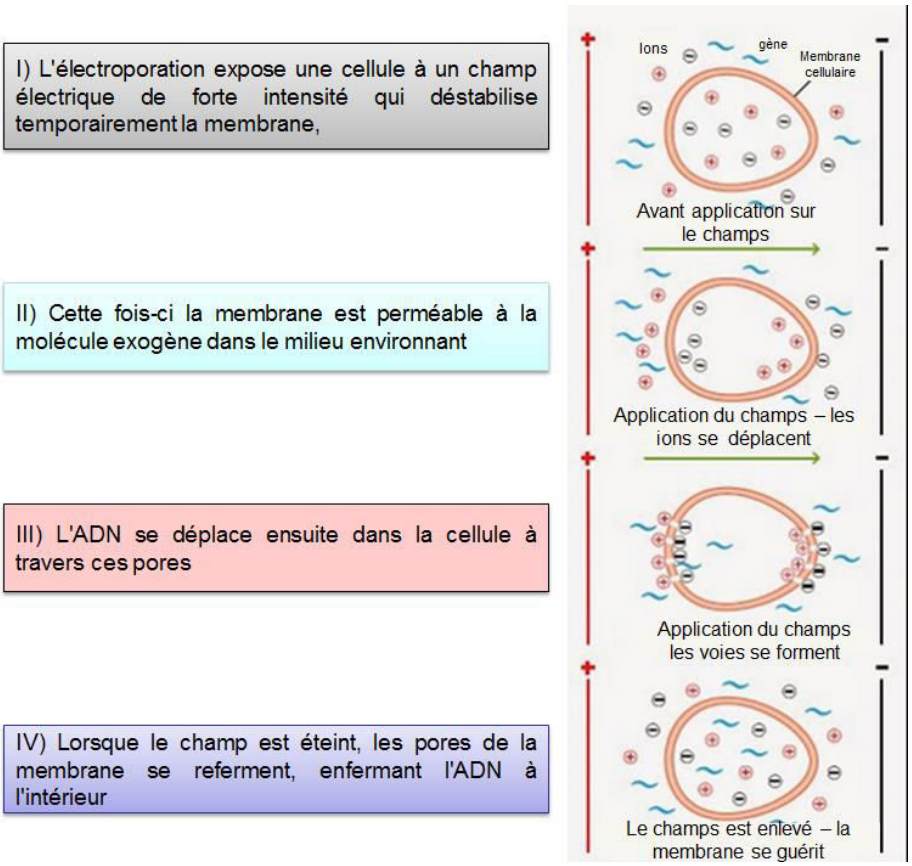


Figure 17 : Le principe de l'électroporation des cellules.
(Adapté de http://www.bio-rad.com/en-gf/applications-technologies/instrument-based-transfection-methods).

bactéries *E. Coli* modifiées produisant de l'ARNdb contenant des séquences du gène cible. Dans les deux méthodes, une fois dans le cytoplasme, l'ARNdb déclenchera le mécanisme d'inactivation par la production des siRNAs qui vont détruire l'ARNm du gène en question. *Le feeding* a montré une remarquable efficacité en particulier pour des analyses fonctionnelles à grande échelle chez *C. elegans* **(Figure 15)**. Comme *le « soaking »*, la technique « *bathing* » consiste à introduire des molécules d'ARNdb par simple contact avec les cellules en culture de la drosophile **(Caplen et al., 2000; Clemens et al., 2000)**.

(b) La transfection : on appelle transfection le processus de transfert d'acide nucléique exogène dans les cellules eucaryotes en utilisant un agent de transfection ou l'électroporation. La transfection par un transfectant concerne la délivrance des siRNAs synthétiques et les constructions plasmidiques contenant un *small hairpin RNA* (shRNA). Un shRNA est une duplication en tandem avec répétition inverse de 50 à 70 nts dont la transcription conduit à la production d'un ARN pseudo double brin sous forme d'épingle à cheveux qui sera clivé par Dicer en siRNA **(Brummelkamp et al., 2002b; McCaffrey et al., 2002; McManus et al., 2002; Xia et al., 2002; Yu et al., 2002)**. L'agent de transfection sert à enrobés l'ADN ou l'ARN dans des nanoparticules **(Gavrilov and Saltzman, 2012; Vorobyeva et al., 2015)** **(Figure 16)** qui protègeront et faciliteront leur passage à travers les membranes cellulaires **(Li and Szoka, 2007)**. L'électroporation, concerne uniquement les constructions plasmidiques, consiste en l'application de fortes impulsions électriques qui augmentent la perméabilité membranaire et permettent de faire pénétrer le plasmide à l'intérieur de la cellule **(Potter and Heller, 2003)** **(Figure 17)**. L'électroporation permet la transfection de cellules primaires ou en suspension et qui rencontrent des difficultés à être transfectées par les agents de transfection. Cette technique peut conduire à l'établissement des clones stables si la construction plasmidique s'intègre dans le génome de la cellule sous l'effet d'une pression sélective assurée par un marqueur de sélection. Cependant, l'électroporation engendre un taux de mortalité cellulaire assez fort. Contrairement aux shRNAs, l'introduction des siRNAs synthétiques ne peut générer que des clones transitoires.

(c) La transduction : la transduction est un processus du transfert d'ADN entre deux cellules *via* des particules virales. C'est particulièrement les rétrovirus et les lentivirus qui sont utilisés pour délivrer des shRNAs dans les cellules eucaryotes ou dans un organisme

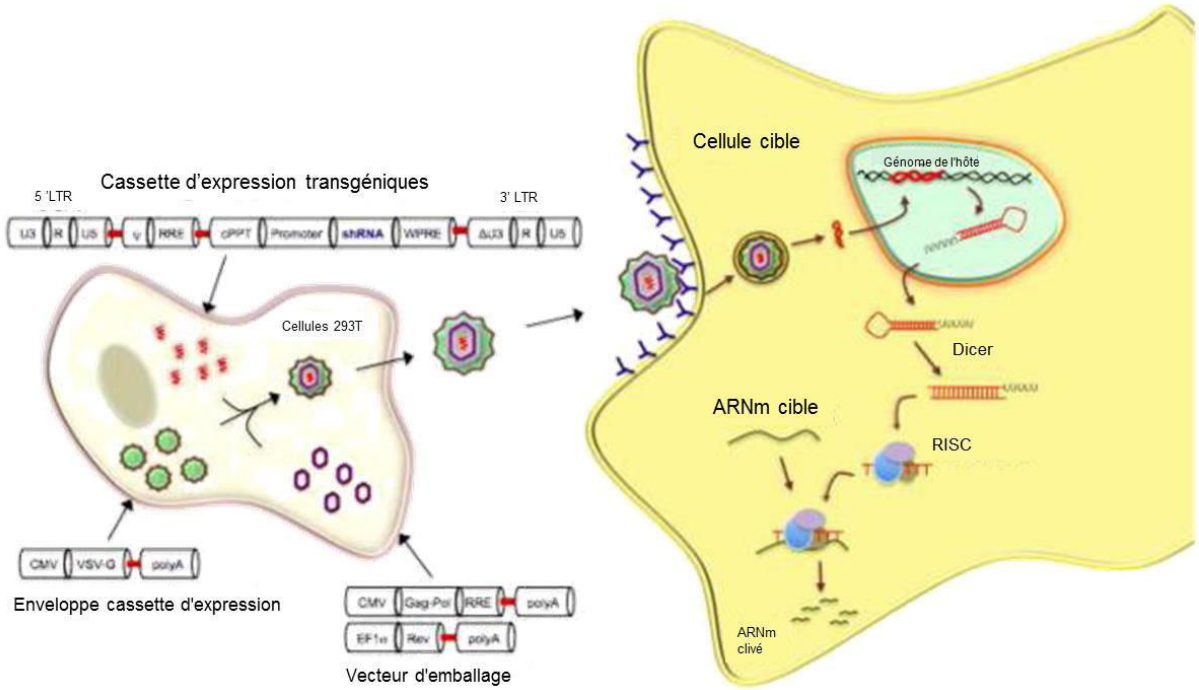


Figure 18 : La transduction des shRNAs via des particules virales.

Des cellules hôtes appropriées (telles que des cellules 293T) sont transfectées avec un mélange de plasmides consistant en i) une cassette d'expression shRNA, ii) une cassette d'emballage et iii) une cassette d'expression d'enveloppe virale hétérologue (VSV-G). Le lentivirus généré est ensuite utilisé pour transduire le type de cellule désiré pour l'expression de shRNA. Parce que seul le vecteur contenant la cassette d'expression shRNA (dépourvue des gènes structurels viraux) s'intègre dans le génome de la cellule hôte dans les cellules transduites, l'ARNm est continuellement exprimé mais le virus infectieux n'est pas produit **(Manjunath et al., 2009)**.

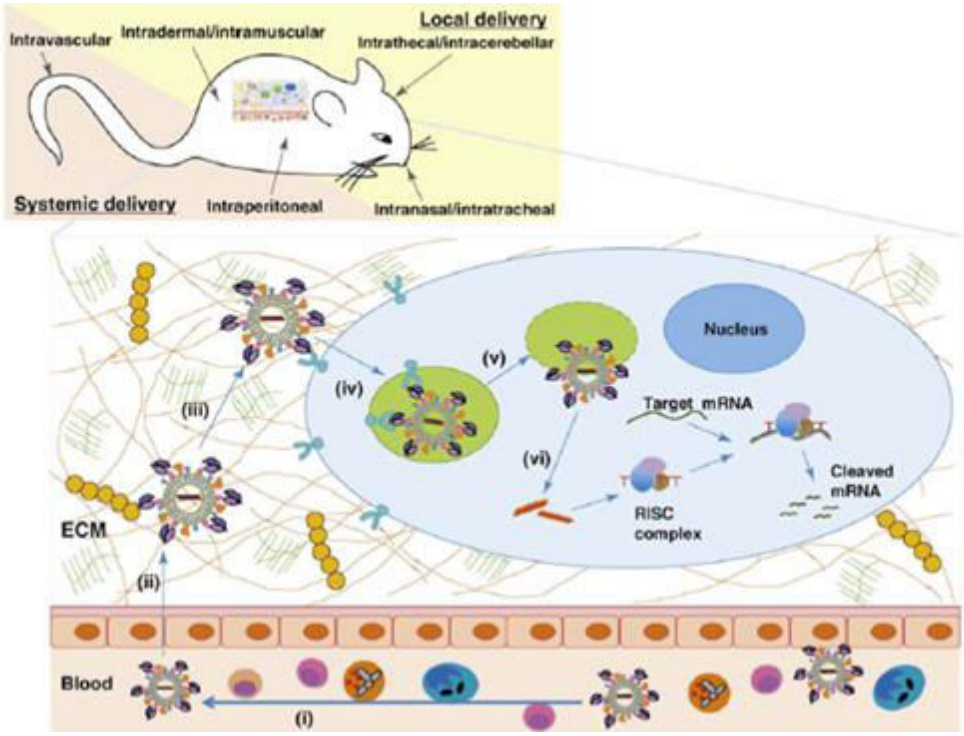


Figure 19 : Les différentes voies d'injection chez la souris.

Des particules virales modifiées contenant un shRNA ou des siRNAs synthétiques sont administrés soit par injection systémique intra péritonéale ou intra nasale, soit par injection locale intra vasculaire, intramusculaire ou intracérébral **(Kim et al., 2009)**.

entier comme la souris **(Barton and Medzhitov, 2002; Brummelkamp et al., 2002a; Devroe and Silver, 2002)**. Cette méthode à médiation virale commence d'abord par la construction de deux types de vecteurs : (i) les plasmides d'empaquetage contenant les gènes nécessaires à l'encapsidation du virus, (ii) le plasmide de transfert contenant le shRNA d'intérêt, le signal d'encapsidation et les deux LTR du virus. Ces plasmides seront par la suite introduits par co-transfection transitoire dans des cellules hôtes appropriées, telles que les cellules 293T, pour la production des virions modifiées. Le surnageant produit est récolté pour être déposé sur les cellules cibles afin de les transduire et permettre ainsi l'intégration d'une façon stable de l'ADN du shRNA dans le génome de la cellule et la production continuelle du shRNA qui va sous exprimer le gène cible **(Figure 18)** **(Manjunath et al., 2009)**.

(d) L'injection : c'est une méthode qui s'utilise particulièrement chez *C. elegans*, la drosophile ou la souris. Selon cette méthode, l'ARNdb synthétisé *in vitro* est injecté dans l'intestin (semble être le plus efficace) des jeunes vers hermaphrodites **(Fire et al., 1998)**; la diffusion de l'ARNdb injecté dans la plupart des cellules du ver, y compris la lignée germinale, permet la transmission de l'ARNdb à la descendance sur laquelle le phénotype de l'inactivation est détecté et étudié **(Grishok, 2001; Grishok et al., 2005)**. Chez la drosophile, l'ARNdb est délivré par micro-injection manuelle dans les embryons et les phénotypes associés à l'inactivation seront étudiés chez l'adulte **(Kennerdell and Carthew, 1998)**. Contrairement à *C. elegans* et la drosophile, c'est des particules virales modifiées contenant un shRNA ou des siRNAs synthétiques qui seront administrés par injection systémique (intra péritonéale ou intra nasale) ou locale (intra vasculaire, intramusculaire ou intracérébral etc.) dans la souris **(Figure 19)** **(Kim et al., 2009)**. Des molécules protectrices, non toxiques et non immunogènes de nature lipidique ou peptidique sont associées aux duplexes siRNAs pour faciliter leur délivrance dans les tissus et les cellules de l'animal tout en évitant le système réticuloendothélial. Ces molécules sont de plus en plus employées aussi bien dans la recherche académique que clinique **(Li and Szoka, 2007; Kim et al., 2009; Gavrilov and Saltzman, 2012; Vorobyeva et al., 2015)**.

III.3. Les contrôles et la quantification de l'effet des siRNAs

L'utilisation de contrôles constitue un outil d'analyse important permettant la validation de la méthode d'inactivation. C'est une étape critique pour le succès de l'application qui veille à établir la relation de cause à effet entre l'inactivation et le siRNA utilisé. Trois types de contrôles sont souvent réalisés : (i) le contrôle négatif permet de vérifier la spécificité de l'effet des siRNAs en effectuant une expérience parallèle avec un contrôle qui ne cible aucun transcrit cellulaire comme un siRNA *scramble* qui correspond aux mêmes bases que le siRNA mais dans un ordre aléatoire, ou un siRNA dirigé contre une séquence inexistante dans la cellule ou l'organisme traité ; (ii) les contrôles positifs sont utilisés afin d'optimiser l'efficacité de la transfection et permettent de contrôler la délivrance du siRNA. Il s'agit en fait de cibler des protéines résidentes hautement exprimées ayant un effet biologique facilement mesurable ; (iii) le contrôle de la transfection qui consiste à effectuer une expérience de transfection contrôle sans siRNA.

Afin de quantifier l'inactivation par l'ARNi, il suffit de vérifier le niveau d'ARNm du gène cible par RT-PCR ou northern-blot, ou le niveau d'expression de la protéine par western-blot. De plus, pour vérifier l'efficacité de la transfection on peut utiliser un siRNA marqué ou un gène rapporteur puis suivre les cellules inactivées par cytométrie en flux ou microscopie à fluorescence pour calculer le taux de la transfection.

IV. Les cribles ARNi à grande échelle

Au cours de ces dernières années, l'ARNi est devenu un outil de choix pour réaliser des cribles à haut débit afin de caractériser simultanément la fonction de plusieurs gènes aussi bien dans des cellules en cultures que dans des organismes modèles. En effet, l'association de la technologie ARNi avec les données de séquences d'ADNc et/ou du génome entier a permis de construire à l'échelle génomique des banques produisant de petits ou de grands ARNs capables d'investiguer le rôle des gènes à haut débit. Depuis quelques années, l'effet le plus important à cet égard est sans doute la capacité d'effectuer ce type de cribles à grande échelle sur des systèmes mammifères **(Echeverri and Perrimon, 2006)**. Les cribles ARNi dans les cellules de mammifères ont déjà conduit à un grand nombre de résultats avec d'importantes retombées dans le domaine de la recherche biomédicale,

comme l'identification de nouveaux oncogènes ou de cibles potentielles pour le développement de traitements thérapeutiques **(Kim et al., 2007; Lee et al., 2014)**. Cependant, l'application des cribles ARNi à grande échelle *in vivo* semble être très compliquée à réaliser chez les mammifères, telle que la souris, à cause du coût et le temps qu'un crible nécessite **(Sharma and Rao, 2009)**.

IV.1 Le nématode *C. elegans*

Deux banques de bactéries « *feeding* » sont disponibles actuellement chez *C. elegans* : « *Ahringer* » et « *ORFeome* », couvrent respectivement 86% et 55% des 20000 gènes codants **(Fraser et al., 2000; Gonczy et al., 2000; Reboul et al., 2003; Lamesch et al., 2004)** et fournissent ensemble près de 94% de couverture génique **(Kim et al., 2005)**. Les deux banques contiennent des bactéries *E. Coli* hébergeant des plasmides recombinants dont les inserts sont des exons ou des ORFs (*open reading frame*) entiers sous forme d'ADNc **(Maeda et al., 2001; Kamath and Ahringer, 2003)**. Les inserts de la banque « *Ahringer* », vérifiées par séquençage, sont disponibles en ligne *via* la base de données *CelRNAi* et constituent une ressource inestimable pour la communauté scientifique travaillant sur *C. elegans* (http://biocompute.bmi.ac.cn/CelRNAi/index.php).

Les premiers cribles ARNi à l'échelle génomique ont été réalisés sur la souche sauvage de *C. elegans* dans le but d'identifier des gènes impliqués dans le développement ou la morphologie ; près de 1700 gènes ont été identifiés dont les deux tiers étaient de fonction inconnue **(Fraser et al., 2000; Kamath et al., 2003)**. Dans le but d'augmenter le nombre de gènes associés à ces phénotypes spécifiques de développement et/ou de morphologie, ce type de cribles a été répété en utilisant des souches hypersensibles à l'ARNi telles que la souche rrf-3 **(Simmer et al., 2003)**. Il est à souligner que les cribles synthétiques, utilisant un fond génétique sensibilisé comme dans le cas du crible ENU (décrit dans le premier chapitre), sont souvent associés à un taux assez important de faux positifs **(Parry et al., 2007)**.

Pour de nombreux phénotypes, une simple inspection visuelle des vers traités par ARNi n'est pas suffisante pour révéler le défaut du Knock-down. Pour cette raison, de nombreux cribles ARNi à grande échelle ont été réalisés grâce des analyses microscopiques très poussées après l'utilisation d'un marquage spécifique des cellules **(Lettre et al., 2004)**, d'un

gène rapporteur **(Sonnichsen et al., 2005)** ou des protéines fluorescentes. Cette dernière approche a été appliquée avec succès à une gamme diversifiée de questions biologiques comme le développement musculaire **(Meissner et al., 2009)**. Ce crible réalisé par Meissner et al., est particulièrement intéressant puisque il a été conçus pour identifier des gènes impliqués dans l'assemblage et la maintenance du sarcomère (l'unité de base des myofibrilles des muscles striés) des muscles de *C. elegans*. Le crible a été effectué en utilisant une banque « *feeding* » sur près de 3600 gènes identifiés par des approches transcriptomiques comme étant exprimés dans le muscle. Le « *feeding* » a été effectué sur des vers exprimant la myosine marquée à la GFP, ce qui a permis d'analyser l'organisation de myofilaments en microscopie à fluorescence. 122 gènes ont été identifiés comme étant impliqués dans l'organisation correcte de myofilaments, dont 108 n'avaient aucun rôle précédemment caractérisé dans la structure ou la fonction musculaire **(Meissner et al., 2009)**.

La technique « *feeding* » montre généralement une efficacité d'inactivation relativement moindre que celle de l'injection ou du « *soaking* », mais elle reste la technique la plus adaptée aux cribles à haut débit en termes de rapidité et de facilité d'utilisation des souches hypersensibles à l'ARNi. Il est à souligner que vu la quantité de travail que nécessite la mise en place des cribles ARNi à grande échelle ainsi que l'annotation phénotypique chez *C. elegans*, plusieurs tentatives d'automatisation basées essentiellement sur l'imagerie et la cytométrie en flux ont pu avoir lieu **(Dimitriadi et al., 2010; Gosai et al., 2010)**.

IV.2. La Drosophile

L'utilisation de l'ARNi comme outil chez la Drosophile a permis d'effectuer des cribles à l'échelle du génome entier non seulement dans des cellules en culture *in vitro* mais aussi dans des tissus spécifiques *in vivo* ; ceci grâce à l'utilisation de banques d'ARNi conçues et générées d'une façon systématique **(Venken and Bellen, 2014)**. Pour les cellules en culture, c'est souvent des molécules d'ARNdb de 50 à 500 pb synthétisées *in vitro* qui sont introduit par transfection ou par « *bathing* » **(Ramadan et al., 2007; Steinbrink and Boutros, 2008)**. Par contre, les études *in vivo* sont réalisées par transgénèse. Les souches transgéniques ont été initialement construites par l'introduction, *via* l'élément transposable P, de transgènes capables de produire de longs ARNdb sous forme d'épingle à cheveux **(Dietzl et al., 2007)**. Actuellement, c'est des shRNAs introduits d'une façon dirigée dans les lignées transgéniques en utilisant

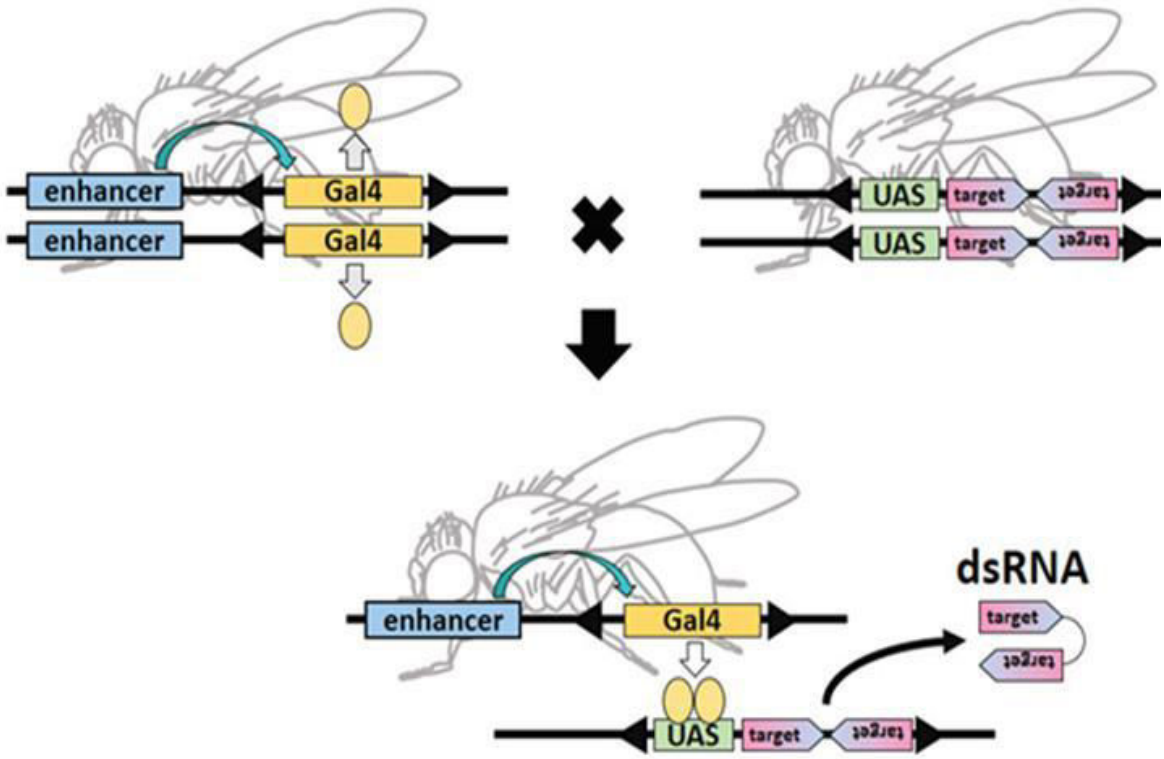


Figure 20 : Expression inductible de l'ARNdb par le système UAS-GAL4.
La première souche exprimant le gène gal4 est obtenue par injection chez l'embryon d'un plasmide contenant le gène gal4 sous le contrôle d'un promoteur tissu spécifique. La protéine GAL4 exprimée ne devient efficace que si elle se fixe sur une séquence UAS (Upstream Activation Sequence), région régulatrice de l'ADN qui fixe les facteurs de transcription. Selon la même méthode de transgénèse, la deuxième souche contenant la séquence UAS est créée, le transgène comporte la séquence d'un shRNA en aval de la séquence UAS. Si on croise ces deux souches transgéniques, les descendants vont avoir les deux transgènes mais le shRNA sera uniquement produit dans les cellules ou les tissus exprimant GAL4 **(Yamamoto-Hino and Goto, 2013)**.

des vecteurs optimisés du système UAS-GAL4 résultant en une inactivation tissu spécifique de n'importe quel tissu même de la lignée germinale **(Ni et al., 2011)**.

Le système UAS-GAL4 permet de diriger spécifiquement la production d'un shRNA dans un territoire donné ou à un moment donné. Le gène gal4, présent chez la levure, code pour un facteur de transcription, GAL4, dont on a montré qu'il est fonctionnel lorsqu'il est exprimé chez la Drosophile. Deux souches transgéniques sont nécessaires pour réaliser l'inactivation. La première souche exprimant le gène gal4 est obtenue par injection chez l'embryon dans la zone des cellules germinales d'un plasmide contenant le gène gal4 sous le contrôle d'un promoteur tissu spécifique. La protéine GAL4 exprimée ne devient efficace que si elle se fixe sur une séquence UAS (*Upstream Activation Sequence*), région régulatrice de l'ADN qui fixe les facteurs de transcription. Selon la même méthode de transgénèse, la deuxième souche contenant la séquence UAS est créée, le transgène comporte la séquence d'un shRNA en aval de la séquence UAS. Si on croise ces deux souches transgéniques, les descendants vont avoir les deux transgènes mais le shRNA sera uniquement produit dans les cellules ou les tissus exprimant GAL4 **(Figure 20)** **(Brand and Perrimon, 1993)**.

Plus d'une centaine de cribles d'ARNi à haut débit sur des cellules en culture ont été réalisés depuis que l'approche a été mise au point il y a plus de dix ans **(Mohr et al., 2010)**. On distingue deux catégories de cribles (i) ceux appelés « *pooled screens* » dans lesquels l'ARNdb est introduit d'une façon aléatoire dans des cellules en masse, les cellules montrant des phénotypes après le traitement sont sélectionnées et les ARNdb responsables de ces phénotypes sont ensuite identifiés par séquençage ou par PCR. (ii) Ceux appelés « *arrayed screens* », pour lesquels les ARNdb sont introduits individuellement dans les puits d'une plaque 96 ou 384 qui constituent des banques d'ARNdb, les cellules inactivées sont ensuite identifiées et analysées. Bien évidemment, le type de cellule idéal pour réaliser un crible ARNi serait celui issu du tissu d'intérêt ; toute fois les cellules S2 restent de loin les cellules les plus utilisées pour les cribles ARNi à base de cellule chez la drosophile **(Mohr et al., 2010)**. Les collections des lignées construites pour l'ARNi *in vivo* à l'échelle du génome entier sont disponibles à partir du « *Transgenic RNAi project* » dans les centres de stockages : « *Bloomington Drosophila Stock Centre* », « *Vienna Drosophila RNAi Center* » et « *National Institute of Genetics* ». Les ressources concernant les lignées et les phénotypes sont

répertoriées en ligne dans « *RNAi Stock Validation and Phenotype* » (www.flyrnai.org/rsvp) et dans la base de donnés *FlyBase* **(Drysdale, 2008)**.

La capacité d'induire l'ARNi d'une façon spécifique dans des tissus et des stades de développement spécifiques ouvre la porte à la réalisation des cribles non seulement sur des embryons et des larves mais aussi sur des animaux adultes et même lorsque le knockdown à un stade précoce est associé à une létalité. En outre, l'utilisation des shRNAs plutôt que de longs ARNdb a permis d'obtenir un niveau important de sous expression des gènes inactivés aussi bien dans les cellules somatiques que germinales **(Ni et al., 2011)**. Depuis quelques années, les cribles ARNi à haut débit et *in vivo* chez la Drosophile sont focalisés sur l'identification et l'étude des gènes liés à des maladies humaines comme la perception de la douleur, l'obésité, la fonction cardiaque, les infections bactériennes, l'auto renouvellement des cellules souches neurales et les maladies neurologiques **(Cronin et al., 2009; Neely et al., 2010; Pospisilik et al., 2010; Fernandes and Rao, 2011; Neumuller et al., 2011)**.

Il ne faut pas oublier que la drosophile a été largement utilisée comme modèle pour l'identification par ARNi des gènes impliqués dans les processus biologiques fondamentaux. Ces cribles ont pu identifier des gènes impliqués dans la croissance **(Boutros et al., 2004)**, la régulation de la viabilité cellulaire **(Boutros et al., 2004)**, la morphologie des cellules **(Kiger et al., 2003)**, la cytocinèse **(Eggert et al., 2004)**, les voies de signalisation comme Hedgehog **(Lum et al., 2003; Aikin et al., 2012)** et Notch **(Mummery-Widmer et al., 2009)**, la glycosylation **(Yamamoto-Hino et al., 2010)**, la suppression des tumeurs **(Avet-Rochex et al., 2010)** et le développement musculaire **(Bai et al., 2008; Schnorrer et al., 2010; Perkins et al., 2014)**. L'étude sur le muscle de Schnorrer et al., est particulièrement intéressante de par son approche *in vivo* et le nombre de gènes analysés. Les auteurs ont en effet étudié les effets de la mise sous silence de façon stable et muscle spécifique de 10461 gènes dans 17759 lignées de drosophile construites selon le système UAS-GAL4. Ils ont ainsi pu établir une liste de 2785 gènes jouant un rôle dans le développement musculaire, classés en sept catégories phénotypiques **(Schnorrer et al., 2010)**.

IV.3. Les Mammifères

L'ARNi a révolutionné l'analyse fonctionnelle des gènes dans les cellules de mammifères. Les cribles à l'échelle du génome entier basés sur l'ARNi peuvent être réalisés par transfection transitoire des siRNA ou par un knockdown stable avec des shRNA introduits

dans les cellules *via* une construction plasmidique ou une infection par particules virales modifiées **(Hu and Luo, 2012)**.

Comme mentionné précédemment, l'ARNdb de grande taille déclenche souvent une réponse immunitaire et une activation de la voie d'interféron dans les cellules de mammifères, entrainant ainsi une régulation négative et non spécifique de l'expression protéique **(Gantier and Williams, 2007)**. Par conséquent, les siRNAs sont devenus le principal outil pour inactiver les gènes par ARNi **(Elbashir et al., 2001)**. De plus, la réalisation des cribles ARNi à grande échelle *in vivo* semble être très fastidieux chez les mammifères, à cause du coût qui peut générer et le temps qui nécessite **(Sharma and Rao, 2009)**. Récemment, une variante intéressante de ces cribles *in vivo* s'applique avec succès sur le modèle murin ; ce type d'étude commence par un criblage *in vitro* par shRNA sur des cellules en culture tumorales qui seront transplantées par xénogreffe dans l'organe cible de la souris afin de voir l'effet des gènes inactivés sur la formation des métastases **(Zender et al., 2008; Bric et al., 2009; Zuber et al., 2011; Livshits and Lowe, 2013; Wuestefeld et al., 2013; Rudalska et al., 2014)**. Les cribles ARNi chez les mammifères ont commencé par l'identification des acteurs de l'apoptose dans les cellules HeLa en utilisant une banque de siRNAs synthétiques ciblant 510 gènes humains **(Aza-Blanc et al.)**. La chimie combinatoire, utilisée pour générer les siRNAs synthétiques, a facilité la réalisation d'une série de cribles à haut débit ou ciblés dans les cellules de mammifères en culture. Ainsi, de nombreux gènes impliqués dans divers processus biologiques ont été identifiés par ces cribles **(Whitehurst et al., 2007; Brass et al., 2008; Krishnan et al., 2008; Brass et al., 2009; Li, 2009; Chia, 2010; Houzet, 2011)**. Ces banques de siRNA synthétiques sont disponibles auprès des industriels comme Ambion, Dharmacon, Qiagen, Sigma-Proligo, etc. Pour limiter le coût de la production chimique des siRNAs dirigés contre le génome entier, aussi bien des scientifiques que des industriels ont pu créer des banques de siRNA préparées par digestion enzymatique (esiRNAs) **(Yang et al., 2002; Kittler et al., 2007; Theis and Buchholz, 2011)**. Toutefois, la transfection des siRNAs synthétiques et des esiRNAs était souvent associée à des effets non spécifiques et/ou toxiques pour la cellule cible. De plus, de nombreuses cellules de mammifères en particulier les cellules primaires sont réfractaires à la transfection par agents chimiques ou impulsions électriques. C'est pour cette raison, les vecteurs viraux hébergeant des constructions shRNAs sont devenus l'outil alternatif vu leur

efficacité pour introduire et produire des siRNA dans les cellules de mammifères **(Couto and High, 2010)**.

De nombreuses banques de shRNA à l'échelle du génome entier ont été construites pour inactiver les gènes de l'homme et de la souris. Les banques shRNA sont plus avantageux que celles de siRNAs parce que (i) elles peuvent être criblées dans des pools, ce qui réduit considérablement le coût du crible ; (ii) elles fournissent un knockdown du gène stable et à long terme et peuvent donc révéler des changements phénotypiques lents dans les cellules inactivées; et (iii) elles peuvent être facilement adaptées pour des cribles *in vivo* dans le modèle murin. Parmi ces banques shRNA, crées par des groupes de recherches ou par des industriels, on trouve : « *the Netherlands Cancer Institute* » **(Berns et al., 2004; Bernards et al., 2006)**, « *the RNAi consortium* » **(Moffat and Sabatini, 2006; Root et al., 2006)**, « *the Hannon-Elledge libraries* » **(Paddison et al., 2004; Silva et al., 2005; Chang et al., 2006)** et « *the System Biosciences* ». Ces banques se diffèrent entre elles au niveau de la taille et de la couverture de la banque ainsi qu'au niveau de la conception de la séquence du shRNA. Une fois introduites dans les cellules hôtes, les constructions shRNA peuvent avoir une expression constitutive ou inductible. Contrairement aux shRNAs constitutifs, ceux à expression inductible fournissent un contrôle temporel et réversible de l'expression du shRNA, ce qui permet l'étude fonctionnelle des gènes essentiels dont le knockdown stable peut conduire à une létalité. En raison de leur efficacité de délivrance élevée, les vecteurs rétroviraux, lentiviraux et adénoviraux sont actuellement les plus utilisés pour introduire les shRNAs dans les cellules cibles. Les vecteurs rétroviraux et lentiviraux sont utilisés essentiellement dans la recherche fondamentale puisqu'ils peuvent fournir une inactivation génique stable grâce à l'intégration de la construction shRNA dans le génome de la cellule hôte **(Cullen and Arndt, 2005)**. En revanche, les vecteurs adénoviraux sont beaucoup plus adaptés pour des expériences *in vivo* ou pour des applications thérapeutiques vu que l'ADN viral modifié ne s'intègre pas dans le génome et n'est pas reproduit pendant les divisions cellulaires **(Grimm et al., 2006)**.

La technologie ARNi a déjà influencé les stratégies de traitement pharmacologique et thérapeutique des maladies humaines. Des essais de traitement pharmacologiques sur de nombreuses maladies comme le cancer, les maladies virales, les agents pathogènes bactériens, l'inflammation, les maladies du système nerveux et d'autres sont en cours

(Seyhan, 2011). Cette technologie permet de mieux comprendre les mécanismes sous-jacents de la maladie ce qui conduit à l'identification de nouveaux facteurs qui peuvent altérer la maladie. Avec l'utilisation des banques ARNi sous différents formats, des cribles fonctionnels et systématiques peuvent être menés dans diverses pathologies pour interroger les fonctions biologiques de gènes, de voies spécifiques, voire du génome entier. L'identification de ces nouveaux facteurs de la réponse cellulaire à la pathologie peut conduire à l'identification de nouvelles cibles pour les médicaments, au développement d'approches de traitement combinatoire ou encore à la mise en place des bio-marqueurs. La thérapeutique par l'ARNi soulève de grands espoirs, approximativement 22 molécules de type siRNA/shRNA ont atteint le test clinique pour le traitement d'au moins 16 maladies **(Lares et al., 2010)**. À l'instar d'autres produits thérapeutiques à base d'ARN, l'efficacité des médicaments par siRNA repose sur l'optimisation de l'administration ciblée, de la stabilité et de la spécificité du siRNA tout en minimisant sa toxicité et/ou sa dégradation hors cible **(Peer and Lieberman, 2011; Burnett and Rossi, 2012)**.

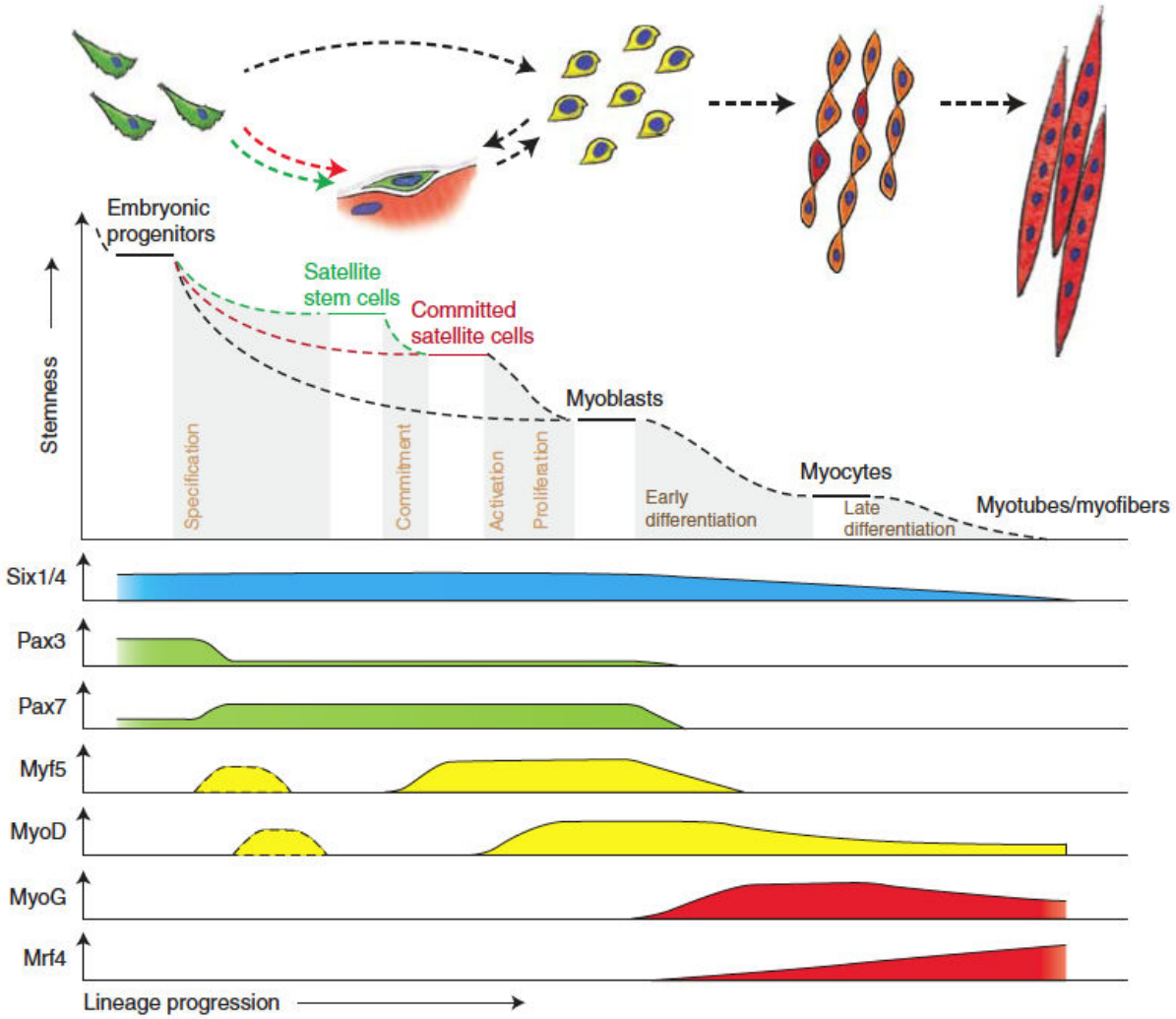


Figure 21 : Hiérarchie des facteurs impliqués dans la myogenèse du muscle squelettique.

Les cellules progénitrices du muscle expriment Pax3 et Six1/4. Pax7 est un marqueur des cellules progénitrices adultes ou cellules satellites. Ces quatre facteurs sont impliqués dans la détermination précoce alors que Myf5 et MyoD sont nécessaires à l'initiation du programme myogénique. La différenciation est réalisée à la fois par la myogénine et MRF4, et conduit à l'expression de la protéine de structure myosine au cours de la maturation des fibres. MRF4, myogenic regulatory factor 4 ; MyoD, myoblast determination protein 1 ; Myf5, myogenic factor 5 ; Pax3, paired box protein 3 ; Six, sine oculis homeobox homolog protein **(Bentzinger et al., 2012)**.

