Analysis of the transcriptional basis of natural transdifferentiation initiation in Y cell using Caenorhabditis elegans as a model organism

Jaime Osuna Luque

Université de Strasbourg

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTÉ

IGBMC, CNRS UMR7104, INSERM U1258, Université de Strasbourg

Thèse présentée par :

Jaime OSUNA LUQUE

Soutenue le : 13 Mai 2020

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Biologie Cellulaire et Moléculaire / Biologie du Développement

Analyse de la base transcriptionnelle de la transdifférenciation naturelle chez Caenorhabditis elegans

THÈSE dirigée par :

Sophie JARRIAULT	DR2, Université de Strasbourg
Peter MEISTER (Co-Directeur)	Professor. Universität Bern

RAPPORTEURS :

Peter ASKJAER	Professor, établissement
Francesca PALLADINO	DR, École Normal Supérieure Lyon

AUTRES MEMBRES DU JURY :

Benoît ZUBER	Prof. Dr. Universität Bern
Laszlo TORA	DR Université de Strasbourg

Université de Strasbourg

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Université de Strasbourg

Jaime OSUNA LUQUE

Analyse de la base transcriptionnelle de la transdifférenciation naturelle chez Caenorhabditis elegans

Résumé

Le développement est considéré comme un processus hiérarchique au cours duquel la grande majorité des cellules restreignent leur potentiel cellulaire au fil du temps pour adopter une identité spécialisée finale. Dans la plupart des cas, cette identité différenciée est maintenue jusqu'à la mort de la cellule. Cela avait initialement conduit à l'idée que l'identité différenciée ne pouvait pas être inversée. Cependant, des études marquantes de la dernière décennie ont montré que non seulement une identité cellulaire différenciée peut être effacée expérimentalement, mais remarquablement, qu'elle peut être naturellement convertie en un autre type de cellule in vivo, un processus appelé transdifférenciation. Ces résultats ont suscité de nouvelles questions sur le maintien et la reprogrammation de l'identité cellulaire. Par exemple, dans quelle mesure l'identité initiale est-elle progressivement effacée? Quels gènes sont activés et désactivés au cours du processus? Peut-on définir des programmes d'expression distincts successifs lors de la conversion? Comment sont-ils contrôlés? Ces questions, au cœur du domaine, sont au cœur de ce projet de thèse. À plus long terme, ces connaissances seront essentielles pour améliorer notre capacité à manipuler l'identité cellulaire et à concevoir des cellules de remplacement sûres pour la médecine régénérative.

Au cours de ce projet de doctorat, j'ai étudié un événement de transdifférenciation naturelle se produisant naturellement in vivo dans une seule cellule en utilisant le ver nématode Caenorhabditis elegans comme organisme modèle. Cette cellule, appelée Y, se différencie d'une identité rectale en une identité moto-neurone appelée PDA. Ce système a fourni des informations clés sur la transition et les étapes cellulaires impliquées dans la transdifférenciation et l'identification des facteurs nucléaires conservés cruciaux pour le démarrage du processus, ou sur l'importance relative et les rôles des facteurs de transcription par rapport aux facteurs de modification des histones pour la dynamique et la robustesse du conversion. Ces preuves suggèrent que de nombreux gènes sont activés ou désactivés. Cependant, la dynamique transcriptionnelle de la transition reste inconnue.

Cet événement de transdifférenciation ne peut pas être étudié in vitro. J'ai donc besoin d'utiliser des animaux entiers pour comprendre comment ce processus fonctionne, dans une seule cellule. Pendant le développement de mon travail de thèse, j'ai mis en place de nouvelles façons d'utiliser une méthodologie appelée DamID, que j'ai implémentée sur des animaux entiers. L'identification de l'ADN adénine méthyltransférase (DamID) utilise une fusion entre une protéine d'intérêt, dans notre cas une sous-unité d'ARN polymérase et une adénine méthyltransférase bactérienne (Dam). La liaison de l'ARN polymérase aux gènes transcrits conduit à la méthylation de l'ADN de leur locus génomique, qui peut ensuite être identifié à l'aide de techniques moléculaires. Afin de limiter l'expression des dam ::fusion à la cellule transdifférenciante, j'ai utilisé différentes méthodologies, y compris les systèmes de recombinaison in vivo et le tri cellulaire activé par fluorescence. Ce travail de thèse a produit une collection de gènes spécifiques Y qui peuvent aider à mieux comprendre comment les initiations de la transdifférenciation Y-PDA et quels acteurs clés moléculaires régulent le début de cet événement de transdifférenciation naturelle chez Caenorhabditis elegans. Ce travail de doctorat a également généré un nouveau pipeline DamID utilisant les technologies Oxford Nanopore qui sera extrêmement utile pour mener des études de biologie moléculaire en utilisant ce nématode ou d'autres organismes modèles.

Résumé en anglais

Development is viewed as a hierarchical process during which the vast majority of cells restrict their cellular potential over time to adopt a final specialised identity. In most cases this differentiated identity is maintained until the cell's death. This had initially led to the idea that the differentiated identity could not be reversed. However, landmark studies in the last decade have shown that not only can a differentiated cell identity be experimentally erased, but remarkably, that it can be naturally converted into a different cell type in vivo, a process called transdifferentiation. Those findings have sparked new questions around the maintenance and reprogramming of the cellular identity. For example, how progressively is the initial identity erased? What genes are switched on and off during the process? Can we define successive distinct expression programs during the conversion? How are they controlled? These questions, core to the field, are at the heart of this PhD project. In the longer run, such knowledge will be key to improve our ability to manipulate the cellular identity and engineer safe replacement cells for regenerative medicine.

During this PhD project, I studied a natural transdifferentiation event naturally occurring in vivo in a single cell using the nematode worm Caenorhabditis elegans as a model organism. This cell, called Y, transdifferentiates from a rectal identity into a moto-neuron identity called PDA. This system has contributed key insights on the transition and cellular steps involved in transdifferentiation and the identification of conserved nuclear factors crucial to the initiation of the process, or the relative importance and roles of transcription factors versus histone modifying factors for the dynamics and robustness of the conversion. Those evidences suggest that many genes are switched ON or OFF. However, the transcriptional dynamics of the transition remains unknown.

This transdifferentiation event cannot be studied in vitro. I therefore need to use entire animals to understand how this process functions, in a single cell. During the development of my PhD work, I setup new ways to use a methodology called DamID, which I implemented on whole animals. DNA adenine methyltransferase identification (DamID) uses a fusion between a protein of interest, in our case an RNA polymerase subunit and a bacterial adenine methyltransferases (Dam). Binding of the RNA polymerase to transcribed genes leads to DNA methylation of their genomic locus, which can be subsequently identified using molecular techniques. In order to restrict expression of the dam::fusions to the transdifferentiating cell, I used different methodologies including in vivo recombination systems and fluorescence activated cell sorting. This PhD work have produced a collection of Y specific genes that can help to get a better understanding of how Y-to-PDA transdifferentiation initiates and what molecular key players are regulating the beginning of this natural transdifferentiation event in the Caenorhabditis elegans. This PhD work also generated a new DamID pipeline using Oxford Nanopore Technologies that will be extremely useful to carry molecular biology studies using this nematode or other model organisms

Es la duda lo que mantiene joven a la gente. La certeza es como un virus maligno. Te contagia de vejez.

-Arturo Pérez-Reverte, El tango de la Guardia Vieja.

Contents

List of figures

List of tables

List of abbreviations

DNA: Deoxyribose Nucleic Acid.

RNA: Ribonucleic Acid.

C. elegans: *Caenorhabditis elegans.*

b.C: Before Christ.

ESCs: Embryonic Stem cells

iPSCs: Induced Pluripotent Stem Cell.

TF: Transcription factor.

PCR: Polymerase chain reaction.

CRISPR-Cas9: clustered regularly interspaced short palindromic repeats.

DNMT: DNA methyltransferase.

JMJD: Jumonji protein domains.

FACS: Fluorescence activated cell sorting.

ChIP-seq: Chromatin immunoprecipitation sequencing.

DamID: DNA adenine methyltransferase identification.

IlluminaDamID: DNA adenine methyltransferase identification using Illumina technologies for sequencing.

NanoDamID: DNA adenine methyltransferase identification using Oxford Nanopore Technologies for sequencing.

GTFs: General transcription factors.

BRE: B recognition element.

DBD: DNA binding domain.

IAA: Indole-3-acetic acid.

NAA: 1-naphtaleneacetic acid.

NGM: Nematode growth medium.

MosSCI: Mos1 mediated single copy insertion.

GFP: Green fluorescence protein.

GO: Gene Ontology.

ONT: Oxford Nanopore Technologies.

AID: Auxin inducible degradation system.

Abstract

Le développement est considéré comme un processus hiérarchique au cours duquel la grande majorité des cellules restreignent leur potentiel cellulaire au fil du temps pour adopter une identité spécialisée finale. Dans la plupart des cas, cette identité différenciée est maintenue jusqu'à la mort de la cellule. Cela avait initialement conduit à l'idée que l'identité différenciée ne pouvait pas être inversée. Cependant, des études marquantes de la dernière décennie ont montré que non seulement une identité cellulaire différenciée peut être effacée expérimentalement, mais remarquablement, qu'elle peut être naturellement convertie en un type de cellule différent in vivo, un processus appelé transdifférenciation. Ces résultats ont suscité de nouvelles questions autour du maintien et de la reprogrammation de l'identité cellulaire. Par exemple, dans quelle mesure l'identité initiale est-elle progressivement effacée? Quels gènes sont activés et désactivés au cours du processus? Peut-on définir des programmes d'expression distincts successifs lors de la conversion? Comment sont-ils contrôlés? Ces questions, au cœur du domaine, sont au cœur de ce projet de thèse. À plus long terme, ces connaissances seront essentielles pour améliorer notre capacité à manipuler l'identité cellulaire et à concevoir des cellules de remplacement sûres pour la médecine régénérative.

Au cours de ce projet de doctorat, j'ai étudié un événement de transdifférenciation naturelle se produisant naturellement in vivo dans une seule cellule en utilisant le ver nématode Caenorhabditis elegans comme organisme modèle. Cette cellule, appelée Y, se différencie d'une identité rectale en une identité moto-neurone appelée PDA. Ce système a fourni des informations clés sur la transition et les étapes cellulaires impliquées dans la transdifférenciation et l'identification des facteurs nucléaires conservés cruciaux pour le démarrage du processus, ou sur l'importance relative et les rôles des facteurs de transcription par rapport aux facteurs de modification des histones pour la dynamique et la robustesse du conversion. Ces preuves suggèrent que de nombreux gènes sont activés ou désactivés. Cependant, la dynamique transcriptionnelle de la transition reste inconnue.

Cet événement de transdifférenciation ne peut pas être étudié in vitro. J'ai donc besoin d'utiliser des animaux entiers pour comprendre comment ce processus fonctionne, dans une seule cellule. Au cours du développement de ma thèse, j'ai mis en place de nouvelles façons d'utiliser une méthodologie appelée DamID, que j'ai implémentée sur des animaux entiers. L'identification de l'ADN adénine méthyltransférase (DamID) utilise une fusion entre une protéine d'intérêt, dans notre cas une sous-unité d'ARN polymérase et une adénine méthyltransférase bactérienne (Dam). La liaison de l'ARN polymérase aux gènes

transcrits conduit à la méthylation de l'ADN de leur locus génomique, qui peut ensuite être identifié à l'aide de techniques moléculaires. Afin de restreindre l'expression des fusions dam :: à la cellule transdifférenciante, j'ai utilisé différentes méthodologies, y compris les systèmes de recombinaison in vivo et le tri cellulaire activé par fluorescence. Ce travail de thèse a produit une collection de gènes spécifiques Y qui peuvent aider à mieux comprendre comment les initiations de la transdifférenciation Y vers PDA et quels acteurs clés moléculaires régulent le début de cet événement de transdifférenciation naturelle chez Caenorhabditis elegans. Ce travail de doctorat a également généré un nouveau pipeline DamID utilisant les technologies Oxford Nanopore qui sera extrêmement utile pour mener des études de biologie moléculaire à l'aide de ce nématode ou d'autres organismes modèles.

1. Introduction

1.1 Biologie moléculaire et organismes modèles.

Les XX et XXI siècles sont considérés comme le début d'un âge d'or en biologie moléculaire. Depuis la découverte de la double hélice de l'acide désoxyribonucléique (ADN) par Watson, Crick et Franklin (1) le nombre de découvertes liées à la génétique et à la biologie moléculaire (découverte du clonage animal, amplification en chaîne par polymérase «PCR», Human Genome Project, etc.) ont connu une croissance exponentielle. Ces découvertes ont souvent été accompagnées de méthodes techniques innovantes qui ont transformé la façon d'étudier le matériel génétique.

Cependant, pour répondre à un problème biologique, la seule utilisation des méthodes techniques ne peut suffire; c'est pourquoi, parallèlement au développement des avancées techniques, la communauté scientifique a développé une collection d'organismes modèles pour découvrir la base moléculaire des processus biologiques.