PROJET DE THESE

Chez les mammifères, le muscle squelettique joue un rôle crucial dans le mouvement, la respiration et le métabolisme. Une modification de cette fonction due à une maladie, une blessure ou le vieillissement peut conduire à des altérations de la mobilité et/ou de la durée de vie. La fonction musculaire correcte dépend de myofibres, qui sont des cellules multinucléés formées pendant le développement, la croissance ou la régénération musculaire par un processus appelé « myogenèse ». La myogenèse, processus hautement conservé au cours de l'évolution et aboutissant à la formation des fibres musculaires contractiles, est initiée par la détermination des précurseurs myogéniques embryonnaires en myoblastes qui à leur tour commencent à proliférer au niveau des bourgeons des membres. Ensuite, elles sortent du cycle cellulaire, fusionnent entre elles et se différencient en myotubes (cellules allongées multinucléées). Ces myotubes subiront ensuite une étape de maturation aboutissant à la formation des fibres musculaires contractiles. Des populations de cellules quiescentes retrouvées à la périphérie des fibres musculaires, appelées cellules satellites, sont capables de fusionner avec les fibres déjà formées ou avec les fibres lésées pour régénérer le muscle endommagé **(Bentzinger et al., 2012)** **(Figure 21)**.

La croissance du muscle squelettique embryonnaire est la résultante d'un équilibre entre les processus de prolifération et de différenciation. La prolifération, qui désigne les divisions cellulaires par mitose des myoblastes, est régulée par de nombreux facteurs, notamment les cyclines et les kinases cyclines-dépendantes (Cdks) **(Tintignac et al., 2000)**. Les myoblastes prolifèrent d'une façon intensive une fois que leur migration jusqu'aux bourgeons des membres est terminée. La diminution de la quantité des facteurs de croissance conduit à la sortie du cycle cellulaire des myoblastes et à l'engagement dans la voie de la différenciation. Cette sortie du cycle requiert la diminution d'expression des cyclines et Cdks et l'augmentation d'expression de leurs inhibiteurs **(Reynaud et al., 2000)**.

La différenciation musculaire, processus par lequel les myoblastes se spécialisent et deviennent des myotubes, fait intervenir plusieurs acteurs représentés par :

(i) Des facteurs de croissance : FGF (*Fibroblast Growth Factor*) et TGF (*Transforming Growth Factor*) qui sont impliqués dans la régulation du développement embryonnaire et le maintien de l'homéostasie tissulaire chez l'adulte

(ii) Des facteurs de transcription spécifiques du muscle : les MRF (*myogenic regulatory factor*) qui sont des facteurs de transcription à domaines bHLH : MyoD **(Davis et al., 1987; Tapscott and Weintraub, 1991)**, Myf5 **(Braun et al., 1989)**, myogénine **(Wright et al., 1989)**, et MRF4 **(Rhodes and Konieczny, 1989)** **(Figure21)**. Ils vont activer la transcription des gènes muscles-spécifiques, favorisant la différenciation musculaire de façon spatio-temporelle **(Punch et al., 2009; Londhe and Davie, 2011; Kusakabe and Inoue, 2015)**.

(iii) Des facteurs MEF (*Myocyte Enhancer Factor*) : des transactivateurs portant un domaine MADS (MCM1, Agamous, Deficiens, SRF) qui interagissent avec certains MRFs sur leurs cibles transcriptionnelles **(Molkentin et al., 1995)**.

Une fois sortis du cycle cellulaire, les myoblastes s'alignent, activent l'expression des protéines contractiles comme la myosine et fusionnent en cellules multinucléées, les myotubes **(Figure 21)**. Le contact cellulaire entre les myoblastes est un facteur très important pour cette étape **(Krauss et al., 2005)**. La fusion myoblastique nécessite un grand nombre d'acteurs impliqués dans les remaniements cellulaires, les réarrangements cytosquelettiques et membranaires, et qui interviennent durant la migration, la reconnaissance et l'adhésion cellulaire **(Schejter, 2016)**. Deux types de fusion myoblastique, primaire et secondaire, existent dans le développement musculaire des mammifères et se distinguent par les protéines cellulaires impliquées **(Kim et al., 2015)**. La fusion primaire permet la fusion de deux myoblastes entre eux pour donner naissance à une cellule multinucléée initiale qui se déroule au cours du développement embryonnaire ou de la régénération. La fusion secondaire correspond à l'addition d'un myoblaste avec un myotube déjà formé ; elle intervient durant les mêmes processus que la fusion primaire, et elle est responsable de l'hypertrophie et du maintien de l'intégrité des fibres musculaires tout au long de la vie adulte **(Abmayr and Pavlath, 2012; Schejter, 2016)**. Les myotubes ainsi formés subissent ensuite une étape de maturation en fibres musculaires qui conduit à une spécialisation des cellules vers leur future fonction contractile grâce à l'assemblage du sarcomère porté par la protéine

de structure myosine et à une croissance de la taille des myotubes ou des myofibres, en réponse à des stimuli environnementaux **(Schejter, 2016)**.

Bien que la myogenèse ait été largement décrite, les mécanismes de régulation qui régissent ce processus complexe sont encore mal connus, notamment les réseaux de gènes et les interactions potentielles entre les voies de signalisation impliquées. Ainsi, lors de ma thèse, je me suis attachée à développer un outil basé sur l'ARNi pour identifier de nouveaux gènes impliqués dans la différenciation myogénique. En mettant en place un nouveau protocole *in vitro*, basé sur la lignée myoblastique C2C12, l'utilisation des siRNA et l'analyse quantitative d'images, j'ai pu réaliser un crible fonctionnel sur une centaine de gènes inconnus pour identifier des gènes jouant un rôle durant la myogenèse. Par une analyse phénotypique quantitative, j'ai pu montrer que six gènes parmi les 100 sont impliqués à la fois dans la prolifération et la différenciation des cellules C2C12 et 13 gènes jouant un rôle uniquement dans l'étape de différenciation.


MATERIELS ET
METHODES

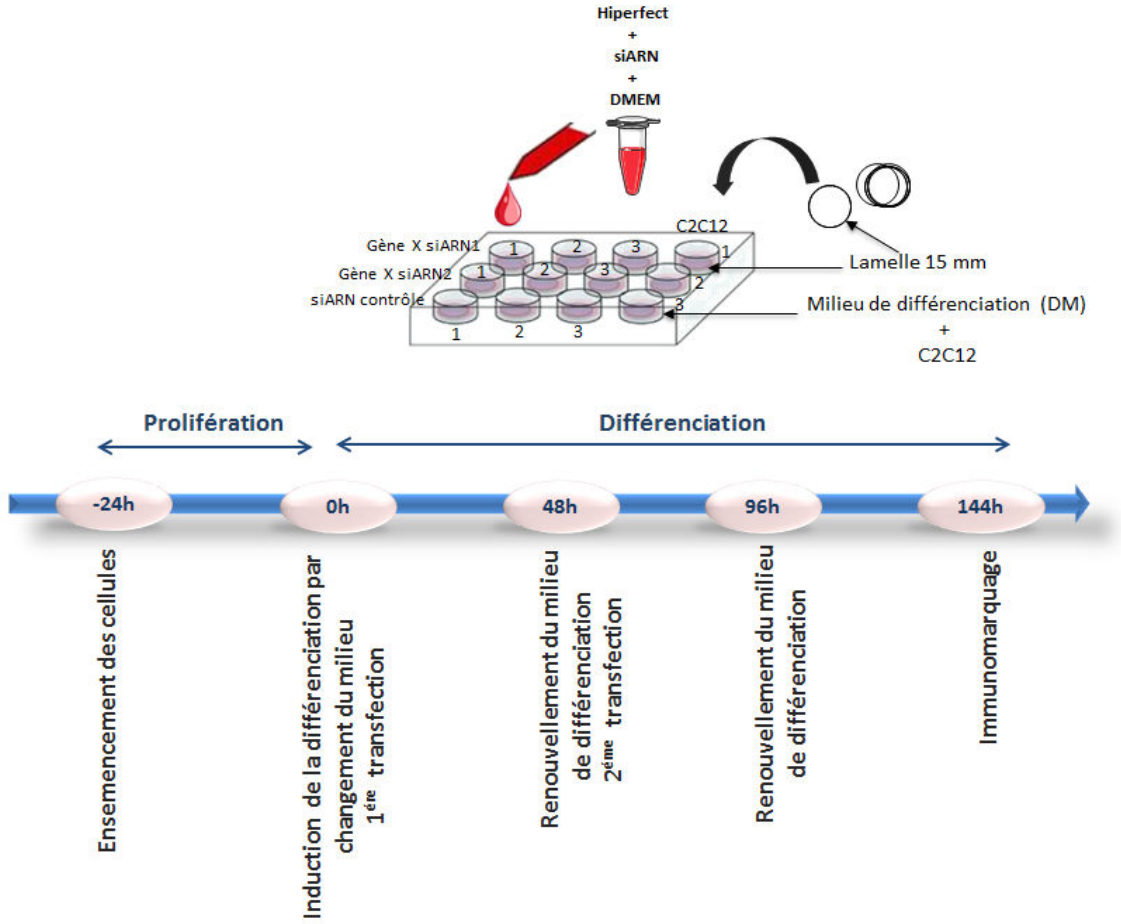


Figure 22 : Démarche expérimentale mise en oeuvre.

Les cellules C2C12 sont ensemencées sur des lamelles de 15mm (1,8 cm^2) dans des plaques à 12 puits. Après 24 heures, le milieu de prolifération est remplacé par le milieu de différenciation (T=0h) ; les cellules à 70 à 80% de confluence, sont transfectées par les siARNs à T= 0h et à T= 48h. A 96h de différenciation le milieu est renouvelé et à 144h, les cellules sont récupérées pour l'immunomarquage.

A. **Culture cellulaire**

Le modèle cellulaire utilisé est la lignée murine myoblastique C2C12 issue de souris à fond génétique C3H, et fournies en passage 2 par *ATCC* (*American Type Culture Collection*) **(Yaffe, 1968; Yaffe and Saxel, 1977)**.

Prolifération : C2C12 sont des cellules adhérentes qui sont ensemencées à raison de 5000 à 10000 cellule/cm2 dans un milieu de prolifération constitué de *Dubelco's Modified Eagle's Medium* (DMEM), supplémenté de 10% (V/V) de sérum de veau fœtal (*Eurobio, Courtaboeuf, France*), 50 Unités/ml de Pénicilline et 50 µg/ml de Streptomycine. Les cellules sont incubées dans une étuve à 37°C à 5% de CO_2. Le milieu est renouvelé toutes les 48 heures et les cellules sont repiquées lorsqu'elles atteignent 70% de confluence. Pour les repiquer, les cellules sont lavées au DPBS (*GiBCO,Dulbecco's Phosphate Buffered Saline, 1X*) puis traitées à la trypsine (*GIBCO*, 2.5 %,10X) diluée 10 fois dans une solution de DPBS/EDTA (0,5mM). Ensuite les cellules sont récupérées, centrifugées 5 minutes à 2500 rpm et comptées sur cellule de *Malassez* avant d'être réensemencées sur de nouvelles boites. Pour le crible de 100 gènes, les cellules sont ensemencées dans des plaques à 12 puits sur des lamelles stériles de 15 mm de diamètre dans 1 mL du milieu de prolifération alors que pour les analyses d'expression par RT-PCR quantitative, les cellules sont ensemencées sur une plaque à 6 puits dans 2 mL du milieu de prolifération (3.10^4 cellules/cm^2 par puits) **(Figure 22)**.

Différenciation : La différenciation est induite lorsque les cellules C2C12 atteignent 80% de confluence, en remplaçant le milieu de prolifération par le milieu de différenciation constitué de DMEM supplémenté par 2 %(V/V) de sérum de cheval (GiBCO) complémenté avec 50 Unités/ml de Pénicilline et 50 µg/ml de Streptomycine. Les cellules sont incubées dans une étuve à 37°C à 5% de CO2 et le milieu est renouvelé toutes les 24 heures.

Congélation et décongélation : Lorsque les cellules atteignent 70% de confluence, une trypsination est réalisée et les cellules sont comptées puis récupérées dans un volume de milieu de congélation (5% DMSO, 95 % GM) adéquat pour une concentration cellulaire de 10^6 cellules/mL. Les cellules sont ensuite aliquotées en fraction de 1 mL dans des cryotubes et sont stockées dans l'azote liquide ou à -80°C. Pour réutiliser ces stocks, les cellules sont

décongelées à 37°C puis centrifugées (5 minute à 2500 rpm) et ensemencées sur une boite 100 mm dans 10 mL de milieu de prolifération.

B. Transfection

La transfection est réalisée en se basant sur le protocole proposé par le fournisseur (*HiperFect transfection Reagent Quick-Start Protocol)*. Pour optimiser la transfection des cellules adhérentes comme C2C12, un mélange de transfection est réalisé pour chaque transfection, il est constitué de 37.5 ng de siRNAs *(Qiagen®)* dilués dans 100 uL de milieu de culture sans sérum (afin d'obtenir une concentration finale de 5 nM) et 3 uL de l'agent de transfection «*HiPerFect®» (HiPerFect® Transfection Reagent, Qiagen)*. Le mélange est ensuite incubé 5 à 10 minute à température ambiante permettant la formation de complexe de transfection, puis ajouté goutte à goutte aux cellules à transfecter et enfin homogénéiser gentiment afin de s'assurer d'avoir une distribution uniforme du complexe transfectant dans le milieu de culture **(Figure 22)**.

Test d'efficacité de la transfection : afin de tester l'efficacité de la transfection, les cellules C2C12 sont transfectées avec un siRNA killer (10uM Qiagen®) ciblant l'ARNm d'un gène indispensable à la survie des cellules. Ce test a été réalisé sur $1{,}9.10^4$ cellules (1.10^4 cellules/cm2) au cours de la prolifération et 7.10^4 cellules (36.10^3 cellules/cm2) au cours de la différenciation. Les cellules sont ensemencées dans des plaques à 24 puits puis transfectées suivant le protocole décrit ci-dessus. Les cellules survivantes sont trypsinées et comptées toutes les 24h pendant 3 jours.

Transfection réalisée pour le crible : Afin de faciliter les étapes d'immunomarquage et les analyses microscopiques, les cellules C2C12 ont été ensemencées, sur des lamelles stériles de 15 mm de diamètre, dans des plaques à 12 puits à raison de 3.10^4 cellules/cm2 (120 000 cellules par puits) et dans 1 mL du milieu de prolifération. 24 heures après, les cellules, à 80% de confluence, sont mises en différenciation par changement de milieu. Les cellules sont ensuite transfectées avec un siRNA dirigé contre un gène cible à l'aide de l'agent de transfection *HiPerFect*. Pour chaque gène candidat, deux siRNAs différents sont utilisés et pour chaque siRNA, trois répliquas techniques sont réalisés **(Figure 22)**. Des cellules non transfectées et des cellules transfectées par un siRNA dirigé contre un

gène qui n'existe pas chez les souris (*3'alexafluor647*, *QIAGEN*, AATTCTCCGAACGTGTCACGT) sont utilisées comme contrôles de l'expérience. Les transfections sont réalisées de façon transitoire au deux temps de la différenciation des C2C12, la première au temps T= 0h et la deuxième au temps T= 48h. Le milieu de différenciation doit être changé toutes les 48 heures, la différenciation est arrêtée au temps T= 144h (Figure 18). Les cellules sont incubées dans une étuve à 37°C et à 5% de CO_2 pendant toute l'expérience

Transfections réalisées pour les analyses par RT-PCRq : Les cellules sont ensemencées sur une plaque à 6 puits dans 2 mL du milieu de prolifération (3.10^5 cellules/cm2 par puits). 24 heures après, les cellules à 80% de confluence sont mises en différenciation par changement de milieu et transfectées, pour chaque gène choisi, deux siRNAs sont utilisées. Au temps T= 48h de différenciation, les plaques sont récupérées, les cellules sont lavées 2 fois au PBS 1X et trypsinées. Les cellules sont ensuite centrifugées 5 minutes à 2500 rpm, récupérées en culot cellulaires et préparées pour l'extraction des ARN totaux.

C. Immunomarquage et coloration des noyaux

Après 144 heures de différenciation, les cellules sont lavées, dans les puits, à trois reprises au PBS 1X puis fixées au paraformaldéhyde 4% pendant 15 minutes à température ambiante. Après trois lavages au PBS 1X, les cellules sont perméabilisées avec du tampon HEPES Triton froid (20mM HEPES; 300mM sucrose; 50mM NaCl; 3mM $Mgcl_2$; TritonX-100 0,5%; pH 7,4) durant 30 minutes à 4°C, puis lavées 3 fois au PBS 1X. Enfin, les cellules sont incubées dans un milieu de saturation (PBS 1X, GS 10%, BSA 1%, TritonX-100 0,1%) pendant 1h à température ambiante et lavées au PBS 1X, BSA 0,2%. L'anticorps primaire anti-myosine anti-MyHC (*Monoclonal Anti-Myosine skeletal, Fast, Clone My-32, Mouse Ascities Fluid, M4276, Sigma-Aldrich*) dilué 400 fois dans PBS1X-BSA 1% est déposé à raison de 20 µL par lamelle et incubé une nuit à l'obscurité à 4°C dans un endroit humide. 24 heures après, 2 lavages au PBS 1X et 2 bains de 5 minutes au PBS 1X, BSA 0,2%, Tween20 0,1% sont réalisés suivis d'une incubation avec l'anticorps secondaire (*Alexa Fluor® 488 Goat Anti-Mouse IgG (H+L Invitrogen, Thermo Fisher Scientific, Watham, MA*) dilué 1000 fois dans du PBS1X-BSA 1%, pendant 15 minutes à 37°C à l'obscurité à raison de 20 µL par lamelle. 2 lavages au PBS 1X et 2 bains de 5 minutes au PBS 1X, BSA 0,2%, Tween20 0,1% sont réalisés et les noyaux

sont marqués au DAPI (1mg/ml à la dilution 1:1000 dans PBS 1X) pendant 5 minutes à température ambiante. Les lamelles sont enfin lavées trois fois 5 minutes au PBS 1X et placées à l'envers sur des lames (3 lamelles par lame) avec du milieu de montage (Mowiol 4-88 9.6%, glycérol 12%, tris HCL pH 8.5 0,1 M), scellées avec du vernis à ongle et stockées à 4°C et à l'obscurité.

D. Microscopie à épi-fluorescence

Les cellules sont observées au microscope à fluorescence inverse (*DMI6000 B Leica*). L'acquisition et le traitement des images sont réalisés à l'aide du logiciel *Leica MM AF Imaging System, powered by MetaMorph*. Les cellules sont observées au grossissent 100X. La quasi-totalité de la lamelle est prise en image grâce à la technologie « *platine scanning* » du microscope. 100 à 150 images sont acquises par lamelle afin de reconstruire une image définitive de l'ensemble de la zone scannée. Pour détecter la fluorescence *Alexa Fluor® 488* et DAPI, 2 longueur d'onde ont été choisies avec ajustement de temps d'exposition et d'intensité lumineuse : pour DAPI (345nm-455nm), 50 ms le temps d'exposition et 150 l'intensité lumineuse ; pour *Alexa Fluor® 488* (499nm-519nm), 1000 ms le temps d'exposition et 150 l'intensité lumineuse. Les images individuelles ainsi que l'image de reconstruction sont sauvegardées sous format « .tif » pour la suite de l'analyse.

E. Analyse et traitement des images

Deux logiciels, *Metamorph* (*Leica microsystems*) et *ImageJ* v1.49 (*Wayne Rasband, national institues of health, USA*), sont utilisés pour traiter les images de reconstruction et quantifier les deux types de fluorescence (*Alexa Fluor® 488 et DAPI)*.

1- Quantification de l'aire de la fluorescence

Les aires de fluorescence *Alexa Fluor® 488* et DAPI sont quantifiées par le logiciel *Metamorph*. Après élimination du bruit de fond de l'image par correction statistique Moyenne et coloration de l'intérieur des myotubes ou des noyaux, l'aire contenant la fluorescence ainsi que son pourcentage par rapport à la zone scannée ont été calculés. Le pourcentage de l'aire de fluorescence de *Alexa Fluor® 488* représente donc la surface de

l'image occupée par les myotubes alors que le pourcentage de l'aire de fluorescence de DAPI représente la surface de l'image occupée par les noyaux.

2- Indice de fusion

L'indice de fusion correspond au rapport du nombre de noyaux dans les myotubes sur le nombre total de noyaux, multiplié par 100. C'est le logiciel ImageJ v1.49 qui a été utilisé pour dénombrer d'une façon semi-automatique aussi bien le nombre total des noyaux que le nombre de noyaux dans les myotubes. Cette analyse a été réalisée sur une section des images de reconstruction DAPI et Alexa Fluor® 488 selon le protocole suivant :

Sur l'image DAPI :

- Ajustement automatique de la luminosité et du contraste,
- Réglage du seuil (*threshold*) par la méthode Li (*Minimum Cross Entropy thresholding method*),
- Réalisation d'un masque binaire,
- Séparation des noyaux collés par *watershed,*
- Dénombrement des noyaux sur toute la zone de l'image.

Sur l'image *Alexa Fluor® 488* :

- Ajustement automatique de la luminosité et du contraste et réglage du seuil de la même façon que pour l'image DAPI,
- Construction d'un filtre médian (taille de pixel = 3),
- Réalisation d'un nouveau masque aussi bien sur l'image *Alexa Fluor® 488* que l'image DAPI,
- Combinaison des deux masques,
- Comptage des noyaux dans les myotubes

L'indice de fusion est calculé suivant la formule suivante : $\frac{nb\ noyaux\ \ dans\ les\ myotubes \times 100}{nb\ totaux\ noyaux}$

3- Détermination de la longueur et de la largeur des Myotubes

Les dimensions des myotubes sont calculées en utilisant le logiciel *ImageJ* sur les images de reconstruction *Alexa Fluor® 488*. Elles sont calculées pour 30 myotubes dispersés sur 15 champs différents de chaque image. La longueur du myotube est déterminée par la mesure de la distance entre les deux extrémités distales du myotubes alors que la largeur est

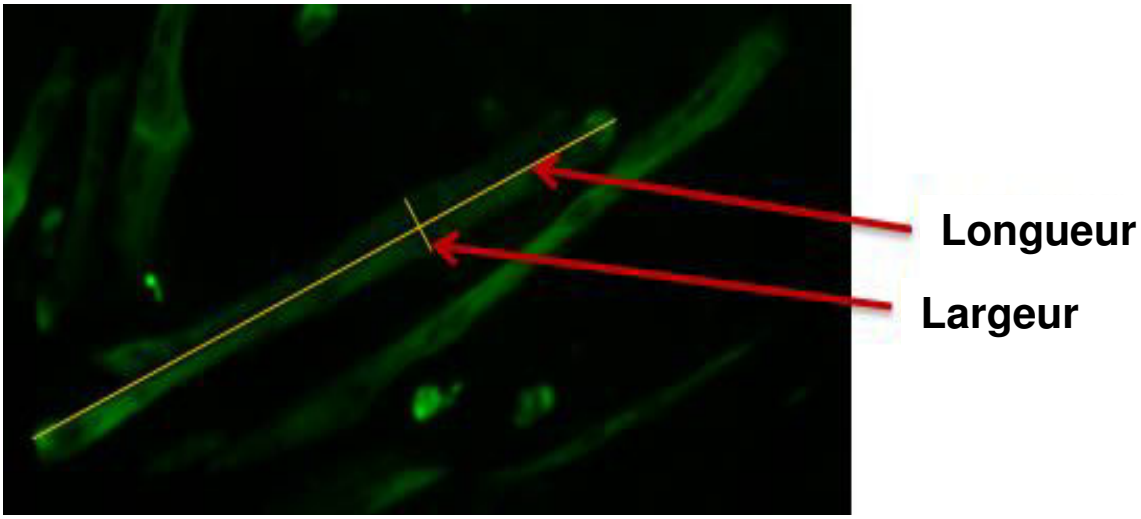


Figure 23 : Calcul de la taille des myotubes.
La longueur et la largeur des myotubes sont déterminées d'une façon semi-automatique sur les images de reconstruction Alexa Fluor® 488 en utilisant le logiciel ImageJ et en dessinant deux lignes droites entre les extrémités distales et proximales de chaque myotube.

déterminée par la mesure de la distance entre les extrémités proximales du myotube **(Figure 23)**.

F. Test statistique

A la fin de l'analyse de chaque paramètre cité ci-dessus (*Alexa Fluor® 488*, DAPI, dimensions de myotubes et indice de fusion), nous obtenons pour chaque inactivation par un siRNA, 3 valeurs correspondantes au 3 réplicas techniques réalisés. La valeur moyenne de ces 3 valeurs est calculée et comparée à la valeur moyenne obtenue à partir des contrôles correspondants. Les comparaisons sont effectuées par le test de *Student* et les correction sont réalisées par la méthode des contrastes de *Dunnett* **(Dunnett, 1955)**. Ces tests ont été réalisés par le logiciel R **(Team, 2011)**. Le test de significativité est déterminé selon les *P-value* suivantes *:* *** p<0,001 ; ** p<0,01 ; * p<0,05.

G. Vérification de l'extinction par RT-PCR quantitative

Pour vérifier si les ARNm des gènes inactivés ont bien diminués après l'extinction par les siRNAs, ces ARNm ont été quantifiés par RT-PCRq. Cette vérification a été uniquement réalisée sur 13 gènes. Pour ce faire, une deuxième série d'inactivation de ces 13 gènes par les siRNAs correspondants a été effectuée.

1- Extraction, dosage et rétrotranscription des ARN totaux

Les ARN totaux sont extraits à partir des culots cellulaires en utilisant le Kit *RNeasy Mini Kit* (*Qiagen*) et en appliquant le protocole fourni par *Qiagen*. La concentration des ARN est déterminée par dosage au *NanoDrop* (*NanoDrop Technologies, Wilmington, USA*). 2 ug d'ARN totaux sont retro-transcrits en ADN complémentaire par le kit *Hight capacity cDNA reverse transcription* (*Applied Biosystem*) et l'enzyme *Multiscribe™ Reverse Transcriptase* (25U par 50 µL) selon le programme suivant : 10 minutes à 25°C suivie de 2 heures à 37°C, 5 minutes à 85°C et 10 min à 4°C.

Gènes	Référence de la sonde
Dhx38	Mm00617713_m1
Set	Mm04243941_g1
Smarcc2	Mm01159912_m1
Srsf4	Mm00491080_m1
Taok1	Mm00522816_m1
Tmem205	Mm00724248_m1
Yipf7	Mm00470162_m1
Otud4	Mm01248651_m1
Pno1	Mm00481372_m1
Pola2	Mm00447142_m1
Prpf6	Mm00467081_m1
Rbm28	Mm00513597_m1
Rpl11	Mm01613214_g1

Tableau 2 : Liste des sondes Taqman utilisées pour quantifier par PCRq les transcrits des gènes analysés.

2- PCR quantitative (PCRq)

Les PCRq ont été réalisées en utilisant la technologie Taqman (*Taqman Gene Expression Master Mix Applied biosystems*) selon le protocole *d'Applied biosystems* dans *Quant studio 3* (*Applied Biosystems by Thermo Fisher Scientific*) ou *7900 HT* (*Applied Biosystems by Thermo Fisher Scientific*). 2 ng d'ADNc ont été utilisés pour chaque qPCR. Deux gènes références ont été utilisés : *GAPDH* (*Glycéraldehyde-3-phosphatase déshydrogénase* Mm99999915_g1, *Applied Biosystems*) et *β2-M* (*β2- microglobuline* Mm 00437762_m1, *Applied Biosystems*). L'ensemble des sondes Taqman utilisées sont représentées dans le **Tableau 2**. L'analyse a été effectuée grâce aux systèmes *QuantStudio® Design and Analysis Software* version 1.1 et *SDS* 2.3 et qui permet de suivre en temps réel le niveau d'expression des ARNm des gènes cibles.

H. Gene Ontology

L'analyse « *gene ontology »* a été réalisée en utilisant « le système de classification PANTHERE » et en suivant les instructions fournies par le site www.pantherdb.org **(Mi et al., 2013)**. La liste de 100 gènes a été analysée par « *PANTHER Overrepresentation* Test » (version du 15-07-2016) en utilisant la banque de données « *GO Ontology database* » (version 22-06-2015-06) et tous les gènes de la souris (*Mus Musculus database*) comme une liste référence. L'ensemble de gènes a été annoté par « *GO molecular function complete* », « *GO cellular component complete* » et « *GO biological process complete* ». Uniquement les résultats avec une surreprésentation significative (*P-value* ≤ 0.001) sont présentés sur les histogrammes.


RESULTATS

Avant de présenter mes résultats, il est nécessaire de décrire le chemin suivi durant mon travail afin de clarifier et de comprendre la nécessité et l'intérêt de chaque étape réalisée. En premier lieu, je présenterai la stratégie adoptée et en second lieu, je détaillerai les phénotypes d'inactivation observés et leurs analyses.

I. Stratégie adoptée

Mon but était de mettre en place un crible fonctionnel afin d'identifier de nouveaux gènes impliqués dans la myogenèse. La démarche suivie pour la réalisation de ce crible est basée sur 5 points essentiels :

Quel est le modèle cellulaire adéquat ?

Quels sont les gènes à inactiver ?

Comment inactiver ces gènes ?

Quels sont les phénotypes à suivre et comment ?

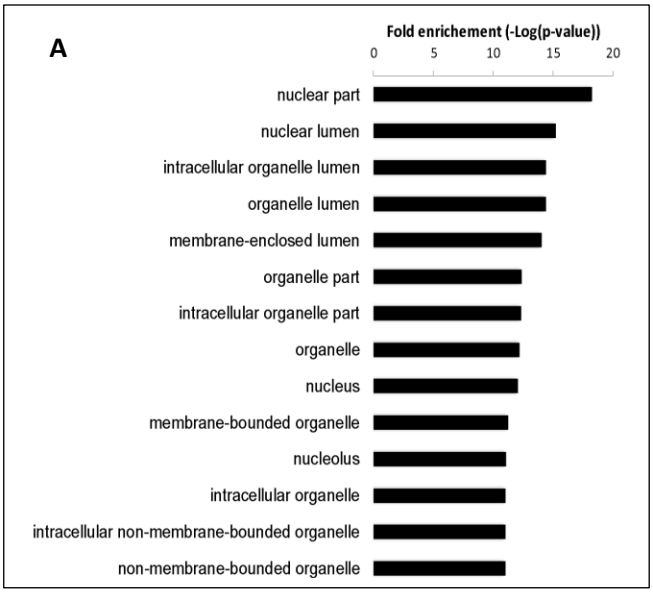
Comment analyser les phénotypes pour classer et sélectionner les gènes ?

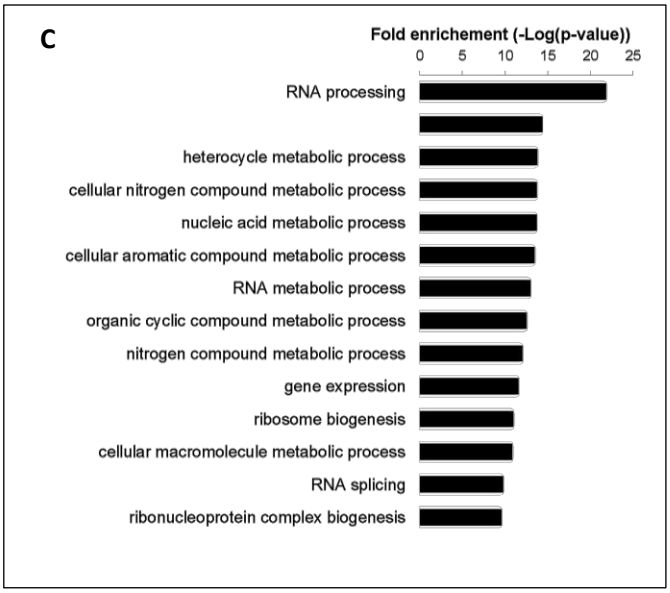
1- Quel est le modèle cellulaire adéquat ?

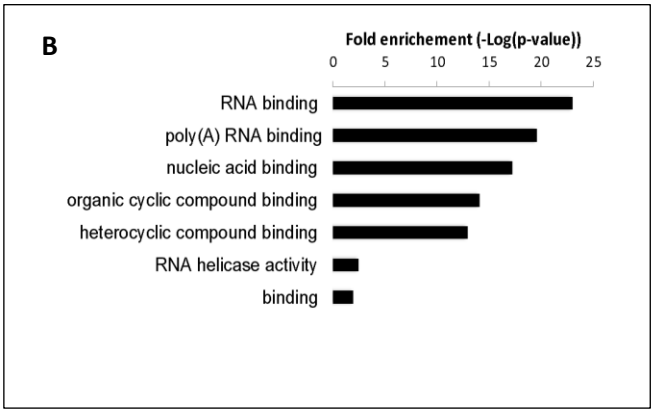
Plusieurs étapes sont nécessaires au déroulement de la myogenèse, ce qui en fait un processus complexe avec des acteurs multiples et divers. Les cribles à grande échelle pour identifier de nouveaux gènes ayant un rôle au cours du développement musculaire ont été réalisés *in vivo* sur des organismes modèles invertébrés comme la drosophile et *c. elegans*. Ce type de crible semble être difficile à appliquer chez les mammifères. Pour ce faire, nous avons décidé d'utiliser un modèle cellulaire tel que les cellules primaires ou les lignées cellulaires. Vu les difficultés associées à l'introduction des acides nucléiques dans les cellules primaires, nous avons choisi d'utiliser une lignée cellulaire. Plusieurs lignées pourraient être utilisées telles que les lignées de rat L6 et L8 **(Yaffe, 1969; Yaffe and Saxel, 1977)** ou les lignées myobalstiques murines comme C2.7 **(Pinset et al., 1988)**, MM14 **(Linkhart et al., 1980)** ou C2C12. C2C12 est la lignée la plus employée pour étudier le muscle, il s'agit d'un sous-clone dérivé de la lignée originelle C2, isolée à partir de muscle de cuisse de souris C3H âgées de deux mois **(Yaffe, 1968; Yaffe and Saxel, 1977)**. Elles correspondent à des cellules satellites du muscle, qui *in vivo* sont impliquées dans la régénération musculaire. La prolifération et la différenciation des myoblastes C2C12 *in vitro* peuvent facilement récapituler les différentes



Figure 24 : Schéma descriptif des différents filtres utilisés pour établir la liste finale des 100 gènes.







<u>Figure 25 : L'analyse des gènes candidats par « gene ontology » (GO).</u>

L'ensemble de gènes a été annoté par « *GO cellular component complete* » (A), « *GO molecular function complete* » (B), et « *GO biological process complete* » (C). Uniquement les résultats avec une surreprésentation significative (P-value ≤ 0.001) sont présentés.

étapes du processus de myogenèse *in vivo*. En effet, leurs comportements en terme de signalisation et de réponse transcriptionnelle sont assez proches de ceux de myoblastes primaires **(Cornelison, 2008)**. Elles sont donc largement utilisées pour disséquer les mécanismes moléculaires sous-jacents la différenciation musculaire. Ces caractéristiques associées à leur facilité de culture et de transfection font de cette lignée un modèle de choix pour l'identification de nouveaux gènes jouant un rôle durant la myogenèse. Pour faciliter les expériences d'Immunomarquage et d'observation microscopique qui suivent les étapes de culture cellulaire et d'inactivation génique, les cellules C2C12 ont été cultivées sur des lamelles mises au fond des puits des plaques de culture. L'adhérence de C2C12 sur ces lamelles a été testée à partir des stocks de cellules ayant différents passages. Le test a montré que les cellules restent adhérées sur les lamelles jusqu'à la fin de l'expérience si leur âge ne dépasse pas le passage 6.

2- Quels sont les gènes à inactiver ?

Plusieurs études d'identification de gènes impliquées dans le développement musculaire ont été réalisées chez plusieurs organismes modèles. L'étude réalisée par Schnorrer et al. semble être la plus complète, les auteurs ont pu mettre en évidence l'implication de 2785 gènes dans le développement musculaire de la drosophile parmi 10461 gènes inactivés **(Schnorrer et al., 2010)**. L'établissement de notre liste de gènes candidats est basé sur la démarche suivante :

(i) en premier lieu, nous avons cherché les orthologues murins des 2785 gènes de la drosophile. Les protéines issues des 2785 gènes ont été récupérées et comparées au protéome de la souris (version NCBIM37.65). Les gènes codant pour les protéines présentant le score d'alignement (hit) le plus haut ont été récupérés. Ceci nous a permis d'avoir une liste de 2425 gènes,

(ii) ce grand nombre de gènes contient vraisemblablement des gènes qui ne sont pas impliqués dans le développement musculaire vu la distance évolutive qui existe entre la drosophile et la souris.

Pour résoudre ce problème et se rapprocher plus de la souris, cette liste a été affinée : (i) nous avons récupéré deux listes de gènes obtenues dans des études de transcriptomique et de protéomique réalisées sur des cellules C2C12 en cours de

1110008P14Rik	Ermp1	Nsmce2	Rnps1	Tagln2
Actl6a	Fam136a	Nsun4	Rpl11	Taok1
Adsl	Fip1l1	Nudt21	Rpl12	Tbc1d23
Brix1	Grcc10	Nvl	Rplp0	Thoc6
cct7	H2afv	Osbp2	RsL1d1	Tmem205
Ctps1	Ints5	Otud4	Rufy3	Top2a
Ctu1	Klhl24	Pabpc4	Sap130	Trappc2l
D10Wsu52e	Mdn1	Phf14	Sass6	Tsr1
Dcakd	Mplkip	Phldb2	Seh1l	U2surp
Ddx18	Mrto4	Pno1	Set	Ufc1
Ddx23	Mstn	Pola2	Sf3a3	Usp39
Ddx47	Myh7	Prim2	Sf3b3	Utp15
Dhx38	Naa10	Prpf4b	Sf3b5	Vars
Dna2	Ncaph	Prpf6	Sfpq	vipas39
Dnajc19	Nhp2	psma1	Slc7a6os	Wdr18
Dnajc2	Nip7	Pwp1	Smarcc2	Wdr43
Dph5	Nmt2	Rars	Srpr	Wdr46
Ebna1bp2	Nol8	Rbm28	Srsf4	Yipf7
Eef1b2	Nop10	Rcc1	Srsf7	Zfp276
Eftud2	Nop2	Rnf20	Svil	Zfp322a

Tableau 3 : Liste des 100 gènes présélectionnés pour le crible.

différenciation **(Tomczak et al., 2004; Kislinger et al., 2005)**. Ces deux études ont identifié respectivement 1806 et 1259 gènes (1880 protéines) **(Figure 24)**. (ii) Pour avoir dans notre liste que des gènes de fonction inconnue, nous avons cherché à éliminer à partir de ces deux listes, les gènes avec fonction attribuée ou avec phénotype associé. C'est grâce à l'outil « *Batch Query* » du site officiel du génome de la souris (*MGI, Mouse Genome Informatics*, http://www.informatics.jax.org/), que nous avons pu réaliser cette élimination. Cette analyse a permis d'avoir respectivement 1052 gènes et 1046 gènes sans fonction et sans phénotype associé à partir des données protéomique et transcriptomique **(Figure 24)**. (iii) La comparaison de ces deux listes de gènes, en gardant uniquement les gènes communs, a permis de sélectionner une liste de 312 gènes. (iv) Le croisement de cette liste de gènes avec celle de 2425 orthologues murins (en gardant que les gènes communs) nous a permis d'établir notre liste finale qui contient désormais 176 gènes **(Figure 24)**. Ces gènes présentent les propriétés suivantes : leur expression varie au cours de la différenciation des C2C12 et leurs orthologues chez la drosophile sont nécessaires au développement musculaire. Ces 176 gènes ont été sélectionnés pour la réalisation du crible. Au cours de ma thèse j'ai pu inactiver et cribler une centaine de gènes parmi cette liste définitive **(Figure 24 Et Tableau 3)**. Dans la suite de la présentation des résultats, nous avons fait le choix de nommer les gènes de G1 à G100.

Analyse par gene ontology (GO) : Dans le but d'avoir une idée concernant les fonctions putatives des 100 gènes présélectionnés pour le crible, une analyse préliminaire par l'annotation fonctionnelle issue de l'ontologie (GO) a été réalisée. L'analyse des termes de « *GO cellular component* » a révélé que notre liste est enrichie en termes comme « *nuclear part* », « *nuclear lumen* » et « *nucleolus* » **(Figure 25A)**. Les termes «*GO molecular function* » les plus surreprésentés dans notre liste sont ceux qui sont liés à « *RNA binding* » and « *RNA helicase activity* » **(Figure 25B)**. Enfin notre liste est enrichie, concernant les termes « *GO biological procès* », en ceux liés aux termes comme « *RNA* », « *ribosome* » et « *splicing* » **(Figure 25C)**. Cette analyse préliminaire montre que notre liste a été enrichie en termes GO en relation avec la synthèse des ARNs.

3- Quel est l'outil d'inactivation à utiliser?

L'ARNi est l'outil de choix pour faire de la génomique fonctionnelle ou pour sélectionner des gènes dans les cribles à haut débit. Pour notre modèle, la lignée C2C12, l'extinction génique

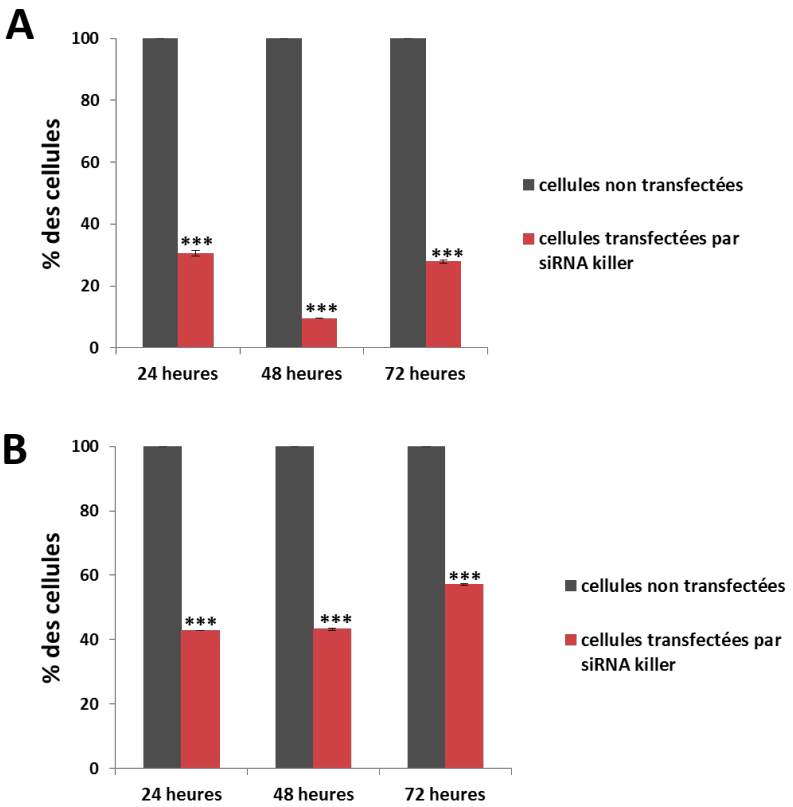


Figure 26 : L'efficacité de la délivrance des siRNAs dans les cellules C2C12.