Initialement, les organismes modèles les plus utilisés étaient les procaryotes et le plus connu était une bactérie appelée Escherichia coli. Après l'étude des organismes bactériens, la plupart des scientifiques se sont concentrés sur les virus bactériens appelés phages pour comprendre la base moléculaire de la réplication, de la transcription et de la traduction de l'ADN. Outre les procaryotes, la communauté scientifique a concentré son attention sur les eucaryotes en tant que prochain organisme modèle et objet d'étude, bien que les procaryotes soient encore utilisés pour d'importantes découvertes récentes telles que la technique d'édition génomique CRISPR / Cas9 (2).

Depuis la période néolithique (5000 avant J.-C.), *Saccharomyces cerevisiae*, une levure en herbe, était la levure la plus utilisée par l'humanité pour fabriquer du vin, du pain et de la bière (3). Dans la seconde moitié du XXe siècle, cette levure en herbe et la levure de fission Schizosaccharomyces pombe sont devenues les organismes modèles eucaryotes les plus utilisés pour étudier de manière intensive le mécanisme moléculaire des fonctions biologiques uniquement présentes dans les cellules eucaryotes. Quelques exemples de processus biologiques découverts utilisant la levure étaient le cycle CDK-cycline / cellule, la méiose et la mitose, l'épissage de l'acide ribonucléique (ARN), de nombreuses approches pour étudier la structure de la chromatine et la première étude sur la dynamique du cytosquelette, qui était également étudié en utilisant la levure comme organisme modèle (4). Les processus et voies hautement conservés comme objets d'étude au cours de l'évolution eucaryote ont ouvert la porte pour envisager la recherche d'organismes modèles applicables et utiles pour comprendre les processus biologiques qui affectaient directement les humains. Cependant, de nouvelles approches étaient nécessaires pour répondre aux

questions liées à l'interaction des cellules et des tissus au cours du développement et de la sénescence et de la façon dont ils sont physiologiquement régulés. La nécessité d'utiliser des animaux multicellulaires a été demandée. Beaucoup d'entre eux ont émergé, surtout après avoir réalisé, que de nombreux gènes et les mécanismes moléculaires dans lesquels ils étaient impliqués sont conservés à travers l'évolution (5). Les métazoaires ont fourni des avantages pour la recherche scientifique afin de comprendre la physiologie et le développement humains. Les raisons de leur utilité sont multiples. L'utilisation d'organismes modèles tels que les mouches des fruits, les vers ou les poissons zèbres ne nécessite pas d'énormes investissements d'argent pour les entretenir et convient à une manipulation expérimentale étant des animaux génétiquement traitables. En outre, ils n'ont pas les fortes considérations éthiques et les limites présentes chez les mammifères, telles que la recherche sur la souris, le chien ou l'homme. De plus, il y a un degré de similitude remarquable à travers l'évolution chez ces animaux. Le génome de Drosophila melanogaster, la mouche des fruits, est homologue à 60% au génome humain (6). D'autres génomes d'organismes modèles, comme le génome du poisson zèbre Danio rerio ou le ver rond Caenorhabditis elegans sont respectivement de 70% (7) et 38% (8). Les connaissances acquises à partir d'organismes modèles pourraient être fréquemment appliquées aux humains pour comprendre le mécanisme de développement de l'embryogenèse ou des pathologies. Les organismes modèles métazoaires les plus connus et les plus utilisés sont trois comme précédemment mentionnés: le ver rond Caenorhabditis elegans, la mouche Drosophila melanogaster, le poisson zèbre Danio rerio et la souris Mus musculus. Tous ont en commun d'être des organismes génétiquement traitables et ont été largement utilisés pour étudier le développement et les bases moléculaires de la vie.

Dans ce travail de doctorat, j'ai utilisé le nématode Caenorhabditis elegans comme organisme modèle pour découvrir la base moléculaire d'un événement de transdifférenciation qui se produit dans une cellule de l'animal à un stade précoce de son développement. Mais d'abord, avant d'aborder ce processus passionnant, je vais expliquer en détail les particularités et la régulation de ce type de processus cellulaires qui se produisent dans la nature.

1.2 Differentiation cellulaire et régénération.

La différenciation cellulaire était classiquement définie comme la capacité d'une cellule souche à changer son état cellulaire totipotent d'origine et à se différencier en un type cellulaire spécifique, un processus essentiel pour la croissance, la reproduction, le développement et la longévité de tous les organismes multicellulaires (9).

Dans les trophoblastes, lors de la différenciation cellulaire, les cellules totipotentes du zygote (morula) génèrent des cellules embryonnaires pluripotentes (blastocystes), qui sont les précurseurs de trois couches germinales fondamentales: l'endoderme, le mésoderme et l'ectoderme (10). Les cellules de ces trois couches vont devenir des cellules multipotentes, générant différentes lignées cellulaires spécialisées et produisant une variété de types cellulaires qui acquerront une morphologie spécialisée et fonctionneront après la différenciation cellulaire (figure 1). Les cellules toti-, pluri- et multipotentes présenteraient une plasticité cellulaire. La plasticité cellulaire est définie comme la capacité d'une cellule à donner naissance à plusieurs cellules différentes dans un organisme multicellulaire (11). De l'état de cellules souches totipotentes à l'état de cellules multipotentes, les cellules perdent leur potentiel de différenciation cellulaire. Le potentiel de différenciation cellulaire est défini comme la gamme de types de cellules qui peuvent être produites par une cellule souche (12).

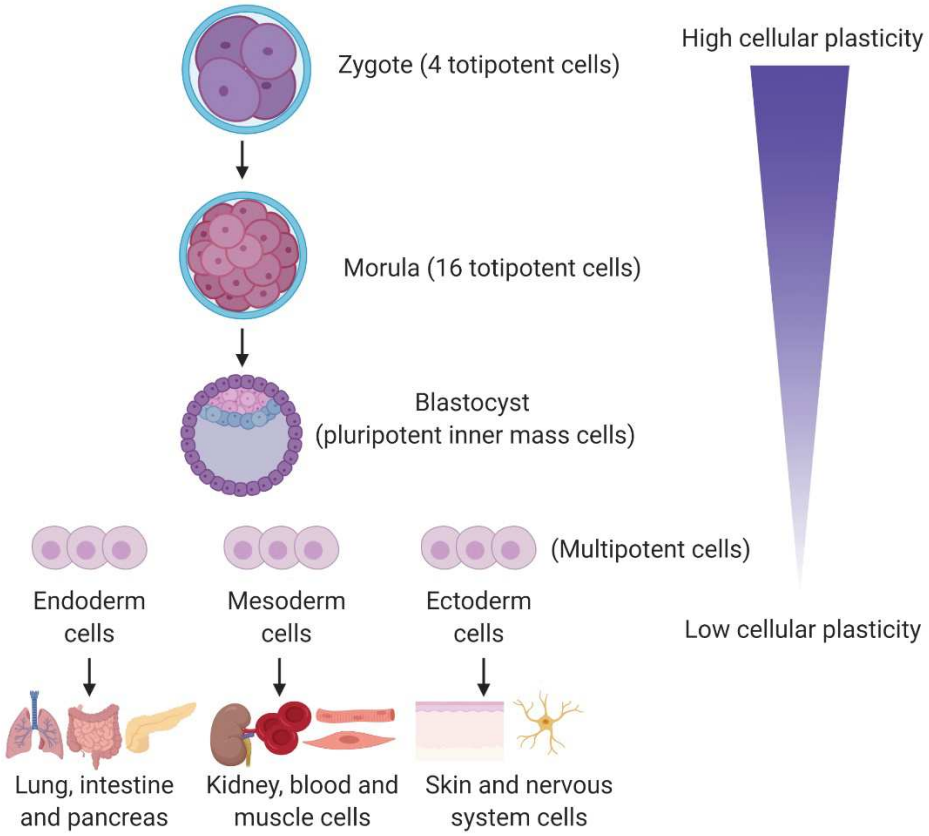


Figure 1. Processus de différenciation cellulaire. La différenciation cellulaire commence avec une cellule totipotente devenant pluripotente. Toutes les couches germinales seront créées et différenciées des cellules de masse interne pluripotentes du blastocyste en différentes lignées cellulaires créant les organes spécialisés de l'organisme multicellulaire et perdant la plasticité cellulaire au cours du processus.

La régulation de la différenciation cellulaire a fait l'objet de discussions et de réexamens au cours des années, définissant et redéfinissant ce qu'est une cellule souche ou une cellule

différenciée (9). À un moment donné, il a fallu se poser une question importante : qu'est-ce que cela signifie pour une cellule d'être fonctionnelle ou différenciée ? La définition moderne de la différenciation cellulaire a commencé en 1893 avec le livre « The Theory of the Germ Plasm » d'August Weismann (13). Weismann a affirmé qu'il y avait deux types différents de cellules produites pendant l'embryogenèse. Un type contient toutes les informations génétiques qui pourraient être transmises à la génération suivante, héritées par la descendance et l'autre type produira tous les différents types cellulaires de l'organisme multicellulaire. Avec cette hypothèse, il a créé une définition précoce des cellules germinales et des cellules somatiques. Selon le postulat de Weismann, les cellules somatiques au cours du développement deviendraient progressivement et irréversiblement restreintes au point que chaque cellule somatique ne pourrait se différencier qu'en un autre type de cellule spécifique tel que les muscles, les neurones, les épithéliums, etc. Cette affirmation était valable jusqu'à l'étude de la régénération cellulaire dans laquelle de nouveaux tissus ont été fabriqués en utilisant des tissus préexistants et selon le postulat de Weismann, les cellules différenciées en phase terminale n'avaient aucune capacité d'inversion du destin cellulaire. Toutefois, la capacité de certaines cellules à revenir en arrière était réelle. Weismann a passé les dernières années de sa vie à écrire une théorie non concluante afin d'expliquer ces phénomènes (13).

Parallèlement à Weismann, et en utilisant des siphonophores comme organismes modèles animaux, Ernst Haeckel a démontré que les cellules embryonnaires, qui avaient déjà acquis un destin cellulaire spécifique dans le cadre d'un tissu spécifique d'un organe, n'étaient pas limitées à cette lignée, mais capables de donner naissance à une autre type de cellule somatique. En amputant des embryons de Siphonophores, Haeckel a prouvé que les fragments embryonnaires pouvaient produire le corps complet de la larve (14). Cette démonstration a été le premier exemple enregistré dans l'histoire naturelle de la pluripotence des cellules embryonnaires. En outre, il y avait d'autres preuves expérimentales qui suggéraient la plasticité du destin de développement de la cellule.

En utilisant l'hydre, Ethel Browne a décrit qu'après avoir coupé une certaine partie du corps de l'hydre, comme la bouche, et l'avoir greffée dans une autre partie de la paroi corporelle d'une autre hydre, le stimulus était suffisant pour induire le développement d'un autre nouvel organisme décrit comme hydranth sur le site de la transplantation (15). Plus tard, Hilde Mangold et Hans Spemann ont mené des expériences similaires en utilisant des embryons de grenouilles. Ils ont greffé un morceau de la lèvre du blastopore sur le flanc d'un autre embryon de gastrula, éloigné du blastopore hôte d'origine, ce qui a entraîné l'induction d'un deuxième axe corporel (16). Toutes ces preuves ont permis de conclure que les cellules

embryonnaires n'avaient pas adopté un destin cellulaire irréversible et que la possibilité de procéder à une reprogrammation cellulaire était réelle.

En 1891, plusieurs expériences de Colucci utilisant des tritons comme système animal modèle ont démontré la capacité de ces amphibiens à régénérer leurs lentilles oculaires après une blessure sévère de l'œil (17). Après cette réalisation, en 1895, Wolff a découvert l'origine de ces nouvelles cellules de lentilles oculaires régénérées : les nouvelles cellules de lentilles oculaires ont été générées à partir des cellules épithéliales pigmentées de l'iris dorsal du triton (18). Récemment, une capacité de régénération similaire des planaires, un animal dont les cellules sont postmitotiques, a été démontrée, après l'amputation de plusieurs fragments du corps (19). Les fragments amputés sans cellules souches caractérisées peuvent effectuer une reprogrammation de leur état transcriptionnel, en activant ou désactivant différents signaux de structuration. Ce type de modulations modifie le sort des fragments amputés, conduisant à réarranger les tissus préexistants et enfin à produire de nouvelles petites planaires. Ceux-ci vont ensuite développer et générer de nouveaux organes, en gardant la taille relative entre eux (allométrie). Ces observations indiquent la présence de mécanismes de reprogrammation dans les cellules différenciées des tissus qui ont été amputés ; par conséquent, il doit y avoir un changement rapide de signalisation et des protéines exprimées régulant la structure corporelle de ces organismes (20) (21).