Pour tester l'efficacité de la délivrance, un siRNA Killer a été transfecté dans les cellules C2C12 aussi bien au cours de la prolifération que la différenciation. Les cellules survivantes sont récupérées et comptées après 24, 48 et 72 heures. Les résultats montrent qu'au cours de la prolifération (A) le nombre des cellules survivantes a diminué de 70 à 95% par rapport aux contrôles. Les cellules en différenciation (B) montrent une diminution d'environ 45 à 60% de nombre de cellules survivantes par rapport aux contrôles.

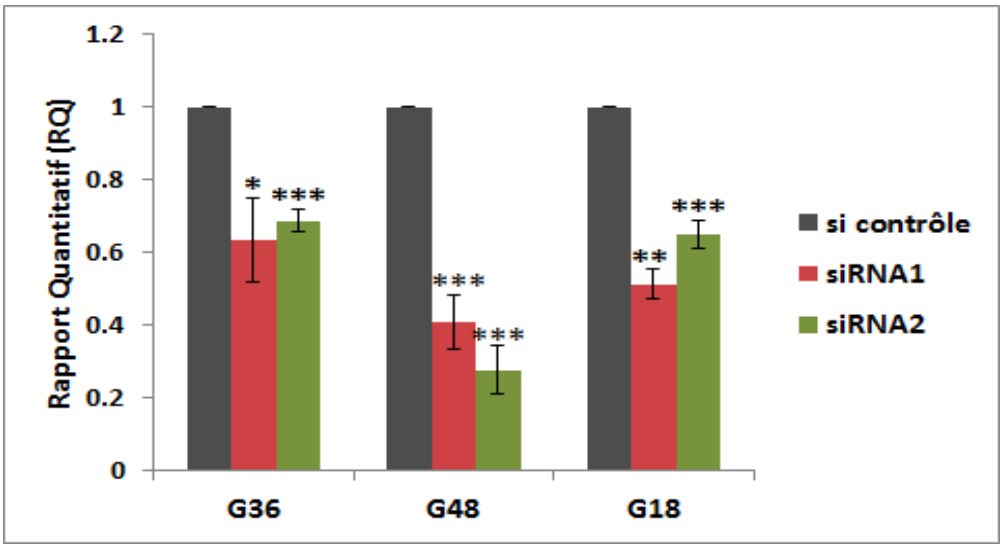


Figure 27 : La quantité relative des transcrits G36, G48 et G18 après inactivation.

L'efficacité de l'extinction génique par les siRNAs a été vérifiée par RT-PCRq sur des cellules en différenciation, inactivées pour les gènes G36, G48 et G18. Après extraction d'ARNs totaux et retrotranscription, les ARNm de ces gènes ont été quantifiés dans les cellules inactivées et les contrôles par PCRq en utilisant la technologie Taqman. Les trois gènes ont montré une diminution significative du niveau d'expression de leur ARNm avec les 2 siRNAs. (Test de t : *** p<0,001 ; ** p<0,01 ; * p<0,05).

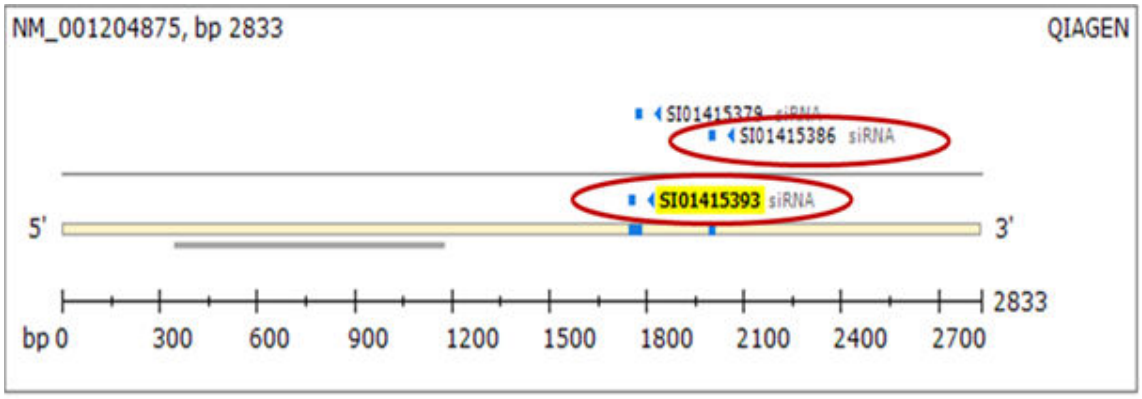
par ARNi peut être réalisé par l'introduction des vecteurs viraux ou plasmidiques mais aussi par la transfection directe et transitoire des duplexes de siRNAs. C'est cette méthode qui a été adoptée et utilisée pour notre crible.

Avant de se lancer dans l'inactivation des 100 gènes présélectionnés précédemment, il a fallu vérifier trois paramètres :

a- L'efficacité de la délivrance des siRNAs dans les cellules C2C12 : Pour tester l'efficacité de la délivrance, un siRNA particulier capable de tuer les cellules C2C12 (*siRNA Killer, AllS tars Mm Cell Death control siRNA, Qiagen*) a été transfecté, en utilisant l'agent de transfection *Hiperfect*, dans les cellules C2C12 aussi bien au cours de la prolifération que la différenciation. Les cellules survivantes sont récupérées et comptées après 24, 48 et 72 heures de culture en présence du siRNA killer. Les résultats montrent qu'au cours de la prolifération le nombre des cellules survivantes a diminué de 70 à 95% par rapport aux contrôles **(Figure 26A)**. De même, pour les cellules au cours de la différenciation, on note une diminution d'environ 45 à 60% de nombre de cellules survivantes par rapport aux contrôles **(Figure 26B)**. D'après le test statistique utilisé, ces variations sont significatives. A l'issu de ces expériences, nous pouvons en déduire que les siRNAs ainsi que l'agent de transfection utilisés semblent être bien efficaces dans les cellules C2C12 et donc pourront être utilisés pour réaliser l'inactivation des gènes candidats de notre crible.

b- L'efficacité de l'extinction génique par les siRNAs : Il est nécessaire de vérifier après l'extinction de chaque gène candidat la diminution de son ARNm. Vu que le nombre de gènes inactivés est très grand, uniquement l'extinction de 13 gènes a été vérifiée par RT-PCRq (figure 29). Ces gènes ont été inactivés par 2 siRNAs différents au cours de la différenciation. Après extraction d'ARNs totaux et retrotranscription, les ARNm de ces 13 gènes ont été quantifiés dans les cellules inactivées et les contrôles par PCRq en utilisant la technologie Taqman. 9 gènes ont montré une diminution significative du niveau d'expression de leur ARNm avec les 2 siRNAs (G25, G36, G95, G20, G18, G88, G32, G48, G52) **(Figure 27)** ; 3 gènes ont montré une diminution de leur ARNm uniquement avec un des deux siRNAs utilisés (G51, G38, G84) ; et un seul gène (G50) a montré des résultats de quantification incohérents puisque le niveau de son ARNm semble augmenter après l'inactivation par les deux siRNAs. Ces résultats

A



B

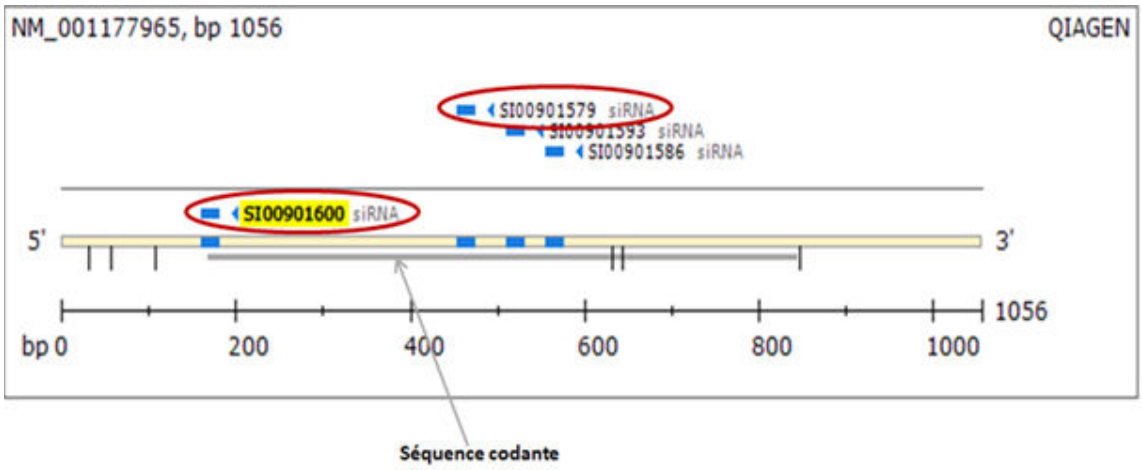


Figure 28 : Sites de fixation des siRNAs choisis sur les ARNm des gènes G63 et G32.

A : pour le gène G32 : les 3 siRNAs proposées par Qiagen se fixent sur la partie 3' UTR du transcrit. Les 2 siRNAs utilisées dans notre étude sont entourés en rouge.

B : pour le gène G63 : les 4 siRNAs proposés par Qiagen se fixent sur la phase codante du transcrit.

montrent pour 12 gènes sur les 13 analysés que les siRNAs utilisés sont capables de diminuer par dégradation la quantité des ARNm des gènes cibles.

c- La durée de l'inactivation : D'après le fournisseur (Qiagen), l'efficacité de l'inactivation par les siRNAs est estimée maximale entre 6 et 72h après transfection. Pour notre crible, nous voulons étudier le Knockdown des gènes candidats pendant six jours de différenciation des cellules C2C12. C'est pour cette raison, nous avons choisis de transfecter les siRNAs deux fois au cours de la différenciation : la première transfection est effectuée au moment d'induire la différenciation et la deuxième transfection 48h plus tard ; de cette façon, les cellules seront en contact avec le siRNA pendant 96 heures **(Figure 22)**.

Chaque gène candidat a été inactivé par deux siRNAs différents en trois réplicas techniques. La majorité des siRNAs utilisés dans le crible, cible la région codante des ARNm des gènes candidats **(Figure 28A)**; toutefois, pour des raisons de spécificité, des siRNAs dirigés contre les parties 3'UTR des ARNm de quelques gènes sont utilisés **(Figure 28B)**. Les 100 gènes ont été inactivés par groupe de 10 gènes en 10 expériences distinctes avec chaque fois les contrôles suivants : des cellules non transfectées et des cellules transfectées par un siRNA dirigé contre un gène qui n'existe pas chez les souris.

4- Quel sont les phénotypes à suivre et comment ?

Pour détecter les éventuels phénotypes après le knockdown des gènes candidats par les siRNAs, il nous faut deux types de marqueurs spécifiques des deux étapes de croissance des cellules C2C12 : la prolifération et la différenciation. Ces marqueurs doivent nous permettre d'abord de visualiser les phénotypes et ensuite de les quantifier, les analyser et les comparer aux contrôles.

L'utilisation du colorant spécifique de l'ADN, le DAPI, montre une très bonne efficacité de marquage sur les cellules C2C12. Avec cet intercalant d'ADN, nous allons pouvoir quantifier la surface occupée par les noyaux et les dénombrer aussi bien dans les cellules prolifératives mononuclées que dans les cellules différenciées multinucléés (myotubes). Ce marqueur nous permettra de mettre en évidence après le knockdown des gènes candidats impliqués dans l'étape de prolifération, des anomalies relatives à la division cellulaire durant cette étape.

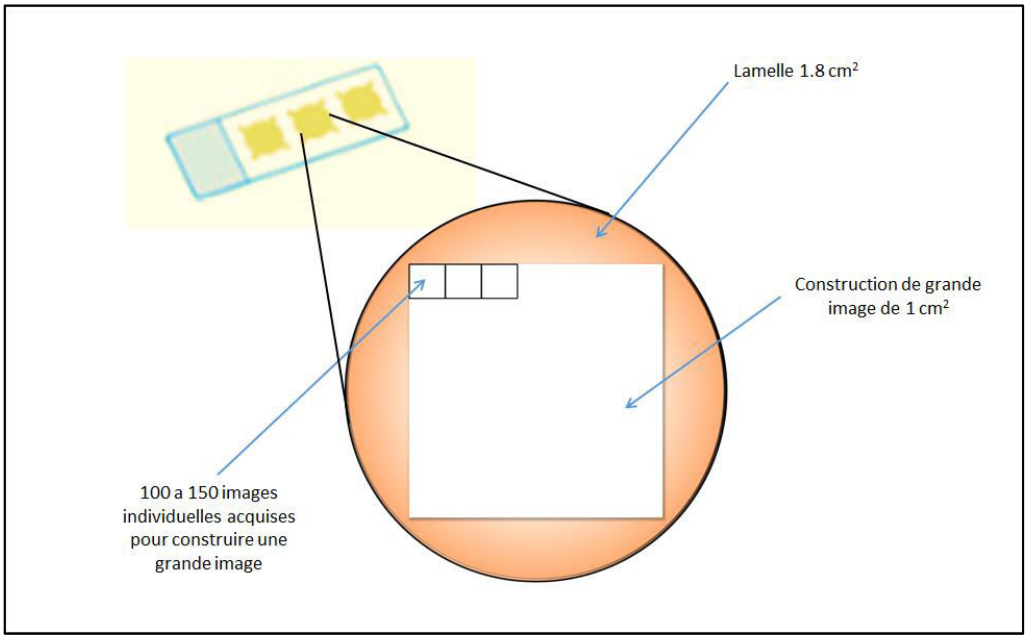


Figure 29 : schéma montrant la reconstruction de la grande image.

La quantification de la surface occupée par chaque marqueur (DAPI et myosine) a été réalisée sur une grande image de reconstruction (1 cm^2) contenant 100 à 150 images individuelles adjacentes qui couvrent la totalité de la zone scannée et prise en photo.

Le deuxième marqueur doit être spécifique de l'étape de la différenciation des cellules C2C12 et capable de visualiser des phénotypes liés à la quantité, la taille et/ou la forme des myotubes. Les protéines qui s'expriment et s'accumulent dans les myotubes matures, telles que la myosine ou la tropomyosine, pourraient jouer ce rôle et être visualisées et quantifiées par étiquetage à une protéine fluorescente comme la GFP (*green fluorescent protein*). Des constructions de lignées stables de C2C12, modifiées dans l'un des gènes codant pour la myosine ou la tropomyosine, ont été tentées au laboratoire. En effet, nous avons voulu remplacer par recombinaison homologue le gène endogène par une construction plasmidique dont le gène codant pour la GFP a été introduit en phase du coté 3' du gène en question. L'étiquetage de la myosine à la GFP par cette méthode était difficilement réalisable vu que cette protéine est codée par plusieurs gènes groupés en cluster sur le génome des cellules C2C12. Toutefois, plusieurs tentatives de construction de telles lignées stables hébergeant la tropomyosine marquée à la GFP ont échoués à cause des difficultés rencontrées et qui sont certainement liées à l'intégration des transgènes par recombinaison homologue dans cette lignée. Pour remédier à ce problème, nous avons eu l'idée de remplacer l'étiquetage à la GFP par l'immunmarquage en utilisant un anticorps anti-myosine pour visualiser et quantifier la surface occupée par la myosine durant l'étape de différenciation. La quantification de cette surface dans les cellules C2C12 différenciées peut nous renseigner sur la quantité, la taille et la forme des myotubes.

Pour visualiser ces deux marqueurs (DAPI et myosine), les cellules C2C12 après six jours de différenciation et inactivation par les siRNAs, sont récupérées pour un immunmarquage par l'anticorps anti-myosine et pour une coloration au DAPI. Ces cellules sont par la suite observées et imagées par microscopie à épi fluorescence. La quantification de la surface occupée par chaque marqueur a été réalisée sur une grande image de reconstruction contenant 100 à 150 images individuelles adjacentes qui couvrent la totalité de la zone scannée et prise en photo **(Figure 29)**. Les données quantitatives obtenues dans les cellules inactivées sont comparées à celles obtenues dans les contrôles.

Le pourcentage de l'aire de fluorescence DAPI par rapport à la surface totale de la grande image, reflète le nombre de noyaux et par conséquent la densité cellulaire. La variation significative de ce pourcentage par rapport à celui obtenu dans les cellules contrôles indique que l'étape de prolifération a été affectée dans les cellules inactivées pour

100				
Sans phénotype: 5 (élimination)	Avec phénotype: 95			
	Critères phénotypiques			
	Densité cellulaire	Quantité de myotube	Indice de fusion	Dimensions et morphologie des myotubes
Phénotypes différents entre les 2 siARNs (elimination)	72	73	77	71
Phénotypes similaires entre les 2 siARNs	23	22	18	24
Phénotype différent entre les 2 siARNs pour au moins un critère phénotypique(élimination)	17	14	10	15
Phénotype similaire entre les 2 siARNs pour un, deux ou trois critères phénotypiques	6	8	8	9
Augmentation	2	2	1	Long : 3
				Petit : 5
Diminution	4	6	7	Fin : 1
				Large : 0
Gènes sélectionnés	19			
Phénotypes observés durant	Prolifération et différentiation : 6	Différentiation : 13		

Tableau 4 : Les étapes de la classification des gènes selon les critères phénotypiques considérés.

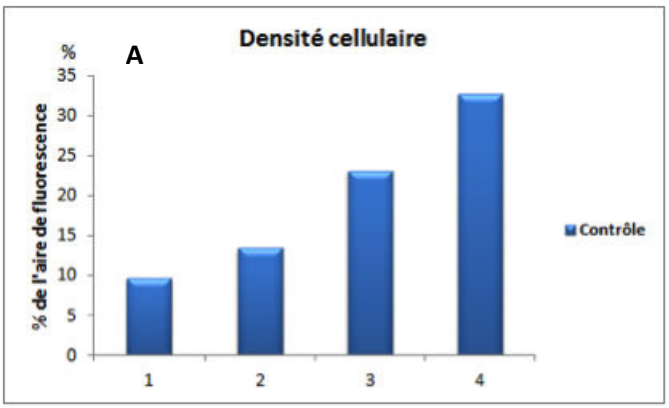
L'analyse phénotypique et la classification des 100 gènes candidats ont été réalisées selon les phénotypes observés avec chacun des deux siRNAs utilisés et en fonction des critères phénotypiques : « densité cellulaire », « quantité de myotubes », « dimensions et morphologie des myotubes » et « indice de fusion ». Trois étapes ont été effectuées : (i) éliminer tous les gènes dont l'inactivation par les deux siRNAs ne montre aucun phénotype ; (ii) analyser le reste des gènes en quatre groupes de gènes en fonction des quatre critères phénotypiques puis exclure dans chaque groupe les gènes pour lesquels l'inactivation par les deux siRNAs ne montre pas le même phénotype ; (iii) regrouper les différents gènes restants dans les quatre groupes (en excluant ceux qui sont dupliqués) en deux ensembles de gènes selon qu'ils sont impliqués dans la prolifération ou la différenciation.

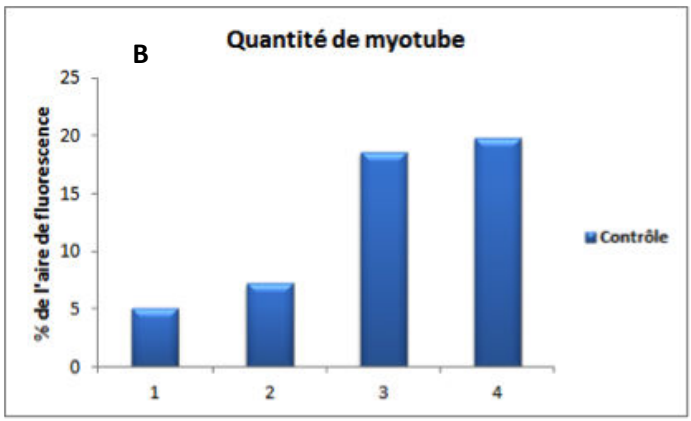
les gènes candidats impliqués dans cette étape. Le pourcentage de l'aire de fluorescence de l'anticorps anti-myosine par rapport à la surface totale de l'image de reconstruction reflète la quantité de myotube dans la zone scannée. Les dimensions et la morphologie de ces myotubes ont été aussi étudiées sur la grande image par détermination de la longueur et la largeur de plusieurs myotubes **(Figure 23)**. La superposition des images anti-myosine et DAPI nous a permis de déterminer le nombre de noyaux dans les myotubes et donc de calculer l'indice de fusion. La variation significative de l'un ou de plusieurs de ces trois paramètres par rapport à ceux obtenus dans les cellules contrôles indique que la fusion cellulaire et/ou la maturation des myotubes ont été affectées au cours de l'étape de différenciation des cellules inactivées pour les gènes candidats.

Les pourcentages des aires de fluorescence (DAPI et myosine), l'indice de fusion ainsi que la taille et la forme des myotubes, ont été calculés d'une façon systématique sur toutes les images de reconstructions prises sur les cellules inactivées des 10 expériences et sur leurs contrôles associés. Toutes les analyses comparatives ont été réalisées entre les données obtenues à partir des cellules inactivées de chaque expérience et celles obtenues des contrôles associés. Ces analyses ont été effectuées en utilisant le test de *student* et la correction de *Dunnett*. Le test de significativité est déterminé selon une *P-value* (voir matériel et méthode).

5- Comment analyser les phénotypes pour classer les gènes ?

L'analyse phénotypique et la classification des 100 gènes candidats ont été réalisées selon les phénotypes observés avec chacun des deux siRNAs utilisés et en fonction des critères phénotypiques décrits dans le paragraphe précédent : « densité cellulaire », « quantité de myotubes », « dimensions et morphologie des myotubes » et « indice de fusion ». Nous avons effectué cette analyse en trois étapes : (i) éliminer tous les gènes dont l'inactivation par les deux siRNAs ne montre aucun phénotype ; (ii) analyser le reste des gènes en quatre groupes de gènes en fonction des quatre critères phénotypiques puis exclure dans chaque groupe les gènes pour lesquels l'inactivation par les deux siRNAs ne montre pas le même phénotype ; (iii) regrouper les différents gènes restants dans les quatre groupes (en excluant ceux qui sont dupliqués) en deux ensembles de gènes selon qu'ils sont impliqués dans la prolifération ou la différenciation **(Tableau 4)**.





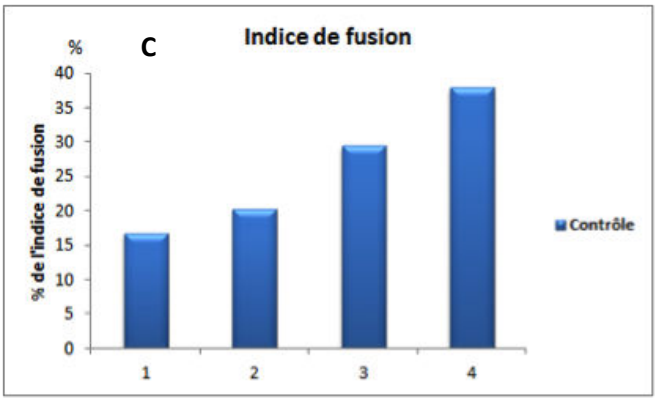


Figure 30 : Variation de la densité cellulaire, de la quantité des myotubes et de l'indice de fusion dans les contrôles.

Les quantifications sur les contrôles montrent que (A) les noyaux recouvrent 10 à 32% de la surface scannée ; (B) les myotubes occupent entre 5 à 20 % de cette surface et (C) l'indice de fusion fluctue entre 16 et 38 %.

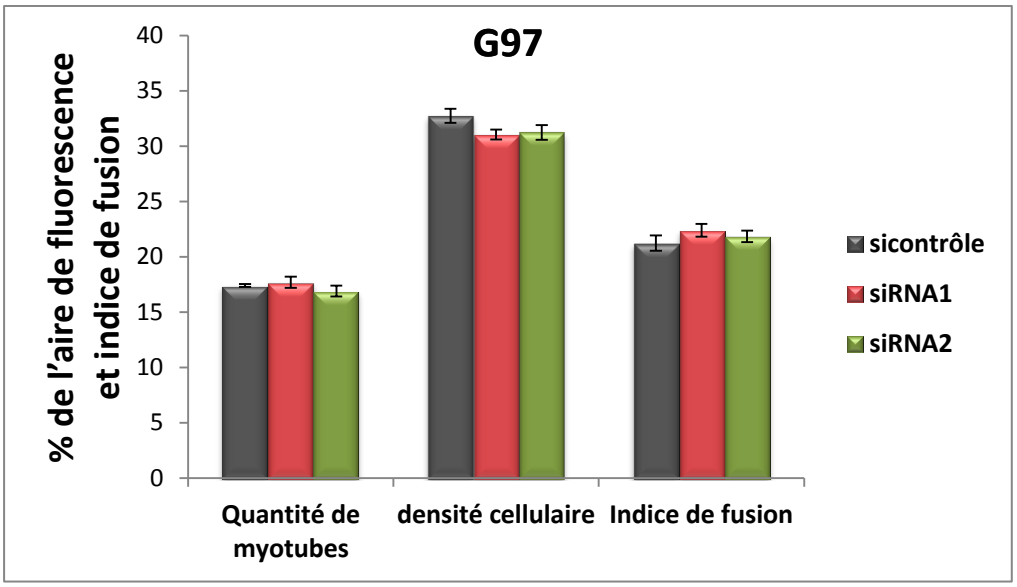
II. Conséquences phénotypiques de l'inactivation des 100 gènes candidats.

Comme mentionné précédemment, les 100 gènes candidats ont été inactivés en 10 expériences différentes, deux siRNAs par gène en trois réplicas techniques avec pour chaque expérience les contrôles nécessaires. Pour chaque expérience, 66 lamelles sont récupérées pour marquage et acquisition d'images. Les pourcentages d'aires de fluorescence (DAPI et myosine), les dimensions et la morphologie des myotubes, et l'indice de fusion ont été calculés d'une façon systématique sur les 60 grandes images correspondantes aux inactivations par les siRNAs et comparées aux données obtenues à partir des six conditions de contrôles. Ceci nous fait pour les 100 gènes candidats du crible, 600 grandes images comparées aux 60 images contrôles.

Les cellules inactivées et les contrôles ont été récupérés à 144 heures de différenciation. À ce stade final, les myotubes différenciés dans les contrôles ont des structures fusiformes, contiennent plusieurs noyaux répartis tout le long du myotube et accumulent de la myosine sur la totalité de la surface. Les quantifications sur ces contrôles montrent que (i) les noyaux recouvrent 10 à 32% **(Figure 30 A)** de la surface scannée ; (ii) les myotubes occupent entre 5 à 20 % de cette surface avec 400 à 600 µm de longueur, de 30 à 50 µm de largeur et 10 et 20 noyaux par myotube **(Figure 30 B)** et (iii) l'indice de fusion quant à lui fluctue entre 16 et 38 % **(Figure 30 C)**. Les variations significatives de toutes ces données dans les myotubes inactivés sont considérées comme des phénotypes dus à l'inactivation des gènes candidats. Pour les pourcentages d'aires de fluorescence et l'indice de fusion, ces variations peuvent être une augmentation ou une diminution de la valeur calculée dans les cellules inactivées par rapport à celle calculée dans les contrôles ; par contre, les variations des dimension et morphologie des myotubes sont classées en quatre groupes phénotypiques en fonction de la taille et la forme du myotube : nous avons pu avoir des myotubes de grande taille, des myotubes longs, des myotubes de petite taille et des myotubes fins.

Trois cas de figures ont été obtenus suite à l'utilisation de deux siRNAs par gène : (i) les deux siRNAs ne montrent aucun phénotype, (ii) les deux siRNAs montrent deux phénotypes différents et (iii) les deux siRNAs montrent le même phénotype. Les gènes qui

A



B

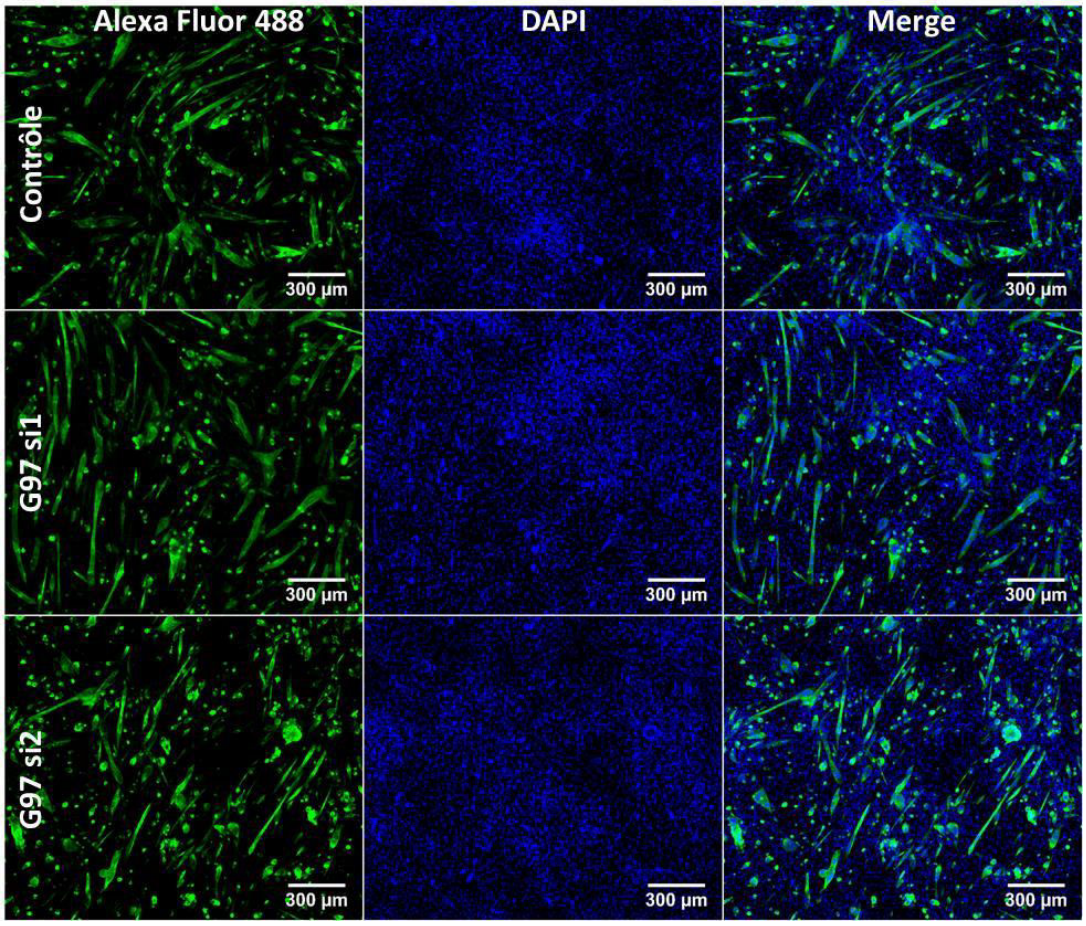


Figure 31 : (A) : Densité cellulaire, quantité de myotubes et indice de fusion des cellules inactivées dans le gène G97 : Le gène ne montre aucune variation significative par rapport aux contrôles (*** p<0,001 ; ** p<0,01 ; * p<0,05).

(B) : Immunomarquage dans des cellules contrôles et des cellules inactivées pour le gène G97 : De la myosine (en vert : Alexa 488), du DAPI (en bleu) et superposition (merge Alexa 488/DAPI). Agrandissement : X100, échelle = 300 µm.

seront sélectionnés par le crible seront ceux pour lesquels les inactivations par les deux siRNAs montrent exactement le même phénotype diffèrent des contrôles.

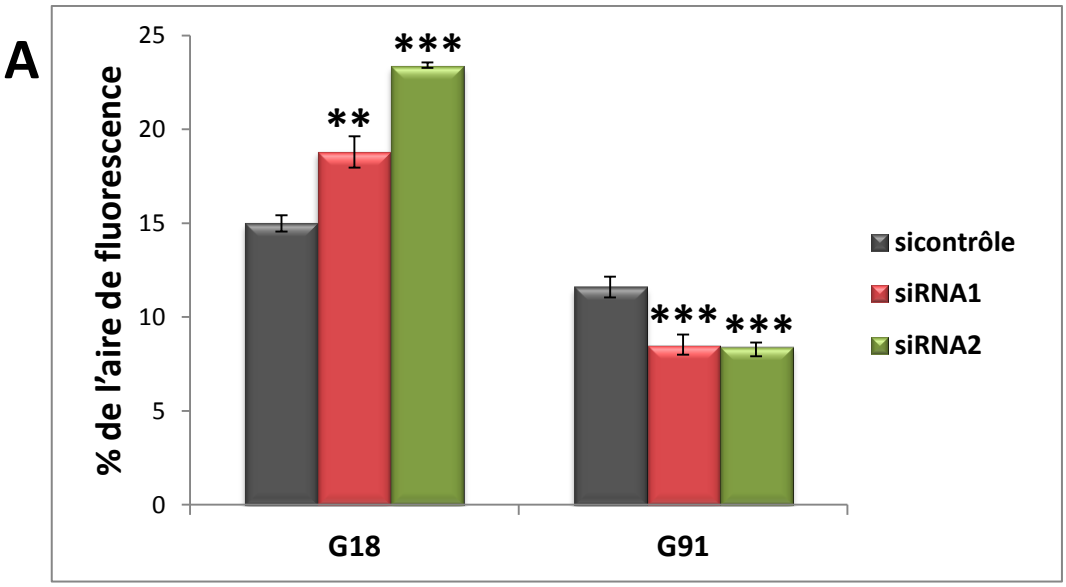
III. Classification des 100 gènes candidats.

Nous avons réalisé la classification des gènes en trois étapes :

La première étape : après l'inactivation des 100 gènes candidats, l'analyse phénotypique nous a révélé cinq gènes dont le Knockdown par les deux siRNAs ne montre aucun phénotype particulier **(Tableau 4)**. Toutes les données quantitatives, obtenues suite à l'extinction de ces cinq gènes, ne montrent aucune différence significative par rapport aux données obtenues à partir des contrôles. Les cinq gènes sont : G31, G69, G51, G97 et G16 **(Figure 31 A et B)**.

La deuxième étape : Les 95 gènes restants ont été analysés critère par critère en quatre groupes. Nous avons exclus les gènes dont l'inactivation par les deux siRNAs montre des phénotypes différents ou sauvages et gardé ceux pour lesquels l'inactivation par les deux siRNAs montre le même phénotype par rapport au critère en question. Après cette opération, nous avons retenu respectivement 23, 22, 24 et 18 gènes pour les critères phénotypiques : « densité cellulaire », « quantité de myotubes », « dimensions et morphologie des myotubes » et « indice de fusion » **(Tableaux 4)**.

La troisième étape : les gènes de chaque groupe, retenus dans l'étape précédente, montrent le même phénotype avec les deux siRNAs par rapport au critère phénotypique analysé. Ces mêmes gènes peuvent montrer des phénotypes différents avec les deux siRNAs quand ils sont analysés par rapport à un autre critère phénotypique. Par exemple, parmi les 23 gènes sélectionnés par le critère phénotypique « densité cellulaire », 17 gènes montrent des phénotypes différents avec les deux siRNAs pour les autres critères phénotypiques tandis que pour les 6 autres gènes, les phénotypes étaient similaires pour les deux siRNAs. Nous avons réalisé la même analyse sur tous les gènes des quatre groupes et nous avons donc retenu respectivement 6, 8, 8 et 9 gènes pour les critères phénotypiques : « densité cellulaire », « quantité de myotubes », « dimensions et morphologie des myotubes », et « indice de fusion ». Nous avons ensuite subdivisé ces groupes de gènes en sous-groupes en fonction de la nature du phénotype observé. Les groupes de gènes relatifs aux critères



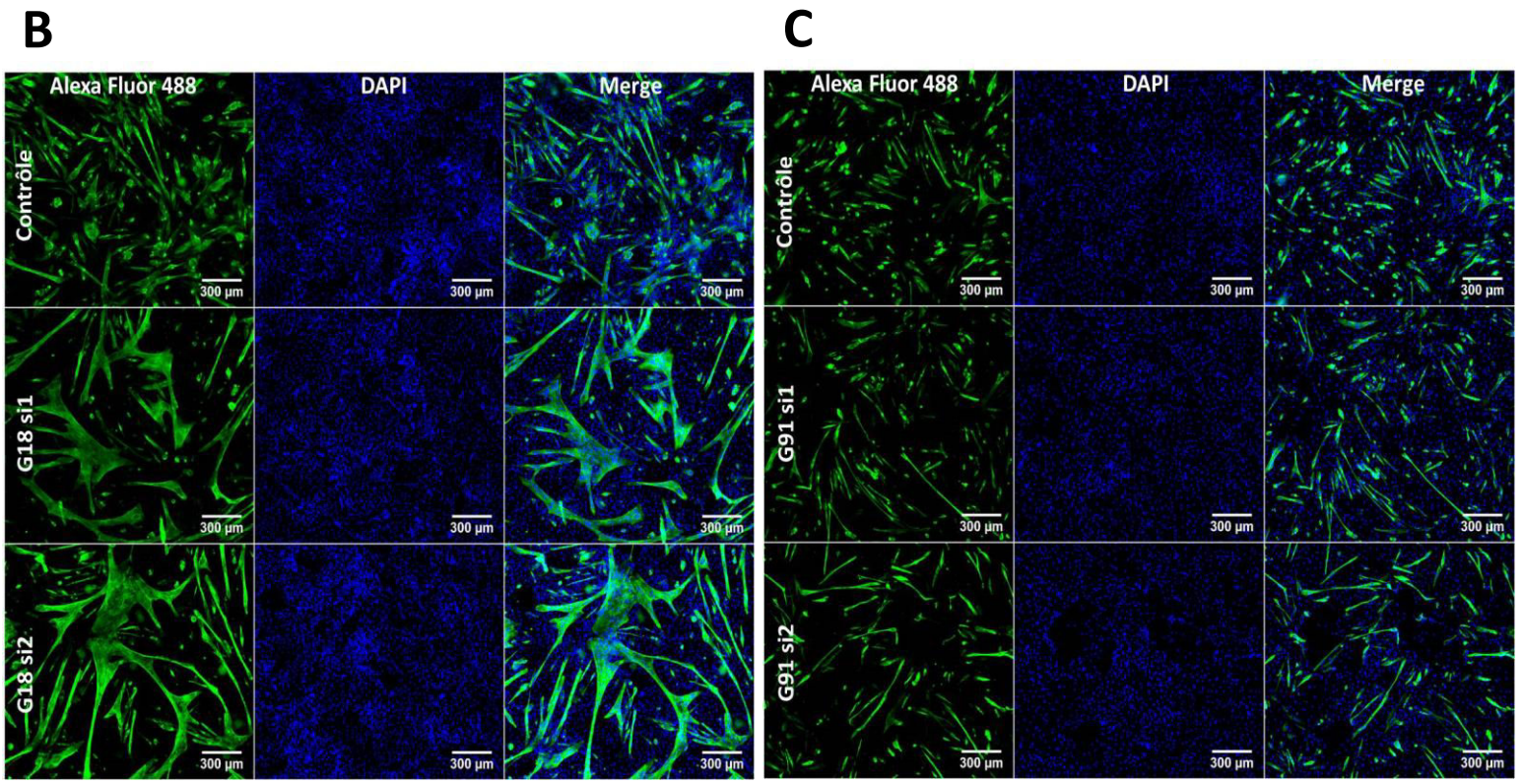


Figure 32 : Analyse de la quantité de myotubes.

Exemple de 2 phénotypes observés (A) : L'extinction du gène G18 montre une augmentation de la quantité de myotubes et celle du gène G91 montre une diminution de la quantité de myotubes. (*** p<0,001 ; ** p<0,01 ; * p<0,05).

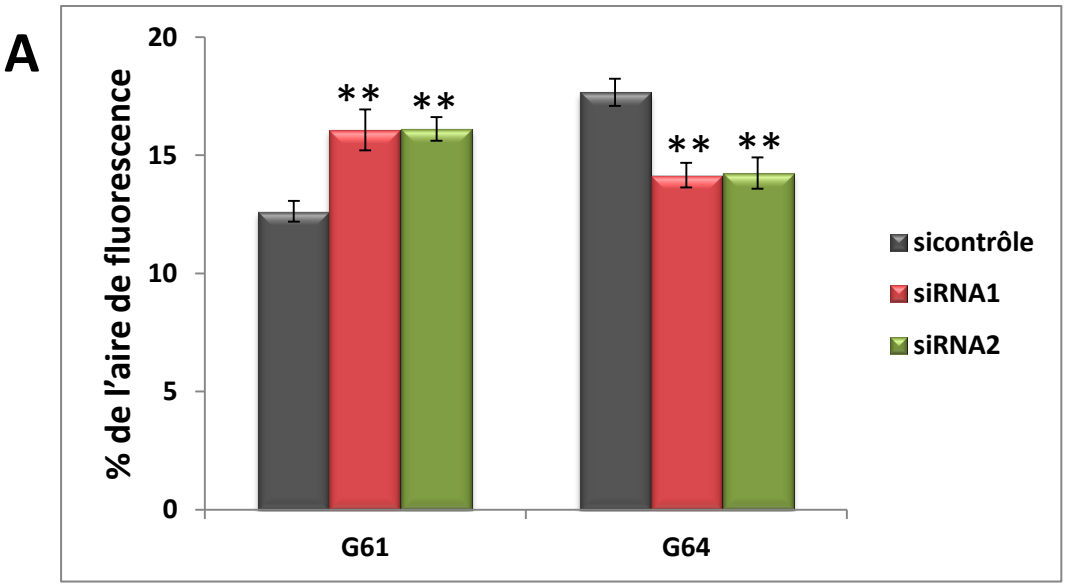
Immunomarquage dans des cellules contrôles et des cellules inactivées pour les gènes G18 (B) et G91 (C) : de la myosine (en vert : Alexa 488), du DAPI (en bleu) et superposition (merge Alexa 488/DAPI). Agrandissement : X100, échelle = 300 µm.

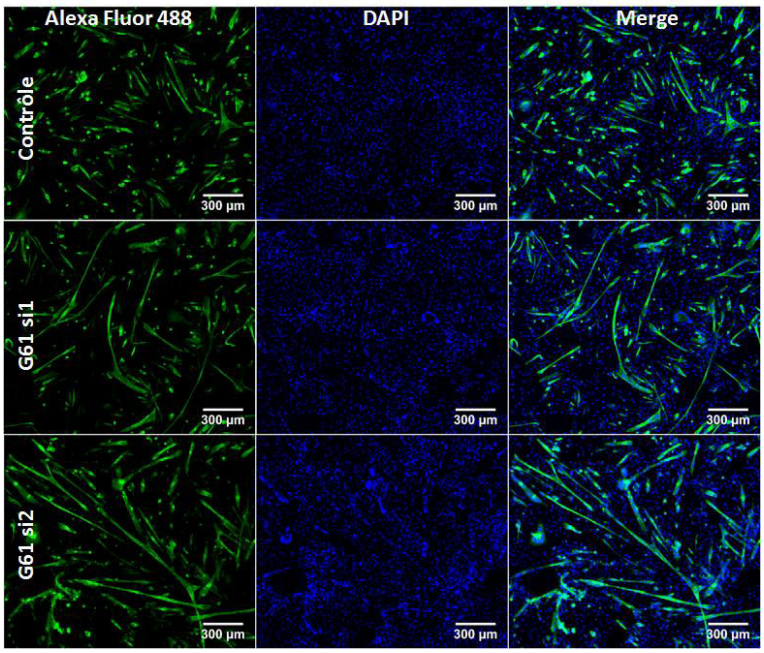
phénotypiques « densité cellulaire », « quantité de myotubes » et « indice de fusion », ont été subdivisés en deux sous-groupes (augmentation ou diminution) **(Tableau 4) (Figure 32, 33, 35)**. Par contre, le groupe de gènes, relatif au critère phénotypique « dimensions et morphologie des myotubes », a été subdivisé en quatre groupes **(Figure 35)**.

Pour avoir la liste finale des gènes sélectionnés parmi les 100 gènes candidats, les gènes retenus dans l'étape précédente ont été groupés tout en éliminant les doublons. Cette étape nous a permis de sélectionner 19 gènes dont le Knockdown par les deux siRNAs montre le même phénotype pour un ou plusieurs critères **(Tableau 4)**.

IV. Le crible révèle 13 gènes impliqués dans la différenciation et six gènes impliqués dans la prolifération et la différenciation

Les gènes pour lesquels l'inactivation montre un phénotype inclus dans le critère « densité cellulaire», sont impliqués dans l'étape de prolifération des cellules C2C12. Par contre, les gènes pour lesquels l'inactivation aboutie à l'apparition des phénotypes faisant partie des critères « quantité de myotubes », « indice de fusion » et/ou « dimension et morphologie des myotubes » sont plutôt impliqués au cours de l'étape de différenciation des C2C12. Enfin, les gènes dont les phénotypes d'inactivation s'inscrivent à la fois dans le critère « densité cellulaire » et au moins dans l'un des trois autres critères auront un rôle durant les deux étapes de développement de la lignée C2C12. Tenant compte de ce type de classification, on obtient à la fin de notre crible six gènes qui peuvent avoir un rôle aussi bien durant la prolifération que la différenciation des C2C12 et 13 gènes qui semblent avoir un rôle uniquement au cours de l'étape de la différenciation **(Tableau 4)**.





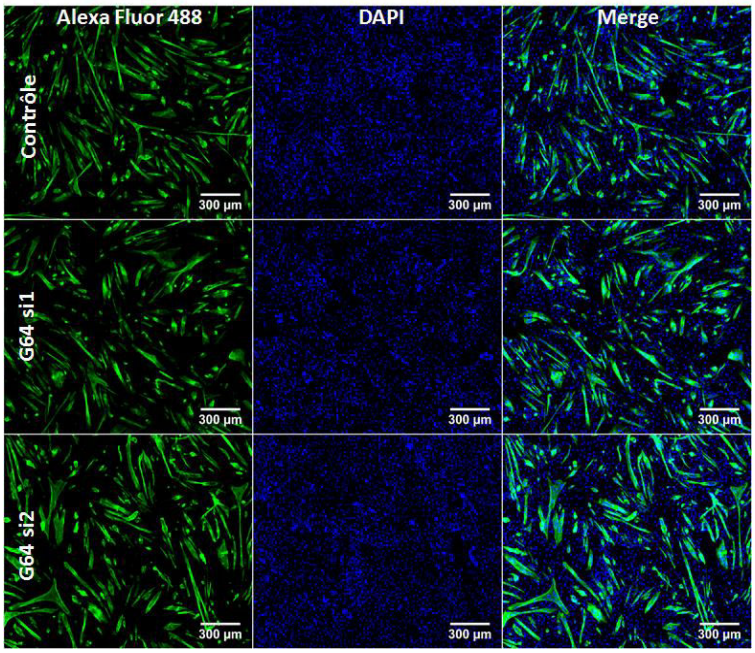


Figure 33 : Analyse de la densité cellulaire.

Exemple de 2 phénotypes observés (A) : L'extinction du gène **G61** montre une augmentation de la densité cellulaire et celle du gène **G64** montre une diminution de la densité cellulaire. (*** p<0,001 ; ** p<0,01 ; * p<0,05).

Immunomarquage dans des cellules contrôles et des cellules inactivées pour les gènes G61(B) et G64 (C) : de la myosine (en vert : Alexa 488), du DAPI (en bleu) et superposition (merge Alexa 488/DAPI). Agrandissement : X100, échelle = 300 µm.

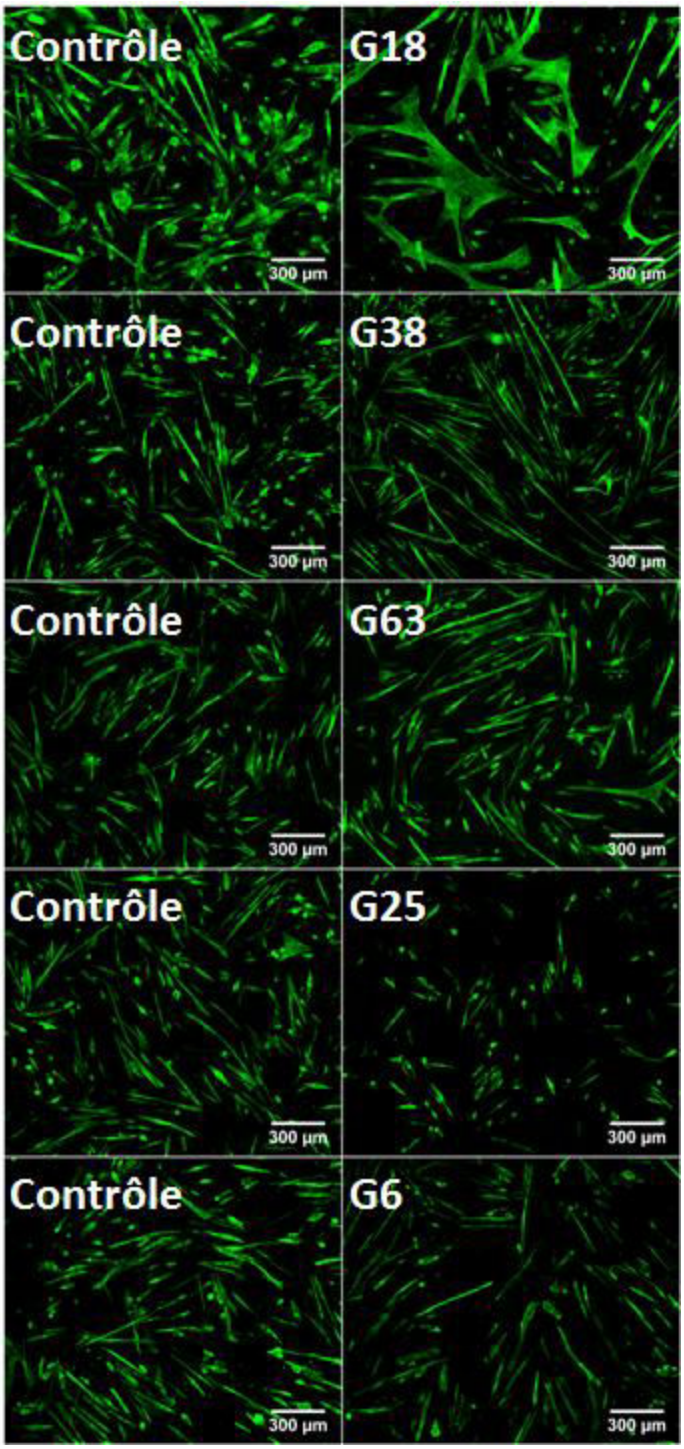
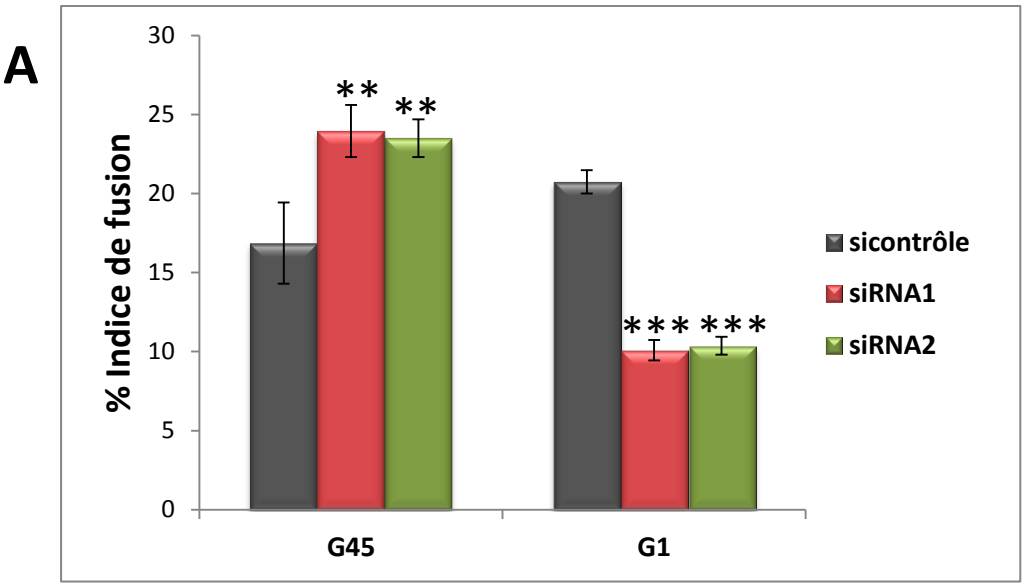


Figure 34 : Analyse de la morphologie des myotubes par immunofluorescence.
L'inactivation des gènes : **G18** : myotubes de grande taille et croisés ; **G38** : myotubes longs et fins ; **G63** : myotubes longs ; **G25** : myotubes petits et ronds ; **G6** : myotubes identiques aux contrôles (Échelle: 300 µm).



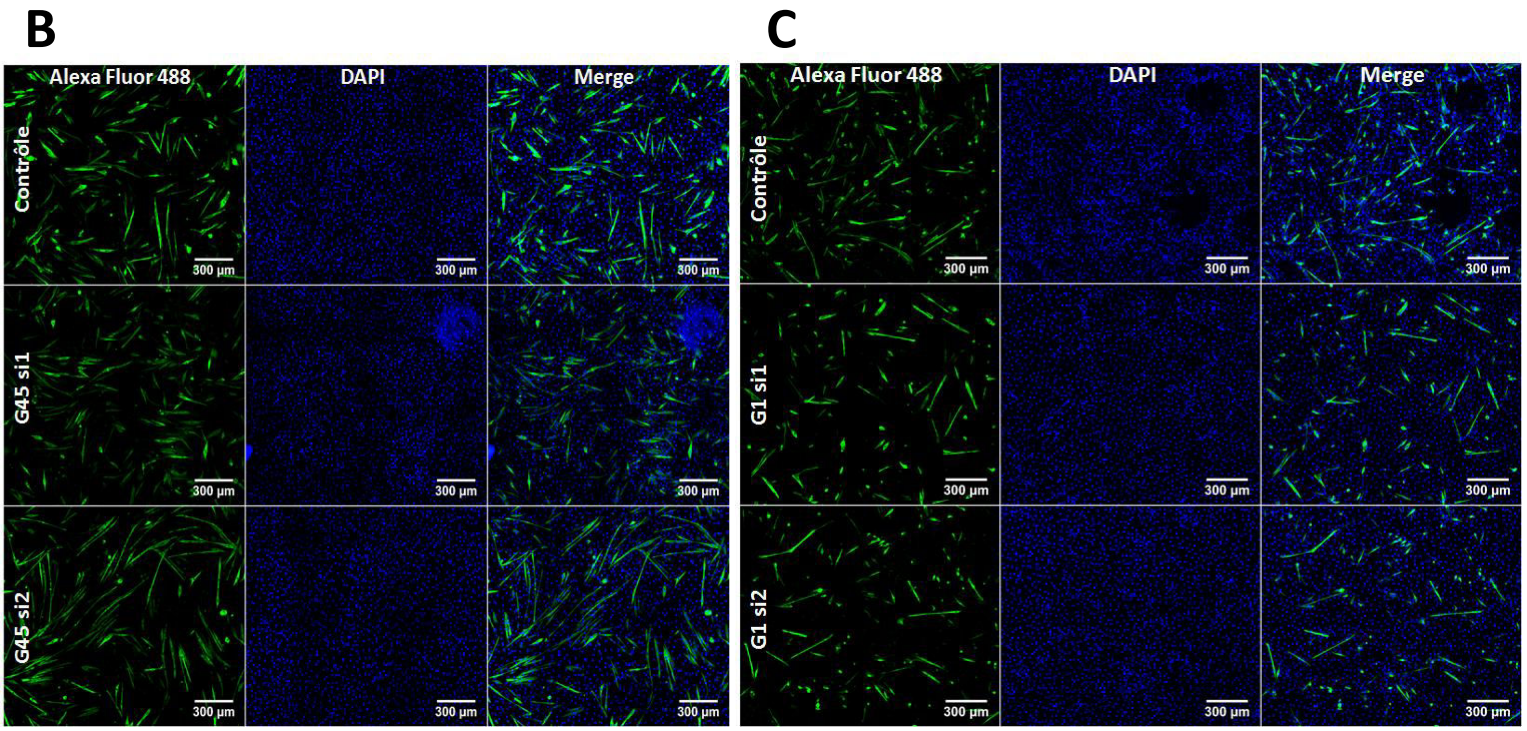


Figure 35 : Analyse de l'indice de fusion.

Exemple de 2 phénotypes observés (A) : le knockdown du gène G45 montre une augmentation de l'indice de fusion par rapport aux contrôles alors que celui du gène G1 montre une diminution. (*** p<0,001 ; ** p<0,01 ; * p<0,05).

Immunomarquage dans des cellules contrôles et des cellules inactivées pour les gènes G45 (B) et G1 (C) : de la myosine (en vert : Alexa 488), du DAPI (en bleu) et superposition (merge Alexa 488/DAPI). Agrandissement : X100, échelle = 300 µm.

DISCUSSION GENERALE ET PERSPECTIVES

Au cours de ma thèse, je me suis attachée à identifier des gènes impliqués dans le développement musculaire en mettant en place un crible fonctionnel basé sur l'ARNi. En effet, le criblage *in vivo* d'un grand nombre de gènes candidats jouant un rôle dans le muscle chez les mammifères apparaît être une tâche expérimentale fastidieuse qui nécessiterait la création d'un grand nombre de lignées pour lesquelles les gènes candidats sont inactivés. Les cellules musculaires primaires, qui ont été largement utilisées dans l'étude des mécanismes moléculaires régissant les processus biologiques liés aux muscles et les altérations pathologiques, peuvent être une alternative. Cependant, leur utilisation est assez restreinte en raison du nombre limité de passages cellulaires, de la difficulté liée à leur production et au degré de pureté des cultures. Malgré ces facteurs limitants, un crible à haut débit à plusieurs étapes capable d'identifier des gènes jouant un rôle durant la myogenèse a été réalisé avec succès sur des cellules primaires murines dérivées du muscle des membres **(Khanjyan et al., 2013)**. Cette étude a été effectuée d'une part, dans le but de mieux comprendre les mécanismes qui régissent la transition des cellules progénitrices en cellules musculaires différenciées et d'autre part, pour développer de nouvelles stratégies qui pourraient améliorer la régénération musculaire. Ces cribles successifs, réalisés en utilisant une banque de 16000 siRNAs conçues pour cibler près de 4000 gènes de la souris, ont permis d'identifier 635 gènes impliqués dans la myogenèse **(Khanjyan et al., 2013)**.

Nous avons fait le choix d'utiliser la lignée myoblastique murine C2C12 qui représente le modèle le plus pertinent pour l'étude des phases de prolifération et/ou de différenciation au cours de la myogenèse. Les cellules C2C12 à confluence à 80% sont engagées dans la différenciation par simple changement de milieu, avec cependant un laps de temps de 12h durant lequel elles continuent à proliférer. Ces cellules sont aussi moins sensibles et moins fragiles que les cellules primaires et facilement transfectables. Afin de faciliter les analyses microscopiques et de quantification, nous avons mis au point un protocole de culture basé sur l'utilisation de lamelles de 1,8 cm^2 de surface placées au fond des puits de plaques de culture (la lamelle représente 50% de la surface du puits) avec un nombre adapté de cellules ensemencées. Ceci nous a permis d'optimiser l'étape d'acquisition des images ainsi que leur quantification. Ceci nous a permis de prendre des images (scanner) pour presque la totalité de la lamelle et, en utilisant le logiciel *Metamorph*,

de construire une grande image de 1cm^2 de surface de façon rapide et valide. Ces grandes images sont utilisées pour quantifier la fluorescence, étudier la morphologie des myotubes et calculer l'indice de fusion. Il est à noter que les cellules C2C12 ont été transfectées avec les siRNAs au moment de l'engagement dans la différenciation et 48h après, l'objectif étant de laisser les cellules dans le milieu d'inactivation durant les dernières 12h de prolifération et toute la phase de différenciation.

Nous avons préféré éteindre les gènes candidats par transfection transitoire de duplex siRNAs à la place de générer des clones stables en utilisant de shRNAs afin d'éviter les difficultés techniques liées au temps de sélection et à l'effet de position associé à l'intégration du plasmide hébergeant le shRNA. Cependant, pour assurer l'efficacité en continu de l'inactivation par les siRNAs durant toute la phase de différenciation, nous avons transfecté les cellules deux fois avec le même siRNA à 48h d'intervalle. Nous avons réalisé l'inactivation des 100 gènes candidats sur 10 expériences distinctes, chacune comprenant 10 gènes et ses propres contrôles afin de minimiser l'effet de la variabilité expérimentale. Pour éliminer les faux positifs, 2 siRNAs différents ciblant soit la région codante, soit la partie 3'UTR ont été utilisés pour chaque gène candidat.

Initialement, nous avons établi une liste de 100 gènes candidats à inactiver remplissant les conditions suivantes: (i) ils n'ont pas de fonction déjà caractérisée dans le muscle chez la souris, (ii) ils présentent un profil d'expression différent entre les étapes de prolifération et de différenciation des cellules C2C12 **(Tomczak et al., 2004; Kislinger et al., 2005)**, (iii) ils sont des orthologues de gènes muscle spécifiques identifiés lors d'un crible ARNi chez la drosophile**(Schnorrer et al., 2010)**. Nous avons mis en place ces filtres successifs afin de premièrement, valider le protocole de notre crible, puis identifier de nouveaux acteurs impliqués dans la prolifération et/ou différenciation musculaire et enfin, caractériser les phénotypes associés à la mise en silence de chaque gène.

La caractérisation phénotypique dans notre crible a été réalisée grâce à l'utilisation de deux marqueurs cellulaires fluorescents (DAPI et anti-myosine) qui décorent respectivement les noyaux et les myotubes. La quantification de la surface fluorescente nous a permis de définir et d'analyser quatre critères phénotypiques : « densité cellulaire », « quantité de myotubes », « index de fusion », et « taille et morphologie des myotubes ». Le

premier critère s'observe pendant la phase de prolifération tandis que les autres se déroulent durant la différenciation.

Nos résultats montrent la pertinence de notre crible puisque nous avons pu associer des phénotypes à plus de la moitié des gènes candidats inactivés (60 gènes). Ces soixante gènes, retenus sur la base d'un phénotype similaire pour les 2 siRNAs, se répartissent comme suit : 11 gènes impliqués dans la phase de prolifération, 37 gènes dans la phase de différenciation et 12 gènes concernent les deux phases. En gardant uniquement les gènes dont le knockdown par les deux siRNAs montre le même phénotype pour un ou plusieurs critères, nous avons identifié 13 gènes spécifiques à l'étape de différenciation et 6 gènes impliqués aussi bien dans la prolifération que la différenciation. L'identification des gènes impliqués dans la prolifération malgré que la transfection a été réalisée au début de la différenciation s'explique par le fait que les cellules C2C12 continuent à se diviser pendant au moins 12 heures après l'induction de la différenciation par le changement du milieu de culture.

Pour les gènes exclus, nous avons 5 gènes montrant aucun phénotype avec les 2 siRNAs et 35 gènes avec uniquement un phénotype avec l'un des deux siRNAs. Trois hypothèses pourraient expliquer cette absence de phénotypes: (i) l'inactivation incomplète du gène cible due aux siRNAs partiellement ou totalement non fonctionnels, (ii) la redondance fonctionnelle entre les gènes candidats et d'autres gènes appartenant à la même famille multigénique, (iii) l'absence de contribution de ces gènes dans la myogenèse malgré les filtres utilisés pour établir la liste des gènes à inactiver.

A notre connaissance, l'étude menée par Ge *et al.* 2013 pour examiner les rôles potentiels des cytokines dans la myogenèse du muscle squelettique, représente le seul crible réalisé sur les cellules C2C12 pour identifier de nouveaux acteurs myogéniques **(Ge et al., 2013)**. Les auteurs ont inactivé 134 gènes codant pour des cytokines au travers d'un crible ARNi « focalisé » pour mettre en évidence leur implication dans la différenciation des myoblastes C2C12. Le crible a été réalisé en deux étapes : (i) dans la première, les 134 gènes ont été inactivés par 2 à 5 shRNAs introduits par infection retrovirale dans les cellules C2C12 ensemencées dans des plaques à 96 puits. Au bout de 3 jours de différenciation, les cellules sont immunomarquées par l'anticorps anti-myosine et analysées au microscope pour des changements dans la taille et le nombre de myotubes. Les gènes avec au moins 2 shRNAs

provoquant ces changements sont sélectionnés pour la deuxième étape du crible. (ii) dans la deuxième étape, le même protocole est appliqué sauf que les cellules, cette fois-ci, sont ensemencées dans des plaques à 6 puits pour pouvoir les quantifier pour trois paramètres: « indice de différenciation » défini comme le pourcentage de noyaux dans les cellules myosine-positive, « indice de fusion » défini comme le pourcentage de noyaux dans des cellules contenant au moins 2 noyaux (myotubes) et « nombre moyen de noyaux par myotube » qui donne une idée sur la taille de myotube. Leurs résultats ont révélé 29 cytokines (22%) comme des candidats solides pour de nouveaux régulateurs myogéniques **(Ge et al., 2013)**. Notre étude est pour l'essentiel comparable à celle de Ge et al., 2013. En effet, les deux études sont basées sur l'utilisation des cellules C2C12, de l'ARNi, de l'analyse microscopique et de l'imagerie. Cependant, notre étude, basée sur un crible non focalisé et moins fastidieux, se montre différente de celle de Ge *et al.* 2013 sur plusieurs points :

(i) Pour mettre en silence nos gènes candidats, nous avons utilisé des siRNAs synthétiques facilement transfectables dans les cellules C2C12. En revanche, Ge et al., ont choisi d'utiliser des shRNAs introduits par transduction retrovirale dans la lignée C2C12. Il est à noter que la transduction des shRNAs nécessite des constructions plasmidiques, une production au préalable dans un autre type cellulaire et une sélection par antibiotique.

(ii) Nous avons examiné l'effet du knockdown des gènes candidats après six jours de différenciation contre trois jours dans le protocole de Ge et al., ceci nous a permis de détecter des phénotypes associés à des gènes agissant durant les étapes précoce et/ou tardive de la différenciation. Par ailleurs, dans l'étude Ge et al., les auteurs ne discutent en aucun moment l'inactivation génique qui peut avoir lieu juste après l'étape de transduction pendant laquelle les cellules sont encore en phase de prolifération.

(iii) Contrairement au crible de Ge et ses collaborateurs, nous avons réalisé notre crible en une seule étape dans des plaques à 12 puits. Il faut savoir que la détection des phénotypes par microscopie est moins efficace en plaque à 96 puits qu'en plaque à six puits. De plus, l'analyse en plaque à 96 puits est souvent

associée à un grand nombre des faux positifs et un faible nombre de phénotypes subtils détectés.

(iv) Pour détecter d'une façon rapide et efficace ces phénotypes subtils, nous avons quantifié systématiquement la fluorescence sur des grandes images de reconstruction de 1 cm^2 grâce à la technologie « *platine scanning* » du microscope. En revanche, l'analyse d'imagerie dans Ge *et al.* a été réalisée uniquement sur cinq champs d'une image individuelle.

(v) Dans Ge *et al.* la quantification de la fluorescence avait pour but uniquement la détermination du nombre de noyaux totaux, dans les myocytes ou dans les myotubes. Par contre, dans notre étude, la quantification de la fluorescence a été utilisée pour dénombrer les noyaux mais aussi pour déterminer la quantité des myotubes ainsi que leur morphologie ; ceci ce nous a permis de détecter des phénotypes liés à l'architecture et/ou la maturation des myotubes.

(vi) Le crible dans l'étude Ge *et al.* était focalisé sur une liste de 134 gènes candidats codant pour des cytokines. Par contre, nous avons criblé 100 gènes candidats présélectionnés sur la base d'une compilation de données génomiques réalisées sur la lignée C2C12 et affinés par une comparaison avec des gènes jouant un rôle dans le muscle de la drosophile.

Tous ces points de différence entre les deux études indiquent que notre crible semble être plus efficace. En effet, notre crible a pu sélectionner 60 % des gènes criblés comme étant impliqués dans la prolifération et/ou la différenciation des cellules C2C12 contre 22 % sélectionnés dans le crible de Ge *et al.* (2013).

Des analyses transcriptomiques ont été également effectuées pour identifier (i) les gènes régulés par l'inhibiteur négatif du développement musculaire, GDF8 **(Wicik et al., 2011)**, (ii) les régulateurs transcriptionnels impliqués dans la transition entre les étapes de prolifération et de différenciation des cellules C2C12 **(Rajan et al., 2012)** ou (iii) les gènes qui s'expriment en présence d'adipocytes dans le microenvironnement des myoblastes **(Grabiec et al., 2016)**. Outre les analyses transcriptomiques, les études protéomiques ont été également conduites pour examiner de manière exhaustive les adaptations au niveau protéique associées à la formation de myocytes tout le long du programme de différenciation myogénique des

cellules C2C12 **(Casadei et al., 2009; Deshmukh et al., 2015)**. Les études basées sur la technique d'immuno-précipitation de la chromatine (ChIP), qui consiste à purifier des complexes ADN-protéine formés dans les cellules et liés de façon covalente, ont été aussi employées. Cette technique associée à une analyse transcriptomique par *microarray* ou séquençage à haut débit a été utilisée pour identifier de nouveaux gènes qui peuvent être soit des cibles d'un activateur transcriptionnel comme dans le cas de MyoD, YY1 ou P38 **(Cao et al., 2010; Lu et al., 2013; Segalés et al., 2016; Sun et al., 2016a)** soit régulés par une marque épigénétique telle que la marque répressive de l'expression génique : la triméthylation de la lysine 27 sur l'histone 3 **(Asp et al., 2011)**.

Comme décrit précédemment, les cribles ENU sont capables par une analyse systématique et simultanée de nombreuses souris, de récupérer un grand nombre de mutations fonctionnelles **(Nguyen et al., 2011)**. Comme précédemment mentionné, avant d'être accidentellement interrompu à la fin de l'année 2009, un crible ENU a été développé au sein de notre laboratoire afin d'identifier de nouveaux gènes impliqués dans la système squeletto-musculaire ; c'était un crible dominant et sensibilisé réalisé à partir de souris knockout pour la myostatine **(Magnol et al., 2011)**. Les auteurs ont pu cribler 3000 souris F1pour identifier 49 phénovariants. Ces lignées présentant des phénotypes liés au taux de créatinine, au développement musculaire et/ou à des anomalies du système squeletto-musculaire. Sept lignées ont été établies et des mutations causales ont été localisées et identifiées pour trois d'entre elles **(Magnol et al., 2011; El Hakam Kamareddin et al., 2015)**. Cependant, ce cible était tout de même un projet très laborieux, demandait un effort organisationnel et logistique assez conséquent et nécessitait énormément de temps. C'est particulièrement l'étape de la cartographie pour localiser les mutations causales, qui représente le processus le plus long du crible. Des essais d'identifier les mutations causales des autres lignées par séquençage à haut débit sont en cours dans le laboratoire. Contrairement aux cribles ARNi, les cribles ENU ne nécessitent aucune connaissance préalable de la fonction ou de la séquence des gènes étudiés et s'appliquent avec plus de facilité sur un organisme entier chez les mammifères **(Oliver and Davies, 2012)**. Les cribles ARNi, générant une sous-expression des gènes candidats (knockdown), conduisent à une suppression partielle des protéines ciblées, ce qui complique l'interprétation des phénotypes ; par contre, les mutagenèses ENU peuvent produire les différents types

d'allèles tels que les substitutions d'acides aminés, les changements de conformation de domaines protéiques ou encore les modifications des produits de l'épissage **(Godinho and Nolan, 2005)**. Les allèles produits par mutagenèse ENU révèlent des fonctions géniques distinctes qui ne sont pas identifiables par l'ARNi. En revanche, avec notre crible ARNi, il a été possible d'inactiver fonctionnellement une centaine de gènes avec une spécificité, efficacité et rapidité remarquable et d'évaluer l'implication fonctionnelle de 60 gènes dans le développement myogénique des cellules C2C12. Avec ces avantages, notre crible basé sur l'outil ARNi semble être plus précis et apparaît comme une solution alternative plus efficace, plus rapide, et moins couteuse à la mutagenèse ENU où les modifications sont en grande partie aléatoires et difficile à identifier.

Notre crible a été validé en identifiant 60 gènes impliqués dans la myogenèse des cellules C2C12, 19 parmi eux sont sélectionnés parce que leur inactivation montre le même phénotype avec les deux siRNAs pour plusieurs critères phénotypiques. Il me semble indispensable maintenant de vérifier le knockdown de ces 19 gènes sélectionnés par quantification de leurs ARNm par RTqPCR et voire même les 60 gènes. Il serait intéressant aussi de refaire les expériences d'inactivation sur tous les gènes pour lesquels celle-ci n'a montré aucun phénotype ou des phénotypes différents avec les deux siRNAs, et ce en employant d'autres siRNAs dirigés contre ces gènes.

Il semble aussi indispensable que les 19 gènes identifiés par notre crible comme étant de nouveaux acteurs de la myogenèse soient invalidés *in vivo* chez la souris pour voir l'effet de leur perte de fonction directement sur le muscle voire sur l'organisme entier. Ce travail est en cours pour un certain nombre des gènes jugés pertinents et repose sur des approches de transgénèse somatique (injection intramusculaire des vecteurs lentiviraux exprimant les siRNA) ou germinale (système *CRISPR/cas9*) (L Helary, travaux de thèse, communication personnelle). Cette validation fonctionnelle *in vivo* d'un certain nombre des gènes identifiés confirmera la puissance de l'approche RNAi que nous avons développée.

PUBLICATION

ALWAN et al., soumis à *Experimental Cell Research*

An siRNA-based screen in C2C12 myoblasts identifies novel genes involved in myogenic differentiation

Rayan Alwan[a], Ange-Line Bruel[b], Anne Da Silva[a], Véronique Blanquet[a]*, Khaled Bouhouche[a]*.

[a] University of Limoges, INRA, UMR 1061, UGMA, F- 87000 Limoges, France

[b] University of Burgundy, INSERM, UMR 1231, GAD, F-1079, Dijon, France.

*These authors contributed equally to this work.

Corresponding Author: e-mail address: veronique.blanquet@unilim.fr

Abstract

Myogenesis is a highly regulated multi-step process involving myoblast proliferation and differentiation. Although studies over the last decades have identified several factors governing these distinct major phases, many of them are not yet known. In order to identify novel genes, we took advantage of the C2C12 myoblastic line to establish a functional siRNA screen combined with quantitative-imaging analysis of a large amount of differentiated myoblasts. We knocked down 100 preselected mouse genes without a previously characterized role in muscle. Using image analysis, we tracked gene-silencing phenotypes by quantitative assessment of cellular density, myotube quantity, myotube morphology and fusion index. Our results have revealed six genes involved in both stages of C2C12 myogenesis and 13 genes specific to the differentiation stage.

These findings prove that our RNAi-based screen could be an efficient tool to detect clear and subtle phenotypes allowing the identification of new myogenic regulators in mammals.

Keywords: Screen; siRNA; C2C12; Proliferation; Differentiation; Myogenesis.

1. Introduction

The completion of several genome sequences along with technological advances have tremendously accelerated and renovated the tools to identify new genes and to study their function. The functional analysis of each of the thousand of protein-coding genes provided by genome sequencing is a major goal but remains a significant technical challenge [1]. Several genome-wide mutagenesis strategies have been applied in mammals, in order to analyse gene function in a genome-scale manner [2]. These strategies include two complementary approaches: (i) the “phenotype-driven approach” such as ethyl-nitroso-urea (ENU) mutagenesis screen [3], which has been successfully used for generating large numbers of new mutants for systematic studies of mammalian gene function [4-6] and for the identification and functional characterisation of hundreds of disease-related genes in mice [7], (ii) the “gene-driven approach” which is a high-throughput method used to introduce insertional mutations across the mammalian genome [8]. Several knockout programs were developed to mutate all protein-encoding genes in mouse embryonic stem cells using a combination of gene trapping and gene targeting techniques to generate mice harboring the corresponding genetic changes [9, 10]. To accelerate progress towards the genetic analysis of all mammalian genes, large-scale mutagenesis [11], knockout [12, 13] and phenotyping [14, 15] consortia were established worldwide to generate a wealth of genomic information which are made publicly available to support diverse research efforts, related to understanding mammalian biology and human diseases [16-19]. Despite their strengths, phenotype- and gene-driven approaches remain an important challenge and an experimentally tedious task [20].

Genomic resources such as transcriptomic and proteomic data enable both the simultaneous measurement of the expression of multiple genes and the identification of novel genes or proteins that contribute to the regulation of cellular processes. For instance, to identify new genes and to investigate the global gene expression program of muscle development, several meta-analysis of transcriptomic [21, 22], proteomic [23-25] and chromatin immuno-precipitation [26, 27] datasets have been performed especially in *in-vitro* mammalian systems. Based only on gene expression patterns, the novel genes identified in genomic resources need further experiments to explore their functions.

The field of functional genomics has been re-energized by the discovery of the RNA interference (RNAi) pathway. RNAi is a regulatory mechanism using small RNAs as

specificity factors, recognizing target sequences by base-pairing interactions inducing target mRNA destruction or translation repression [28]. RNAi is currently a valuable gene-silencing tool both in cell culture and in living organisms [29], also used for large-scale screens that can systematically knockdown every single gene within a genome in a relatively short period of time [30].

Driven by genomic sequences, RNAi silencing has been used in high-throughput screening aiming at the identification of muscular-specific genes in lower eukaryotes such as *Drosophila and worms* [31-33]. Currently, aside from some RNAi screening in cell transplant–based models of cancer in mouse [34], the *in-vivo* high-throughput screens seem to be difficult to practice generally in mammals and especially in muscle mostly because of low efficiency and high expense [30]. Nevertheless, high-throughput RNAi screens could be used on mammalian primary cultures [35] and cell lines [36] such as C2C12 myoblasts [37, 38] which is often used for deciphering the basic molecular mechanisms of muscle-related biological processes. Differentiation of C2C12 myoblasts *in vitro* can easily recapitulate the different steps of *in-vivo* myogenesis [39], a multi-step process enabling generation of muscle tissue during embryonic development and postnatal life [40]. The major myogenic phases, proliferation, differentiation and fusion of myoblasts into myotubes, are well orchestrated by the myogenic regulatory factors and the molecules regulating them via different signalling cascades [41].

Despite all the knowledge acquired on the molecular mechanisms regulating myogenesis in mammalian pre- and postnatal development, many of the genes involved are still unknown. In the present study, we have used a newly designed protocol based on a cellular system (C2C12 myoblasts), RNAi technology and quantitative image analysis to develop an siRNA-mediated functional screen, aiming identification of novel genes involved in the myogenesis process. We applied the screen on 100 preselected genes, which were transiently knocked down in proliferating and differentiating C2C12 cells. Phenotypic analysis and gene classification show that six genes are involved during both stages of the C2C12 myogenesis and 13 genes are implicated only during the differentiation stage.

2. Materials and methods

2.1 Cell line

C2C12 cell line (ATCC - CRL-1772) is a subclone produced through serial passages of myoblasts cultured from the thigh muscle of C3H mice after a crush injury [42]. They readily proliferate in high-serum conditions and differentiate and fuse in low-serum conditions, forming contractile myotubes and producing characteristic muscle proteins. C2C12 cells were incubated (at 37°C in a 5% CO_2 in air atmosphere) during proliferation in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% (V/V) heat-inactivated foetal calf serum (FBS), 50 units.ml^{-1} penicillin and 50 µg.ml^{-1} streptomycin (Growth Medium, GM). To induce C2C12 differentiation, cells at 80% confluence were shifted to DMEM containing 2% (V/V) horse serum (Invitrogen) (Differentiating Medium, DM). The medium was changed every 48 hours.

For the screen, C2C12 cells were not directly seeded in the well but on a sterile 1.8 cm^2-coverslips placed at the bottom of the well and covering nearly half of its surface. Cells were seeded at 4.10^4 cells/cm^2 in 1 mL GM for 24 hours. Next day, cells were shifted to DM and cultivated for 144 hours. During this differentiation time, cells were transfected twice by siRNA (at 0 and 48 hours) and medium was changed three times (48, 96 and 120 hours). At 144 hours, the coverslips with cells were recovered for immuno-staining.

2.2 siRNA Transfection

All siRNAs used in our study are mouse-specific siRNAs designed and produced by Qiagen, delivered in C2C12 cells with Hiperfect transfection reagent (Qiagen) and target either coding sequence or 3'UTR of the mRNA to be knocked down. 100 genes were silenced in 10 groups of 10 genes with two siRNAs for each gene in three technical replicates (Supplementary data, Table S1). For each group of 10 genes, cells transfected with one siRNA (3'alexafluor647, QIAGEN, AATTCTCCGAACGTGTCACGT) without any mouse target as well as untransfected cells were used as controls. The efficiency of siRNA delivery and silencing was tested by RT-PCR and cell viability assays. We verified that the introduction of these synthetic siRNAs into C2C12 cells can trigger highly efficient gene silencing through degradation of the target mRNA (Supplementary data, Fig. S1). C2C12 transfection was performed according to the manufacturers' protocol (Fast-forward transfection of adherent cells, Qiagen). Briefly, siRNAs were diluted at final concentration of 5 nM in

100 µL of culture medium without serum containing 3 µL of Hiperfect transfection reagent. After 10 min incubation at room temperature, the transfection complex was added drop-wise onto cells with fresh medium. Plates were swirled gently, to ensure uniform distribution of the transfection complex, and incubated under normal conditions.