Pour finir, l'exemple le plus remarquable de reprogrammation cellulaire découvert dans la nature est celui présent dans les méduses *Turritopsis dohrnii et Laodicea undulata*. Ces animaux peuvent ramener leur corps adulte complet à un état larvaire antérieur. Les méduses reprogramment des parties du corps après avoir subi un stress chimique, physique ou thermique ou simplement vieillir. Dans ces cas, l'animal revient dans son cycle de vie en transformant les cellules adultes à des stades de développement antérieurs (22) (23) (24) (25). Le potentiel élevé de différenciation et de dédifférenciation des tissus hydrozoaires est la preuve finale et la plus concrète que les cellules de ces organismes ne sont pas différenciées de manière irréversible. On pourrait penser que la simplicité anatomique de la méduse pourrait être importante car elle faciliterait la reprogrammation. Cependant, ces animaux sont composés de trois couches différentes : un épiderme, une couche intermédiaire appelée mésoglée et la couche interne appelée gastrodermis. Ils ont également un « filet nerveux » agissant comme un système nerveux élémentaire leur permettant de détecter différents stimuli. Néanmoins, les Hydrozoaires n'ont pas de cerveau, de cœur et d'autres organes alambiqués. La seule structure complexe est une cavité digestive agissant à la fois comme un estomac et un intestin avec une seule ouverture qui fonctionne à la fois comme une bouche et comme un anus (26). Cependant, les autres

cas mentionnés ci-dessus chez les planaires ou les vertébrés tels que les axolotls (27) démontrent qu'une régénération chez les animaux anatomiquement complexes est également possible.

1.3 Reprogrammation cellulaire, régénération inductible et naturelle.

1.3.1 Reprogrammation cellulaire, transplantation de noyaux et cellules souches pluripotentes induites

Au cours du siècle dernier, la possibilité d'induire artificiellement des changements dans l'identité cellulaire a été découverte. Une expérience remarquable a été réalisée en 1943 par Johannes Holtfreter. Au cours de ses expériences, des cellules de grenouille ont été cultivées pour adopter un destin de cellules de peau; cependant, après avoir modifié le pH en ajoutant une solution acide (pH bas), les cellules de peau de grenouille ont adopté un nouveau sort cellulaire et sont devenues des cellules neuronales créant du tissu cérébral (28) (29). Des études supplémentaires ont développé une méthode pour transplanter avec succès un noyau d'une cellule de têtards dans un ovocyte de grenouille énucléé. Dans une proportion importante des transplantations, ces ovocytes se sont développés en embryons de grenouilles complets prouvant que le développement d'un animal complet à partir d'une seule cellule somatique était possible (figure 2a). La méthode a été étendue aux noyaux transplantés de l'intestin de grenouille dans des ovocytes produisant des grenouilles clonales dérivées des cellules somatiques d'autres grenouilles (30) (31). Suivant cette méthodologie, en 1996, Dolly le mouton a été le premier mammifère cloné par transplantation nucléaire de noyaux de cellules somatiques adultes dans un ovocyte (32) suivi par d'autres mammifères tels que la chèvre, le bétail, la souris ou le porc (33).

Après ces réalisations, l'étape suivante a été une reprogrammation cellulaire sans transplantation nucléaire, en modifiant le programme de transcription des cellules. L'un des faits saillants récents dans le domaine de la reprogrammation cellulaire a été la génération d'iPSC (cellules souches pluripotentes induites) par Shinya Yamanaka. Il a émis l'hypothèse qu'il y avait des gènes importants pour la fonction des cellules souches embryonnaires et que ces gènes pouvaient induire un état embryonnaire dans les cellules adultes. Il a sélectionné jusqu'à vingt-quatre gènes candidats, potentiellement capables d'induire une reprogrammation, et les a livrés à des cellules de fibroblastes de souris à l'aide de rétrovirus. Plus tard, pour sélectionner les bons, il a supprimé un gène à la fois de son pool génétique d'origine. En suivant cette méthodologie, il a identifié quatre gènes codant pour des facteurs de transcription qui sont suffisants pour la génération de cellules souches embryonnaires à partir de fibroblastes (34). Le fibroblaste peut être reprogrammé par l'introduction simultanée de ces quatre gènes différents : SOX2, OCT3 / 4, KLF4 et c-Myc.

Les fibroblastes ont été transformés en iPSC, un état similaire à l'état d'origine des cellules souches pluripotentes embryonnaires (ESC) (35) (figure 2b). Les iPSC peuvent se différencier en un autre type de cellule d'intérêt car ce sont des cellules pluripotentes comme la masse cellulaire interne du blastocyste, capables de se différencier dans de nombreux types de cellules.

Depuis la découverte de la capacité des cellules somatiques à changer leur identité cellulaire par une réversion induite en état de cellules souches embryonnaires en tant qu'iPSC (35), les efforts déployés par la communauté scientifique ont augmenté de façon exponentielle la capacité à induire une reprogrammation cellulaire. La reprogrammation peut être obtenue en manipulant des voies de signalisation ou en manipulant des facteurs qui médient la conversion de différents types de cellules (36) (37). En utilisant ces approches, les cellules pancréatiques exocrines peuvent être converties en cellules pancréatiques β, tandis que les fibroblastes peuvent être adoptés dans différents destins cellulaires, même en dehors de la lignée de fibroblastes en développement comme les neurones, le pancréas ou le foie (38) (39) (40).

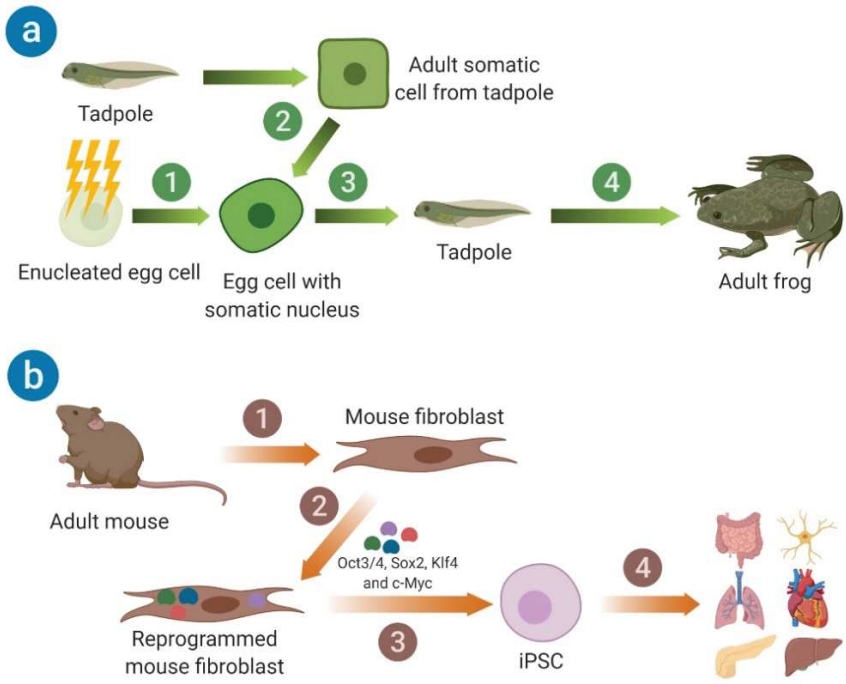


Figure 2. Reprogrammation cellulaire induite pour générer la pluripotence. a) John B Gurdon a éliminé le noyau d'un ovule (1) pour introduire le noyau d'une cellule somatique adulte prélevée sur un têtard (2). L'œuf modifié s'est développé en un têtard (3) qui devient une grenouille adulte (4). b) Shinya Yamanaka a extrait des cellules de fibroblastes de souris adulte (1) et transféré des gènes précédemment sélectionnés qui étaient importants pour réactiver la fonction des cellules souches (2). Après avoir introduit quatre facteurs de transcription, le fibroblaste de souris reprogrammé est devenu iPSC (3) capable de donner naissance à tous les types cellulaires d'une souris adulte (4). Références: (9) (31).

1.3.2 Régénération par dédifférenciation naturelle et transdifférenciation naturelle.

Le remplacement des cellules endommagées par des cellules saines dérivées de cellules adultes redifférenciées du patient est l'objectif principal et le défi de la médecine régénérative. La restauration de la pluripotence cellulaire dans un état complètement différencié est possible comme décrit ci-dessus. De plus, dans la nature, il est également possible d'induire la prolifération cellulaire après dédifférenciation d'une cellule ou de passer directement à un autre type de cellule par transdifférenciation avec ou sans prolifération cellulaire (12) (41), comme cela sera décrit ci-dessous. Dans ce chapitre, je présente quelques exemples de processus de dédifférenciation naturelle et de transdifférenciation naturelle. La différence entre la dédifférenciation naturelle et la transdifférenciation est définie par les événements survenus après la perte de l'identité adulte d'origine. Dans les événements de dédifférenciation naturelle, les cellules adultes se dédifférencieront et prolifèreront pour se redifférencier plus tard dans le même type de cellule initiale, par conséquent, l'identité initiale et finale sera toujours la même (figure 3). Dans les événements de transdifférenciation naturels, une cellule adulte se dédifférenciera avant la prolifération, mais après que les cellules de redifférentiation adoptent une identité différente, ainsi, l'identité initiale et l'identité finale sont différentes (figure 4).

Plusieurs espèces de vertébrés non mammifères ont des capacités de régénération pour restaurer des parties de leur corps. Dans la majorité des cas, la régénération est possible grâce à la dédifférenciation des cellules matures. Par exemple, le poisson zèbre (Danio rerio) peut régénérer complètement une partie de son cœur après avoir amputé le ventricule. Après amputation, les cardiomyocytes encore présents dans cet organe, commencent à se différencier, à démonter leur structure de sarcomère. Ils commencent alors à proliférer afin de restaurer le ventricule, en reconstruisant un nouveau (41) (42). Ce type de régénération a cependant une limite car il est nécessaire qu'au moins 20% du ventricule reste dans l'animal, sinon la dédifférenciation ne se produit pas (43) (44) (figure 3a). Les cardiomyocytes subissant une prolifération et une redifférenciation par dédifférenciation sont appelés blastèmes (27) (45). De même, la régénération des membres de tritons comme l'axolotl (Ambystoma mexicanum) est également possible par des cellules qui se différencient et prolifèrent. Ces animaux ont la capacité de régénérer leur membre après l'avoir perdu en raison d'une attaque de prédateur et de son amputation. Une fois le bras coupé, les cellules commencent à proliférer au site de coupe, générant une nouvelle masse cellulaire (blastème) constituée de cellules dédifférenciées avec une capacité de prolifération élevée qui se redifférencient finalement pour créer les tissus et les structures du membre perdu (45) (46) (figure 3b).

Chez les mammifères, la régénération par dédifférenciation naturelle est également possible. La régénération des cellules de Schwann a été réalisée : les cellules de Schwann qui perdent le contact avec l'axone peuvent redevenir précurseurs. Ce processus est observé in vivo après une lésion nerveuse, montrant les cellules de Schwann dénervées phénotypiquement similaires aux cellules de Schwann immatures avant la myélinisation. Les cellules de Schwann dénervées subissent une régression dans leur développement et se différencient en perdant leur propre myéline et prolifèrent ensuite. Au cours du processus de dédifférenciation et de prolifération, les macrophages et les monocytes éliminent les débris cellulaires produits par la démyélinisation des cellules de Schwann (47) (48). Une fois que les cellules de Schwann dédifférenciées établissent de nouveaux contacts avec les axones, elles redifférencient et myélinisent le nerf (49) (figure 3c).