2.3 Immuno-staining

Cells were washed three times in PBS 1X, fixed during 15 min in 4% paraformaldehyde, washed again three times in PBS 1X and permeabilized with HEPES/Triton X-100 buffer (20 mM HEPES, 300 mM sucrose, 50 mM NaCl2, 3 mM MgCl2 and 0.5% Triton X-100 pH7.4). Then, cells were blocked for 1 hour at room temperature in blocking buffer (10% goat serum, 1% bovine serum albumin (BSA) and 0.1% Triton X-100 in PBS). After a quick wash in 1X PBS containing 0,2% BSA, cells were incubated overnight at 4°C with primary antibody anti-MyHC (Monoclonal Anti-Myosin skeletal, Fast, Clone My-32, Mouse Ascities Fluid, M4276, Sigma-Aldrich) diluted (1/400) in 1X PBS containing 1% BSA. After two washes in BSA 0.2% Tween-20 0.1% PBS 1X, coverslips were incubated for 15 min at 37°C with secondary antibody conjugated to a fluorescent dye (Alexa Fluor® 488 Goat Anti-Mouse IgG (H+L), Invitrogen) diluted (1/1000) in BSA 1% PBS 1X. The staining was completed with three washes in BSA 0,2% PBS 1X Tween-20 0,1%, incubation for 5 min at room temperature in DAPI (1mg/mL diluted in PBS 1X) and three washes in PBS 1X. Coverslips were finally placed back to front on slides with the mounting solution (Mowiol 4-88 12%, glycerol 24%, tris HCL pH8 0,125 M), sealed with nail polish and stored at 4°C in the dark.

2.4 Image acquisition and analysis

Images were acquired and analyzed with the automated Leica DMI6000B inverted epifluorescence microscope using the *MetaMorph* software (Molecular Devices, Sunnyvale, USA). Per coverslip, automated 100 to 150 fluorescent images (DAPI or Alexa Fluor 488) were taken in 100X magnification with the aim to reconstruct a large image of the whole scanned zone, the surface of which can reach 1-cm^2. Alexa Fluor and DAPI acquisition parameters were, respectively: wavelengths spectrum 499-519nm and 345-455nm, exposure time 1000 ms and 50 ms and light intensity 150. DAPI and Alexa Fluor fluorescence quantifications were done on the corresponding

reconstructed images, it consists of the percentage of fluorescent area compared to the whole scanned zone. The total number of nuclei and the number of nuclei inside myotubes were counted, on a section of the overlaid reconstructed images, using *ImageJ* software (version 1.48q) [43] (http://rsbweb.nih.gov.gate1.inist.fr/ij/). The fusion index of C2C12 cells was calculated as the ratio of the number of nuclei in myotubes to the total number of nuclei. The *ImageJ* software was also used to assess myotubes size by measuring length and width of 30 myotubes from 15 randomly chosen fields of the Alexa Fluor 488 reconstructed images.

2.5 Statistical analysis

Three replicates were analyzed for each experiment presented. Data obtained (percentage of fluorescent area, fusion index and myotube size) in siRNA-treated cells were compared to those obtained in control cells using Student's test analysis and Dunnett's test correction [44]. Statistical significance is attained when a *P*-value is less than the significance levels: *** p<0,001 ; ** p<0,01 ; * p<0,05.

2.6 Gene ontology

Gene ontology (GO) analysis was performed using PANTHER classification system according to the instructions provided at www.pantherdb.org [45]. Our gene list was analysed by PANTHER Overrepresentation Test (released 2016-07-15) using GO Ontology database (released at 2015-06-22) and all the genes of *Mus Musculus* database as a reference list. The data set was annotated by "GO molecular function complete", "GO cellular component complete" and "GO biological process complete".

3. Results

3.1 The experimental design of the siRNA-based screen

To search for new genes involved in myoblast differentiation, we developed a functional screen based on a novel protocol combining cell culture and staining, RNAi, microscopy and quantitative-image analysis. Cells were cultured on coverslips placed at the bottom of the plate. After a proliferation step lasting 24 hours, cells were induced to undergo differentiation and transiently transfected twice by synthetic siRNA duplexes at the beginning and at the middle of differentiation (Fig. 1). Six days later, differentiated myotubes adhered on the coverslips were immuno-stained by an anti-myosin antibody that recognizes all isoforms of myosin heavy chain, stained by

DAPI, mounted on a microscopic slide and then scanned and imaged under a fluorescent-automated microscope. Per coverslip, 100 to 150 individual fluorescent images were taken in order to reconstruct a nearly 1-cm^2 large image covering the whole scanned zone (Fig. 1).

Quantifications of myosin (Alexa Fluor 488) and DAPI fluorescent areas were then performed on these reconstructed-images; they correspond respectively to the percentage of the fluorescent surface occupied by myotubes or nuclei compared to the total coverslip-scanned zone. Myosin reconstructed images were used to evaluate the size and morphology of myotubes by measuring the length and the width of many myotubes. Myosin and DAPI reconstructed-images were superposed in order to calculate the fusion index (Fig. 2). Percentages of fluorescent area (DAPI and myosin), myotube size and morphology and fusion index were defined as phenotypic criteria, which will be used to classify candidate genes (Fig. 1). For each gene, quantitative data obtained from silencing performed with the two siRNAs were systematically compared to those obtained from associated controls. Finally, we retained genes for which quantitative data are similar between the two siRNAs and significantly different from controls.

To validate our screen, we built a panel of 100 mouse genes without a previously characterized function in mouse muscle and reported to change expression between proliferation and differentiation stages of the C2C12 cells. To establish this preselected gene list, we made use of a previously identified gene lists described in (i) C2C12 transcriptomic and proteomic studies, which classified respectively 2895 and 1052 genes with different expression level between proliferating and differentiating C2C12 cells [21, 23]; (ii) *Drosophila* RNAi screening, through which authors identified 2785 specific-muscle genes [31]. We started by eliminating housekeeping genes and genes with already assigned functions from C2C12 lists [46] and then compared the remaining genes to each other and to the 2425 mouse orthologs of *Drosophila* genes. 176 genes were recovered and only 100 of them were knocked down in our siRNA-based screen (Table 1 and Supplementary data Table S2). GO annotation analysis on the 100 candidate genes revealed that our list is enriched in GO terms related to RNA synthesis and maturation. The 100 candidate genes, denoted G1 to G100, were silenced in groups of 10 genes with controls for each group.

3.2 RNAi revealed 23 genes involved in C2C12 proliferation stage

In order to identify genes involved in the C2C12 proliferation stage, we compared the cellular density of silenced cells to controls. The use of the nuclear DAPI label allowed us to evaluate the cellular density and therefore the cell proliferation. This cellular density was estimated to represent 21% of the total surface of the reconstructed-image. This percentage was calculated by averaging all control values of the 10 groups. Cellular density was significantly different with each of the two siRNAs for 23 genes among the 100 candidate genes. These 23 genes seem to be involved in C2C12 cell cycle during the proliferation stage (Table 2 and Fig. 3).

3.3 RNAi revealed 49 genes involved in C2C12 differentiation stage

To identify genes involved in the differentiation stage, we calculated myotube quantity, myotube size and morphology, and fusion index in silenced myotubes. Inactivated cells and controls were recovered for staining at 144 hours of differentiation. At this final stage, differentiated myotubes in controls have fusiform structures, 400 to 600 µm of length, 30 to 50 µm of width, contain 10 to 20 nuclei centrally localized and distributed all along of myotubes, and accumulate myosin over the whole surface area of the myotube (Fig. 2). By averaging all control values, myotube quantity and fusion index were found to be 12,5% and 27%, respectively. Among the 100 candidate genes, significant differences were observed with each of the two siRNAs for 22 genes for the “myotube quantity” criterion (Table 2 and Fig. 4), 24 genes for the “myotube size and morphology” criterion (Table 2 and Fig. 5) and 18 genes for the “fusion index” criterion (Table 2 and Fig. 6). By removing any redundancy from these three gene groups, we retained 49 genes, which appear to be essential to cell fusion and/or myotube maturation during C2C12 differentiation stage.

3.4 The siRNA-based screen selected 19 genes essential to C2C12 myogenesis

By performing phenotypic analysis criterion by criterion and by eliminating duplicate genes, we obtained a subset of 60 genes (Table 2). A second filter was used, taking into account this time all phenotypic criteria; we retained 19 genes for which silencing with each of the two siRNAs exhibited the same phenotype for one, two or three phenotypic criteria and excluded 41 genes displaying the same phenotype only for one criterion (Table 2). This subset of 19 genes was considered as the final list of genes selected by our siRNA-based screen. 13 genes among them are specifically

involved tin the differentiation stage and the remaining 6 genes are implicated in both stages of the C2C12 myogenesis (Table 2). According to the silencing effect, which is either to promote or to inhibit the phenotype, we have further divided the retained 19 genes into two categories: positive- and negative regulator (Table 2 and Fig. 3-6). We used the PANTHER classification system to infer the putative biological function of the 19 selected genes. Gene products could be classified essentially in nucleic acid binding family and then in transferase, chaperone, enzyme modulator, hydrolase or ligase families.

Taken together, our siRNA-based screen led to the identification of 13 novel genes as potential regulators of *in-vivo* myogenic differentiation.

4. Discussion

In this study, thanks to our new *in-vitro* siRNA-based functional screening, we identified 19 genes involved in C2C12 myogenesis, which could be strong candidates for novel *in-vivo* myogenic regulators. Several key features made our siRNA-based screen particularly reliable. We set up an optimized protocol for C2C12 cell culture, RNAi gene silencing and phenotypic characterization based on a fine image analysis. The advantage of this protocol compared to others is the improvement done to analyse a maximal number of cells and myotubes. This is essential for detecting and quantifying both clear and subtle phenotypes during proliferation and differentiation stages of C2C12 cells. Screening 100 candidate genes directly in skeletal muscle of living mammals is tricky because of the great number of knocked-down lines that have to be generated [30]. Mammalian primary muscle cells which have been extensively used in the investigation of the molecular mechanisms governing muscle-related biological processes and pathological alterations, could be an alternative to living organisms. But their use is quite restricted because of limited life span, poor availability and purity, and difficulties related to gene silencing reproducibility. C2C12 myoblasts, because of the ease of their culture and transfection, and because of the rapidity of differentiation made them our tool of choice for the siRNA-based screen. To optimize myotubes immuno-staining and image acquisition, C2C12 cells were cultivated, stained and imaged on 1.8-cm^2 coverslips instead of on the bottom well plate. In addition, we used an automated-scan microscope capable of very fast image acquisition and the reconstruction of 1-cm^2 large images. This protocol

allowed us to perform fluorescence quantification and to track easily four types of phenotypes mediated by the silencing of candidate genes.

We preferred to silence the candidate genes by transient transfection of siRNA duplexes rather than by generating stables expressing short hairpin RNAs (shRNAs) to prevent technical difficulties related to time selection and position effect associated with the establishment of hundreds of shRNA stable clones. However, to ensure the continued efficiency of siRNA silencing during the whole differentiation stage, we transfected cells twice with the same siRNA. Assays were done in 10 separate experiments to knockdown 10 groups of 10 genes in order to minimize the effect of experimental variability. To eliminate false positives, two distinct siRNAs targeting either the coding region or the 3'UTR were used for each candidate gene. The considered phenotypes must be similar for the two siRNAs on one hand, and different from that of the associated controls on the other hand. Among the 100 silenced genes, we excluded 5 genes showing no phenotype with the two siRNAs and 35 genes displaying a phenotype with only one siRNA. Three hypotheses could explain this absence of phenotype: (i) the incomplete target gene suppression due to non- or partially functional siRNAs, (ii) the functional redundancy that can occur between candidate genes and their paralogues belonging to the same multigenic family, (iii) no contribution of these genes in C2C12 myogenesis despite a change in expression levels during C2C12 differentiation switch and the involvement of their *Drosophila* orthologues in muscle [21, 23].

Phenotypic analysis in our protocol was based on two cellular fluorescent markers (DAPI and myosin) decorating nuclei and myotubes respectively. Quantification of the decorated surface allowed us to define and analyse four phenotypic criteria (cellular density, myotube quantity, fusion index, myotube size and morphology). The first one occurs during the proliferation stage while the others take place during the differentiation stage. We reasoned that with this set of phenotypic criteria, we were able to track associated phenotypes of candidate-gene silencing during the whole C2C12 myogenic process. By scoring these phenotypic changes, we could identify 11 genes involved in the proliferation stage, 37 genes in the differentiation stage and 12 genes in both stages.

We have used in our siRNA-based screen a refined list of 100 candidate genes that fulfil the following conditions: they don't have a previously characterized function in mouse muscle and display a modified expression pattern between proliferating and

differentiating C2C12 cells [21, 23]. In order to screen genes with evolutionary conserved physiological function in muscle, we chose only orthologues of specific muscle genes identified in a *Drosophila* RNAi screen [31]. It should be emphasized that we had to use these successive filters for three reasons: to validate the screen, to identify new myogenic regulators and to characterize the phenotype associated with silencing of each gene. Indeed, thanks to this choice of genes, our siRNA-based screen proved its ability to associate knockdown phenotypes to more than half of candidate genes (60 genes) and gave more accurate idea of their functions during C2C12 myogenesis. We used a two-step gene classification based on the above-described phenotypic criteria. For the first step, the analysis was carried out criterion by criterion leading to the selection of 60 genes. Among them, 19 genes were selected with our multiple criteria analysis. The remaining 41 genes were not retained because we decided to score phenotypic changes only when silencing with each of the two siRNAs showed the same effect for one or several criteria.

Ge et al., 2013 have conducted a similar screen in C2C12 to examine the potential roles of 129 genes coding for cytokines in skeletal muscle myogenesis [38]. Their results led to the identification of 29 genes (22%) as novel actors. In comparison, our study seems to be more efficient since we identified 60 genes (60%) involved in proliferation and/or differentiation of C2C12 cells. Our newly designed protocol could explain this difference, in particular the systematic quantification of fluorescence on large images. It is, however, difficult to compare the percentage of candidate genes identified in the two screens since Ge et al focused their study on the cytokine super-family.

5. Conclusions

This functional siRNA-based screen demonstrated the effectiveness of our new silencing and analysis protocol and allowed us to assign silencing-associated phenotypes to at least 19 of the 100 candidate genes. 6 of them are essential for both stages of C2C12 myogenesis and 13 genes are specifically involved in the differentiation stage. These genes represent strong candidates for *in vivo* functional studies.

Authors' contributions

Conceived and designed the experiments: KB VB. Performed the experiments: RA AB KB. Analyzed and discussed the data: KB RA ADS VB. Wrote the paper: KB VB. All authors have read and approved the manuscript.

Acknowledgments

We are grateful to Olivier ARNAIZ for the search of murine orthologs. We thank Eric MEYER for his careful work on the English editing of the manuscript.

This project was co-financed by the European Union, the Limousin Regional Council and the French National Institute for Agricultural Research. RA was supported by a PhD fellowship from Lebanese Association for Scientific Research (LASeR).

References

[1] B.E. Housden, M. Muhar, M. Gemberling, C.A. Gersbach, D.Y. Stainier, G. Seydoux, S.E. Mohr, J. Zuber, N. Perrimon, Loss-of-function genetic tools for animal models: cross-species and cross-platform differences, Nat Rev Genet (2016).

[2] T. Mashimo, K. Yanagihara, S. Tokuda, B. Voigt, A. Takizawa, R. Nakajima, M. Kato, M. Hirabayashi, T. Kuramoto, T. Serikawa, An ENU-induced mutant archive for gene targeting in rats, Nat Genet 40 (2008) 514-515.

[3] R. Stottmann, D. Beier, ENU mutagenesis in the mouse, Curr Protoc Mouse Biol 4 (2014) 25-35.

[4] P.M. Nolan, J. Peters, M. Strivens, D. Rogers, J. Hagan, N. Spurr, I.C. Gray, L. Vizor, D. Brooker, E. Whitehill, R. Washbourne, T. Hough, S. Greenaway, M. Hewitt, X. Liu, S. McCormack, K. Pickford, R. Selley, C. Wells, Z. Tymowska-Lalanne, P. Roby, P. Glenister, C. Thornton, C. Thaung, J.A. Stevenson, R. Arkell, P. Mburu, R. Hardisty, A. Kiernan, A. Erven, K.P. Steel, S. Voegeling, J.L. Guenet, C. Nickols, R. Sadri, M. Nasse, A. Isaacs, K. Davies, M. Browne, E.M. Fisher, J. Martin, S. Rastan, S.D. Brown, J. Hunter, A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse, Nat Genet 25 (2000) 440-443.

[5] M.H. Hrabe de Angelis, H. Flaswinkel, H. Fuchs, B. Rathkolb, D. Soewarto, S. Marschall, S. Heffner, W. Pargent, K. Wuensch, M. Jung, A. Reis, T. Richter, F. Alessandrini, T. Jakob, E. Fuchs, H. Kolb, E. Kremmer, K. Schaeble, B. Rollinski, A. Roscher, C. Peters, T. Meitinger, T. Strom, T. Steckler, F. Holsboer, T. Klopstock, F. Gekeler, C. Schindewolf, T. Jung, K. Avraham, H. Behrendt, J. Ring, A. Zimmer, K. Schughart, K. Pfeffer, E. Wolf, R. Balling, Genome-wide, large-scale production of mutant mice by ENU mutagenesis, Nat Genet 25 (2000) 444-447.

[6] G. Caignard, M.M. Eva, R. van Bruggen, R. Eveleigh, G. Bourque, D. Malo, P. Gros, S.M. Vidal, Mouse ENU Mutagenesis to Understand Immunity to

Infection: Methods, Selected Examples, and Perspectives, Genes (Basel) 5 (2014) 887-925.
[7] P.K. Potter, M.R. Bowl, P. Jeyarajan, L. Wisby, A. Blease, M.E. Goldsworthy, M.M. Simon, S. Greenaway, V. Michel, A. Barnard, C. Aguilar, T. Agnew, G. Banks, A. Blake, L. Chessum, J. Dorning, S. Falcone, L. Goosey, S. Harris, A. Haynes, I. Heise, R. Hillier, T. Hough, A. Hoslin, M. Hutchison, R. King, S. Kumar, H.V. Lad, G. Law, R.E. MacLaren, S. Morse, T. Nicol, A. Parker, K. Pickford, S. Sethi, B. Starbuck, F. Stelma, M. Cheeseman, S.H. Cross, R.G. Foster, I.J. Jackson, S.N. Peirson, R.V. Thakker, T. Vincent, C. Scudamore, S. Wells, A. El-Amraoui, C. Petit, A. Acevedo-Arozena, P.M. Nolan, R. Cox, A.M. Mallon, S.D. Brown, Novel gene function revealed by mouse mutagenesis screens for models of age-related disease, Nat Commun 7 (2016) 12444.
[8] E.J. Michaud, C.T. Culiat, M.L. Klebig, P.E. Barker, K.T. Cain, D.J. Carpenter, L.L. Easter, C.M. Foster, A.W. Gardner, Z.Y. Guo, K.J. Houser, L.A. Hughes, M.K. Kerley, Z. Liu, R.E. Olszewski, I. Pinn, G.D. Shaw, S.G. Shinpock, A.M. Wymore, E.M. Rinchik, D.K. Johnson, Efficient gene-driven germ-line point mutagenesis of C57BL/6J mice, BMC Genomics 6 (2005) 164.
[9] C. Guan, C. Ye, X. Yang, J. Gao, A review of current large-scale mouse knockout efforts, Genesis 48 (2010) 73-85.
[10] Y. Gondo, R. Fukumura, T. Murata, S. Makino, Next-generation gene targeting in the mouse for functional genomics, BMB Rep 42 (2009) 315-323.
[11] A. Yoshiki, F. Ike, K. Mekada, Y. Kitaura, H. Nakata, N. Hiraiwa, K. Mochida, M. Ijuin, M. Kadota, A. Murakami, A. Ogura, K. Abe, K. Moriwaki, Y. Obata, The mouse resources at the RIKEN BioResource center, Exp Anim 58 (2009) 85-96.
[12] M. Ringwald, V. Iyer, J.C. Mason, K.R. Stone, H.D. Tadepally, J.A. Kadin, C.J. Bult, J.T. Eppig, D.J. Oakley, S. Briois, E. Stupka, V. Maselli, D. Smedley, S. Liu, J. Hansen, R. Baldock, G.G. Hicks, W.C. Skarnes, The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium, Nucleic Acids Res 39 (2011) D849-855.
[13] A. Bradley, K. Anastassiadis, A. Ayadi, J.F. Battey, C. Bell, M.C. Birling, J. Bottomley, S.D. Brown, A. Burger, C.J. Bult, W. Bushell, F.S. Collins, C. Desaintes, B. Doe, A. Economides, J.T. Eppig, R.H. Finnell, C. Fletcher, M. Fray, D. Frendewey, R.H. Friedel, F.G. Grosveld, J. Hansen, Y. Herault, G. Hicks, A. Horlein, R. Houghton, M. Hrabe de Angelis, D. Huylebroeck, V. Iyer, P.J. de Jong, J.A. Kadin, C. Kaloff, K. Kennedy, M. Koutsourakis, K.C. Lloyd, S. Marschall, J. Mason, C. McKerlie, M.P. McLeod, H. von Melchner, M. Moore, A.O. Mujica, A. Nagy, M. Nefedov, L.M. Nutter, G. Pavlovic, J.L. Peterson, J. Pollock, R. Ramirez-Solis, D.E. Rancourt, M. Raspa, J.E. Remacle, M. Ringwald, B. Rosen, N. Rosenthal, J. Rossant, P. Ruiz Noppinger, E. Ryder, J.Z. Schick, F. Schnutgen, P. Schofield, C. Seisenberger, M. Selloum, E.M. Simpson, W.C. Skarnes, D. Smedley, W.L. Stanford, A.F. Stewart, K. Stone, K. Swan, H. Tadepally, L. Teboul, G.P. Tocchini-Valentini, D. Valenzuela, A.P. West, K. Yamamura, Y. Yoshinaga, W. Wurst, The mammalian gene function resource: the International Knockout Mouse Consortium, Mamm Genome 23 (2012) 580-586.
[14] A. Ayadi, M.C. Birling, J. Bottomley, J. Bussell, H. Fuchs, M. Fray, V. Gailus-Durner, S. Greenaway, R. Houghton, N. Karp, S. Leblanc, C. Lengger, H. Maier, A.M. Mallon, S. Marschall, D. Melvin, H. Morgan, G. Pavlovic, E. Ryder, W.C. Skarnes, M. Selloum, R. Ramirez-Solis, T. Sorg, L. Teboul, L. Vasseur,

A. Walling, T. Weaver, S. Wells, J.K. White, A. Bradley, D.J. Adams, K.P. Steel, M. Hrabe de Angelis, S.D. Brown, Y. Herault, Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project, Mamm Genome 23 (2012) 600-610.

[15] A.M. Mallon, V. Iyer, D. Melvin, H. Morgan, H. Parkinson, S.D. Brown, P. Flicek, W.C. Skarnes, Accessing data from the International Mouse Phenotyping Consortium: state of the art and future plans, Mamm Genome 23 (2012) 641-652.

[16] S.D. Brown, M.W. Moore, The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping, Mamm Genome 23 (2012) 632-640.

[17] J.K. White, A.K. Gerdin, N.A. Karp, E. Ryder, M. Buljan, J.N. Bussell, J. Salisbury, S. Clare, N.J. Ingham, C. Podrini, R. Houghton, J. Estabel, J.R. Bottomley, D.G. Melvin, D. Sunter, N.C. Adams, D. Tannahill, D.W. Logan, D.G. Macarthur, J. Flint, V.B. Mahajan, S.H. Tsang, I. Smyth, F.M. Watt, W.C. Skarnes, G. Dougan, D.J. Adams, R. Ramirez-Solis, A. Bradley, K.P. Steel, Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes, Cell 154 (2013) 452-464.

[18] M. Hrabe de Angelis, G. Nicholson, M. Selloum, J.K. White, H. Morgan, R. Ramirez-Solis, T. Sorg, S. Wells, H. Fuchs, M. Fray, D.J. Adams, N.C. Adams, T. Adler, A. Aguilar-Pimentel, D. Ali-Hadji, G. Amann, P. Andre, S. Atkins, A. Auburtin, A. Ayadi, J. Becker, L. Becker, E. Bedu, R. Bekeredjian, M.C. Birling, A. Blake, J. Bottomley, M.R. Bowl, V. Brault, D.H. Busch, J.N. Bussell, J. Calzada-Wack, H. Cater, M.F. Champy, P. Charles, C. Chevalier, F. Chiani, G.F. Codner, R. Combe, R. Cox, E. Dalloneau, A. Dierich, A. Di Fenza, B. Doe, A. Duchon, O. Eickelberg, C.T. Esapa, L.E. Fertak, T. Feigel, I. Emelyanova, J. Estabel, J. Favor, A. Flenniken, A. Gambadoro, L. Garrett, H. Gates, A.K. Gerdin, G. Gkoutos, S. Greenaway, L. Glasl, P. Goetz, I.G. Da Cruz, A. Gotz, J. Graw, A. Guimond, W. Hans, G. Hicks, S.M. Holter, H. Hofler, J.M. Hancock, R. Hoehndorf, T. Hough, R. Houghton, A. Hurt, B. Ivandic, H. Jacobs, S. Jacquot, N. Jones, N.A. Karp, H.A. Katus, S. Kitchen, T. Klein-Rodewald, M. Klingenspor, T. Klopstock, V. Lalanne, S. Leblanc, C. Lengger, E. le Marchand, T. Ludwig, A. Lux, C. McKerlie, H. Maier, J.L. Mandel, S. Marschall, M. Mark, D.G. Melvin, H. Meziane, K. Micklich, C. Mittelhauser, L. Monassier, D. Moulaert, S. Muller, B. Naton, F. Neff, P.M. Nolan, L.M. Nutter, M. Ollert, G. Pavlovic, N.S. Pellegata, E. Peter, B. Petit-Demouliere, A. Pickard, C. Podrini, P. Potter, L. Pouilly, O. Puk, D. Richardson, S. Rousseau, L. Quintanilla-Fend, M.M. Quwailid, I. Racz, B. Rathkolb, F. Riet, J. Rossant, M. Roux, J. Rozman, E. Ryder, J. Salisbury, L. Santos, K.H. Schable, E. Schiller, A. Schrewe, H. Schulz, R. Steinkamp, M. Simon, M. Stewart, C. Stoger, T. Stoger, M. Sun, D. Sunter, L. Teboul, I. Tilly, G.P. Tocchini-Valentini, M. Tost, I. Treise, L. Vasseur, E. Velot, D. Vogt-Weisenhorn, C. Wagner, A. Walling, M. Wattenhofer-Donze, B. Weber, O. Wendling, H. Westerberg, M. Willershauser, E. Wolf, A. Wolter, J. Wood, W. Wurst, A.O. Yildirim, R. Zeh, A. Zimmer, A. Zimprich, C. Holmes, K.P. Steel, Y. Herault, V. Gailus-Durner, A.M. Mallon, S.D. Brown, Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics, Nat Genet 47 (2015) 969-978.

[19] M.A. Haendel, N. Vasilevsky, M. Brush, H.S. Hochheiser, J. Jacobsen, A. Oellrich, C.J. Mungall, N. Washington, S. Kohler, S.E. Lewis, P.N. Robinson, D. Smedley, Disease insights through cross-species phenotype comparisons, Mamm Genome 26 (2015) 548-555.
[20] F. Scavizzi, E. Ryder, S. Newman, M. Raspa, D. Gleeson, H. Wardle-Jones, L. Montoliu, A. Fernandez, M.L. Dessain, V. Larrigaldie, Z. Khorshidi, R. Vuolteenaho, R. Soininen, P. Andre, S. Jacquot, Y. Hong, M.H. de Angelis, R. Ramirez-Solis, B. Doe, Blastocyst genotyping for quality control of mouse mutant archives: an ethical and economical approach, Transgenic Res 24 (2015) 921-927.
[21] K.K. Tomczak, V.D. Marinescu, M.F. Ramoni, D. Sanoudou, F. Montanaro, M. Han, L.M. Kunkel, I.S. Kohane, A.H. Beggs, Expression profiling and identification of novel genes involved in myogenic differentiation, FASEB J 18 (2004) 403-405.
[22] K. Grabiec, A. Majewska, Z. Wicik, M. Milewska, M. Blaszczyk, K. Grzelkowska-Kowalczyk, The effect of palmitate supplementation on gene expression profile in proliferating myoblasts, Cell Biol Toxicol 32 (2016) 185-198.
[23] T. Kislinger, A.O. Gramolini, Y. Pan, K. Rahman, D.H. MacLennan, A. Emili, Proteome dynamics during C2C12 myoblast differentiation, Mol Cell Proteomics 4 (2005) 887-901.
[24] L. Casadei, L. Vallorani, A.M. Gioacchini, M. Guescini, S. Burattini, A. D'Emilio, L. Biagiotti, E. Falcieri, V. Stocchi, Proteomics-based investigation in C2C12 myoblast differentiation, Eur J Histochem 53 (2009) e31.
[25] A.S. Deshmukh, M. Murgia, N. Nagaraj, J.T. Treebak, J. Cox, M. Mann, Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors, Mol Cell Proteomics 14 (2015) 841-853.
[26] Y. Cao, Z. Yao, D. Sarkar, M. Lawrence, G.J. Sanchez, M.H. Parker, K.L. MacQuarrie, J. Davison, M.T. Morgan, W.L. Ruzzo, R.C. Gentleman, S.J. Tapscott, Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming, Dev Cell 18 (2010) 662-674.
[27] J. Segales, A.B. Islam, R. Kumar, Q.C. Liu, P. Sousa-Victor, F.J. Dilworth, E. Ballestar, E. Perdiguero, P. Munoz-Canoves, Chromatin-wide and transcriptome profiling integration uncovers p38alpha MAPK as a global regulator of skeletal muscle differentiation, Skelet Muscle 6 (2016) 9.
[28] S.E. Mohr, J.A. Smith, C.E. Shamu, R.A. Neumuller, N. Perrimon, RNAi screening comes of age: improved techniques and complementary approaches, Nat Rev Mol Cell Biol 15 (2014) 591-600.
[29] S. Crotty, M.E. Pipkin, In vivo RNAi screens: concepts and applications, Trends Immunol 36 (2015) 315-322.
[30] S. Sharma, A. Rao, RNAi screening: tips and techniques, Nat Immunol 10 (2009) 799-804.
[31] F. Schnorrer, C. Schonbauer, C.C. Langer, G. Dietzl, M. Novatchkova, K. Schernhuber, M. Fellner, A. Azaryan, M. Radolf, A. Stark, K. Keleman, B.J. Dickson, Systematic genetic analysis of muscle morphogenesis and function in Drosophila, Nature 464 (2010) 287-291.
[32] J. Bai, R. Binari, J.Q. Ni, M. Vijayakanthan, H.S. Li, N. Perrimon, RNA interference screening in Drosophila primary cells for genes involved in muscle assembly and maintenance, Development 135 (2008) 1439-1449.

[33] K.J. Simpson, G.M. Davis, P.R. Boag, Comparative high-throughput RNAi screening methodologies in C. elegans and mammalian cells, N Biotechnol 29 (2012) 459-470.
[34] G. Livshits, S.W. Lowe, Accelerating cancer modeling with RNAi and nongermline genetically engineered mouse models, Cold Spring Harb Protoc 2013 (2013).
[35] M.V. Khanjyan, J. Yang, R. Kayali, T. Caldwell, C. Bertoni, A high-content, high-throughput siRNA screen identifies cyclin D2 as a potent regulator of muscle progenitor cell fusion and a target to enhance muscle regeneration, Hum Mol Genet 22 (2013) 3283-3295.
[36] C.J. Echeverri, N. Perrimon, High-throughput RNAi screening in cultured cells: a user's guide, Nat Rev Genet 7 (2006) 373-384.
[37] S. Rajan, H. Chu Pham Dang, H. Djambazian, H. Zuzan, Y. Fedyshyn, T. Ketela, J. Moffat, T.J. Hudson, R. Sladek, Analysis of early C2C12 myogenesis identifies stably and differentially expressed transcriptional regulators whose knock-down inhibits myoblast differentiation, Physiol Genomics 44 (2012) 183-197.
[38] Y. Ge, R.J. Waldemer, R. Nalluri, P.D. Nuzzi, J. Chen, RNAi screen reveals potentially novel roles of cytokines in myoblast differentiation, PLoS One 8 (2013) e68068.
[39] S. Burattini, P. Ferri, M. Battistelli, R. Curci, F. Luchetti, E. Falcieri, C2C12 murine myoblasts as a model of skeletal muscle development: morpho-functional characterization, Eur J Histochem 48 (2004) 223-233.
[40] T. Braun, M. Gautel, Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis, Nat Rev Mol Cell Biol 12 (2011) 349-361.
[41] C.F. Bentzinger, Y.X. Wang, M.A. Rudnicki, Building muscle: molecular regulation of myogenesis, Cold Spring Harb Perspect Biol 4 (2012).
[42] D. Yaffe, O. Saxel, Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle, Nature 270 (1977) 725-727.
[43] M.D. Abramoff, P.J. Magalhaes, S.J. Ram, Image processing with ImageJ, Biophotonics Int 11 (2014) 36-42.
[44] C.W. Dunnett, R. Crisafio, The operating characteristics of some official weight variation tests for tablets, J Pharm Pharmacol 7 (1955) 314-327.
[45] H. Mi, A. Muruganujan, J.T. Casagrande, P.D. Thomas, Large-scale gene function analysis with the PANTHER classification system, Nat Protoc 8 (2013) 1551-1566.
[46] C.J. Bult, J.T. Eppig, J.A. Blake, J.A. Kadin, J.E. Richardson, Mouse genome database 2016, Nucleic Acids Res 44 (2016) D840-847.

Figure captions

Figure 1: Experimental design flowchart of the siRNA-based screen.

C2C12 myoblasts were seeded in growth medium on sterile coverslips putted at the bottom of 12 well plates. After switch to the differentiation medium, 200 siRNAs

targeted 100 genes were used to transfect twice these cells during the differentiation stage. Cells were cultivated during six days, at the end of which they were fixed and stained by myosin antibody and DAPI. Scan-slide image acquisition, fluorescence quantification, phenotypic analysis and gene classification were then performed.

Figure 2: Wild type C2C12 cells imaged after myotube- and nuclei staining.
At six days of differentiation, myotubes and nuclei were stained by anti-myosin and DAPI, respectively. (A) Phase contrast, (B) Myosin immuno-staining (green), (C) DAPI labelling (blue) and (D) Merge images of the same field. (B) and (C) images were used to quantify area occupied by myotubes and nuclei, respectively and to evaluate the size and the morphology of myotubes. Superposed myosin and DAPI image (D) was used to calculate the fusion index. Magnification was X100 and scale bar = 300 µm

Figure 3: siRNA-silencing analysis of G11, G61 and G64 genes according to "cellular density" phenotypic criterion.
(A) The pie chart shows gene number for each gene category (B) The bar chart shows cellular density variation in wild type (control) and G11, G61 and G64 siRNA-silenced cells. No significant difference was observed for G11 siRNA silencing and significant differences were observed for G61 and G64 siRNA silencing (C) Images represent nuclear DAPI labelling (in blue) of wild type C2C12 (control) and G11, G61 and G64 siRNA knocked-down cells. Standard errors were calculated using triplicate and statistical tests determined significant differences between control and siRNA silenced cells *** $p<0.001$; ** $p<0.01$; * $p<0.05$. Magnification was X100 and scale bar = 300 µm

Figure 4: siRNA-silencing analysis of G4, G18 and G91 genes according to "myotube quantity" phenotypic criterion.
(A) The pie chart shows gene number for each gene category (B). The bar chart shows myotube quantity variation in wild type (control) and G4, G18 and G91 siRNA-silenced cells. No significant difference was observed for G4 siRNA silencing and significant differences were observed for G18 and G91 siRNA silencing (C) Images represent myosin immuno-staining (in green Alexa 488) of wild type C2C12 (control) and G4, G18 and G91 siRNA knocked-down cells. Images show obvious

modifications in myotube size and morphology for the siRNA silencing of the three genes. Standard errors were calculated using triplicate and statistical tests determined significant differences between control and siRNA silenced cells.
*** p<0.001 ; ** p<0.01 ; * p<0.05. Magnification was X100 and scale bar = 300 µm.

Figure 5: siRNA-silencing analysis of G18, G38, G63, G25 and G6 genes according to "myotube size and morphology" phenotypic criterion.

(A) The pie chart shows gene number for each gene category and for each phenotype related to the myotube morphology. (B) Measurements of myotube length and width in control cells and in G25, G38 and G18 siRNA silenced cells. (C) Images represent myosin immuno-staining (in green Alexa 488) of wild type C2C12 (control) and G18 (large myotubes), G38 (long myotubes), G63 (thin myotubes), G25 (small myotubes) and G6 (without phenotype) siRNA knocked-down cells. Standard errors were calculated using triplicate and statistical tests determined significant differences between control and siRNA silenced cells *** p<0.001 ; ** p<0.01 ; * p<0.05. Magnification was X100 and scale bar = 300 µm.

Figure 6: siRNA-silencing analysis of G98, G45 and G1 genes according to "fusion index" phenotypic criterion.

(A) The pie chart shows gene number for each gene category (B) The bar chart shows fusion index variation in wild type (control) and G98, G45 and G1 siRNA-silenced cells. No significant difference was observed for G98 siRNA silencing and significant differences were observed for G45 and G1 siRNA silencing (C) Merged images represent nuclear DAPI labelling (in blue) and myosin immuno-staining (in green Alexa 488) of wild type C2C12 (control) and G98, G45 and G1 siRNA knocked-down cells. Standard errors were calculated using triplicate and statistical tests determined significant differences between control and siRNA silenced cells.
*** p<0.001 ; ** p<0.01 ; * p<0.05. Magnification was X100 and scale bar = 300 µm.

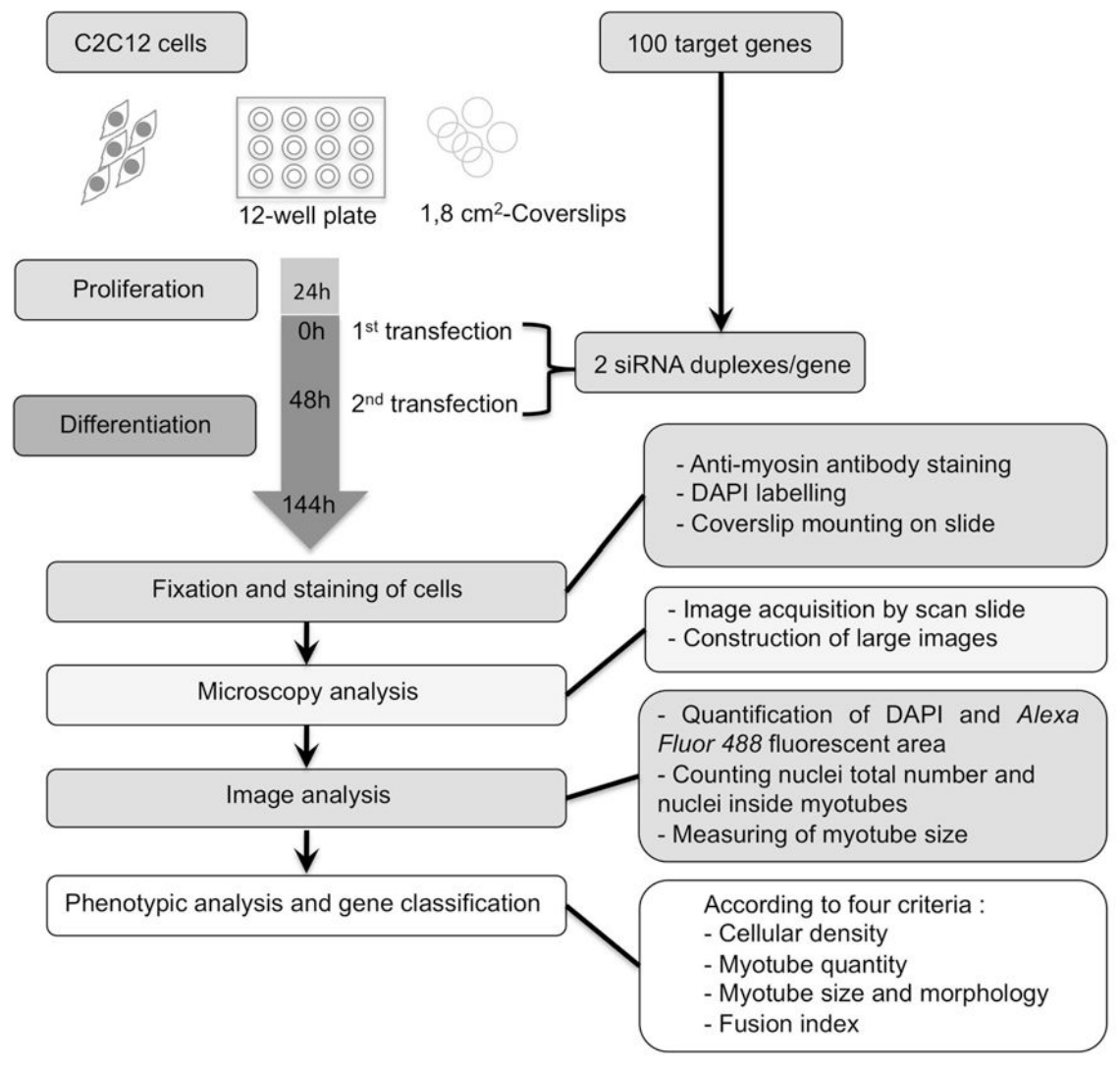
Table 1: Gene names of the 100 preselected genes for the siRNA-based screen.