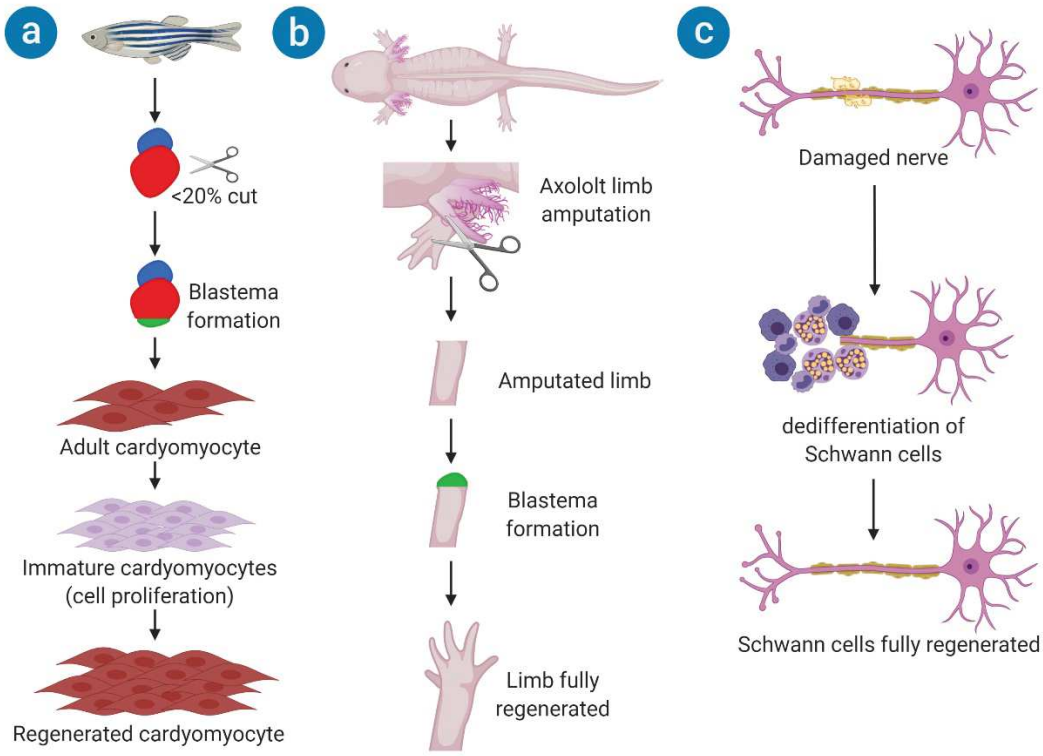


Figure 3. Dédifférenciation naturelle et prolifération cellulaire chez le poisson zèbre (Danio rerio), l'axolotl (Ambystoma mexicanum) et les cellules de Schwann de mammifères. a) La régénération du cœur du poisson zèbre n'est possible que si l'amputation est <20% de la taille totale du cœur, dans ces cas, les cardiomyocytes adultes présents dans le site de coupe commenceront à former un blastème (vert) capable de se différencier en cardiomyocytes immatures. Ces derniers prolifèrent avant de se redifférencier en se transformant en nouvelles cellules de cardiomyocytes adultes, régénérant ainsi le ventricule cardiaque complet. b) L'amputation du membre dans les axolotls a produit une masse hétérogène de cellules appelée blastème (45) (vert) capable de régénérer les muscles, les os et le cartilage, pour finalement faire repousser complètement le membre de l'animal. c) Lorsqu'un nerf est blessé, les cellules de Schwann se dédifférencient en précurseurs des cellules de Schwann immatures avant de proliférer pour former de nouvelles cellules de Schwann. Après prolifération et redifférenciation en nouvelles cellules de Schwann, l'axone nerveux sera totalement recouvert de myéline.

D'un autre côté, la transdifférenciation naturelle est une autre option pour régénérer la perte ou l'endommagement de parties du corps par la conversion de tissus différenciés existants ou de cellules individuelles en d'autres types de cellules. L'un des exemples classiques de transdifférenciation chez les vertébrés est la régénération du cristallin chez le triton (salamandres) : ces animaux peuvent transdifférencier les cellules de leurs yeux lorsqu'ils sont blessés. Les cellules épithéliales pigmentées de l'iris peuvent se différencier pour régénérer les cellules du cristallin. Tout d'abord, les cellules pigmentées se dédifférencient et prolifèrent, puis, les cellules deviendront des cellules cristallines, adoptant un nouveau destin et redonnant ainsi la vision du triton (41) (50) (figure 4).

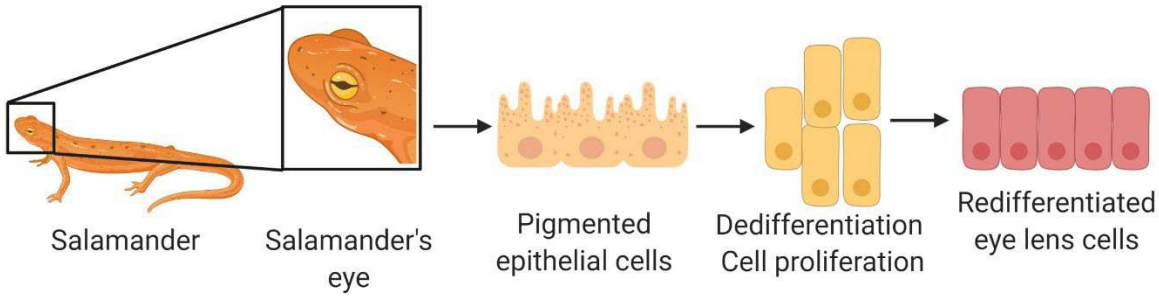


Figure 4. Transdifférenciation naturelle de la salamandre pour régénérer le cristallin. Les cellules épithéliales pigmentées de l'œil commencent à se différencier en modifiant leur morphologie et en adoptant un nouveau destin cellulaire en se différenciant des cellules du cristallin adulte.

1.4 Le rôle des facteurs de transcription dans la reprogrammation.

Toutes ces découvertes ont changé le modèle statique traditionnel dans lequel seules les cellules souches donnent lieu à différentes identités cellulaires par différenciation, tandis que ces identités sont fixées jusqu'à la mort de la cellule. La découverte de la base moléculaire pour induire une reprogrammation cellulaire par Gurdon et Yamanaka a mis fin à ce modèle (31) (35) (51). La découverte des facteurs de transcription de Yamanaka (35) a été une étape importante pour démontrer l'importance de la transcription dans la reprogrammation cellulaire. Les travaux de Yamanaka ont conduit à un nouveau modèle de plasticité cellulaire, dans lequel la plasticité cellulaire n'implique aucune directionnalité de / vers les cellules souches et les cellules différenciées, et où la modification de l'expression des gènes peut changer la nature d'une cellule.

1.4.1 Le rôle des facteurs de transcription dans la reprogrammation pluripotente.

Trois facteurs de transcription ont été largement étudiés pour leur rôle dans le maintien de la pluripotence et l'activation / inactivation de l'expression des gènes (41) (52). Ces facteurs de transcription sont OCT4, SOX2 et NANOG (41). Tous sont associés à l'activation et au maintien de la pluripotence pouvant activer ou désactiver l'expression des gènes (53) (54)

(55). Ces facteurs de transcription ont plusieurs rôles dans le développement, certains d'entre eux étant limités au stade des cellules souches pluripotentes. NANOG est uniquement exprimé dans les cellules pluripotentes du blastocyste, dans la couche cellulaire de masse interne. C'est un acteur important dans l'acquisition de la pluripotence mais il n'est pas nécessaire de maintenir le destin des cellules souches (56). De plus, lorsque le NANOG n'est pas exprimé dans les cellules pluripotentes de la masse cellulaire interne du blastocyste, les embryons ne sont pas en mesure de se développer correctement (57). En revanche, SOX2 et OCT4 sont directement impliqués dans la différenciation. Le facteur de transcription SOX2 n'est pas limité au stade pluripotent des cellules souches, il est également présent à d'autres stades adultes différenciés comme les cellules de la plaque neurale (58). Le facteur de transcription OCT4 est un acteur central du maintien de la pluripotence par son rôle dans l'activation ou la répression de différents gènes (54) (55). SOX2 a besoin d'OCT4 pour interagir avec l'ADN, agissant ainsi comme un cofacteur (59) et les trois facteurs de transcription agissent ensemble comme des régulateurs clés de la pluripotence, coopérant à de nombreux niveaux moléculaires pour maintenir le caractère souche. En outre, OCT4, NANOG et SOX2 peuvent réguler leurs propres niveaux de transcription en se liant à leurs propres séquences promotrices. Cela a été suggéré comme étant la clé du maintien pluripotent du destin (60). De plus, SOX2 et OCT4 régulent l'expression de NANOG, expliquant pourquoi NANOG n'est pas l'un des quatre facteurs Yamanaka requis pour la génération d'iPSC (9) (35) (61). Il a été démontré que des variations de la combinaison de facteurs de transcription précédemment mentionnée utilisée par Yamanaka (OCT4, SOX2, KLF4 et c-Myc) induisent la pluripotence dans d'autres contextes. Par exemple, des fibroblastes reprogrammés sans c-Myc ont été décrits (62). De plus, il est possible de générer des iPSC à partir de fibroblastes de souris uniquement en utilisant OCT4 et SOX2 (63) et de reprogrammer des cellules de sang de cordon humain en utilisant uniquement ces deux facteurs de transcription (64). Les cellules souches neurales, en raison de leur expression élevée de SOX2, ont été décrites comme les cellules les plus faciles à reprogrammer, ne nécessitant que de l'OTCT4 (65). La seule certitude pour induire une reprogrammation pluripotente est la nécessité d'une OCT4 fournie endogène ou exogène (66) pour effectuer une reprogrammation cellulaire (41).

1.4.2 Le rôle des facteurs de transcription dans la reprogrammation directe.

Les exemples décrits ci-dessus sont applicables à une reprogrammation dans laquelle une population cellulaire doit passer par un état de cellule souche pluripotent avant de se redifférencier en un autre. Cependant, certaines cellules procèdent à une reprogrammation directe, convertissant les cellules ou la population cellulaire en un destin différent par une régulation négative simultanée du programme transcriptionnel original et une régulation

positive du nouveau programme transcriptionnel correspondant à la nouvelle identité cellulaire (67). Un bon exemple de ce type de mécanisme de reprogrammation cellulaire est la transdifférenciation des cellules B en macrophages (67). Au cours du développement, la différenciation des cellules B des précurseurs hématopoïétiques est initiée par deux facteurs de transcription différents EBF et E2A. Les deux facteurs de transcription induisent l'expression d'un facteur d'engagement appelé PAX5 qui conduit à une régulation positive de nombreux gènes spécifiques des cellules B permettant à la cellule progénitrice de se différencier en cellule B. D'un autre côté, l'engagement des macrophages est induit par deux facteurs de transcription appelés CeBpα et CeBpβ. De plus, un autre facteur de transcription entraîne l'engagement des deux types de cellules, il s'agit du facteur de transcription est PU.1. Les expériences ont montré que les souris défectueuses pour PU.1 n'avaient pas de cellules B ni de macrophages (68). L'expression induite de CeBpα et CeBpβ dans les cellules B a conduit à une régulation négative de Pax5 affectant le reste de l'expression des gènes spécifiques des cellules B. De plus, la régulation positive des gènes spécifiques des macrophages est médiée par CeBpα et CeBpβ qui restent dédifférenciés les cellules B dans un état intermédiaire dans lequel de faibles niveaux de gènes spécifiques B et spécifiques des macrophages sont exprimés en même temps. Après la redifférenciation, les cellules B ont acquis expérimentalement la transdifférenciation de l'identité des macrophages (67).

1.5 Épigénétique: les marques de méthylation pourraient affecter l'expression des gènes et la reprogrammation cellulaire.

L'épigénétique joue un rôle remarquable dans la reprogrammation cellulaire. Cette discipline a été définie comme «la branche de la biologie qui étudie les interactions causales entre les gènes et leurs produits qui font naître le phénotype» par Conrad Waddington (69) et aujourd'hui comme «l'étude des changements de la fonction des gènes qui sont mitotiques et / ou héréditaire méiotique et qui n'entraînent pas de modification de la séquence d'ADN " (70). Les modifications épigénétiques affectent l'expression des gènes et la mise en silence des gènes par des modifications de l'architecture de la chromatine. Ces modifications de la chromatine jouent un rôle important dans l'expression des gènes, maintenant et stabilisant ainsi le destin cellulaire (71) (72). La méthylation de l'ADN et les modifications post-traductionnelles des histones, deux exemples de modifications épigénétiques, peuvent induire des changements conformationnels régulant l'accessibilité à l'ADN et la régulation des gènes (72) (73) (74). La méthylation de l'ADN est catalysée par les ADN méthyltransférases (DNMT). D'un autre côté, les modifications des histones peuvent être catalysées par les histone méthyltransférases (HMT).

Les ADN méthyltransférases sont des enzymes qui catalysent l'addition de groupes méthyle aux résidus de cytosine dans les îles CpG qui sont abondantes dans les régions promotrices de nombreux gènes chez les vertébrés. Cette méthylation peut entraver l'activité de liaison du facteur de transcription ou agir comme un site de reconnaissance pour le recrutement de facteurs épigénétiques répressifs. Il est important de noter que la méthylation médiée par les DNMT en cinquième position de la cytosine (5mC) est présente chez les mammifères, où la méthylation sert à réprimer la différenciation dans les cellules pluripotentes (75). Cette méthylation n'est pas présente dans l'organisme modèle utilisé pour ce travail de doctorat, *C. elegans*. Cependant, des études récentes ont montré la présence de méthylation endogène chez *C. elegans* en sixième position de l'adénine (6mA), concrètement dans les motifs GAGG et AGAA de la séquence d'ADN (76). Cette méthylation est médiée par DNMT-1, une ADN méthyltransférase qui a été identifiée dans la même étude en utilisant le nématode comme organisme modèle (76). Néanmoins, la fonction de cette méthylation reste inconnue mais pourrait être potentiellement impliquée dans la transmission d'informations épigénétiques d'une génération à l'autre.

Les histone méthyltransférases (HMT) sont des enzymes qui catalysent le transfert de plusieurs groupes méthyle (trois, deux ou un) aux résidus lysine des protéines histones. Le compactage de la chromatine dépend fortement de la méthylation des histones par HMT et d'autres modifications post-traductionnelles des histones (77). La méthylation sur les histones peut soit activer soit réprimer la transcription selon le site de méthylation. H3K9me sont enrichis dans la région promotrice des gènes et sa fonction est d'empêcher une expression génique élevée associée à des retards de transition du cycle cellulaire ou à la prolifération cellulaire (78) (79). H3K9me est également impliqué dans l'arrêt du développement et est lié au pluripotent au silençage génique (73). De plus, H3K9me2 et H3K9me3 sont associés à la répression transcriptionnelle avec une importance critique dans la reprogrammation cellulaire (79). Il a été démontré qu'une réduction de la marque d'histone répressive H3K9me2 / 3 améliorera la reprogrammation cellulaire, ouvrant la chromatine et activant l'expression des gènes (73). Une autre modification épigénétique, la méthylation H3K27, est également impliquée dans le compactage de la chromatine et la formation d'hétérochromatine (80). H3K27me est une modification répressive des histones associée au maintien de la pluripotence et à l'auto-renouvellement (81) (82). La déméthylation de H3K27me3 est nécessaire pour effectuer une reprogrammation cellulaire : une reprogrammation du fibroblaste cardiaque en myocytes a été réalisée avec succès afin de régénérer le myocarde lésé (83). Pour terminer, la méthylation de H3K4 est associée à une chromatine transcriptionnelle ouverte et active avec une activité d'expression génique intense (84) et régule la puissance et le maintien des cellules souches (85).

Les domaines des protéines Jumonji ont été caractérisés comme des lysines déméthylasses (KDM) (86). Deux des domaines JMJD les plus étudiés sont 2C et 1A, tous deux responsables de la déméthylation H3K9me3 / 2 et de la déméthylation H3K9me2 / 1, respectivement (87). L'expression de ces deux domaines Jumonji est régulée par le facteur de transcription OCT4 (87). En utilisant des sites de liaison OCT4 précédemment cartographiés (60), il a été démontré par des tests Chip-seq que OCT4 se lie à l'expression du gène activant jmjd2c et jmjd1a. De plus, la déplétion d'Oct4 en utilisant l'ARNi a diminué l'expression de JMJD2C et JMJD1A (87). Ces preuves ont confirmé que l'OCT4 régule l'expression de ces gènes d'histone déméthylasse. De plus, lors de la reprogrammation des fibroblastes pour obtenir des iPSC, une augmentation de l'expression des deux déméthylasses a été détectée (35) en corrélation positive avec l'état pluripotent des cellules souches embryonnaires et des iPSCs.

1.6 *Caenorhabditis elegans* as a model organism.

C. elegans est un ver non parasite, un nématode, d'une taille approximative de 1 mm de longueur à l'âge adulte. L'alimentation du ver dépend de l'action d'un organe appelé pharynx qui pompe les micro-organismes de la bouche vers l'intestin (88). Dans la nature, cet animal peut vivre dans des habitats influencés par l'homme tels que des tas de compost, des jardins et des vergers végétaux ou botaniques, et dans des endroits sauvages tels que des fruits et des plantes pourris ou du sol sous différentes latitudes, étant tous des environnements riches en microbes (89). Dans des conditions de laboratoire, les nématodes se développent dans une boîte de Pétri nourrie avec Escherichia coli comme source de nourriture.

Il existe deux sexes de *C. elegans*, hermaphrodite et mâle, présentant un fort dimorphisme sexuel entre eux. L'hermaphrodite avec 959 cellules somatiques et deux chromosomes X peut s'auto-féconder et ne peut pas féconder d'autres hermaphrodites. Le mâle, avec 1031 cellules somatiques et un seul chromosome X, peut fertiliser les hermaphrodites. Ceci est avantageusement utilisé pour introduire de nouveaux allèles par croisements, par exemple entre différentes souches transgéniques et mutants, fournissant un outil puissant pour les études génétiques. Les mâles représentent 0,1 à 0,2% (90) (91) de la population naturelle, le reste de la population est hermaphrodite.

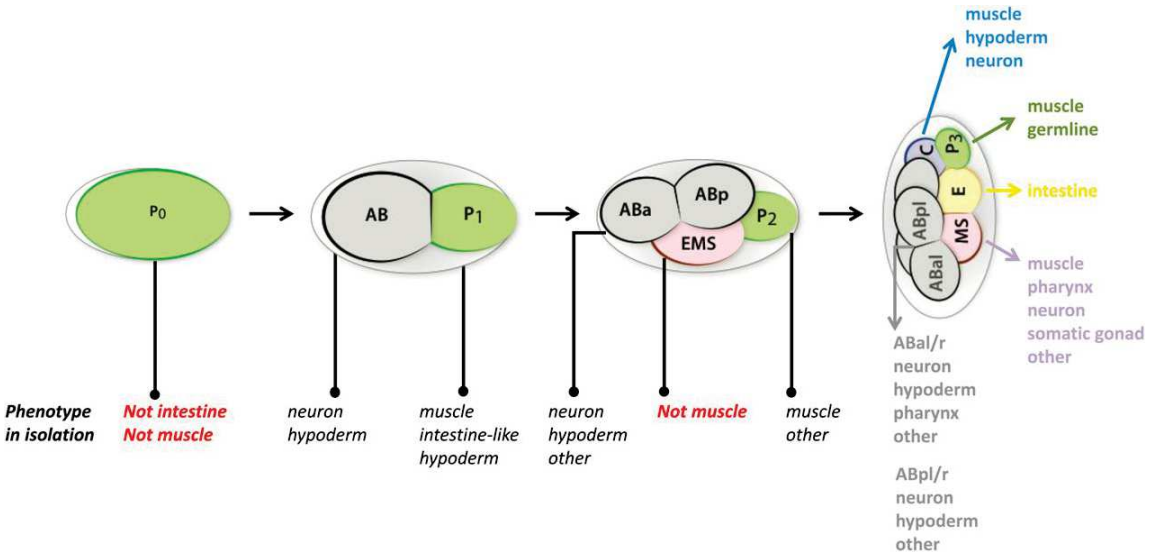


Figure 5. Early embryonic *C. elegans* lineage. Source: (12)

The development of *Caenorhabditis elegans* starts after the fecundation of the oocyte, then, the fertilized egg called P_0 divides into two different blastomeres AB and P_1 after an asymmetric division. AB divides again to form ABa and ABp that are going to follow different divisions patterns. P_1 will divide asymmetrically in EMS and P_2 blastomeres. P_2 is going to divide to C and P_3 blastomeres, on the other hand EMS are going to divide to MS and E. Each blastomere are going to contribute to create different tissues, MS are going to differentiate into muscles, pharynx, neurons and gonad cells E are going to form the gut, C are going to gives muscles, neurons and epithermal cells. The AB lineages, ABa and ABp are going to create neurons, skin tissue, muscle and pharynx (Figure 6). The morphogenesis of the worm starts at the 550-cell stage. Worms at the first larval stage have 558 cells 12 hours post fertilization at 20-22ºC(92)(93).

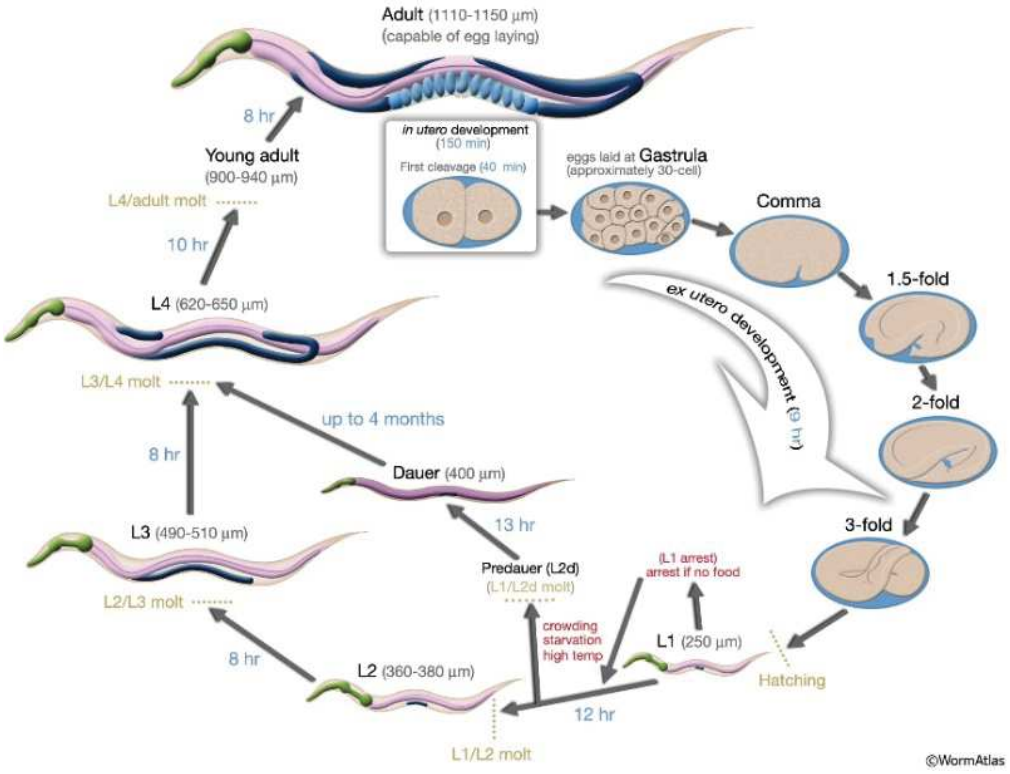


Figure 6. *C. elegans* life cycle. Source: WormAtlas.

There are four larval stages (L1, L2, L3, L4) in the life cycle (Figure 6) of *C. elegans* punctuated by four different molts before animals reach adulthood. The adult stage goes from young adult to gravid adult, which is a self-fertilizing worm able to lay eggs which will produce progeny during up to four days. Animals then live 10 to 15 days once the adult state is reached. (94). The time to complete the life cycle in optimal laboratory conditions with a temperature at 20ºC is 3.5 days. However, the life cycle might be accelerated or slowed down by changing the temperature from 20ºC to 25ºC or 15ºC, respectively, (94) producing variations in development in the timing of egg-hatching, larval molting and egg laying (95) (Figure 7).

DEVELOPMENT AT DIFFERENT TEMPERATURES

	"16°C" (16.0 ± 0.3°C)	"20°C" (19.5 ± 0.5°C)	"25°C" (25.0 ± 0.2°C)
Egg laid	0 hr	0 hr	0 hr
Egg hatches	16-18 hr	10-12 hr	8-9 hr
First-molt lethargus	36.5 hr	26 hr	18.0 hr
Second-molt lethargus	48 hr	34.5 hr	25.5 hr
Third-molt lethargus	60.0 hr	43.5 hr	31 hr
Fourth-molt lethargus	75 hr	56 hr	39 hr
Egg-laying begins	~90 hr	~65 hr	~47 hr
Egg-laying maximal	~140 hr	~96 hr	~62 hr
Egg-laying ends	~180 hr	~128 hr	~88 hr
Length at first molt	360 µm	370 µm	380 µm
Length at second molt	490 µm	480 µm	510 µm
Length at third molt	650 µm	640 µm	620 µm
Length at fourth molt	900 µm	850 µm	940 µm
Length at egg-laying onset	1150 µm	1060 µm	1110 µm
Maximal egg-laying rate	5.4/hr	9.1/hr	8.1/hr
Total eggs laid	275	280	170

Figure 7. Different developmental timing changes of *C. elegans* according to the growth temperature. Source: (95)

Additionally, when the environmental conditions become harsh (crowding, starvation or high temperatures), *C. elegans* larvae can undergo Dauer larvae arrest during the L2 stage. Animals develop into an intermediate stage called L2d before entering Dauer. Dauer larva can resume the normal life cycle if favorable conditions are restored, for example when food is present and temperature becomes lower, passing briefly through the L3 stage and entering in L4 stage to complete the life cycle (Wormbook).

The genome of *C. elegans* was sequenced in 1998, being the first pluricellular organism to get its genome fully sequenced (96). The genome of *C. elegans* has a size of 100 Mbp with 20.512 protein coding genes, 21.500 noncoding genes and 1484 pseudogenes (Wormbook). 83% of the *C. elegans* proteome has homology with human proteins making this model organism an excellent system to study the basis of many disease-relevant molecular processes (97). Furthermore, the nematode offers the advantage of being transparent during all the life cycle, making it possible the study and observe all its development and internal anatomy as well as gene expression patterns using fluorescent markers (98).

1.7 Cellular reprogramming in *C. elegans.*

Cellular reprogramming was observed in *Caenorhabditis elegans* either using inducible cell conversion or natural transdifferentiation (12). These facts showed the potential of this worm to study cellular *in vivo* reprogramming processes.