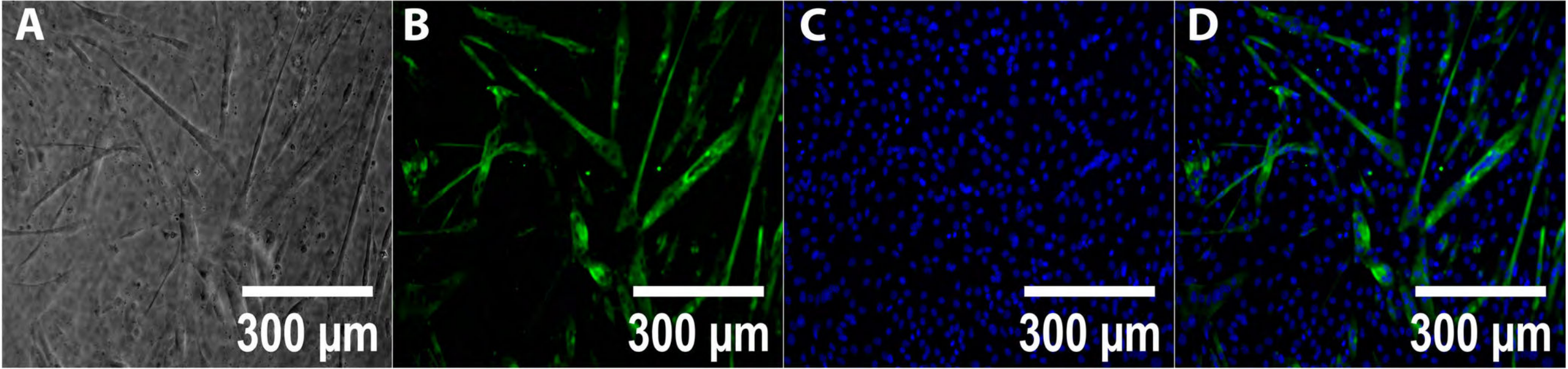
Table 2: Phenotypic analysis and gene classification of the 100 preselected genes according to the elimination method.

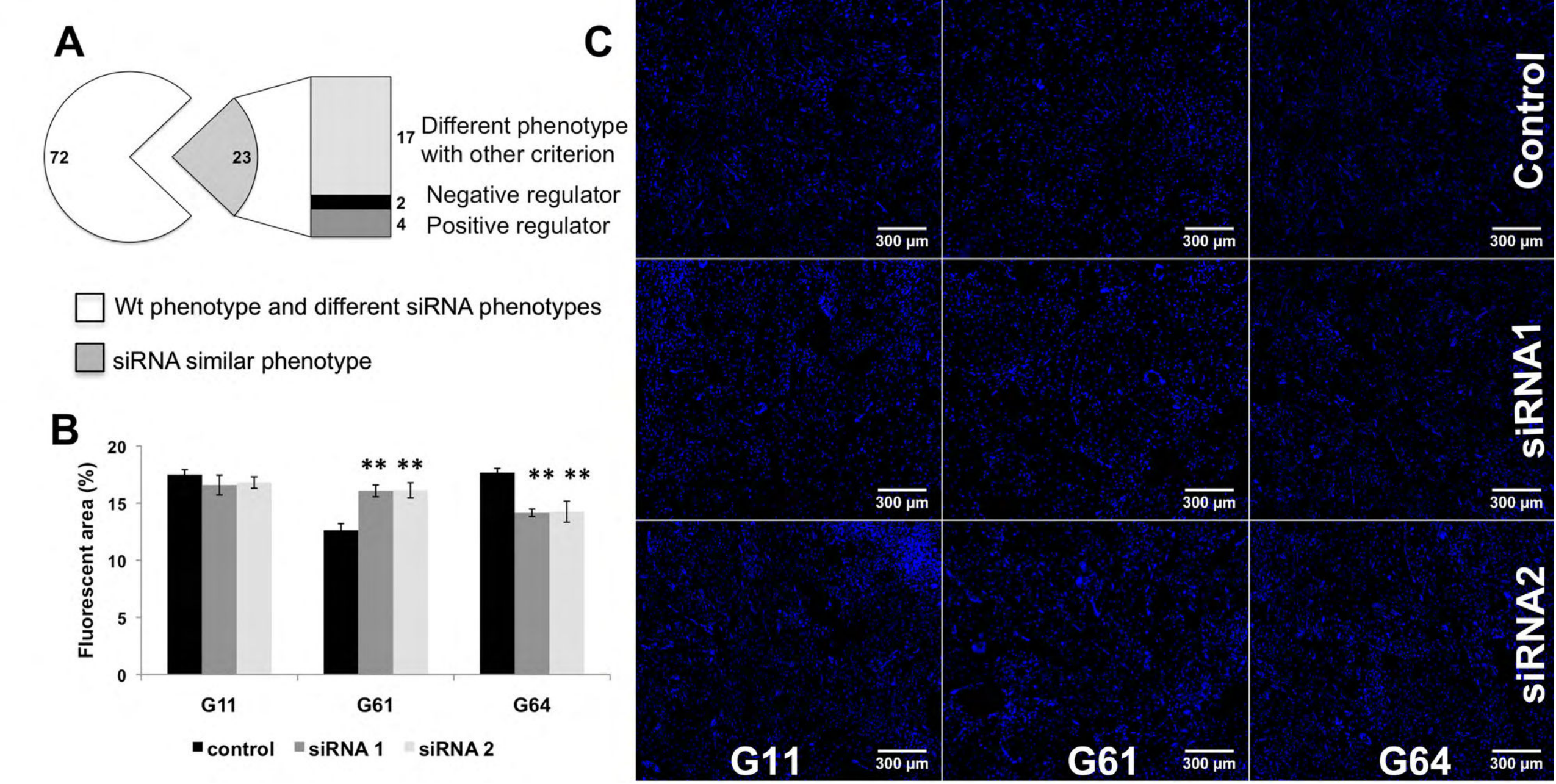
Firstly, 5 genes for which the two-siRNA inhibition doesn't show any particular phenotype were eliminated. Secondly, the phenotypes of silencing by the two siRNA of the remaining 95 genes were analysed criterion by criterion according to the four phenotypic criteria: the "myotube quantity", the "cellular density", the "fusion index" and the "myotube size and morphology". Thirdly, genes for which the silencing by each of the two siRNA displayed different phenotypes were excluded. 19 genes were selected, for which the inactivation by the two siRNA displayed the same phenotype for one, two or three criteria. Six of them are involved in the C2C12 proliferation and differentiation stages, while the other 13 genes are specifically involved in the differentiation stage. Finally, the remaining genes were divided into two subgroups: negative- and positive regulator. The numbers presented in the table represent the gene number at each step.

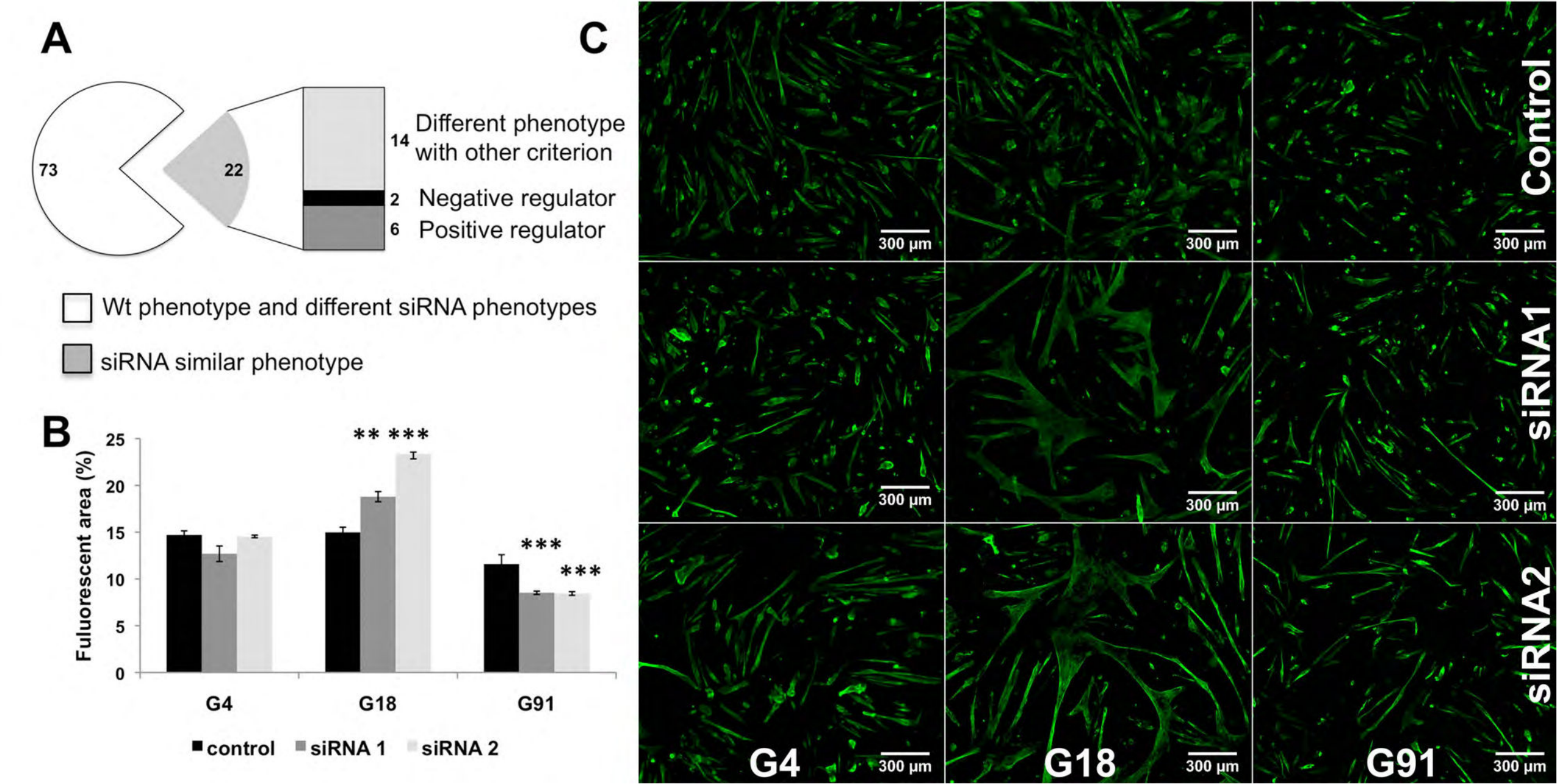
Figure S1: Evaluation of siRNA efficiency delivery and silencing

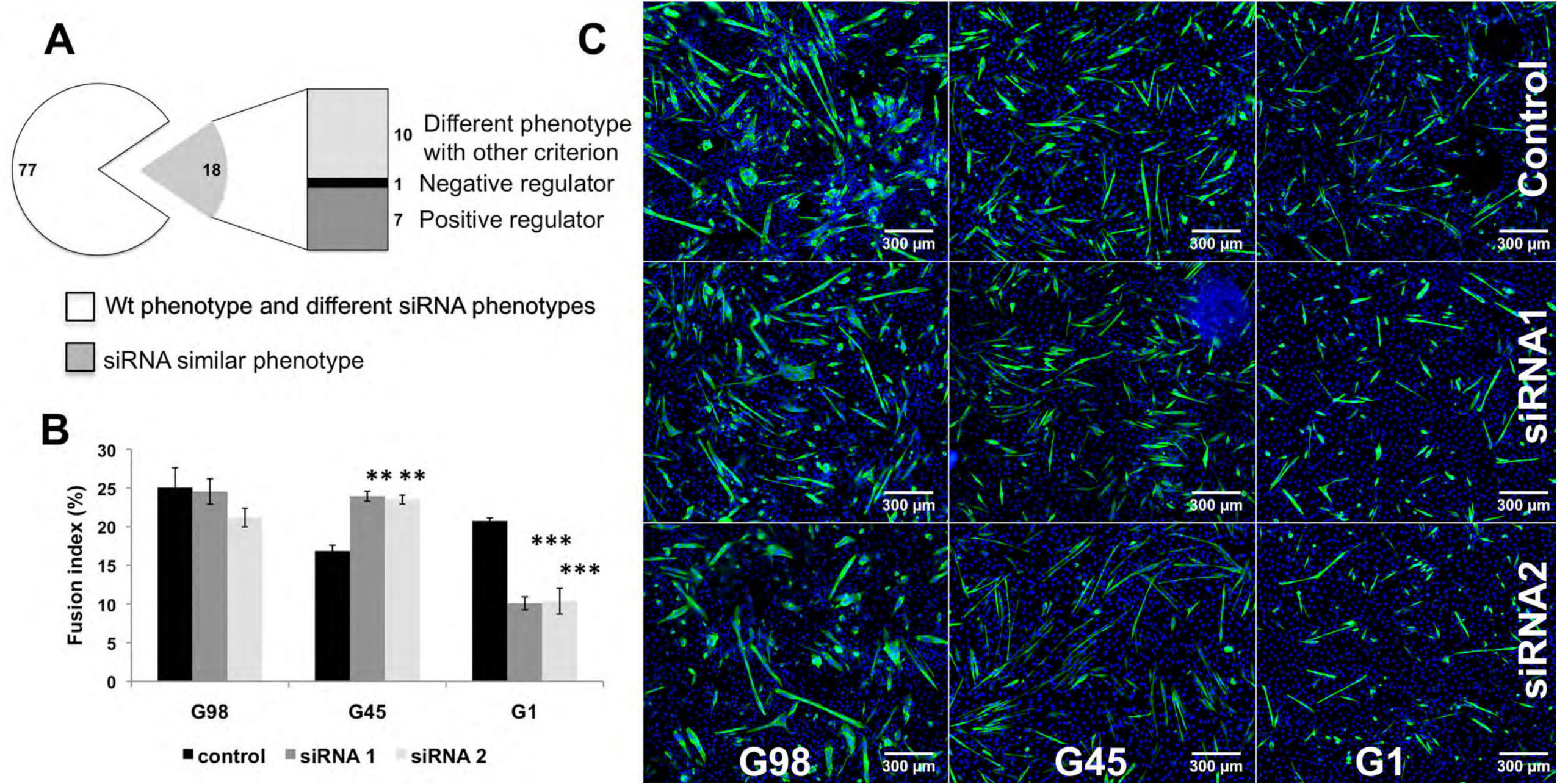
(A) The siRNA delivery efficiency was evaluated by transfecting C2C12 cells with an siRNA killer that targets ubiquitous cell survival genes leading to C2C12 cell death. Cell viability assay shows that only 10% of proliferating cells and 40% of differentiating cells survive after siRNA killer treatment. (B) The siRNA-silencing efficiency was checked on nine targeted genes, randomly selected among the 100 candidate genes. Transcripts were quantified by quantitative RT-PCR on transfected cells by two siRNAs and untransfected cells. A significant decrease in siRNA-treated cells was observed for all the tested genes. *** $p<0.001$; ** $p<0.01$; * $p<0.05$

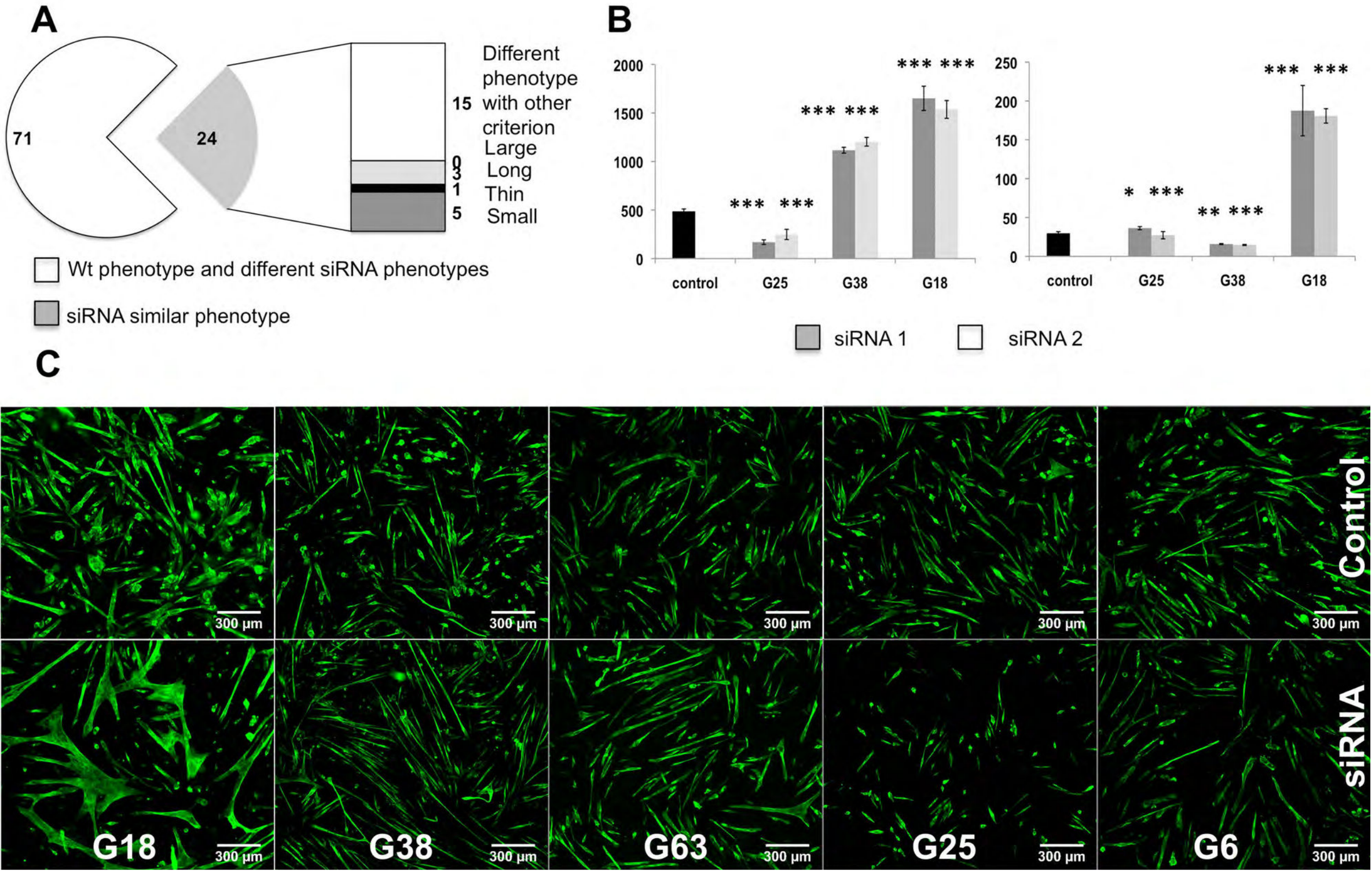

C2C12 cells
100 target genes
12-well plate
1,8 cm2-Coverslips
Proliferation
24h
0h
1st transfection
2 siRNA duplexes/gene
48h
2nd transfection
Differentiation
144h
- Anti-myosin antibody staining
- DAPI labelling
- Coverslip mounting on slide
Fixation and staining of cells
- Image acquisition by scan slide
- Construction of large images
Microscopy analysis
- Quantification of DAPI and Alexa Fluor 488 fluorescent area
- Counting nuclei total number and nuclei inside myotubes
- Measuring of myotube size
Image analysis
Phenotypic analysis and gene classification
According to four criteria :
- Cellular density
- Myotube quantity
- Myotube size and morphology
- Fusion index


A
300 µm
B
300 µm
C
300 µm
D
300 µm


A
72
23
17
Different phenotype with other criterion
2
Negative regulator
4
Positive regulator
Wt phenotype and different siRNA phenotypes
siRNA similar phenotype
B
Fluorescent area (%)
20
15
10
5
0
** **
** **
G11
G61
G64
control
siRNA 1
siRNA 2
C
Control
siRNA1
siRNA2
G11
G61
G64
300 µm


A
73
22
14
Different phenotype with other criterion
2
Negative regulator
6
Positive regulator
Wt phenotype and different siRNA phenotypes
siRNA similar phenotype
B
Fuluorescent area (%)
25
20
15
10
5
0
** ***
***
***
G4
G18
G91
control
siRNA 1
siRNA 2
C
300 µm
Control
siRNA1
siRNA2
G4
G18
G91


A
77
18
10
Different phenotype with other criterion
1
Negative regulator
7
Positive regulator
Wt phenotype and different siRNA phenotypes
siRNA similar phenotype
B
Fusion index (%)
30
25
20
15
10
5
0
** **
***
***
G98
G45
G1
control
siRNA 1
siRNA 2
C
Control
siRNA1
siRNA2
G98
G45
G1
300 µm


A
71
24
15
Different phenotype with other criterion
0
Large
3
Long
1
Thin
5
Small
Wt phenotype and different siRNA phenotypes
siRNA similar phenotype
B
2000
1500
1000
500
0
*** ***
*** ***
*** ***
control
G25
G38
G18
250
200
150
100
50
0
* ***
** ***
*** ***
control
G25
G38
G18
siRNA 1
siRNA 2
C
300 µm
Control
siRNA
G18
G38
G63
G25
G6

1110008P14Rik	Ddx23	Ermp1	Mstn	Nsmce2	Pola2	Rnps1	Sf3a3	Tagln2	Usp39
Actl6a	Ddx47	Fam136a	Myh7	Nsun4	Prim2	Rpl11	Sf3b3	Taok1	Utp15
Adsl	Dhx38	Fip1l1	Naa10	Nudt21	Prpf4b	Rpl12	Sf3b5	Tbc1d23	Vars
Brix1	Dna2	Grcc10	Ncaph	Nvl	Prpf6	Rplp0	Sfpq	Thoc6	vipas39
cct7	Dnajc19	H2afv	Nhp2	Osbp2	psma1	RsL1d1	Slc7a6os	Tmem205	Wdr18
Ctps1	Dnajc2	Ints5	Nip7	Otud4	Pwp1	Rufy3	Smarcc2	Top2a	Wdr43
Ctu1	Dph5	Klhl24	Nmt2	Pabpc4	Rars	Sap130	Srpr	Trappc2l	Wdr46
D10Wsu52e	Ebna1bp2	Mdn1	Nol8	Phf14	Rbm28	Sass6	Srsf4	Tsr1	Yipf7
Dcakd	Eef1b2	Mplkip	Nop10	Phldb2	Rcc1	Seh1l	Srsf7	U2surp	Zfp276
Ddx18	Eftud2	Mrto4	Nop2	Pno1	Rnf20	Set	Svil	Ufc1	Zfp322a

100	
Without phenotype ⇩ *Eliminated (5)*	**With phenotype 95**

		Phenotypic criteria			
		Cellular density	**Myotube quantity**	**Fusion index**	**Myotube size and morphology**
Phenotypes observed with the two-siRNA	**Different** ⇩ *Eliminated (35)*	*72*	*73*	*77*	*71*
	similar	**23**	**22**	**18**	**24**
			49		
		60			

Phenotypes observed with the two-siRNA for the other criteria	**Different** ⇩ *Eliminated (41)*	*17*	*14*	*10*	*15*
	similar	**6**	**8**	**8**	**9**
		19			

	Proliferation and differentiation 6	**Differentiation 13**
Negative regulator	**G32, G46**	**G48, G50**
Positive regulator	**G36, G55, G74, G75**	**G1, G7, G8, G40, G43, G57, G60, G73, G78, G82, G100**

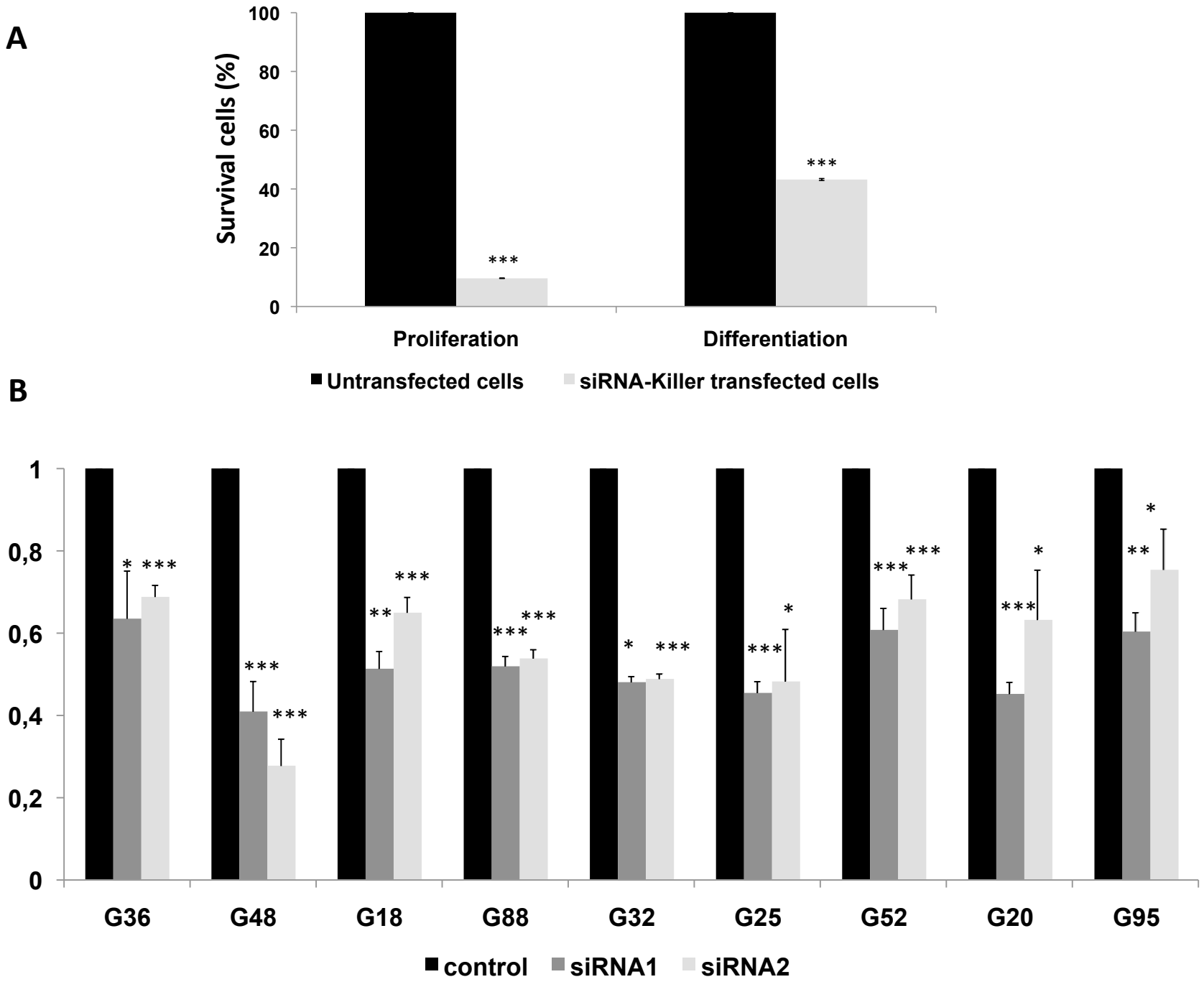

A
Survival cells (%)
100
80
60
40
20
0
***
***
Proliferation
Differentiation
Untransfected cells
siRNA-Killer transfected cells
B
Relative mRNA quantity
1
0,8
0,6
0,4
0,2
0
G36
G48
G18
G88
G32
G25
G52
G20
G95
control
siRNA1
siRNA2

Table 1 supp data: The 200 siRNAs target sequences of the 100 preselected genes for the siRNA-based screen with product ID and gene name.

NCBI gene symbc	siRNA Target Sequence	Product Id
1110008P14Rik	CCGGTGGATGCGGGCACGGAA	SI00783958
1110008P14Rik	CTCCATGTTGGATCAGATCAA	SI00783965
2810021B07Rik	ATGACTAATAGTAATCTCAAA	SI00825342
2810021B07Rik	CTGGGAGTGTGTATCATAGAA	SI00825356
2610101N10Rik	CAGGGAGATTCTCCAACGAAA	SI00821807
2610101N10Rik	AAGAATGAACCCATATTTAAA	SI00821828
Actl6a	ATGGAGATTGATGGTGACAAA	SI02694097
Actl6a	AGAGTTGATGTTTGAGCACTA	SI04909282
Adsl	CACAAGATTTGCACTGACATA	SI00186382
Adsl	CACTGCAGATACTATATTAAA	SI00186375
Brix1	ATGGATCGTAAAGATAAATTA	SI00787451
Brix1	TACGTTCACCATTCTGGATAA	SI00787465
Ctps	CAGTGTGGGCACAATACTTAA	SI00962115
Ctps	CACAGGGAAGATATACCAGTA	SI00962129
Ctu1	CTGGAGAAAGTTTGTTCTGAA	SI00916958
Ctu1	CAACATGAACCTTGAAGTATA	SI00916965
D10Wsu52e	CACCATGAGTCGTAACTACAA	SI00965461
D10Wsu52e	CACGTGGTGGATGGGAAGGAA	SI00965468
Dcakd	CACCCTGAGATCCGCAAGGAA	SI00865998
Dcakd	CACAGTGGTAGTGTACTGTGA	SI00866012
Ddx18	AAGAAGAATTAAAGCAAATTA	SI00975982
Ddx18	CAAATCCTACATAAGAGCATA	SI05171341
Ddx23	GGCCGTGGTATTGATATCCAA	SI00976129
Ddx23	CCCGGAGAAGATGTTGGCCAA	SI02771503
Ddx47	TCGGATCCCTTAATAAGTTTA	SI00976465
Ddx47	TACCTGGTTTATATTCTCAAT	SI00976472
Dhx38	CAGCAGTGAGTTTGCCAAGAA	SI00979118
Dhx38	CAGGCAAATCAACGAGGACAA	SI00979132
Dna2	AACGCTGGAGTCGCAATCTAA	SI00980707
Dna2	ATGGTTGAGGTTAACACAGTA	SI00980714
Dnajc19	CCGGATTATGCTATTAAATCA	SI00806155
Dnajc19	CAGAGGTGGATTTGAACCCAA	SI00806176
Dnajc2	CAGGAACTGGAAGATAAGAAA	SI00981463
Dnajc2	GAGGAAGAAGTTAGACAGCAA	SI00981477
Dph5	CACAGTGATCTTATTCTGAGA	SI00816942
Dph5	TCGCGTGTATCTGGAAGCCTA	SI00816956
Ebna1bp2	ATGGAGAAGTTTGCACATTAA	SI00989303
Ebna1bp2	AAGACAGAAACTGAAGAGTAA	SI00989317
Eef1b2	TACGTTGGTATAATCACATCA	SI00990052
Eef1b2	CCAGTGGGATATGGAATTAAA	SI00990031
Eftud2	CCGGCTGATCTTGGAGCTGAA	SI01306095
Eftud2	TTCGTAGATTGTTTAATAGAA	SI01306116
Ermp1	ATGGGTTTGATTATACTGGAA	SI00968219
Ermp1	TACAAGGATGTTTGTAATACA	SI00968233
Fam136a	ATCCTAAATATTCATATGCAA	SI00808395
Fam136a	CTGAATCTACAGAGCCTTATA	SI00808416
Fip1l1	AAGGATCTAGATGAGAATGAA	SI01003205
Fip1l1	CTGGATAATGTTTAAGAAATA	SI01003191

Grcc10	ATGGACGAGGCTAGAGACAAT	SI01058064
Grcc10	CAGCCTGTCAGGCAAGCTGAA	SI01058057
H2afv	CACGTCACTTACAGCTTGCAA	SI01061291
H2afv	CAGTGGGAAGGCCAAGGCTAA	SI01061305
Hspe1	CAAAGTAGTTCTAGATGACAA	SI00174076
Hspe1	TCGACTGAAATCACTGTTGAA	SI02708951
Ints5	CGAGTGCATGCAGGGACTTTA	SI00786877
Ints5	ACCATCATCTCCCGAGTCCTA	SI00786884
Klhl24	AAGGTCCAATCTTATGACCCA	SI04941440
Klhl24	CAGGATCATCCCATGCTGAAA	SI04941461
Mdn1	CTGGAATTCGGTGAACCCGTA	SI01303092
Mdn1	CCCTGACGACCTGACGTTCAA	SI01303078
Mki67ip	TACCTTCAACACTTTCTGAAA	SI01313011
Mki67ip	GACTTCTATCATGTAAATTTA	SI01313025
Mrto4	TCCTGCTGTTTGGGTATGAGA	SI00820253
Mrto4	CACCAAGTATACAGAAATGGA	SI00820260
Mstn	AACAATTACATTTATATAATA	SI01010975
Mstn	TACACTTATATTTGTATTGTA	SI01010982
Myh7	AAAGAAGGACTTTGAGTTAAA	SI01321383
Myh7	CTCAAGCTGCTCAGCAATCTA	SI01321404
Naa10	CGCCGCGATGAACATCCGCAA	SI00901600
Naa10	AGCCATGATAGAGAACTTCAA	SI00901579
Ncaph	CACACAGATTACAGAACATTA	SI00931007
Ncaph	AAGAAAGACTTTGAAATTAAT	SI00931021
Nhp2	TGGGATAGAAGTAAATGCTAA	SI01329223
Nhp2	CACTTGTGTGATCATGGTGAA	SI01329237
Nip7	TTGGGTTCTTATCAACAGAAA	SI00784546
Nip7	CAGGCTACTTGGAGAGTTCAA	SI00784553
Nmt2	ATGGGTGGTGATACTATGGAA	SI00178633
Nmt2	ATGGATAGATTTAAGAATGTA	SI02709833
Nol8	AAAGAGCTAAATGAAGATAAA	SI01329174
Nol8	TAGGGATGATTTCAAATCAAA	SI01329167
Nop10	CGGCCTATTATCAGCAACCAA	SI01329265
Nop10	TCGCGTTTATACGCTGAAGAA	SI01329272
Nop2	CAGGAGCATGAGCGGATCTTA	SI01328978
Nop2	AAGACGAACAAGGATGAGAAA	SI01328992
Nsmce2	CAGGTTCTACCCGTTACATAT	SI00784266
Nsmce2	AACTCTGATGCCGACTTTAAA	SI05184655
Nsun4	ACGGGTTGTTGCCGTAATCTA	SI00826273
Nsun4	CGCGGAGATATCGATATAAGA	SI00826280
Nudt21	CTGGTTCAGCTTCAAGAGAAA	SI00958720
Nudt21	CACCCTGGAGCGCACCATTAA	SI04414746
Nvl	AACGATTTCATTGTTGCTCTA	SI01333423
Nvl	CACGCCATTGCTGACGTCTAA	SI01333430
Osbp2	AAGGCTATCCGTGTGATGAAA	SI01366554
Osbp2	CGAGTCCAAGCTGGTGCCAAA	SI01366568
Otud4	TTGGATCATAATGGGAAATTA	SI00973154
Otud4	CAGTATGAAGTTGGAGATAAA	SI00973147
Pabpc4	CCGAAGCCATGCTGTATGAAA	SI01368066

Pabpc4	CAGCGCGAAGGTGATGCTAGA	SI01368073
Phf14	CTGAATGAAGATATTCAAGTA	SI01376907
Phf14	TTTGGAGATATTGATAAATTA	SI01376928
Phldb2	CCCAGAGTTCAGAGCACTTTA	SI01377446
Phldb2	AAAGAGAAGAACAATCTAATA	SI01377453
Pno1	TAGGATAATTTAAATATGATA	SI00803187
Pno1	TACGCTTTAACCTGAAATCAA	SI00803208
Pola2	ACCAGAGACATTGTTTCTATA	SI01383368
Pola2	CAGATTCAGCCGAGTCCTTAA	SI01383361
Prim2	CTCGGTGTGAGCTATGTGAAA	SI01388184
Prim2	AAGATCTTAGACGATGGTTTA	SI01388177
Prpf4b	CAGCAGTTTATTATTAATGTA	SI01389927
Prpf4b	CCGGAGATCATTATAGGTAAA	SI01389934
Prpf6	TACAAGTAGCTCGAAACCTTA	SI00821289
Prpf6	CTGGAGTCAGAGAACAATGAA	SI00821296
Psma1	CCAGACTGTTATGCAACTTTA	SI02672131
Psma1	TCCATTGGAATCGTTGGTAAA	SI02739590
Pwp1	CTGGAATAAGACAGTCAGAAA	SI00814681
Pwp1	CAAGCCAATCTTTACACTTAA	SI00814688
Rars	AACAATAAATATATTGACAAA	SI02674987
Rars	CTGGGTGTTCTGTGCCATTAA	SI00223972
Rbm28	AAGAAAGGCTAAGGAGAATAA	SI01398292
Rbm28	AAGGCTGAGAACTTAGCTTAA	SI04750417
Rcc1	CTGCATGGATTCGGAAGGGAA	SI00949725
Rcc1	CTGGGCCGTGTACCTGAATTA	SI00949732
Rnf20	TAGAGTGAACATATACTATTA	SI02678543
Rnf20	CCGGAAGTTTGAGGAAATGAA	SI02699459
Rnps1	AAGGATCGATCTAAAGATAAA	SI01404347
Rnps1	AAGGATCATATCATGGAAATA	SI01404354
Rpl11	CAGACTGACCCGGGCAGCCAA	SI01405047
Rpl11	CGGGAGTATGAGTTGCGGAAA	SI01405054
Rpl12	TGCCTCAGCCCTGATCATCAA	SI01405082
Rpl12	CTGGGTCTGTCTCCGAAGAAA	SI01405089
Rplp0	CCGGTCTGGATTTATTTAGTT	SI00901502
Rplp0	TAGCTAGTCTGGGACAAAGAT	SI00901516
Rsl1d1	AAGCAGAAAGTTACAGAGGAA	SI00816606
Rsl1d1	CAGACAGAAGGTTTCTATAAA	SI00816620
Rufy3	ATGGAACGAGTTAAAGAAGAA	SI00970802
Rufy3	AACAACAGAATTATTACCTTA	SI00970816
Sap130	CAGGAAATAGCTAATCAGAAA	SI00822507
Sap130	CCGGGTCAAAGAGGAGAAGAA	SI00822521
Sass6	TAGCAGGATGTTTGAAATGTA	SI00827855
Sass6	CTGAATGTGGTTGATAGACTA	SI00827862
Seh1l	TAGGTTGTGGAAAGCTAATTA	SI01413720
Seh1l	CAGCATCGCGGCGGACCACAA	SI04957722
Set	AAGTTGCTTTGTTACAATAAA	SI01415386
Set	ATGTTTAATTGTGTGGTTTAA	SI01415393
Sf3a3	CAGTTGTTTGACATTCCTAAA	SI01415547
Sf3a3	CGGCCTGAATATCAACTACAA	SI01415554

Sf3b3	CTCGGAGATGATGATGAAGAA	SI01415638
Sf3b3	TACGTACAAGCTTGTGAACAA	SI01415652
Sf3b5	CGCGGACAAGCCCGAGGAGAA	SI01415694
Sf3b5	CGGCGCGATGACGGACCGGTA	SI01415701
Sfpq	CTCGGAGGGATTTAAAGCCAA	SI05783848
Sfpq	TCGCGGAAGATCTACAGGGAA	SI05783876
Slc7a6os	CCGAGGATGGATCCCAAGTAA	SI00815682
Slc7a6os	GAAGATATTTATGAAGATGAA	SI00815689
Smarcc2	CTGGATGAACAGGACGATGAA	SI00861406
Smarcc2	CCGATAGTTGATCCTGAGAAA	SI00861413
Srpr	CACGCTGACCTATGTAGACAA	SI01433551
Srpr	ATGGCACATTTGATTTCCAAA	SI01433565
Srsf4	CACCTCCAAATCCAAACCCAA	SI01416170
Srsf4	CCCAGCCGCCATGACAGTAAA	SI01416177
Srsf7	AACAACCAATTTAATACCAAA	SI01416247
Srsf7	CCCGCTCAAGGTCGAGATCAA	SI01416268
Svil	AGGCGCTTAGAAGGAATAGAA	SI02721817
Svil	AGCGGAAAGGATCGCAAGATA	SI02677381
Tagln2	CCCAACTGGTTTCCTAAGAAA	SI01441251
Tagln2	AAGCTTATTAATTCACTGTAT	SI01441272
Taok1	CTGGATGATAACATCAATTTA	SI01441391
Taok1	ATGGAGAAACTTATTAAGAAA	SI01441398
Tbc1d23	CAGGAGGTTGTGATCTTGAAA	SI00224448
Tbc1d23	AAGGGAGATTCTTTCAAGGAA	SI02695658
Thoc6	TGGGATCTCCGTATAGCCAAA	SI00998410
Thoc6	AGGGTTGGCTTTGGGCAGAAA	SI00998424
Tmem205	ACGCCGGAAGTTCTACCACTA	SI00917462
Tmem205	CAAGGTCTTCCCAGTCTATTT	SI00917476
Top2a	CACTGTGAAGTTTGTCATAAA	SI01453697
Top2a	ATCCCTGTTGTTGAACACAAA	SI01453704
Trappc2l	CACCGAGGACTACAAGGTATA	SI00804286
Trappc2l	CCGGAAGCTGCATAACTCTTA	SI00804300
Tsr1	CACGGTGATTTCCAAATGAAT	SI00910329
Tsr1	CAGGATGGGAGTGGTGAGGAA	SI00910336
Ufc1	CCGGTATGTGGAGAACAACAA	SI00785155
Ufc1	ATGAAAGATTTGGGCCTTGAA	SI00785176
Usp39	CCCGTACTTGGATACCATTAA	SI01463777
Usp39	CAAGAGATTTACTAAGAACAA	SI01463784
Utp15	AACGGAATACTGAGTGTTAAA	SI00910210
Utp15	ATGGGATATTCCAAACTCTAA	SI00910217
Vars	CAGGATTTACCACCAATTAAA	SI01468831
Vars	CAGGGCCTCTATGACCAACTA	SI01468838
Wdr18	CGCCTGGGAGCTGCAGCGTAA	SI01471239
Wdr18	CAGATGTTGGGTGGAGCCATA	SI01471260
Wdr43	TGGGACGAAGATGAAGACAAA	SI01471834
Wdr43	AAGAGGAGAGTTGCACAGTAA	SI01471848
Wdr46	CTGGCTTTATATCTACGACAA	SI00217504
Wdr46	ATGGGTGATGTGGTCAACATA	SI00217518
Yipf7	CTAGAGGAACTTGGAATTAAT	SI00812357

Yipf7	**CTGCTTGAAGTGACTTTGCAA**	**SI00812364**
Zfp276	**TGGCATGAAGAAGCACATTAA**	**SI01477854**
Zfp276	**GAAGATGAGAGTGATAAGAAA**	**SI01477868**
Zfp322a	**TTGGAGGTACCAAGTAAGAAA**	**SI01478526**
Zfp322a	**CTGAATGAGGAAAGTTTATTA**	**SI01478533**

Table 2 supp data: The list of 100 genes preselected for the siRNA-based screen with name, symbol, and MGI and Ensembl ID for each gene.