1.7.1 Induced transdifferentiation in *C. elegans*.

Induced cell conversion in early *C. elegans* blastomeres is possible by the ectopic expression of "master" genes under the control of a heat shock promoter. Overexpression of transcription factor HLH-1 homologue of the human transcription factor MyoD, leads reprogramming of many blastomeres into muscles in embryo populations (99)(100). A similar induced event was observed after overexpression of the transcription factor END-1, homolog of GATA in humans: blastomeres became intestinal cells (101). Surprisingly, more induced events were identified by gene overexpression: LIN-26, an hypodermal transcription factor converted blastomeres into hypodermal cells (102). ELT-1 and ELT-3, both epidermal transcription factors converted blastomeres into epidermal cells (103) and to finish, PHA-4, a transcription factor involved in pharyngeal development converted blastomeres into pharyngeal cell (104). However, it has been described that induction have a limited plasticity window: Blastomere conversion was observed from 50 cell stage until 150 cell stage (105).

1.7.2 Natural transdifferentiation in *C. elegans*.

Natural reprogramming in *C. elegans* has been largely studied in the last decades identifying two extensive and one conceivably natural reprogramming events. The two extensive reprograming events are the natural transdifferentiation from rectal-to-neuronal identity (Y-to-PDA) (12)(106)(107) and the reprogramming of a glial cell into neuron (AMso-to-MCM). There is another putative reprogramming event is the conversion of neuron into another neuron type (G1/G2-to-neuron) (12)(108).

1.7.2.1 Y-to-PDA as a natural transdifferentiation system.

The Y-to-PDA system (Figure 8) consists in the transdifferentiation of a rectal cell into a motoneuron from the beginning of the larval development during L1 stage until the middle larval stage called L3. *Caenorhabditis elegans* rectum is formed by six different cells: K, K', U, F, B and Y (Figure 8) that are born approximately 300 minutes after the fertilization (93). The identity of most of these cells remains unaltered since this time point until the death of the animal, except for the Y cell, which has an extraordinary behavior: by the end of the larval stage L1 the cell will undergo a dedifferentiation abandoning its rectal identity and migrating out of the rectum. This transient state is called Y_0 which is a dedifferentiated state in which the Y_0 cell does not make any cell division. In the L3 stage, the same cell becomes

motoneuron called PDA, extending a long ventral axon that crosses to the dorsal side of the and extends to the head of the worm, synapsing to other dorsal muscles (106)(109). This transdifferentiation event was described by Sophie Jarriault (106) as a *bona fide* natural cell fate change of an adult cell into another adult cell type with a totally different nature using differential interference contrast microscopy, electronic microscopy and expressing specific molecular fluorescence markers in Y and PDA (106). In comparison to other natural transdifferentiation events previously described like the transdifferentiation of pigmented cells into eye lens cells in salamanders (50) (Figure 4; Chapter 1.3.2), this transdifferentiation event can be followed at the single cell level and occurs without any cellular division or proliferation.

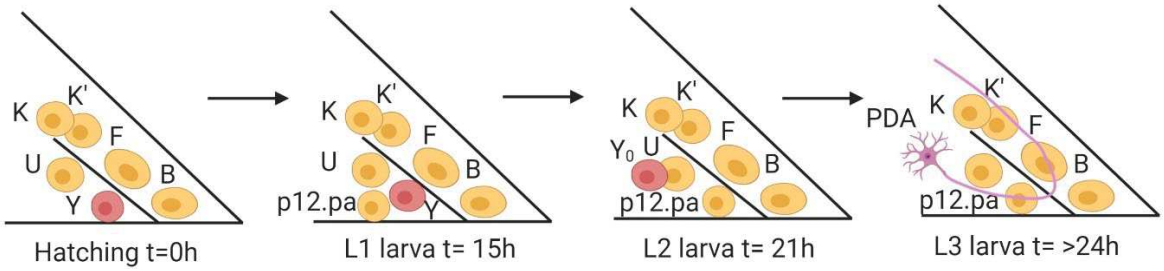


Figure 8. The Y-to-PDA system. Y cell is born during the embryogenesis and remains with an epithelial rectal cell identity until the L2 larval stage when it becomes Y_0, a dedifferentiation state and later adopts a motoneuron identity called PDA.

Additionally, the influence of the neighboring cells was tested by ablation experiments suggesting that the rest of the rectal cells are not influencing this transdifferentiation event. The team further showed that the migration of the rectal Y cell and the change of microenvironment from the rectum to another position inside the worm were not required to carry out the dedifferentiation and the redifferentiation from an epithelial cell to a motoneuron. This hypothesis was confirmed by inhibition of the migration of the cell: a non-migrated Y_0 cell is going to re-differentiate as well in PDA (106).

To determine and identify the inherent and extrinsic mechanism involved in this natural transdifferentiation event, several genetic screens were carried out at the single cell level. Different screen designs using transgenic strains where a fluorescent protein is expressed in the initial or the final identity allowed to target all the cellular events in Y-to-PDA transdifferentiation, from the inability to initiate transdifferentiation to the inability to re-differentiate and produce a PDA neuron (110)(111).

The Y-to-PDA transdifferentiation is a transition between two adult cell types in successive discrete steps. In some reprogramming events, transcription factors can induce direct cellular reprogramming passing through one cell type to another without completely

switching off the specific markers of the first cell type before inducing the new ones, producing a transient mixed cell type instead of a fully dedifferentiated intermediate (12)(112). During *C. elegans* Y-to-PDA transdifferentiation, there is no evidence of an intermediate mixed identity displaying Y and PDA phenotypes at the same time (107). Indeed, the analysis of the Y-to-PDA process suggested that the Y cell transits through a dedifferentiated "Y_0" cell state without gaining cellular potential or entering cellular proliferation (12)(107). Different analyses also revealed that once the cell enters into the dedifferentiation phase, the redifferentiation occurs in a separate step resembling developmental neuronal differentiation and requiring the UNC-3/EBF transcription factor (107). Additionally, the role for the heterochronic pathway in Y-to-PDA could not be observed: there are no evidences that the pathway could be affecting the timing or occurrence of Y transdifferentiation (106).

As introduced earlier in this PhD manuscript, transcription factors play a central role to carry out cellular reprogramming. One good example, already stated, are the four Yamanaka factors previously introduced (Chapter 1.3) used to reprogram fibroblasts into iPSCs recovering pluripotency by inducing dedifferentiation (9)(35). The Y-to-PDA transdifferentiation system is also driven by many different transcription factors that were identified by RNAi screens for early steps of Y-to-PDA transdifferentiation. Those transcription factors are: EGL-27, SEM-4, CEH-6 and SOX-2. Individual knockdown experiments were carried out and for all of them, the Y cell displays a transdifferentiation arrest phenotype. The Y cell is either incapable to migrate or to adopt a PDA phenotype. All those four transcription factors are conserved in mammals: *Mta4/egl-27 Sall/sem-4*, *Sox2/sox-2* and *Oct4/ceh-6* are the closest functional homologues. Three of them are part of the NODE complex, essential to maintain pluripotency of embryonic stem cells in mammals, and associated with SOX2 (12)(110). In addition, it has been shown that loss of *egl-5, a* posterior hox gene, leads to similar early defect in Y-to-PDA reprogramming (106), because of this, the relationship between *egl-27* and *egl-5* was investigated. It was showed that there was not Y-to-PDA arrest in *egl-5* mutants, on the other side in *egl-27* mutants there was at least 40% of Y-to-PDA arrest and in *egl-5/egl-27* mutants there was >60% Y-to-PDA arrest (110). To finish, when EGL-5 is ectopically expressed, this does not affect transdifferentiation in wild-type animals, yet it rescues Y-to-PDA transdifferentiation arrest in *egl-27* mutants. That demonstrates that EGL-5 is acting downstream of EGL-27 and also is necessary for the initiation of the Y-to-PDA transdifferentiation (110).

All evidences suggest that nuclear events control natural cell plasticity processes in the Y-to-PDA transdifferentiation in worms (27)(35)(50). The transcription factors previously described are necessary to allow Y to swap identity, yet they are expressed in all rectal cells.

However, these cells do not transdifferentiate like Y transdifferentiates into PDA. This fact suggests that these factors are permissive, and that, in order to enter transdifferentiation, Y needs an inductive signal. This signal is the LIN-12/Notch signaling pathway that give competence to the Y cell to start Y-to-PDA during the embryogenesis (106).

1.7.2.2 Histone modifications and transcription factor activity in Y-to-PDA.

Robustness and high reproducibility independent of the environmental growing conditions are a must to define a natural transdifferentiation process. In *Caenorhabditis elegans*, the Y-to-PDA transdifferentiation occurs in 100% of wild type animals (106). In order to elucidate the mechanisms which ensure the efficiency and robustness of the natural conversion of differentiated adult cells, a collection of weakly penetrant mutants for Y-to-PDA was analyzed (113). Mutations in different histone modifier-encoding genes were identified using low penetrance worm mutant strains without cause redundancy with other known molecular pathways or producing only partial loss of function. Lysine demethylase JMJD-3.1 and the methyl transferase SET-1 were identified acting on different lysine residues of H3, impacting on the cell conversion (113). These histone modification activities are involved dynamically during Y-to-PDA regulating the different steps of this natural transdifferentiation event. Genetic screens and different biochemical analyses complemented with photoconvertible fluorescence reporters showed that the nuclear levels of JMJD-3.1 protein are actively regulated in a time-controlled manner, a phenomenon which underlies the sequential activity of these histone modification enzymes during Y transdifferentiation. In addition, the association between the SET-1 complex or JMJD-3.1 and step-specific transcription factors is critical for progression through transdifferentiation (110)(113) (Figure 10).

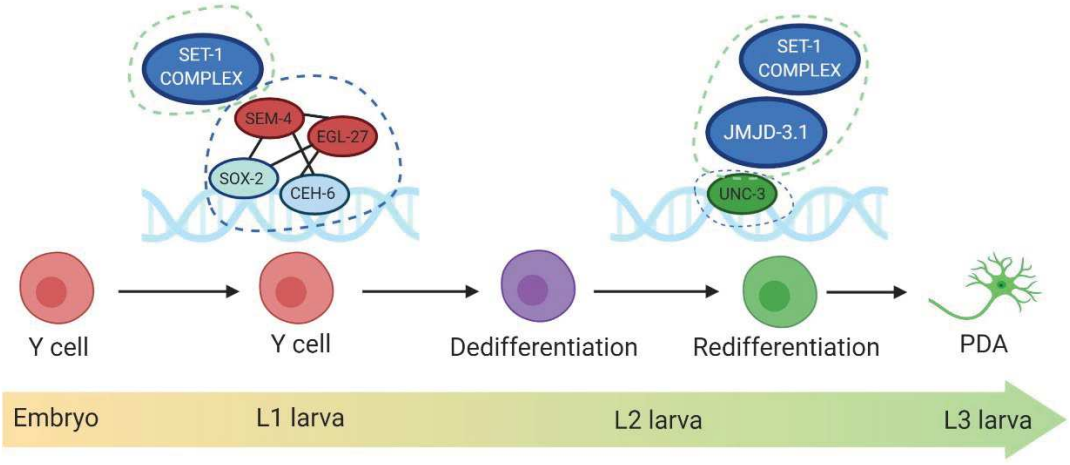


Figure 9. Histone modifications on Y-to-PDA. During the transdifferentiation of a Y cell into PDA cell several combination of histone modifiers complexes are interacting with different transcription factors involved in Y-to-PDA natural transdifferentiation.

It is remarkable that those associations are conserved in mammals: WDR5, a member of SET1 complex in mammals, is also associated with OCT4, a transcription factor involved in cell dedifferentiation during the initial phases of a pluripotent reprogramming event (114). Furthermore, JMJD3 is necessary for the neuronal cell differentiation, a process which has many similarities with the redifferentiation phase from the Y_0 state to a PDA neuron (115). Finally, JMJD3 and SET1 dependent methylations on H3K27me3/me2 and H3K4me3, respectively were not only necessary for an optimal transdifferentiation event, they are also critical to provide protection against all the variations that the nematode can face living in natural environment: each single histone modifying activity proved critical to ensure the robustness of this cell conversion event in non-optimal conditions such as mild heat stress, ultraviolet radiation, oxidative stress (paraquat) and cytotoxicity (DMSO) (113). The result showed that mutants for each transcription factor were completely incapable of initiating transdifferentiation, however, mutants for each single histone modifier showed a very lowly penetrant defect. Altogether, these data suggest that transcription factors are key drivers of this transdifferentiation event when histone modifications are acting as a regulators to promote invariance (110).

1.7.2.3 Other cell reprogramming events in *C. elegans*.

1.7.2.3.1 AMso-to-MCM.

In addition to Y-to-PDA, other putative transdifferentiation events have been reported in the nematode. Another putative transdifferentiation event happens in the MCM neurons (108). Those neurons appear during the L3 larval stage in males and are fully differentiated when the worms reach the L4 stage, being necessary for sex learning behavior during the sexual maturation of the male worm. Ablation experiments and screenings using S-phase reporters showed that the origin of the MCMs cells comes from the asymmetric division of support cells of the head sensilla called AMso cells (amphid sockets cells) that form the junction with the hypodermis on both sides of the worm (108)(116)(117). This cell division does not occur in hermaphrodites but is present in males. In males, AMso cell divide in two cells, one of them remains as an AMso cell but the other becomes an MCM neuron. In that case, and contrary to Y-to-PDA, this is a prospective transdifferentiation event that requires asymmetric cell division. This fact opened the door to new questions and hypothesis to understand how shared and common mechanism in *C. elegans* transdifferentiation events are regulated. However, the molecular mechanisms that are underlying the AMso-to-MCM cell conversion still remain unknown.

1.7.2.3.2 G1/G2-to-neuron and K-to-DVB.

The G1 is an excretory cell pore which born during the embryogenesis and starts to wrap forming a unicellular tube establishing junctions to perform its excretory function. However, after embryogenesis, during the L1 stage, G1 undergoes delamination and cellular division to generate two neurons (92)(93)(118) losing it excretory function, taken over by the G2 excretory pore cell. Later, during the L2 larval stage, G2 will also undergo cell division, and the posterior daughter G2.p will perform the excretory pore function (93)(119). The specialization of those cells, their change of function and the expression of different markers are suggesting that G1, G2 and G2.p can change their fate from an epithelial fate to a neuronal one after cell division in a similar fashion to the previously described transdifferentiation events that request cellular division.

Additionally, another putative transdifferentiation event which requires division and happens in the rectum of the worm as Y-to-PDA is the K-to-DVB transdifferentiation. K cell, which is part of the rectum, divides during the beginning of L2 larval stage producing two different daughter cells, one of them will remain with a rectal identity and function and the other daughter cell will change it fate for a neuronal identity becoming neuron called DVB prospectively under a transdifferentiation process (92) (Jarriault Lab, unpublished).

Considering that only Y-to-PDA had been characterized as a natural transdifferentiation event in comparison with AMso-to-MCM, G1/G2-to-neuron and K-to-DVB and knowing that transcription factors are playing an important role during reprogramming events, we considered that the key to understand Y-to-PDA would be in the transcriptional dynamics during all the Y-to-PDA transdifferentiation. For that reason, in order to elucidate the different transcriptional profiles of this fascinating reprogramming event, we planned a strategy to carry out *in vivo* gene expression footprints in every step of Y-to-PDA natural transdifferentiation to identify the active/inactive genes during Y-to-PDA. This methodology is called DamID.

1.8 DamID: DNA adenine methyltransferase identification.

During the last years, gene expression, genome architecture and organization have been a vibrant subject of study for molecular biologists. Many tools were developed in the last decades to study those three subjects and determine the molecular basis of many biological processes.

Initially in our project, we wanted to check how many and which genes were transcribing in each step of the Y-to-PDA from Y birth during embryogenesis, during transdifferentiation at the L1 to the L3 stage until the PDA motoneuron fate. A priori, the simplest way to proceed

had been carry out in RNA-Seq experiments on Y, Y_0 and PDA isolating cells by Fluorescence Activated Cell Sorting (FACS). However, this approach presented several problems. If Y cell sorting as an epithelial cell is possible, the sorting of Y_0 and specially PDA, a motoneuron with a long axon was not possible using this methodology. In addition, a dual fluorescence *C. elegans* strain was created in Jarriault laboratory allowing us to carry out the sorting in Y but no marker or marker combination specific for Y_0 or PDA was (and still is) available. Another suitable methodology could be the chromatin immunoprecipitation method (120)(121). Chromatin immunoprecipitation (ChIP-Seq) is a methodology to study genome-protein interactions by reversible crosslinking using formaldehyde. However, this method presented the same kind of problem: there is one cell per animal, and we do not have the possibility to isolate Y_0 and PDA cell populations. Because of all mentioned limitations, work in *in vivo* conditions became a must.

A powerful method to study *in vivo* genome-protein interactions is the DNA adenine methyltransferase identification also known as DamID. This methodology was described in 2000 by Bas Van Steensel and Steven Henikoff (122) as an alternative to avoid crosslinking and chromatin immunoprecipitation (ChIP). DamID is based on the fusion between DNA adenine methyltransferase (Dam) from *Escherichia coli* and a protein of interest that either binds chromatin directly or indirectly as part of a protein complex. When the fusion protein is expressed *in vivo*, *Dam* will methylate adenines in the immediate vicinity of the binding sites of the chromatin-associated protein. Recently it has been showed that Dam methylates only GATCs which are closer than approximately 300 base pairs from the binding site of the fusion protein (123). Dam recognizes GATC motifs and creates a unique adenine methylation mark (G^mATC) at the N6 position of the adenine, a methylation normally rarely found in the DNA of eukaryotes. Particularly in *C. elegans*, adenines from GATC motifs are always unmethylated (124)(125). The only endogenous methylation identified in the *C. elegans* is present in GAGG and AGAA motifs (76) which do not interfere with the DamID technique. Genomic methylated DNA by Dam can be extracted from animals expressing Dam fused to the protein of interest and from control animals, which are expressing a fusion of Dam to GFP. This control experiment is later used for normalization of the Dam fusion to the protein of interest (126). Methylated sites are then cut with a restriction enzyme, *DpnI* which recognizes specifically G^mATC. *DpnI* cut sites are then ligated to an adapter before cutting non-methylated GATC sites with *DpnII* (122). Later, PCR using the adapter as a primer binding site amplifies fragments between G^mATC sites. Sequencing of amplicons reveals the ratio of methylation between the assay and the control and their localization on the genome (122)(126)(127).

DamID is highly versatile, allowing researchers to study many nuclear proteins of interest, whether it is a nuclear lamins to study perinuclear chromatin organization, a transcription factor to elucidate its binding sites or RNA polymerase II to study gene expression activity, or histone deacetylases or other chromatin remodelers to study chromatin state or non-DNA binding proteins which do not interact directly with DNA but can impact proximal loci (128). We have sought to take advantage of several of these amenities in my project.

1.9 The importance of the transcription in natural transdifferentiation.

After choosing DamID as a method to study gene expression in the different stages of Y-to-PDA, we had to decide what could be the proper protein of interest to profile actively transcribing genes. One option is to use a component of the transcription preinitiation complex.

The transcription pre-initiation complex (Figure 11) is a complex of proteins needed to start the transcription. The pre-initiation complex is present in eukaryotes and archaea (129)(130). The complex has different functions, it is helping with RNA Polymerase II attachment on the promoter region and subsequent DNA melting and in the positioning and finding of the transcriptional start site (130). The pre-initiation complex includes RNA polymerase II and the six general transcription factors (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TIIF, TFIIH.

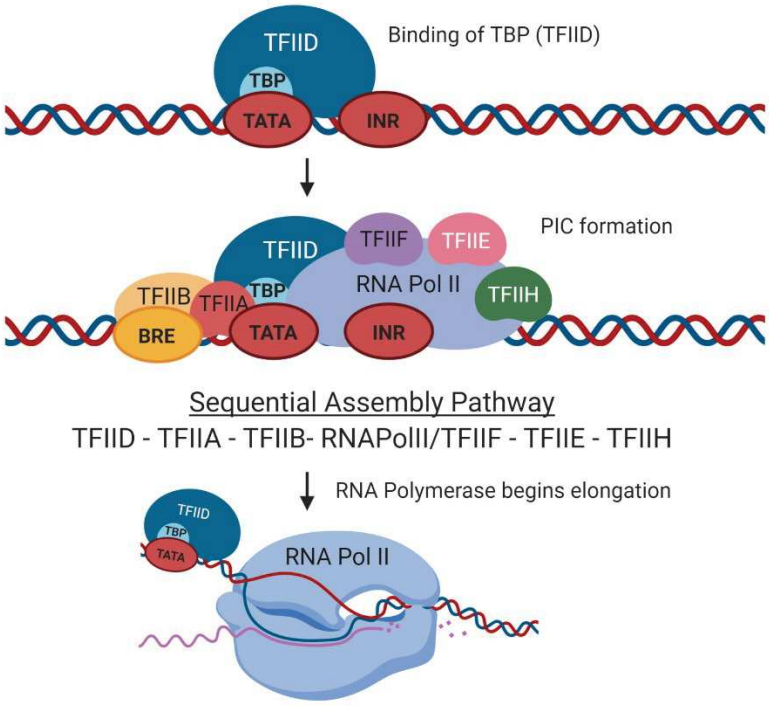


Figure 10. Transcription pre-initiation complex formation. TBP recognizes the TATA-box starting the formation of the preinitiation complex. The rest of the complex is formed in a sequential assembly pathway allowing the RNA Polymerase II starts elongation and gene transcription.

TFIID is the first transcription factor which binds the promoter during early formation of the transcription pre-initiation complex. Initially, TFIID binds the TATA box through its subunit TBP (TATA binding protein) serving as a scaffold for the assembly of the rest of components of the transcription pre-initiation complex (131). The next transcription factor which binds is TFIIA. TFIIA interacts with TBP subunit of the TFIID aiding TBP subunit to bind the TATA-box. This TFIIA-TBP(TFIID) interaction facilitates and stabilizes the pre-initiation complex (129)(131). Following the TFIID-TFIIA, TFIIB is now the next transcription factor to be recruited by the complex. TFIIB recognizes the B recognition element (BRE) on the promoter region and establish a protein-protein interaction with the TBP subunit of the TFIID and later with the RNA polymerase II (132)(133). Once the TFIID-TFIIA-TFIIB complex is formed, RNA polymerase II and TFIIF will be recruited. TFIIF is bound to RNA polymerase II and contacts with TFIIB and TBP(TFIID) stabilizing the TFIID-TFIIA-TFIIB complex (134). To finish, TFIIE and TFIIH are recruited finishing the pre-initiation complex formation. TFIIH will phosphorylate RNA Polymerase II by kinase activity of one of its subunits, then, RNA Polymerase II is released from the promoter, initiating transcription. During the elongation only TFIID remains bound to the core promoter supporting the re-initiation of the process (135). Considering this information, TFIID could be a potential candidate to be protein of interest to be fused with Dam. However, it is important to note that as was mentioned, TFIID remains bound to the core promoter. The total number of GATC motifs in the *C. elegans* is 269’036. The median and the mean distances between motifs are 206 and 370 base pairs respectively, having three exceptions: there are two GATC motifs separated by 3.3 Kb in chromosome II, by 9.9 Kb in chromosome IV and by 32.1 Kb in chromosome X (126). For this reason, with *dam::tbp-1*, Dam will only methylate the neighboring GATC motifs leaving long genes unmethylated because of the distance between the TBP-1 binding site and the rest of the gene. Because TATA boxes are only present in 6% of the *C. elegans* genes (136), we concluded that TBP-1 is not a good candidate protein to be fused with Dam. Finally, we decided to use one of the subunits of the RNA polymerase II, an enzyme which is formed by 12 different subunits in *C. elegans* (Table 1). We choose the RNA polymerase II subunit A *ama-1*, which is largely expressed in the nucleus of all the cells of the body (137)(138). Additionally, ChIP-seq data is available at the worm wide level that could be used as a control to validate DamID-Seq vs ChIP-Seq RNA polymerase II profile using this subunit (139). Another RNA polymerase subunit used in that PhD work was *rpb-6,* which is the subunit F of the RNA polymerase II. We considered to use this subunit as a second option after *ama-1* because its little size made the subunit a good candidate to be integrated in the worm genome by MosSCI. Additionally, as a RNA polymerase II subunit, we can expect an active gene profiling activity when is fused with Dam.

C. elegans gene	RNA polymerase*	Size* (base pairs)
ama-1	II	10370
rpb-2	II	4838
rpb-3	II	1368
rpb-4	II	3438
rpb-5	I, II, III	948
rpb-6	I, II, III	590
rpb-7	II	1729
rpb-8	I, II, III	905
rpb-9	II	1133
rpb-10	I, II, III	394
rpb-11	II	540
rpb-12	I, II, III	533

Table 1. Genes encoding RNA polymerase subunits. *Information related to *C. elegans* only. Source: Wormbase

1.10 Temporal controlled expression of *dam::fusion* proteins.

Since the beginning of the embryogenesis, *dam::ama-1* will be marking active genes during all of embryonic development until the L3 larval, the stage in the life cycle in which Y-to-PDA is finished, continuing until the end of the life of the worm. For this reason a gene control expression of *dam::fusion* proteins in time and space must be achieved if we want to map the active and the inactive genes during Y-to-PDA transdifferentiation. Recently, the Dernburg laboratory developed a wormified version of the Auxin Inducible Degradation system (AID) of *Arabidopsis thaliana*. This system consists of a tag added to the protein of interest with a sequence called degron; in presence of auxin, this degron will be ubiquitinated by the F-box TIR1 and degraded in the proteasome (Figure 12). This process is reversible, allowing to change the worms from auxin-containing plates to plates without auxin at the proper moment of development, activating/inactivating the expression of *dam* fusions to protein of interest and getting therefore temporally different transcription profiles.