MGI Gene/Marker ID	Symbol	Name	Ensembl ID
MGI:1920987	1110008P14Rik	RIKEN cDNA 1110008P14 gene	ENSMUSG00000039195
MGI:1861453	Actl6a	actin-like 6A	ENSMUSG00000027671
MGI:103202	Adsl	adenylosuccinate lyase	ENSMUSG00000022407
MGI:1915082	Brix1	BRX1, biogenesis of ribosomes	ENSMUSG00000022247
MGI:107184	Cct7	chaperonin containing Tcp1, subunit 7 (eta)	ENSMUSG00000030007
MGI:1858304	Ctps	cytidine 5'-triphosphate synthase	ENSMUSG00000028633
MGI:2385277	Ctu1	cytosolic thiouridylase subunit 1	ENSMUSG00000038888
MGI:106379	Rtcb	RNA 2',3'-cyclic phosphate and 5'-OH ligase	ENSMUSG00000001783
MGI:1915337	Dcakd	dephospho-CoA kinase domain containing	ENSMUSG00000020935
MGI:1914192	Ddx18	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18	ENSMUSG00000001674
MGI:1921601	Ddx23	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	ENSMUSG00000003360
MGI:1915005	Ddx47	DEAD (Asp-Glu-Ala-Asp) box polypeptide 47	ENSMUSG00000030204
MGI:1927617	Dhx38	DEAH (Asp-Glu-Ala-His) box polypeptide 38	ENSMUSG00000037993
MGI:2443732	Dna2	DNA replication helicase/nuclease 2	ENSMUSG00000036875
MGI:1914963	Dnajc19	DnaJ heat shock protein family (Hsp40) member C19	ENSMUSG00000027679
MGI:99470	Dnajc2	DnaJ heat shock protein family (Hsp40) member C2	ENSMUSG00000029014
MGI:1916990	Dph5	diphthamide biosynthesis 5	ENSMUSG00000033554
MGI:1916322	Ebna1bp2	EBNA1 binding protein 2	ENSMUSG00000028729
MGI:1929520	Eef1b2	eukaryotic translation elongation factor 1 beta 2	ENSMUSG00000025967
MGI:1336880	Eftud2	elongation factor Tu GTP binding domain containing 2	ENSMUSG00000020929
MGI:106250	Ermp1	endoplasmic reticulum metallopeptidase 1	ENSMUSG00000046324
MGI:1913738	Fam136a	family with sequence similarity 136, member A	ENSMUSG00000057497
MGI:1914149	Fip1l1	FIP1 like 1 (S. cerevisiae)	ENSMUSG00000029227
MGI:1315201	Grcc10	gene rich cluster, C10 gene	ENSMUSG00000072772
MGI:1924855	H2afv	H2A histone family, member V	ENSMUSG00000041126
MGI:1923578	Ints5	integrator complex subunit 5	ENSMUSG00000071652
MGI:1923035	Klhl24	kelch-like 24	ENSMUSG00000062901
MGI:1926159	Mdn1	midasin AAA ATPase 1	ENSMUSG00000058006
MGI:1913558	Mplkip	M-phase specific PLK1 intereacting protein	ENSMUSG00000012429
MGI:1917152	Mrto4	mRNA turnover 4, ribosome maturation factor	ENSMUSG00000028741
MGI:95691	Mstn	myostatin	ENSMUSG00000026100
MGI:2155600	Myh7	myosin, heavy polypeptide 7, cardiac muscle, beta	ENSMUSG00000053093
MGI:1915255	Naa10	N(alpha)-acetyltransferase 10, NatA catalytic subunit	ENSMUSG00000031388
MGI:2444777	Ncaph	non-SMC condensin I complex, subunit H	ENSMUSG00000034906
MGI:1098547	Nhp2	NHP2 ribonucleoprotein	ENSMUSG00000001056
MGI:1913414	Nip7	NIP7, nucleolar pre-rRNA processing protein	ENSMUSG00000031917
MGI:1202298	Nmt2	N-myristoyltransferase 2	ENSMUSG00000026643
MGI:1918180	Nol8	nucleolar protein 8	ENSMUSG00000021392
MGI:1913431	Nop10	NOP10 ribonucleoprotein	ENSMUSG00000027133
MGI:107891	Nop2	NOP2 nucleolar protein	ENSMUSG00000038279
MGI:1915751	Nsmce2	NSE2/MMS21 homolog, SMC5-SMC6 complex SUMO ligase	ENSMUSG00000059586
MGI:1919431	Nsun4	NOL1/NOP2/Sun domain family, member 4	ENSMUSG00000028706
MGI:1915469	Nudt21	nudix (nucleoside diphosphate linked moiety X)-type motif 21	ENSMUSG00000031754
MGI:1914709	Nvl	nuclear VCP-like	ENSMUSG00000026516
MGI:1921559	Osbp2	oxysterol binding protein 2	ENSMUSG00000020435
MGI:1098801	Otud4	OTU domain containing 4	ENSMUSG00000036990
MGI:2385206	Pabpc4	poly(A) binding protein, cytoplasmic 4	ENSMUSG00000011257
MGI:1923539	Phf14	PHD finger protein 14	ENSMUSG00000029629
MGI:2444981	Phldb2	pleckstrin homology like domain, family B, member 2	ENSMUSG00000033149
MGI:1913499	Pno1	partner of NOB1 homolog	ENSMUSG00000020116
MGI:99690	Pola2	polymerase (DNA directed), alpha 2	ENSMUSG00000024833
MGI:97758	Prim2	DNA primase, p58 subunit	ENSMUSG00000026134
MGI:109584	Prpf4b	pre-mRNA processing factor 4B	ENSMUSG00000021413
MGI:1922946	Prpf6	pre-mRNA splicing factor 6	ENSMUSG00000002455
MGI:1347005	Psma1	proteasome (prosome, macropain) subunit, alpha type 1	ENSMUSG00000030751
MGI:1914735	Pwp1	PWP1 homolog, endonuclein	ENSMUSG00000001785
MGI:1914297	Rars	arginyl-tRNA synthetase	ENSMUSG00000018848
MGI:2655711	Rbm28	RNA binding motif protein 28	ENSMUSG00000029701
MGI:1913989	Rcc1	regulator of chromosome condensation 1	ENSMUSG00000028896
MGI:1925927	Rnf20	ring finger protein 20	ENSMUSG00000028309
MGI:97960	Rnps1	ribonucleic acid binding protein S1	ENSMUSG00000034681
MGI:1914275	Rpl11	ribosomal protein L11	ENSMUSG00000059291
MGI:98002	Rpl12	ribosomal protein L12	ENSMUSG00000038900
MGI:1927636	Rplp0	ribosomal protein, large, P0	ENSMUSG00000067274
MGI:1913659	Rsl1d1	ribosomal L1 domain containing 1	ENSMUSG00000005846
MGI:106484	Rufy3	RUN and FYVE domain containing 3	ENSMUSG00000029291
MGI:1919782	Sap130	Sin3A associated protein	ENSMUSG00000024260
MGI:1920026	Sass6	SAS-6 centriolar assembly protein	ENSMUSG00000027959
MGI:1919374	Seh1l	SEH1-like (S. cerevisiae	ENSMUSG00000079614
MGI:1860267	Set	SET nuclear oncogene	ENSMUSG00000054766
MGI:1922312	Sf3a3	splicing factor 3a, subunit 3	ENSMUSG00000028902
MGI:1289341	Sf3b3	splicing factor 3b, subunit 3	ENSMUSG00000033732
MGI:1913375	Sf3b5	splicing factor 3b, subunit 5	ENSMUSG00000078348
MGI:1918764	Sfpq	splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated)	ENSMUSG00000028820
MGI:1916951	Slc7a6os	solute carrier family 7, member 6 opposite strand	ENSMUSG00000033106
MGI:1915344	Smarcc2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 2	ENSMUSG00000025369
MGI:1914648	Srpr	signal recognition particle receptor ('docking protein')	ENSMUSG00000032042
MGI:1890577	Srsf4	serine/arginine-rich splicing factor 4	ENSMUSG00000028911
MGI:1926232	Srsf7	serine/arginine-rich splicing factor 7	ENSMUSG00000024097
MGI:2147319	Svil	supervillin	ENSMUSG00000024236
MGI:1312985	Tagln2	transgelin 2	ENSMUSG00000026547
MGI:1914490	Taok1	TAO kinase 1	ENSMUSG00000017291
MGI:1914831	Tbc1d23	TBC1 domain family, member 23	ENSMUSG00000022749
MGI:2677480	Thoc6	THO complex 6	ENSMUSG00000041319
MGI:3045495	Tmem205	transmembrane protein 205	ENSMUSG00000040883
MGI:98790	Top2a	topoisomerase (DNA) II alpha	ENSMUSG00000020914
MGI:1916295	Trappc2l	trafficking protein particle complex 2-like	ENSMUSG00000015013
MGI:2144566	Tsr1	TSR1 20S rRNA accumulation	ENSMUSG00000038335
MGI:1915208	U2surp	U2 snRNP-associated SURP domain containing	ENSMUSG00000032407
MGI:1913405	Ufc1	ubiquitin-fold modifier conjugating enzyme 1	ENSMUSG00000062963
MGI:107622	Usp39	ubiquitin specific peptidase 39	ENSMUSG00000056305
MGI:2145443	Utp15	UTP15 small subunit processome component	ENSMUSG00000041747
MGI:90675	Vars	valyl-tRNA synthetase	ENSMUSG00000007029
MGI:2144805	Vipas39	VPS33B interacting protein, apical-basolateral polarity regulator, spe-39 homolog	ENSMUSG00000021038
MGI:2158400	Wdr18	WD repeat domain 18	ENSMUSG00000035754
MGI:1919765	Wdr43	WD repeat domain 43	ENSMUSG00000041057
MGI:1931871	Wdr46	WD repeat domain 46	ENSMUSG00000024312
MGI:1922831	Yipf7	Yip1 domain family, member 7	ENSMUSG00000029158
MGI:1888495	Zfp276	zinc finger protein (C2H2 type) 276	ENSMUSG00000001065
MGI:2442566	Zfp322a	zinc finger protein 322A	ENSMUSG00000046351


REFERENCES
BIBLIOGRAPHIQUES

REFERENCES BIBLIOGRAPHIQUES

A

Abmayr, S.M., and Pavlath, G.K. (2012). Myoblast fusion: lessons from flies and mice. Development *139*, 641-656.

Abuin, A., Hansen, G.M., and Zambrowicz, B. (2007). Gene Trap Mutagenesis. In Conditional Mutagenesis: An Approach to Disease Models, R. Feil, and D. Metzger, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 129-147.

Acampora, D., Gulisano, M., and Simeone, A. (1999). OtxGenes and the Genetic Control of Brain Morphogenesis. Molecular and Cellular Neuroscience *13*, 1-7.

Aikin, R., Cervantes, A., D'Angelo, G., Ruel, L., Lacas-Gervais, S., Schaub, S., and Therond, P. (2012). A genome-wide RNAi screen identifies regulators of cholesterol-modified hedgehog secretion in Drosophila. PloS one *7*, e33665.

Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM(®)), an online catalog of human genes and genetic disorders. Nucleic Acids Research *43*, D789-D798.

Asp, P., Blum, R., Vethantham, V., Parisi, F., Micsinai, M., Cheng, J., Bowman, C., Kluger, Y., and Dynlacht, B.D. (2011). Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. Proceedings of the National Academy of Sciences of the United States of America *108*, E149-E158.

Austin CP, B.J., Bradley A, Bucan M, Capecchi M, Collins FS, Dove, WF, D.G., Dymecki S, Eppig JT, Grieder FB, Heintz N, Hicks G,, Insel TR, J.A., Koller BH, Lloyd KC, Magnuson T, Moore MW,, Nagy A, P.J., Roses AD, Sands AT, Seed B, Skarnes WC,, Snoddy J, S.P., Stewart DJ, Stewart F, Stillman B, Varmus H,, Varticovski L, V.I., Vogt TF, von Melchner H, Witkowski J,, and Woychik RP, W.W., Yancopoulos GD, Young SG, Zambrowicz B. (2004). The Knockout Mouse Project. Nature genetics *36*, 921-924.

Avet-Rochex, A., Boyer, K., Polesello, C., Gobert, V., Osman, D., Roch, F., Auge, B., Zanet, J., Haenlin, M., and Waltzer, L. (2010). An in vivo RNA interference screen identifies gene networks controlling Drosophila melanogaster blood cell homeostasis. BMC developmental biology *10*, 65.

Ayadi, A., Birling, M.-C., Bottomley, J., Bussell, J., Fuchs, H., Fray, M., Gailus-Durner, V., Greenaway, S., Houghton, R., Karp, N.*, et al.* (2012). Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. Mammalian Genome *23*, 600-610.

Aza-Blanc, P., Cooper, C.L., Wagner, K., Batalov, S., Deveraux, Q.L., and Cooke, M.P. Identification of Modulators of TRAIL-Induced Apoptosis via RNAi-Based Phenotypic Screening. Molecular cell *12*, 627-637.

B

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes & development *22*, 2773-2785.

Bai, J., Binari, R., Ni, J.Q., Vijayakanthan, M., Li, H.S., and Perrimon, N. (2008). RNA interference screening in Drosophila primary cells for genes involved in muscle assembly and maintenance. Development *135*, 1439-1449.

Barbaric, I., Wells, S., Russ, A., and Dear, T.N. (2007). Spectrum of ENU-induced mutations in phenotype-driven and gene-driven screens in the mouse. Environmental and Molecular Mutagenesis *48*, 124-142.

Barton, G.M., and Medzhitov, R. (2002). Retroviral delivery of small interfering RNA into primary cells. Proceedings of the National Academy of Sciences of the United States of America *99*, 14943-14945.
Becker, L., Kling, E., Schiller, E., Zeh, R., Schrewe, A., Hölter, S.M., Mossbrugger, I., Calzada-Wack, J., Strecker, V., Wittig, I.*, et al.* (2014). MTO1-Deficient Mouse Model Mirrors the Human Phenotype Showing Complex I Defect and Cardiomyopathy. PloS one *9*, e114918.
Bentzinger, C.F., Wang, Y.X., and Rudnicki, M.A. (2012). Building Muscle: Molecular Regulation of Myogenesis. Cold Spring Harbor perspectives in biology *4*, a008342.
Bernards, R., Brummelkamp, T.R., and Beijersbergen, R.L. (2006). shRNA libraries and their use in cancer genetics. Nat Meth *3*, 701-706.
Berns, K., Hijmans, E.M., Mullenders, J., Brummelkamp, T.R., Velds, A., Heimerikx, M., Kerkhoven, R.M., Madiredjo, M., Nijkamp, W., Weigelt, B.*, et al.* (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature *428*, 431-437.
Blewitt, M.E., Vickaryous, N.K., Hemley, S.J., Ashe, A., Bruxner, T.J., Preis, J.I., Arkell, R., and Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proceedings of the National Academy of Sciences of the United States of America *102*, 7629-7634.
Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Consortium, H.F.A., Paro, R., and Perrimon, N. (2004). Genome-Wide RNAi Analysis of Growth and Viability in Drosophila Cells. Science *303*, 832-835.
Bradley A, R.-S.R., Zheng H, Hasty P, Davis A (1992). Genetic manipulation of the mouse via gene targeting in embryonic stem cells. Ciba Foundation Symposium.
Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.
Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen. Science *319*, 921-926.
Brass, A.L.H., I. Chueh Benita, Yair John, Sinu P. Krishnan, Manoj N. Feeley, Eric M., Ryan, B.J., Weyer, J.L., van der Weyden, L., Fikrig, E., Adams, D.J., Xavier, R.J., Farzan, M., and Elledge, S.J. (2009). The IFITM Proteins Mediate Cellular Resistance to Influenza A H1N1 Virus, West Nile Virus, and Dengue Virus. Cell *139*, 1243-1254.
Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E., and Arnold, H.H. (1989). A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. The EMBO journal *8*, 701-709.
Bric, A., Miething, C., Bialucha, C.U., Scuoppo, C., Zender, L., Krasnitz, A., Xuan, Z., Zuber, J., Wigler, M., Hicks, J.*, et al.* (2009). Functional identification of tumor-suppressor genes through an in vivo RNA interference screen in a mouse lymphoma model. Cancer cell *16*, 324-335.
Brodersen, P., and Voinnet, O. (2009). Revisiting the principles of microRNA target recognition and mode of action. Nature reviews Molecular cell biology *10*, 141-148.
Brown SDM, N.P. (1998). Mouse mutagenesis - systematic studies of mammalian gene function. Human Molecular Genetics *7*, 1627-1633.
Brummelkamp, T.R., Bernards, R., and Agami, R. (2002a). Stable suppression of tumorigenicity by virus-mediated RNA interference. Cancer cell *2*, 243-247.
Brummelkamp, T.R., Bernards, R., and Agami, R. (2002b). A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. Science *296*, 550-553.
Buchovecky, C.M., Turley, S.D., Brown, H.M., Kyle, S.M., McDonald, J.G., Liu, B., Pieper, A.A., Huang, W., Katz, D.M., Russell, D.W.*, et al.* (2013). A suppressor screen in mouse Mecp2 implicates cholesterol metabolism in Rett Syndrome. Nature genetics *45*, 1013-1020.
Bull, K.R., Rimmer, A.J., Siggs, O.M., Miosge, L.A., Roots, C.M., Enders, A., Bertram, E.M., Crockford, T.L., Whittle, B., Potter, P.K.*, et al.* (2013). Unlocking the Bottleneck in Forward Genetics Using Whole-Genome Sequencing and Identity by Descent to Isolate Causative Mutations. PLoS genetics *9*, e1003219.

Burnett, John C., and Rossi, John J. (2012). RNA-Based Therapeutics: Current Progress and Future Prospects. Chemistry & Biology *19*, 60-71.
Bushati, N., and Cohen, S.M. (2007). microRNA functions. Annual review of cell and developmental biology *23*, 175-205.

C

Camus, A.B., J; Babinet, C (1998). A la recherche de gènes impliqués dans le développement embryonnaire de la souris : le "piégeage" de gènes - See more at: http://ipubli-inserm.inist.fr/handle/10608/933#sthash.11beoTW3.dpuf. Med Sci (Paris) *Vol. 14*.
Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., MacQuarrie, K.L., Davison, J., Morgan, M.T., Ruzzo, W.L.*, et al.* (2010). Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Developmental cell *18*, 662-674.
Capecchi, M. (1989). Altering the genome by homologous recombination. Science *244*, 1288-1292.
Caplen, N.J., Fleenor, J., Fire, A., and Morgan, R.A. (2000). dsRNA-mediated gene silencing in cultured Drosophila cells: a tissue culture model for the analysis of RNA interference. Gene *252*, 95-105.
Casadei, L., Vallorani, L., Gioacchini, A.M., Guescini, M., Burattini, S., D'Emilio, A., Biagiotti, L., Falcieri, E., and Stocchi, V. (2009). Proteomics-based investigation in C2C12 myoblast differentiation. European Journal of Histochemistry : EJH *53*, e31.
Castel, S.E., and Martienssen, R.A. (2013). RNA interference (RNAi) in the Nucleus: roles for small RNA in transcription, epigenetics and beyond. Nature reviews Genetics *14*, 100-112.
Chang, K., Elledge, S.J., and Hannon, G.J. (2006). Lessons from Nature: microRNA-based shRNA libraries. Nat Meth *3*, 707-714.
Chen, X. (2008). A silencing safeguard: links between RNA silencing and mRNA processing in Arabidopsis. Developmental cell *14*, 811-812.
Chen, Y.-T., Liu, P., and Bradley, A. (2004). Inducible Gene Trapping with Drug-Selectable Markers and Cre/loxP To Identify Developmentally Regulated Genes. Molecular and cellular biology *24*, 9930-9941.
Chia, N.-Y.C., Yun-Shen Feng, Bo Lu, Xinyi Orlov, Yuriy L. Moreau, Dimitri Kumar, Pankaj Yang, Lin Jiang, Jianming Lau, Mei-Sheng Huss, Mikael Soh, Boon-Seng Kraus, Petra Li, Pin Lufkin, Thomas Lim, Bing Clarke, Neil D. Bard, Frederic Ng, Huck-Hui (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316-320.
Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A., and Dixon, J.E. (2000). Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proceedings of the National Academy of Sciences *97*, 6499-6503.
Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U., and Macino, G. (1996). Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. The EMBO journal *15*, 3153-3163.
collins, F.S. (2007). A Mouse for All Reasons. Cell *128*, 9-13.
Collins, F.S., Finnell, R.H., Rossant, J., and Wurst, W. (2007). A New Partner for the International Knockout Mouse Consortium. Cell *129*, 235.
Cordes, S.P. (2005). N-Ethyl-N-Nitrosourea Mutagenesis: Boarding the Mouse Mutant Express. Microbiology and Molecular Biology Reviews *69*, 426-439.
Cornelison, D.D. (2008). Context matters: in vivo and in vitro influences on muscle satellite cell activity. Journal of cellular biochemistry *105*, 663-669.
Couto, L.B., and High, K.A. (2010). Viral vector-mediated RNA interference. Current Opinion in Pharmacology *10*, 534-542.

Cox, B.J., Vollmer, M., Tamplin, O., Lu, M., Biechele, S., Gertsenstein, M., van Campenhout, C., Floss, T., Kühn, R., Wurst, W.*, et al.* (2010). Phenotypic annotation of the mouse X chromosome. Genome Research *20*, 1154-1164.
Cronin, S.J., Nehme, N.T., Limmer, S., Liegeois, S., Pospisilik, J.A., Schramek, D., Leibbrandt, A., Simoes Rde, M., Gruber, S., Puc, U.*, et al.* (2009). Genome-wide RNAi screen identifies genes involved in intestinal pathogenic bacterial infection. Science *325*, 340-343.
Cullen, L.M., and Arndt, G.M. (2005). Genome-wide screening for gene function using RNAi in mammalian cells. Immunology and cell biology *83*, 217-223.
Curtis, D.J., Hall, M.A., Van Stekelenburg, L.J., Robb, L., Jane, S.M., and Begley, C.G. (2004). SCL is required for normal function of short-term repopulating hematopoietic stem cells. Blood *103*, 3342-3348.

D

Davidson, B.L., and McCray, P.B. (2011). Current prospects for RNA interference-based therapies. Nature reviews Genetics *12*, 329-340.
Davis, A.P., and Justice, M.J. (1998). Mouse alleles: if you've seen one, you haven't seen them all. Trends in Genetics *14*, 438-441.
Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.
De-Zolt, S., Schnütgen, F., Seisenberger, C., Hansen, J., Hollatz, M., Floss, T., Ruiz, P., Wurst, W., and von Melchner, H. (2006). High-throughput trapping of secretory pathway genes in mouse embryonic stem cells. Nucleic Acids Research *34*, e25-e25.
de Angelis, M.H., and Balling, R. (1998). Large scale ENU screens in the mouse: genetics meets genomics. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *400*, 25-32.
de Angelis, M.H., Flaswinkel, H., Fuchs, H., Rathkolb, B., Soewarto, D., Marschall, S., Heffner, S., Pargent, W., Wuensch, K., Jung, M.*, et al.* (2000). Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nature genetics *25*, 444-447.
de Angelis, M.H., Nicholson, G., Selloum, M., White, J.K., Morgan, H., Ramirez-Solis, R., Sorg, T., Wells, S., Fuchs, H., Fray, M.*, et al.* (2015). Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. Nature genetics *47*, 969-978.
Deshmukh, A.S., Murgia, M., Nagaraj, N., Treebak, J.T., Cox, J., and Mann, M. (2015). Deep Proteomics of Mouse Skeletal Muscle Enables Quantitation of Protein Isoforms, Metabolic Pathways, and Transcription Factors. Molecular & cellular proteomics : MCP *14*, 841-853.
Devroe, E., and Silver, P.A. (2002). Retrovirus-delivered siRNA. BMC Biotechnology *2*, 15-15.
Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H.*, et al.* (2016). High-throughput discovery of novel developmental phenotypes. Nature *537*, 508-514.
Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S.*, et al.* (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature *448*, 151-156.
Dimitriadi, M., Sleigh, J.N., Walker, A., Chang, H.C., Sen, A., Kalloo, G., Harris, J., Barsby, T., Walsh, M.B., Satterlee, J.S.*, et al.* (2010). Conserved Genes Act as Modifiers of Invertebrate SMN Loss of Function Defects. PLoS genetics *6*, e1001172.
Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576-578.
Drysdale, R. (2008). FlyBase. In Drosophila: Methods and Protocols, C. Dahmann, ed. (Totowa, NJ: Humana Press), pp. 45-59.
Dunnett, C.W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control. Journal of the American Statistical Association *50*, 1096-1121.

E

Echeverri, C.J., and Perrimon, N. (2006). High-throughput RNAi screening in cultured cells: a user's guide. Nature reviews Genetics *7*, 373-384.
Eggert, U.S., Kiger, A.A., Richter, C., Perlman, Z.E., Perrimon, N., Mitchison, T.J., and Field, C.M. (2004). Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. PLoS biology *2*, e379.
El Hakam Kamareddin, C., Magnol, L., and Blanquet, V. (2015). A new Otogelin ENU mouse model for autosomal-recessive nonsyndromic moderate hearing impairment. SpringerPlus *4*, 730.
Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.
Eppig, J.T., Richardson, J.E., Kadin, J.A., Ringwald, M., Blake, J.A., and Bult, C.J. (2015a). Mouse Genome Informatics (MGI): reflecting on 25 years. Mammalian Genome *26*, 272-284.
Eppig, J.T., Richardson, J.E., Kadin, J.A., Smith, C.L., Blake, J.A., Bult, C.J., and the, M.G.D.T. (2015b). Mouse Genome Database: From sequence to phenotypes and disease models. genesis *53*, 458-473.
Eulalio, A., Behm-Ansmant, I., and Izaurralde, E. (2007). P bodies: at the crossroads of post-transcriptional pathways. Nature reviews Molecular cell biology *8*, 9-22.
Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

F

Favor, J. (1998). The mutagenic activity of ethylnitrosourea at low doses in spermatogonia of the mouse as assessed by the specific-locus test. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *405*, 221-226.
Fernandes, C., and Rao, Y. (2011). Genome-wide screen for modifiers of Parkinson's disease genes in Drosophila. Molecular Brain *4*, 17-17.
Fernandez, L., Marchuk, D.A., Moran, J.L., Beier, D.R., and Rockman, H.A. (2009). An N-ethyl-N-nitrosourea mutagenesis recessive screen identifies two candidate regions for murine cardiomyopathy that map to chromosomes 1 and 15. Mammalian genome : official journal of the International Mammalian Genome Society *20*, 296-304.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.
Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature *408*, 325-330.
Friedel, R.H., Seisenberger, C., Kaloff, C., and Wurst, W. (2007). EUCOMM—the European Conditional Mouse Mutagenesis Program. Briefings in functional genomics & proteomics *6*, 180-185.
Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes & development *5*, 1513-1523.

G

Galvani, A., and Sperling, L. (2002). RNA interference by feeding in Paramecium. Trends in Genetics *18*, 11-12.
Gantier, M.P., and Williams, B.R. (2007). The response of mammalian cells to double-stranded RNA. Cytokine & growth factor reviews *18*, 363-371.
Gavrilov, K., and Saltzman, W.M. (2012). Therapeutic siRNA: Principles, Challenges, and Strategies. The Yale Journal of Biology and Medicine *85*, 187-200.

Ge, Y., Waldemer, R.J., Nalluri, R., Nuzzi, P.D., and Chen, J. (2013). RNAi Screen Reveals Potentially Novel Roles of Cytokines in Myoblast Differentiation. PloS one *8*, e68068.
Geng, Y., Whoriskey, W., Park, M.Y., Bronson, R.T., Medema, R.H., Li, T., Weinberg, R.A., and Sicinski, P. (1999). Rescue of Cyclin D1 Deficiency by Knockin Cyclin E. Cell *97*, 767-777.
Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nature reviews Genetics *10*, 94-108.
Godinho, S.I.H., and Nolan, P.M. (2005). The role of mutagenesis in defining genes in behaviour. Eur J Hum Genet *14*, 651-659.
Gogos, J.A., Thompson, R., Lowry, W., Sloane, B.F., Weintraub, H., and Horwitz, M. (1996). Gene trapping in differentiating cell lines: regulation of the lysosomal protease cathepsin B in skeletal myoblast growth and fusion. The Journal of cell biology *134*, 837-847.
Gonczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S.J., Copley, R.R., Duperon, J., Oegema, J., Brehm, M., Cassin, E.*, et al.* (2000). Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. Nature *408*, 331-336.
Gondo, Y. (2008). Trends in large-scale mouse mutagenesis: from genetics to functional genomics. Nature reviews Genetics *9*, 803-810.
Gosai, S.J., Kwak, J.H., Luke, C.J., Long, O.S., King, D.E., Kovatch, K.J., Johnston, P.A., Shun, T.Y., Lazo, J.S., Perlmutter, D.H.*, et al.* (2010). Automated High-Content Live Animal Drug Screening Using C. elegans Expressing the Aggregation Prone Serpin α1-antitrypsin Z. PloS one *5*, e15460.
Gossler, A., Joyner, A., Rossant, J., and Skarnes, W. (1989). Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science *244*, 463-465.
Gow, M., Mirembe, D., Longwe, Z., and Pickard, B.S. (2013). A gene trap mutagenesis screen for genes underlying cellular response to the mood stabilizer lithium. Journal of Cellular and Molecular Medicine *17*, 657-663.
Grabiec, K., Majewska, A., Wicik, Z., Milewska, M., Błaszczyk, M., and Grzelkowska-Kowalczyk, K. (2016). The effect of palmitate supplementation on gene expression profile in proliferating myoblasts. Cell Biology and Toxicology *32*, 185-198.
Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F., and Kay, M.A. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature *441*, 537-541.
Grishok, A. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23-34.
Grishok, A., Sinskey, J.L., and Sharp, P.A. (2005). Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. Genes & development *19*, 683-696.
Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611-620.

H

Haendel, M.A., Vasilevsky, N., Brush, M., Hochheiser, H.S., Jacobsen, J., Oellrich, A., Mungall, C.J., Washington, N., Köhler, S., Lewis, S.E.*, et al.* (2015). Disease insights through cross-species phenotype comparisons. Mammalian Genome *26*, 548-555.
Hall, B., Limaye, A., and Kulkarni, A.B. (2009). Overview: Generation of Gene Knockout Mice. Current protocols in cell biology / editorial board, Juan S Bonifacino [et al] *CHAPTER*, Unit-19.1217.
Hammell, C.M., and Hannon, G.J. (2012). Inducing RNAi in C. elegans by Feeding with dsRNA-expressing E. coli. Cold Spring Harbor Protocols *2012*, pdb.prot072348.
Hanks, M., Wurst, W., Anson-Cartwright, L., Auerbach, A., and Joyner, A. (1995). Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2. Science *269*, 679-682.
Hanks, M.C., Loomis, C.A., Harris, E., Tong, C.X., Anson-Cartwright, L., Auerbach, A., and Joyner, A. (1998). Drosophila engrailed can substitute for mouse Engrailed1 function in mid-hindbrain, but not limb development. Development *125*, 4521-4530.

Hansen, J., Floss, T., Van Sloun, P., Füchtbauer, E.-M., Vauti, F., Arnold, H.-H., Schnütgen, F., Wurst, W., von Melchner, H., and Ruiz, P. (2003). A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. Proceedings of the National Academy of Sciences of the United States of America *100*, 9918-9922.
Hardisty-Hughes, R.E., Parker, A., and Brown, S.D.M. (2010). A hearing and vestibular phenotyping pipeline to identify mouse mutants with hearing impairment. Nat Protocols *5*, 177-190.
Harno, E., Cottrell, Elizabeth C., and White, A. (2013). Metabolic Pitfalls of CNS Cre-Based Technology. Cell Metabolism *18*, 21-28.
Herron, B.J., Lu, W., Rao, C., Liu, S., Peters, H., Bronson, R.T., Justice, M.J., McDonald, J.D., and Beier, D.R. (2002). Efficient generation and mapping of recessive developmental mutations using ENU mutagenesis. Nature genetics *30*, 185-189.
Hicks, G.G., Shi, E.G., Li, X.M., Li, C.H., Pawlak, M., and Ruley, H.E. (1997). Functional genomics in mice by tagged sequence mutagenesis. Nature genetics *16*, 338-344.
Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E., and Prydz, H. (2002). Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. Nucleic Acids Research *30*, 1757-1766.
Horner, V.L., and Caspary, T. (2011). Creating a "hopeful monster": Mouse forward genetic screens. Methods in molecular biology (Clifton, NJ) *770*, 10.1007/1978-1001-61779-61210-61776_61712.
Houzet, L.J., Kuan-Teh (2011). Genome-wide screening using RNAi (RNA interference) to study host factors in viral replication and pathogenesis. Experimental biology and medicine (Maywood, NJ) *236*, 962-967.
Hu, G., and Luo, J. (2012). A primer on using pooled shRNA libraries for functional genomic screens. Acta Biochimica et Biophysica Sinica *44*, 103-112.
Hutvagner, G. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

I

Infrafrontier Consortium (2015). INFRAFRONTIER—providing mutant mouse resources as research tools for the international scientific community. Nucleic Acids Research *43*, D1171-D1175.
Inoue, M., Sakuraba, Y., Motegi, H., Kubota, N., Toki, H., Matsui, J., Toyoda, Y., Miwa, I., Terauchi, Y., Kadowaki, T.*, et al.* (2004). A series of maturity onset diabetes of the young, type 2 (MODY2) mouse models generated by a large-scale ENU mutagenesis program. Human Molecular Genetics *13*, 1147-1157.
International Gene Trap, C., Skarnes, W.C., von Melchner, H., Wurst, W., Hicks, G., Nord, A.S., Cox, T., Young, S.G., Ruiz, P., Soriano, P.*, et al.* (2004). A public gene trap resource for mouse functional genomics. Nature genetics *36*, 543-544.
Ishiguro, T., Avila, H., Lin, S.Y., Nakamura, T., Yamamoto, M., and Boyd, D.D. (2010). Gene trapping identifies chloride channel 4 as a novel inducer of colon cancer cell migration, invasion and metastases. British Journal of Cancer *102*, 774-782.
Ivey, K.N., and Srivastava, D. MicroRNAs as Regulators of Differentiation and Cell Fate Decisions. Cell stem cell *7*, 36-41.

J

Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J.M., Lim, L., Karpilow, J., Nichols, K.*, et al.* (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. RNA *12*, 1197-1205.
Jenkins, N.A., and Copeland, N.G. (1985). High frequency germline acquisition of ecotropic MuLV proviruses in SWR/J-RF/J hybrid mice. Cell *43*, 811-819.

Johan Auwerx1, P.A., Richard Baldock3, Andrea Ballabio4, Rudi Balling5, Mariano Barbacid6, Anton Berns7, Allan Bradley8, Steve Brown9, 33, Peter Carmeliet10, Pierre Chambon2, Roger Cox9, Duncan Davidson3, Kay Davies11, Denis Duboule12, Jiri Forejt13, Francesca Granucci14, Nick Hastie3, Martin Hrabé de Angelis15, Ian Jackson3, Dimitris Kioussis16, George Kollias17, Mark Lathrop18, Urban Lendahl19, Marcos Malumbres20, Harald von Melchner21, Werner Müller22, Juha Partanen23, Paola Ricciardi-Castagnoli14, Peter Rigby24, Barry Rosen25, Nadia Rosenthal26, Bill Skarnes25, A Francis Stewart27, 33, Janet Thornton28, Glauco Tocchini-Valentini29, Erwin Wagner30, Walter Wahli31 & Wolfgang Wurst32, 33 (2004). The European dimension for the mouse genome mutagenesis program. Nature genetics *36*, 925-927.
Johan Auwerx, P.A., Richard Baldock3, Andrea Ballabio4, Rudi Balling5, Mariano Barbacid6, Anton Berns7, Allan Bradley8, Steve Brown9, 33, Peter Carmeliet10, Pierre Chambon2, Roger Cox9, Duncan Davidson3, Kay Davies11, Denis Duboule12, Jiri Forejt13, Francesca Granucci14, Nick Hastie3, Martin Hrabé de Angelis15, Ian Jackson3, Dimitris Kioussis16, George Kollias17, Mark Lathrop18, Urban Lendahl19, Marcos Malumbres20, Harald von Melchner21, Werner Müller22, Juha Partanen23, Paola Ricciardi-Castagnoli14, Peter Rigby24, Barry Rosen25, Nadia Rosenthal26, Bill Skarnes25, A Francis Stewart27, 33, Janet Thornton28, Glauco Tocchini-Valentini29, Erwin Wagner30, Walter Wahli31 & Wolfgang Wurst32, 33 (2004). The European dimension for the mouse genome mutagenesis program. Nature genetics *36*, 925-927.
Justice, M.J., Noveroske, J.K., Weber, J.S., Zheng, B., and Bradley, A. (1999). Mouse ENU Mutagenesis. Human Molecular Genetics *8*, 1955-1963.

K

Kakugawa, K., Yasuda, T., Miura, I., Kobayashi, A., Fukiage, H., Satoh, R., Matsuda, M., Koseki, H., Wakana, S., Kawamoto, H.*, et al.* (2009). A Novel Gene Essential for the Development of Single Positive Thymocytes. Molecular and cellular biology *29*, 5128-5135.
Kamath, R.S., and Ahringer, J. (2003). Genome-wide RNAi screening in Caenorhabditis elegans. Methods *30*, 313-321.
Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M.*, et al.* (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231-237.
Kauppi, M., Murphy, J.M., de Graaf, C.A., Hyland, C.D., Greig, K.T., Metcalf, D., Hilton, A.A., Nicola, N.A., Kile, B.T., Hilton, D.J.*, et al.* (2008). Point mutation in the gene encoding p300 suppresses thrombocytopenia in Mpl(−/−) mice. Blood *112*, 3148-3153.
Kawasaki, H., Suyama, E., Iyo, M., and Taira, K. (2003). siRNAs generated by recombinant human Dicer induce specific and significant but target site-independent gene silencing in human cells. Nucleic Acids Research *31*, 981-987.
Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell *95*, 1017-1026.
Khanjyan, M.V., Yang, J., Kayali, R., Caldwell, T., and Bertoni, C. (2013). A high-content, high-throughput siRNA screen identifies cyclin D2 as a potent regulator of muscle progenitor cell fusion and a target to enhance muscle regeneration. Human Molecular Genetics *22*, 3283-3295.
Khoriaty, R., Everett, L., Chase, J., Zhu, G., Hoenerhoff, M., McKnight, B., Vasievich, M.P., Zhang, B., Tomberg, K., Williams, J.*, et al.* (2016). Pancreatic SEC23B deficiency is sufficient to explain the perinatal lethality of germline SEC23B deficiency in mice. Scientific Reports *6*, 27802.
Kiger, A.A., Baum, B., Jones, S., Jones, M.R., Coulson, A., Echeverri, C., and Perrimon, N. (2003). A functional genomic analysis of cell morphology using RNA interference. J Biol *2*, 27.
Kile, B.T., and Hilton, D.J. (2005). The art and design of genetic screens: mouse. Nature reviews Genetics *6*, 557-567.

Kim, J.-K., Kim, E., Baek, I.-C., Kim, B.-K., Cho, A.-R., Kim, T.-Y., Song, C.-W., Seong, J.K., Yoon, J.-B., Stenn, K.S.*, et al.* (2010). Overexpression of Hr links excessive induction of Wnt signaling to Marie Unna hereditary hypotrichosis. Human Molecular Genetics *19*, 445-453.
Kim, J.H., Jin, P., Duan, R., and Chen, E.H. (2015). Mechanisms of myoblast fusion during muscle development. Current Opinion in Genetics & Development *32*, 162-170.
Kim, J.K., Gabel, H.W., Kamath, R.S., Tewari, M., Pasquinelli, A., Rual, J.-F., Kennedy, S., Dybbs, M., Bertin, N., Kaplan, J.M.*, et al.* (2005). Functional Genomic Analysis of RNA Interference in C. elegans. Science *308*, 1164-1167.
Kim, S.-S., Garg, H., Joshi, A., and Manjunath, N. (2009). Strategies for targeted nonviral delivery of siRNAs in vivo. Trends in molecular medicine *15*, 491-500.
Kim, S.I., Shin, D., Choi, T.H., Lee, J.C., Cheon, G.J., Kim, K.Y., Park, M., and Kim, M. (2007). Systemic and specific delivery of small interfering RNAs to the liver mediated by apolipoprotein A-I. Mol Ther *15*, 1145-1152.
Kim, S.K. (2001). http://C.Elegans: Mining the functional genomic landscape. Nature reviews Genetics *2*, 681-689.
Kislinger, T., Gramolini, A.O., Pan, Y., Rahman, K., MacLennan, D.H., and Emili, A. (2005). Proteome dynamics during C2C12 myoblast differentiation. Molecular & cellular proteomics : MCP *4*, 887-901.
Kittler, R., Surendranath, V., Heninger, A.-K., Slabicki, M., Theis, M., Putz, G., Franke, K., Caldarelli, A., Grabner, H., Kozak, K.*, et al.* (2007). Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. Nat Meth *4*, 337-344.
Koscielny, G., Yaikhom, G., Iyer, V., Meehan, T.F., Morgan, H., Atienza-Herrero, J., Blake, A., Chen, C.-K., Easty, R., Di Fenza, A.*, et al.* (2014). The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. Nucleic Acids Research *42*, D802-D809.
Krauss, R.S., Cole, F., Gaio, U., Takaesu, G., Zhang, W., and Kang, J.-S. (2005). Close encounters: regulation of vertebrate skeletal myogenesis by cell-cell contact. Journal of cell science *118*, 2355-2362.
Krishnan, M.N., Ng, A., Sukumaran, B., Gilfoy, F.D., Uchil, P.D., Sultana, H., Brass, A.L., Adametz, R., Tsui, M., Qian, F.*, et al.* (2008). RNA interference screen for human genes associated with West Nile virus infection. Nature *455*, 242-245.
Kühn, R., and Schwenk, F. (1997). Advances in gene targeting methods. Current Opinion in Immunology *9*, 183-188.
Kuhnert, F., and Stuhlmann, H. (2004). Identifying Early Vascular Genes Through Gene Trapping in Mouse Embryonic Stem Cells. In Current topics in developmental biology (Academic Press), pp. 261-281.
Kurreck, J. (2009). RNA Interference: From Basic Research to Therapeutic Applications. Angewandte Chemie International Edition *48*, 1378-1398.
Kusakabe, R., and Inoue, K. (2015). Developmental regulation and evolution of muscle-specific microRNAs. Seminars in cell & developmental biology.

L

Lambe, T., Simpson, R.J., Dawson, S., Bouriez-Jones, T., Crockford, T.L., Lepherd, M., Latunde-Dada, G.O., Robinson, H., Raja, K.B., Campagna, D.R.*, et al.* (2009). Identification of a Steap3 endosomal targeting motif essential for normal iron metabolism. Blood *113*, 1805-1808.
Lamesch, P., Milstein, S., Hao, T., Rosenberg, J., Li, N., Sequerra, R., Bosak, S., Doucette-Stamm, L., Vandenhaute, J., Hill, D.E.*, et al.* (2004). C. elegans ORFeome Version 3.1: Increasing the Coverage of ORFeome Resources With Improved Gene Predictions. Genome Research *14*, 2064-2069.
Lares, M.R., Rossi, J.J., and Ouellet, D.L. (2010). RNAi and small interfering RNAs in human disease therapeutic applications. Trends in Biotechnology *28*, 570-579.

Lee, H.S., Lee, K., Jang, H.J., Lee, G.K., Park, J.L., Kim, S.Y., Kim, S.B., Johnson, B.H., Zo, J.I., Lee, J.S., *et al.* (2014). Epigenetic silencing of the non-coding RNA nc886 provokes oncogenes during human esophageal tumorigenesis. Oncotarget *5*, 3472-3481.
Leighton, P.A., Mitchell, K.J., Goodrich, L.V., Lu, X., Pinson, K., Scherz, P., Skarnes, W.C., and Tessier-Lavigne, M. (2001). Defining brain wiring patterns and mechanisms through gene trapping in mice. Nature *410*, 174-179.
Lettre, G., Kritikou, E.A., Jaeggi, M., Calixto, A., Fraser, A.G., Kamath, R.S., Ahringer, J., and Hengartner, M.O. (2004). Genome-wide RNAi identifies p53-dependent and -independent regulators of germ cell apoptosis in C. elegans. Cell Death Differ *11*, 1198-1203.
Leung, R.K.M., and Whittaker, P.A. (2005). RNA interference: From gene silencing to gene-specific therapeutics. Pharmacology & Therapeutics *107*, 222-239.
Li, Q.B., Abraham L. Ng, Aylwin Hu, Zongyi Liang, T. Jake Elledge, Stephen J. (2009). A genome-wide genetic screen for host factors required for hepatitis C virus propagation. Proceedings of the National Academy of Sciences *106*, 16410-16415.
Li, W., and Szoka, F.C. (2007). Lipid-based Nanoparticles for Nucleic Acid Delivery. Pharmaceutical Research *24*, 438-449.
Li, Z., Mericskay, M., Agbulut, O., Butler-Browne, G., Carlsson, L., Thornell, L.E., Babinet, C., and Paulin, D. (1997). Desmin is essential for the tensile strength and integrity of myofibrils but not for myogenic commitment, differentiation, and fusion of skeletal muscle. J Cell Biol *139*, 129-144.
Linkhart, T.A., Clegg, C.H., and Hauschka, S.D. (1980). Control of mouse myoblast commitment to terminal differentiation by mitogens. J Supramol Struct *14*, 483-498.
Livshits, G., and Lowe, S.W. (2013). Accelerating Cancer Modeling with RNAi and Nongermline Genetically Engineered Mouse Models. Cold Spring Harbor Protocols *2013*, pdb.top069856.
Lloyd, D.J., Wheeler, M.C., and Gekakis, N. (2010). A Point Mutation in Sec61α1 Leads to Diabetes and Hepatosteatosis in Mice. Diabetes *59*, 460-470.
Londhe, P., and Davie, J.K. (2011). Gamma Interferon Modulates Myogenesis through the Major Histocompatibility Complex Class II Transactivator, CIITA. Molecular and cellular biology *31*, 2854-2866.
Lu, L., Sun, K., Chen, X., Zhao, Y., Wang, L., Zhou, L., Sun, H., and Wang, H. (2013). Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. The EMBO journal *32*, 2575-2588.
Lum, L., Yao, S., Mozer, B., Rovescalli, A., Von Kessler, D., Nirenberg, M., and Beachy, P.A. (2003). Identification of Hedgehog Pathway Components by RNAi in Drosophila Cultured Cells. Science *299*, 2039-2045.
Luteijn, M.J., van Bergeijk, P., Kaaij, L.J.T., Almeida, M.V., Roovers, E.F., Berezikov, E., and Ketting, R.F. (2012). Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. The EMBO journal *31*, 3422-3430.

M

Maeda, I., Kohara, Y., Yamamoto, M., and Sugimoto, A. (2001). Large-scale analysis of gene function in Caenorhabditis elegans by high-throughput RNAi. Current biology : CB *11*, 171-176.
Magnol, L., Monestier, O., Vuillier-Devillers, K., Wagner, S., Cocquempot, O., Chevallier, M.C., and Blanquet, V. (2011). A sensitised mutagenesis screen in the mouse to explore the bovine genome: study of muscle characteristics. animal *5*, 663-671.
Mak, T.W. (2007). Gene Targeting in Embryonic Stem Cells Scores a Knockout in Stockholm. Cell *131*, 1027-1031.
Manjunath, N., Haoquan, W., Sandesh, S., and Premlata, S. (2009). Lentiviral delivery of short hairpin RNAs. Advanced drug delivery reviews *61*, 732-745.

Mansour, S.L., Thomas, K.R., and Capecchi, M.R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature *336*, 348-352.
Marschall, S., Boersma, A., and de Angelis, M.H. (2009). Sperm Cryopreservation and In Vitro Fertilization. In Gene Knockout Protocols: Second Edition, W. Wurst, and R. Kühn, eds. (Totowa, NJ: Humana Press), pp. 407-420.
Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634-7638.
McCaffrey, A.P., Meuse, L., Pham, T.-T.T., Conklin, D.S., Hannon, G.J., and Kay, M.A. (2002). Gene expression: RNA interference in adult mice. Nature *418*, 38-39.
McManus, M.T., Petersen, C.P., Haines, B.B., Chen, J., and Sharp, P.A. (2002). Gene silencing using micro-RNA designed hairpins. RNA *8*, 842-850.
Meissner, B., Warner, A., Wong, K., Dube, N., Lorch, A., McKay, S.J., Khattra, J., Rogalski, T., Somasiri, A., Chaudhry, I.*, et al.* (2009). An Integrated Strategy to Study Muscle Development and Myofilament Structure in Caenorhabditis elegans. PLoS genetics *5*, e1000537.
Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. Nature reviews Genetics *14*, 447-459.
Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Research *41*, D377-D386.
Michael, S.K., Brennan, J., and Robertson, E.J. (1999). Efficient gene-specific expression of Cre recombinase in the mouse embryo by targeted insertion of a novel IRES-Cre cassette into endogenous loci. Mechanisms of Development *85*, 35-47.
Michailidou, Z., Carter, R.N., Marshall, E., Sutherland, H.G., Brownstein, D.G., Owen, E., Cockett, K., Kelly, V., Ramage, L., Al-Dujaili, E.A.S.*, et al.* (2008). Glucocorticoid receptor haploinsufficiency causes hypertension and attenuates hypothalamic-pituitary-adrenal axis and blood pressure adaptions to high-fat diet. The FASEB Journal *22*, 3896-3907.
Miller, G., Neilan, M., Chia, R., Gheryani, N., Holt, N., Charbit, A., Wells, S., Tucci, V., Lalanne, Z., Denny, P.*, et al.* (2010). ENU Mutagenesis Reveals a Novel Phenotype of Reduced Limb Strength in Mice Lacking Fibrillin 2. PloS one *5*, e9137.
Mittal, V. (2004). Improving the efficiency of RNA interference in mammals. Nature reviews Genetics *5*, 355-365.
Moffat, J., and Sabatini, D.M. (2006). Building mammalian signalling pathways with RNAi screens. Nature reviews Molecular cell biology *7*, 177-187.
Mohr, S., Bakal, C., and Perrimon, N. (2010). Genomic Screening with RNAi: Results and Challenges. Annual Review of Biochemistry *79*, 37-64.
Molkentin, J.D., Black, B.L., Martin, J.F., and Olson, E.N. (1995). Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. Cell *83*, 1125-1136.
Moore, K.J., Rayner, K.J., Suárez, Y., and Fernández-Hernando, C. (2010). microRNAs and cholesterol metabolism. Trends in endocrinology and metabolism: TEM *21*, 699-706.
Mummery-Widmer, J.L., Yamazaki, M., Stoeger, T., Novatchkova, M., Bhalerao, S., Chen, D., Dietzl, G., Dickson, B.J., and Knoblich, J.A. (2009). Genome-wide analysis of Notch signalling in Drosophila by transgenic RNAi. Nature *458*, 987-992.
Munroe, R.J., Bergstrom, R.A., Zheng, Q.Y., Libby, B., Smith, R., John, S.W.M., Schimenti, K.J., Browning, V.L., and Schimenti, J.C. (2000). Mouse mutants from chemically mutagenized embryonic stem cells. Nature genetics *24*, 318-321.

N

Nagy, A., and Mar, L. (2001). Creation and Use of a Cre Recombinase Transgenic Database. In Gene Knockout Protocols, M.J. Tymms, and I. Kola, eds. (Totowa, NJ: Humana Press), pp. 95-106.
Nagy, A., Mar, L., and Watts, G. (2009). Creation and Use of a Cre Recombinase Transgenic Database. In Gene Knockout Protocols: Second Edition, W. Wurst, and R. Kühn, eds. (Totowa, NJ: Humana Press), pp. 365-378.
Nakagata, N., and Yamamura, K.-i. (2009). Current Activities of CARD as an International Core Center for Mouse Resources. Experimental Animals *58*, 343-350.
Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. The Plant cell *2*, 279-289.
Naqvi, A.R., Islam, M.N., Choudhury, N.R., and Haq, Q.M. (2009). The fascinating world of RNA interference. International journal of biological sciences *5*, 97-117.
Neely, G.G., Hess, A., Costigan, M., Keene, A.C., Goulas, S., Langeslag, M., Griffin, R.S., Belfer, I., Dai, F., Smith, S.B.*, et al.* (2010). A genome-wide Drosophila screen for heat nociception identifies alpha2delta3 as an evolutionarily conserved pain gene. Cell *143*, 628-638.
Neumuller, R.A., Richter, C., Fischer, A., Novatchkova, M., Neumuller, K.G., and Knoblich, J.A. (2011). Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. Cell stem cell *8*, 580-593.
Ngô, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proceedings of the National Academy of Sciences *95*, 14687-14692.
Nguyen, N., Judd, L.M., Kalantzis, A., Whittle, B., Giraud, A.S., and van Driel, I.R. (2011). Random mutagenesis of the mouse genome: a strategy for discovering gene function and the molecular basis of disease. American Journal of Physiology - Gastrointestinal and Liver Physiology *300*, G1-G11.
Ni, J.-Q., Zhou, R., Czech, B., Liu, L.-P., Holderbaum, L., Yang-Zhou, D., Shim, H.-S., Tao, R., Handler, D., Karpowicz, P.*, et al.* (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nat Meth *8*, 405-407.
Niwa, H., Araki, K., Kimura, S., Taniguchi, S., Wakasugi, S., and Yamamura, K. (1993). An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. Journal of biochemistry *113*, 343-349.
Nobukuni, Y., Kohno, K., and Miyagawa, K. (2005). Gene Trap Mutagenesis-based Forward Genetic Approach Reveals That the Tumor Suppressor OVCA1 Is a Component of the Biosynthetic Pathway of Diphthamide on Elongation Factor 2. Journal of Biological Chemistry *280*, 10572-10577.
Nolan, P.M., Hugill, A., and Cox, R.D. (2002). ENU mutagenesis in the mouse: application to human genetic disease. Briefings in functional genomics & proteomics *1*, 278-289.
Nolan, P.M., Peters, J., Strivens, M., Rogers, D., Hagan, J., Spurr, N., Gray, I.C., Vizor, L., Brooker, D., Whitehill, E.*, et al.* (2000). A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nature genetics *25*, 440-443.
Nord, A.S., Chang, P.J., Conklin, B.R., Cox, A.V., Harper, C.A., Hicks, G.G., Huang, C.C., Johns, S.J., Kawamoto, M., Liu, S.*, et al.* (2006). The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse. Nucleic Acids Research *34*, D642-D648.

O

O'Kane, C.J., and Gehring, W.J. (1987). Detection in situ of genomic regulatory elements in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *84*, 9123-9127.
Oelgeschlager, M., Larrain, J., Geissert, D., and De Robertis, E.M. (2000). The evolutionarily conserved BMP-binding protein Twisted gastrulation promotes BMP signalling. Nature *405*, 757-763.

Oliver, P.L., and Davies, K.E. (2012). New insights into behaviour using mouse ENU mutagenesis. Human Molecular Genetics *21*, R72-R81.
Ostermeier, G.C., Wiles, M.V., Farley, J.S., and Taft, R.A. (2008). Conserving, Distributing and Managing Genetically Modified Mouse Lines by Sperm Cryopreservation. PloS one *3*, e2792.

P

Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K.*, et al.* (2004). A resource for large-scale RNA-interference-based screens in mammals. Nature *428*, 427-431.
Paroo, Q.L.a.Z. (2010). Biochemical Principles of Small RNA Pathways. Annual Review of Biochemistry *79*, 295-319.
Parry, D.H., Xu, J., and Ruvkun, G. (2007). A whole-genome RNAi screen for C. elegans miRNA pathway genes. Current biology : CB *17*, 2013-2022.
Peer, D., and Lieberman, J. (2011). Special delivery: targeted therapy with small RNAs. Gene Ther *18*, 1127-1133.
Perkins, A.D., Lee, M.J.J., and Tanentzapf, G. (2014). The systematic identification of cytoskeletal genes required for Drosophila melanogaster muscle maintenance. Scientific Data *1*, 140002.
Pinset, C., Montarras, D., Chenevert, J., Minty, A., Barton, P., Laurent, C., and Gros, F. (1988). Control of myogenesis in the mouse myogenic C2 cell line by medium composition and by insulin: characterization of permissive and inducible C2 myoblasts. Differentiation; research in biological diversity *38*, 28-34.
Pospisilik, J.A., Schramek, D., Schnidar, H., Cronin, S.J.F., Nehme, N.T., Zhang, X., Knauf, C., Cani, P.D., Aumayr, K., Todoric, J.*, et al.* (2010). Drosophila Genome-wide Obesity Screen Reveals Hedgehog as a Determinant of Brown versus White Adipose Cell Fate. Cell *140*, 148-160.
Potter, H., and Heller, R. (2003). Transfection by Electroporation. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *CHAPTER*, Unit-9.3.
Punch, V.G., Jones, A.E., and Rudnicki, M.A. (2009). Transcriptional networks that regulate muscle stem cell function. Wiley Interdisciplinary Reviews: Systems Biology and Medicine *1*, 128-140.

Q

Quintero-Rivera, F., Xi, Q.J., Keppler-Noreuil, K.M., Lee, J.H., Higgins, A.W., Anchan, R.M., Roberts, A.E., Seong, I.S., Fan, X., Lage, K.*, et al.* (2015). MATR3 disruption in human and mouse associated with bicuspid aortic valve, aortic coarctation and patent ductus arteriosus. Human Molecular Genetics *24*, 2375-2389.
Quwailid, M., Hugill, A., Dear, N., Vizor, L., Wells, S., Horner, E., Fuller, S., Weedon, J., McMath, H., Woodman, P.*, et al.* (2004). A gene-driven ENU-based approach to generating an allelic series in any gene. Academic Journal *Vol. 15*, p585-591.

R

Raess, M., de Castro, A.A., Gailus-Durner, V., Fessele, S., Hrabě de Angelis, M., and the, I.C. (2016). INFRAFRONTIER: a European resource for studying the functional basis of human disease. Mammalian Genome *27*, 445-450.
Rajan, S., Dang, H.C.P., Djambazian, H., Zuzan, H., Fedyshyn, Y., Ketela, T., Moffat, J., Hudson, T.J., and Sladek, R. (2012). Analysis of early C2C12 myogenesis identifies stably and differentially expressed transcriptional regulators whose knock-down inhibits myoblast differentiation. Physiological Genomics *44*, 183-197.

Ramadan, N., Flockhart, I., Booker, M., Perrimon, N., and Mathey-Prevot, B. (2007). Design and implementation of high-throughput RNAi screens in cultured Drosophila cells. Nat Protocols *2*, 2245-2264.
Reboul, J., Vaglio, P., Rual, J.-F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R.s.*, et al.* (2003). C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nature genetics *34*, 35-41.
Reynaud, E.G., Leibovitch, M.P., Tintignac, L.A., Pelpel, K., Guillier, M., and Leibovitch, S.A. (2000). Stabilization of MyoD by direct binding to p57(Kip2). The Journal of biological chemistry *275*, 18767-18776.
Rhodes, S.J., and Konieczny, S.F. (1989). Identification of MRF4: a new member of the muscle regulatory factor gene family. Genes & development *3*, 2050-2061.
Rinchik, E.M. (1991). Chemical mutagenesis and fine-structure functional analysis of the mouse genome. Trends in Genetics *7*, 15-21.
Robertson, E., Bradley, A., Kuehn, M., and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature *323*, 445-448.
Robles-Oteiza, C., Taylor, S., Yates, T., Cicchini, M., Lauderback, B., Cashman, C.R., Burds, A.A., Winslow, M.M., Jacks, T., and Feldser, D.M. (2015). Recombinase-based conditional and reversible gene regulation via XTR alleles. Nature Communications *6*, 8783.
ROCHA-MARTINS, M., CAVALHEIRO, G.R., MATOS-RODRIGUES, G.E., and MARTINS, R.A.P. (2015). From Gene Targeting to Genome Editing: Transgenic animals applications and beyond. Anais da Academia Brasileira de Ciências *87*, 1323-1348.
Rogers, D.C., Fisher, E.M.C., Brown, S.D.M., Peters, J., Hunter, A.J., and Martin, J.E. (1997). Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. Mammalian Genome *8*, 711-713.
Roma, G., Cobellis, G., Claudiani, P., Maione, F., Cruz, P., Tripoli, G., Sardiello, M., Peluso, I., and Stupka, E. (2007). A novel view of the transcriptome revealed from gene trapping in mouse embryonic stem cells. Genome Research *17*, 1051-1060.
Roma, G., Sardiello, M., Cobellis, G., Cruz, P., Lago, G., Sanges, R., and Stupka, E. (2008). The UniTrap resource: tools for the biologist enabling optimized use of gene trap clones. Nucleic Acids Research *36*, D741-D746.
Root, D.E., Hacohen, N., Hahn, W.C., Lander, E.S., and Sabatini, D.M. (2006). Genome-scale loss-of-function screening with a lentiviral RNAi library. Nat Meth *3*, 715-719.
Rosen, B., Schick, J., and Wurst, W. (2015). Beyond knockouts: the International Knockout Mouse Consortium delivers modular and evolving tools for investigating mammalian genes. Mammalian Genome *26*, 456-466.
Rudalska, R., Dauch, D., Longerich, T., McJunkin, K., Wuestefeld, T., Kang, T.-W., Hohmeyer, A., Pesic, M., Leibold, J., von Thun, A.*, et al.* (2014). In vivo RNAi screening identifies a mechanism of sorafenib resistance in liver cancer. Nature medicine *20*, 1138-1146.
Ruiz, F., Vayssié, L., Klotz, C., Sperling, L., and Madeddu, L. (1998). Homology-dependent Gene Silencing in Paramecium. Molecular biology of the cell *9*, 931-943.
Russell, W.L., Kelly, E.M., Hunsicker, P.R., Bangham, J.W., Maddux, S.C., and Phipps, E.L. (1979). Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proceedings of the National Academy of Sciences of the United States of America *76*, 5818-5819.

S

Saga, Y. (1998). Genetic rescue of segmentation defect in MesP2-deficient mice by MesP1 gene replacement. Mechanisms of Development *75*, 53-66.
Sato, H., Suzuki, T., Ikeda, K., Masuya, H., Sezutsu, H., Kaneda, H., Kobayashi, K., Miura, I., Kurihara, Y., Yokokura, S.*, et al.* (2010). A monogenic dominant mutation in Rom1 generated by N-ethyl-N-nitrosourea mutagenesis causes retinal degeneration in mice. Molecular Vision *16*, 378-391.

Schejter, E.D. (2016). Myoblast fusion: Experimental systems and cellular mechanisms. Seminars in cell & developmental biology *60*, 112-120.
Schneider-Maunoury, S., Topilko, P., Seitanidou, T., Levi, G., Cohen-Tannoudji, M., Pournin, S., Babinet, C., and Charnay, P. (1993). Disruption of Krox-20 results in alteration of rhombomeres 3 and 5 in the developing hindbrain. Cell *75*, 1199-1214.
Schnieke, A., Harbers, K., and Jaenisch, R. (1983). Embryonic lethal mutation in mice induced by retrovirus insertion into the [alpha]1(I) collagen gene. Nature *304*, 315-320.
Schnorrer, F., Schonbauer, C., Langer, C.C., Dietzl, G., Novatchkova, M., Schernhuber, K., Fellner, M., Azaryan, A., Radolf, M., Stark, A.*, et al.* (2010). Systematic genetic analysis of muscle morphogenesis and function in Drosophila. Nature *464*, 287-291.
Schnütgen, F., De-Zolt, S., Van Sloun, P., Hollatz, M., Floss, T., Hansen, J., Altschmied, J., Seisenberger, C., Ghyselinck, N.B., Ruiz, P.*, et al.* (2005). Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. Proceedings of the National Academy of Sciences of the United States of America *102*, 7221-7226.
Segalés, J., Perdiguero, E., and Muñoz-Cánoves, P. (2016). Regulation of Muscle Stem Cell Functions: A Focus on the p38 MAPK Signaling Pathway. Frontiers in Cell and Developmental Biology *4*, 91.
Seyhan, A.A. (2011). RNAi: a potential new class of therapeutic for human genetic disease. Human Genetics *130*, 583-605.
Sharma, S., and Rao, A. (2009). RNAi screening: tips and techniques. Nat Immunol *10*, 799-804.
Shenoy, A., and Blelloch, R.H. (2014). Regulation of microRNA function in somatic stem cell proliferation and differentiation. Nature reviews Molecular cell biology *15*, 565-576.
Shigeoka, T., Kawaichi, M., and Ishida, Y. (2005a). Suppression of nonsense-mediated mRNA decay permits unbiased gene trapping in mouse embryonic stem cells. Nucleic Acids Research *33*, e20-e20.
Shigeoka, T., Kawaichi, M., and Ishida, Y. (2005b). Suppression of nonsense-mediated mRNA decay permits unbiased gene trapping in mouse embryonic stem cells. Nucleic Acids Res *33*, e20.
Silva, J.M., Li, M.Z., Chang, K., Ge, W., Golding, M.C., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R.*, et al.* (2005). Second-generation shRNA libraries covering the mouse and human genomes. Nature genetics *37*, 1281-1288.
Silver, J., and Keerikatte, V. (1989). Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. Journal of Virology *63*, 1924-1928.
Simmer, F., Moorman, C., van der Linden, A.M., Kuijk, E., van den Berghe, P.V.E., Kamath, R.S., Fraser, A.G., Ahringer, J., and Plasterk, R.H.A. (2003). Genome-Wide RNAi of C. elegans Using the Hypersensitive rrf-3 Strain Reveals Novel Gene Functions. PLoS biology *1*, e12.
Simon, M.M., Mallon, A.-M., Howell, G.R., and Reinholdt, L.G. (2012). High throughput sequencing approaches to mutation discovery in the mouse. Mammalian genome : official journal of the International Mammalian Genome Society *23*, 499-513.
Singla, V., Hunkapiller, J., Santos, N., Seol, A., Norman, A.R., Wakenite, P., Skarnes, W.C., and Reiter, J.F. (2010). Floxin, a resource for genetically engineering mouse embryonic stem cells. Nature methods *7*, 50-52.
Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nature reviews Molecular cell biology *12*, 246-258.
Skarnes, W.C., Auerbach, B.A., and Joyner, A.L. (1992). A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes & development *6*, 903-918.
Skarnes, W.C., Moss, J.E., Hurtley, S.M., and Beddington, R.S. (1995). Capturing genes encoding membrane and secreted proteins important for mouse development. Proceedings of the National Academy of Sciences of the United States of America *92*, 6592-6596.
Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T.*, et al.* (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature *474*, 337-342.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal [beta]-globin locus by homologous recombination. Nature *317*, 230-234.
Sonnichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Bettencourt, P., Cassin, E.*, et al.* (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature *434*, 462-469.
Spence, S.E., Gilbert, D.J., Swing, D.A., Copeland, N.G., and Jenkins, N.A. (1989). Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic Srev lines of mice. Molecular and cellular biology *9*, 177-184.
Stanford, W.L., Cohn, J.B., and Cordes, S.P. (2001). Gene-trap mutagenesis: past, present and beyond. Nature reviews Genetics *2*, 756-768.
Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila MicroRNA targets. PLoS Biol *1*, E60.
Steel, M., Moss, J., Clark, K.A., Kearns, I.R., Davies, C.H., Morris, R.G., Skarnes, W.C., and Lathe, R. (1998). Gene-trapping to identify and analyze genes expressed in the mouse hippocampus. Hippocampus *8*, 444-457.
Stein, C.A. (2001). Antisense that comes naturally. Nat Biotech *19*, 737-738.
Steinbrink, S., and Boutros, M. (2008). RNAi Screening in Cultured Drosophila Cells. In Drosophila: Methods and Protocols, C. Dahmann, ed. (Totowa, NJ: Humana Press), pp. 139-153.
Stryke, D., Kawamoto, M., Huang, C.C., Johns, S.J., King, L.A., Harper, C.A., Meng, E.C., Lee, R.E., Yee, A., L'Italien, L.*, et al.* (2003). BayGenomics: a resource of insertional mutations in mouse embryonic stem cells. Nucleic Acids Res *31*, 278-281.
Suda, Y., Nakabayashi, J., Matsuo, I., and Aizawa, S. (1999). Functional equivalency between Otx2 and Otx1 in development of the rostral head. Development *126*, 743-757.
Sul, J.-W., Park, M.-Y., Shin, J., Kim, Y.-R., Yoo, S.-E., Kong, Y.-Y., Kwon, K.-S., Lee, Y.H., and Kim, E. (2013). Accumulation of the parkin substrate, FAF1, plays a key role in the dopaminergic neurodegeneration. Human Molecular Genetics *22*, 1558-1573.
Sun, K., Zhou, L., Zhao, Y., Wang, H., and Sun, H. (2016a). Genome-wide RNA-seq and ChIP-seq reveal Linc-YY1 function in regulating YY1/PRC2 activity during skeletal myogenesis. Genomics Data *7*, 247-249.
Sun, X., Brieño-Enríquez, M.A., Cornelius, A., Modzelewski, A.J., Maley, T.T., Campbell-Peterson, K.M., Holloway, J.K., and Cohen, P.E. (2016b). FancJ (Brip1) loss-of-function allele results in spermatogonial cell depletion during embryogenesis and altered processing of crossover sites during meiotic prophase I in mice. Chromosoma *125*, 237-252.
Sundberg, J.P., Sundberg, B.A., and Schofield, P. (2008). Integrating Mouse Anatomy and Pathology Ontologies into a Phenotyping Database: Tools for Data Capture and Training. Mammalian genome : official journal of the International Mammalian Genome Society *19*, 413-419.
Svoboda, P., Stein, P., Hayashi, H., and Schultz, R.M. (2000). Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. Development *127*, 4147-4156.

T

Tabara, H., Grishok, A., and Mello, C.C. (1998). RNAi in C. elegans: soaking in the genome sequence. Science *282*, 430-431.
Tajbakhsh, S., Bober, E., Babinet, C., Pournin, S., Arnold, H. and Buckingham, M. (1996). Gene targeting the myf-5 locus with nlacZ reveals expression of this myogenic factor in mature skeletal muscle fibres as well as early embryonic muscle. . Developmental Dynamics *206*, 291-300.
Tapscott, S.J., and Weintraub, H. (1991). MyoD and the regulation of myogenesis by helix-loop-helix proteins. Journal of Clinical Investigation *87*, 1133-1138.
Team, R.D.C. (2011). R: A Language and Environment for Statistical Computing. (Vienna, Austria: R Foundation for Statistical Computing).

Theis, M., and Buchholz, F. (2011). High-throughput RNAi screening in mammalian cells with esiRNAs. Methods *53*, 424-429.
Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.
Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854.
Tintignac, L.A., Leibovitch, M.P., Kitzmann, M., Fernandez, A., Ducommun, B., Meijer, L., and Leibovitch, S.A. (2000). Cyclin E-cdk2 phosphorylation promotes late G1-phase degradation of MyoD in muscle cells. Experimental cell research *259*, 300-307.
Tolia, N.H., and Joshua-Tor, L. (2007). Slicer and the Argonautes. Nat Chem Biol *3*, 36-43.
Tomczak, K.K., Marinescu, V.D., Ramoni, M.F., Sanoudou, D., Montanaro, F., Han, M., Kunkel, L.M., Kohane, I.S., and Beggs, A.H. (2004). Expression profiling and identification of novel genes involved in myogenic differentiation. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *18*, 403-405.
Toye, A.A., Moir, L., Hugill, A., Bentley, L., Quarterman, J., Mijat, V., Hough, T., Goldsworthy, M., Haynes, A., Hunter, A.J.*, et al.* (2004). A New Mouse Model of Type 2 Diabetes, Produced by N-Ethyl-Nitrosourea Mutagenesis, Is the Result of a Missense Mutation in the Glucokinase Gene. Diabetes *53*, 1577-1583.
Trimborn, M., Ghani, M., Walther, D.J., Dopatka, M., Dutrannoy, V., Busche, A., Meyer, F., Nowak, S., Nowak, J., Zabel, C.*, et al.* (2010). Establishment of a Mouse Model with Misregulated Chromosome Condensation due to Defective Mcph1 Function. PloS one *5*, e9242.

V

Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W., Economides, A.N., Auerbach, W., Poueymirou, W.T., Adams, N.C., Rojas, J., Yasenchak, J.*, et al.* (2003). High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nat Biotech *21*, 652-659.
van Bürck, L., Blutke, A., Kautz, S., Rathkolb, B., Klaften, M., Wagner, S., Kemter, E., Hrabé de Angelis, M., Wolf, E., Aigner, B.*, et al.* (2010). Phenotypic and pathomorphological characteristics of a novel mutant mouse model for maturity-onset diabetes of the young type 2 (MODY 2). American Journal of Physiology - Endocrinology And Metabolism *298*, E512-E523.
Vaucheret, H., and Fagard, M. (2001). Transcriptional gene silencing in plants: targets, inducers and regulators. Trends in genetics : TIG *17*, 29-35.
Venken, K.J.T., and Bellen, H.J. (2014). Chemical Mutagens, Transposons, and Transgenes to Interrogate Gene Function in Drosophila melanogaster. Methods (San Diego, Calif) *68*, 15-28.
Vidal, F., Lopez, P., Lopez-Fernandez, L.A., Ranc, F., Scimeca, J.C., Cuzin, F., and Rassoulzadegan, M. (2001). Gene trap analysis of germ cell signaling to Sertoli cells: NGF-TrkA mediated induction of Fra1 and Fos by post-meiotic germ cells. Journal of cell science *114*, 435-443.
von Melchner, H., DeGregori, J.V., Rayburn, H., Reddy, S., Friedel, C., and Ruley, H.E. (1992). Selective disruption of genes expressed in totipotent embryonal stem cells. Genes & development *6*, 919-927.
Vorobyeva, M., Timoshenko, V., Vorobjev, P., and Venyaminova, A. (2015). Aptamers Against Immunologic Targets: Diagnostic and Therapeutic Prospects. Nucleic Acid Therapeutics *26*, 52-65.

W

Wada, H., Yasuda, T., Miura, I., Watabe, K., Sawa, C., Kamijuku, H., Kojo, S., Taniguchi, M., Nishino, I., Wakana, S.*, et al.* (2009). Establishment of an Improved Mouse Model for Infantile Neuroaxonal Dystrophy That Shows Early Disease Onset and Bears a Point Mutation in Pla2g6. The American Journal of Pathology *175*, 2257-2263.

Wagner, E.F., Covarrubias, L., Stewart, T.A., and Mintz, B. (1983). Prenatal lethalities in mice homozygous for human growth hormone gene sequences integrated in the germ line. Cell *35*, 647-655.
Wang, Y., Waters, J., Leung, M.L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H.*, et al.* (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature *512*, 155-160.
White, Jacqueline K., Gerdin, A.-K., Karp, Natasha A., Ryder, E., Buljan, M., Bussell, James N., Salisbury, J., Clare, S., Ingham, Neil J., Podrini, C.*, et al.* (2013). Genome-wide Generation and Systematic Phenotyping of Knockout Mice Reveals New Roles for Many Genes. Cell *154*, 452-464.
Whitehurst, A.W., Bodemann, B.O., Cardenas, J., Ferguson, D., Girard, L., Peyton, M., Minna, J.D., Michnoff, C., Hao, W., Roth, M.G.*, et al.* (2007). Synthetic lethal screen identification of chemosensitizer loci in cancer cells. Nature *446*, 815-819.
Wicik, Z., Sadkowski, T., Jank, M., and Motyl, T. (2011). The transcriptomic signature of myostatin inhibitory influence on the differentiation of mouse C2C12 myoblasts. In Polish Journal of Veterinary Sciences, pp. 643.
Wilkinson, P., Sengerova, J., Matteoni, R., Chen, C.-K., Soulat, G., Ureta-Vidal, A., Fessele, S., Hagn, M., Massimi, M., Pickford, K.*, et al.* (2010). EMMA—mouse mutant resources for the international scientific community. Nucleic Acids Research *38*, D570-D576.
Wilson, R.C., Tambe, A., Kidwell, M.A., Noland, C.L., Schneider, C.P., and Doudna, J.A. (2015). Dicer–TRBP complex formation ensures accurate mammalian microRNA biogenesis. Molecular cell *57*, 397-407.
Wiltshire, T., Pletcher, M.T., Batalov, S., Barnes, S.W., Tarantino, L.M., Cooke, M.P., Wu, H., Smylie, K., Santrosyan, A., Copeland, N.G.*, et al.* (2003). Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. Proceedings of the National Academy of Sciences of the United States of America *100*, 3380-3385.
Wright, L.M., Yong, S., Picken, M.M., Rockey, D., and Fimmel, C.J. (2009). Decreased Survival and Hepato-Renal Pathology in Mice with C-Terminally Truncated GP73 (GOLPH2). International Journal of Clinical and Experimental Pathology *2*, 34-47.
Wright, W.E., Sassoon, D.A., and Lin, V.K. (1989). Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. Cell *56*, 607-617.
Wuestefeld, T., Pesic, M., Rudalska, R., Dauch, D., Longerich, T., Kang, T.W., Yevsa, T., Heinzmann, F., Hoenicke, L., Hohmeyer, A.*, et al.* (2013). A Direct in vivo RNAi screen identifies MKK4 as a key regulator of liver regeneration. Cell *153*, 389-401.
Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A., Hill, D.P., Guillemot, F., Gasca, S., Cado, D., Auerbach, A.*, et al.* (1995). A Large-Scale Gene-Trap Screen for Insertional Mutations in Developmentally Regulated Genes in Mice. Genetics *139*, 889-899.

X

Xia, H., Mao, Q., Paulson, H.L., and Davidson, B.L. (2002). siRNA-mediated gene silencing in vitro and in vivo. Nat Biotech *20*, 1006-1010.
Xin, H.B., Deng, K.Y., Shui, B., Qu, S., Sun, Q., Lee, J., Greene, K.S., Wilson, J., Yu, Y., Feldman, M.*, et al.* (2005). Gene trap and gene inversion methods for conditional gene inactivation in the mouse. Nucleic Acids Res *33*, e14.

Y

Yaffe, D. (1968). Retention of differentiation potentialities during prolonged cultivation of myogenic cells. Proceedings of the National Academy of Sciences of the United States of America *61*, 477-483.
Yaffe, D. (1969). Cellular aspects of muscle differentiation in vitro. Current topics in developmental biology *4*, 37-77.

Yaffe, D., and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature *270*, 725-727.
Yamamoto-Hino, M., and Goto, S. (2013). In Vivo RNAi-Based Screens: Studies in Model Organisms. Genes *4*, 646-665.
Yang, D., Buchholz, F., Huang, Z., Goga, A., Chen, C.-Y., Brodsky, F.M., and Bishop, J.M. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *99*, 9942-9947.
Yang, W., Musci, T.S., and Mansour, S.L. (1997). Trapping genes expressed in the developing mouse inner ear. Hearing research *114*, 53-61.
Ylikallio, E., Pöyhönen, R., Zimon, M., De Vriendt, E., Hilander, T., Paetau, A., Jordanova, A., Lönnqvist, T., and Tyynismaa, H. (2013). Deficiency of the E3 ubiquitin ligase TRIM2 in early-onset axonal neuropathy. Human Molecular Genetics *22*, 2975-2983.
Yoshida, Y., Makita, Y., Heida, N., Asano, S., Matsushima, A., Ishii, M., Mochizuki, Y., Masuya, H., Wakana, S., Kobayashi, N.*, et al.* (2009). PosMed (Positional Medline): prioritizing genes with an artificial neural network comprising medical documents to accelerate positional cloning. Nucleic Acids Research *37*, W147-W152.
Yoshiki, A., Ike, F., Mekada, K., Kitaura, Y., Nakata, H., Hiraiwa, N., Mochida, K., Ijuin, M., Kadota, M., Murakami, A.*, et al.* (2009). The Mouse Resources at the RIKEN BioResource Center. Experimental Animals *58*, 85-96.
Yu, J.-Y., DeRuiter, S.L., and Turner, D.L. (2002). RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *99*, 6047-6052.

Z

Zambrowicz, B.P., Abuin, A., Ramirez-Solis, R., Richter, L.J., Piggott, J., BeltrandelRio, H., Buxton, E.C., Edwards, J., Finch, R.A., Friddle, C.J.*, et al.* (2003). Wnk1 kinase deficiency lowers blood pressure in mice: A gene-trap screen to identify potential targets for therapeutic intervention. Proceedings of the National Academy of Sciences *100*, 14109-14114.
Zambrowicz, B.P., and Friedrich, G.A. (1998). Comprehensive mammalian genetics: history and future prospects of gene trapping in the mouse. The International journal of developmental biology *42*, 1025-1036.
Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.
Zender, L., Xue, W., Zuber, J., Semighini, C.P., Krasnitz, A., Ma, B., Zender, P., Kubicka, S., Luk, J.M., Schirmacher, P.*, et al.* (2008). An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. Cell *135*, 852-864.
Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M.*, et al.* (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature *478*, 524-528.

RESUME

La myogénèse est un processus multi-étapes hautement régulé impliquant la prolifération et la différenciation de myoblastes. Bien que la myogenèse ait été largement décrite, les mécanismes de régulation qui régissent ce processus complexe sont encore mal connus, notamment les réseaux de gènes et les interactions potentielles entre les voies de signalisation impliquées. Afin d'identifier de nouveaux gènes jouant un rôle dans la différenciation myogénique, j'ai mis en place un nouveau protocole in vitro, basé sur la lignée myoblastique C2C12, l'utilisation de l'ARN interférence et l'analyse quantitative d'images d'une grande quantité de myoblastes différenciés. J'ai pu inactiver une centaine de gènes et par une analyse quantitative de la densité cellulaire, de la quantité de myotubes, de la morphologie du myotube et de l'indice de fusion, j'ai pu montrer que six gènes parmi les 100 sont impliqués à la fois dans la prolifération et la différenciation des cellules C2C12 et 13 gènes jouant un rôle uniquement dans l'étape de différenciation. Nos résultats montrent que notre crible peut être un outil efficace pour détecter aussi bien les phénotypes subtils permettant l'identification de nouveaux régulateurs myogéniques chez les mammifères.

Mots clés : Crible, ARNi, Prolifération, Différenciation myogénique

ABSTRACT

Myogenesis is a highly regulated multi-step process involving myoblast proliferation and differentiation. Although studies over the last decades have identified several factors governing these distinct major phases, many of them are not yet known. In order to identify novel genes, we took advantage of the C2C12 myoblastic line to establish a functional siRNA screen combined with quantitative-imaging analysis of a large amount of differentiated myoblasts. We knocked down 100 mouse-preselected genes without a previously characterized role in muscle. Using image analysis, we tracked gene-silencing phenotypes by quantitative assessment of cellular density, myotube quantity, myotube morphology and fusion index. Our results have revealed six genes involved in both stages of C1C12 myogenesis and 13 genes specific to the differentiation stage. These findings prove that our RNAi-based screen could be an efficient tool to detect clear and subtle phenotypes allowing the identification of new myogenic regulators in Mammals.

Keywords: Screen, RNAi, Proliferation, Myogenic differentiation