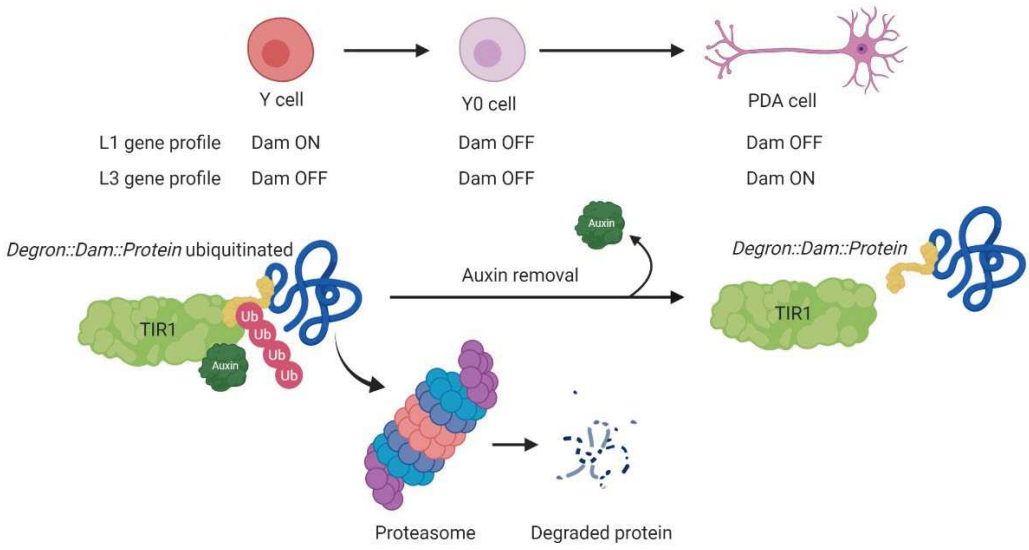


Figure 11. Auxin Inducible Degradation system as a temporal control system. The AID system allows rapid and reversible degradation of target proteins in response to auxin and enables us to generate a reliable time control system of our d*am::fusions* by removing the worms at the proper moment of the development in order to get different expression profiles by DamID.

1.11 Tissue specific expression of *dam::fusion* proteins.

It is important to highlight that the *Dam* fusions cannot be fused to tissue-specific promoters as this would lead to high-level methylation which is toxic. The *dam* fusions need to be expressed at trace levels. We thus use the heat-shock promoter leakiness, in absence of heat shock, to ensure expression of trace amount of the *dam::fusion* protein of interest. Since the heat shock promoter is expressed ubiquitously, to achieve spatial control, the use of a tissue specific system is compulsory. However, the expression in Y cell is not easy as the specific expression in the Y cell cannot be driven only by one single promoter. For that reason, we developed different approaches to express *dam* in the desired cell type. In those approaches, we always combined two different cell specific promoters overlapping only in Y cell creating a double gate system. The promoters envisioned for developmental tissue specific expression during Y cell transdifferentiation are two: *hlh-16* and *egl-5.* Those two promoters are expressed in Y since its birth allowing us to study the development since the very beginning of Y cell development.

1.11.1 Double gate recombination cascade for Y/PDA-specific gene expression.

The first approach we had planned was a double gate system combining two different site-specific recombinases methodologies. The first part of the cascade is going to be under the control of a specific rectal cell promoter expressing a FLP recombinase of the FLP/FRT recombination system from *Saccharomyces cerevisiae* recently adapted to *C. elegans* (140). The second part of the cascade is going to be under the control of a neuronal promoter

expressing CRE recombinase from the recombination system *Cre/lox* from bacteriophage P1 adapted in *C. elegans* (141). The combination of those two different promoters will allow us to express both recombinases only in the Y cell, *i.e.* the *dam::fusion* proteins will only be expressed in this cell *in vivo* conditions establishing a double gate system never tested before in *C. elegans* (Figure 13).

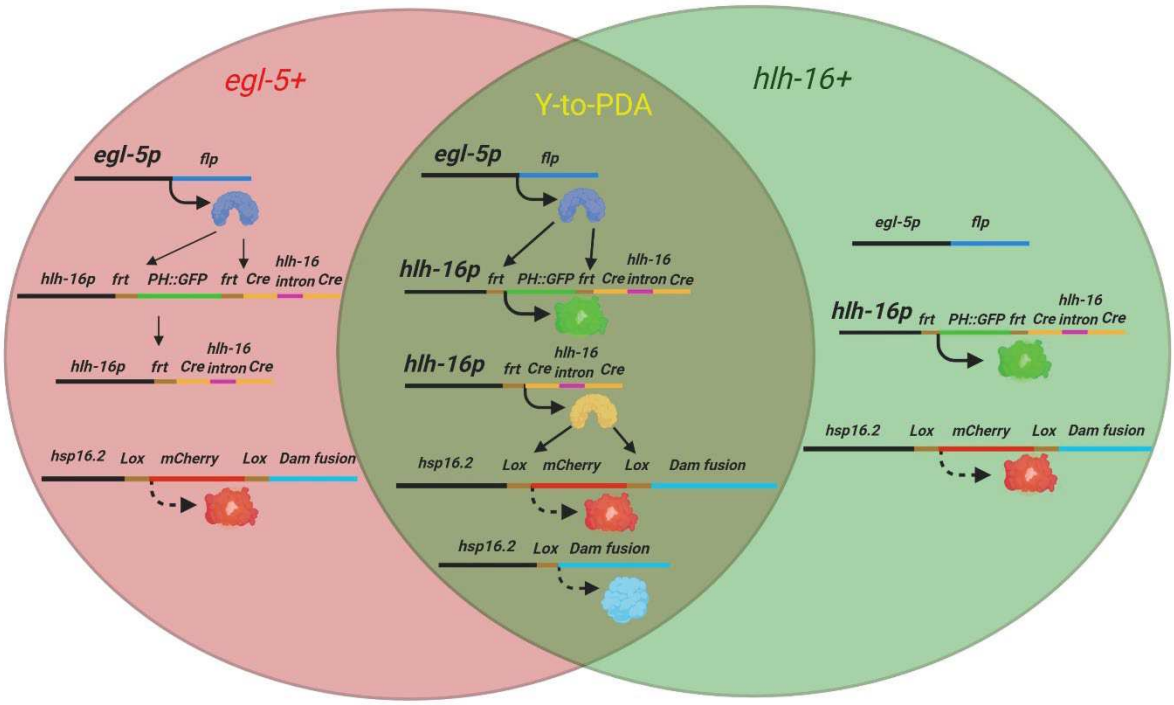


Figure 12. Tissue specificity expression in Y/PDA cells requires the action of two different promoters to create a cascade with two recombination systems. We combined two different promoters, one which is active in rectal cells during all the life cycle of *Caenorhabditis elegans;* and another which only active from the embryogenesis to the L1 larval state in Y and in some head cells. The combination of both promoters will allow us to have tissue specificity (Y only) creating a recombination cascade with two recombination systems.

1.11.2 Double gate split cGAL- CRE recombination system.

The second tissue specific approach was a combination of the split cGAL system from *Drosophila melanogaster* recently adapted in *C. elegans* (142) with the *Cre/lox* recombination system. The split cGAL system is based on the expression of two halves of the GAL4 transcriptional activator under tissue-specific promoters. One part is the DNA binding domain (DBD) and the other part is the transcriptional activation domain being assembled by inteins during gene expression. Inteins are protein domains that can establish covalent bonds between proteins. The DNA binding domain is co-expressed with the N-terminal part of the intein and the transcription activation domain with the C-terminal part. In cells in which both parts of the cGAL system are expressed, the cGAL transcription factor is reconstituted by the intein binding activity. For us, the overlap between the expression patterns of these promoters is uniquely in the Y cell (as for the double recombination

cascade), and thus an active GAL4 is only present in the Y cell. The CRE recombinase is then placed under the transcriptional control of an upstream activation sequence (UAS sequence) targeted by cGAL, and uniquely expressed in Y.

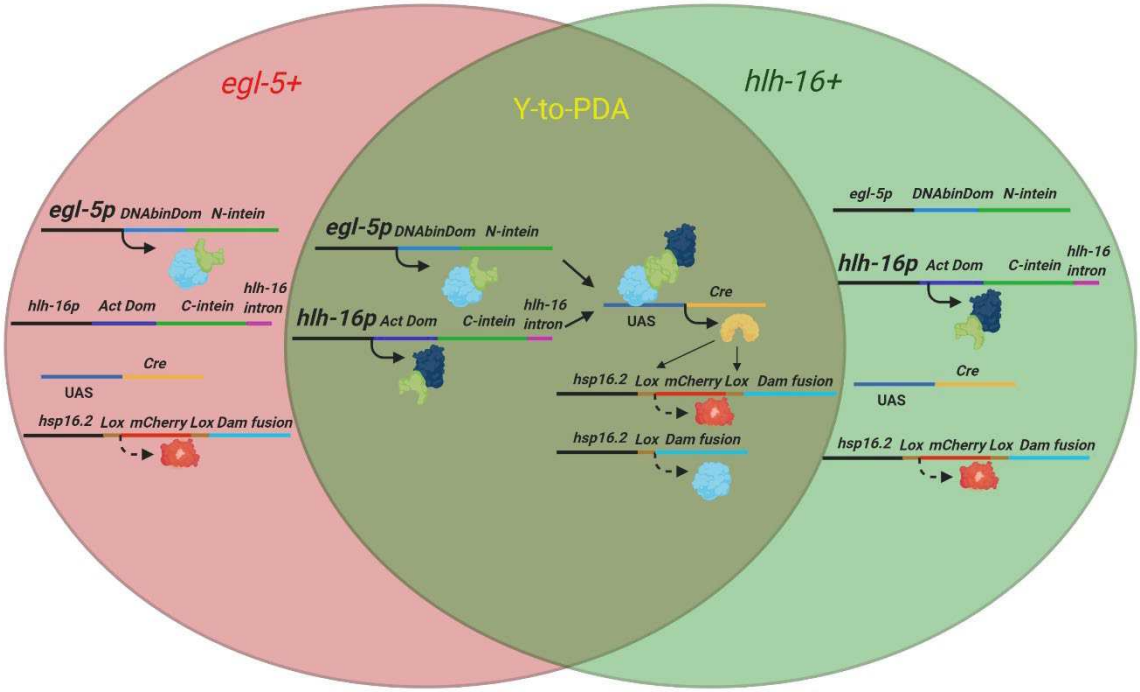


Figure 13. Split cGAL system using a split intein to create a refined spatiotemporal control system in Y/PDA. To achieve tissue (Y only) specificity, we are going to use a dual promoter double gate strategy combined with *Cre/lox* recombination system.

1.11.3 Double gate to create different fluorescent cell populations able for cell sorting.

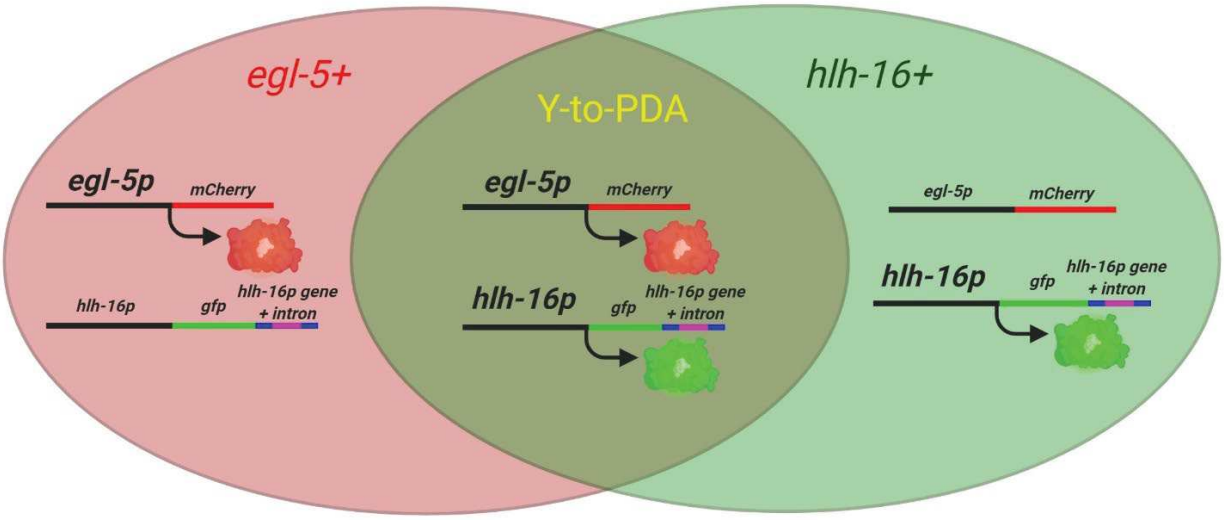


Figure 14. Double fluorescent gate using different fluorescent proteins which converged in Y.

The third approach developed in our laboratory such as the two previous strategies, was a combination of promoters that converged, but only expressing red and green fluorophores. Using this strategy, we are going to have three different cell fluorescent populations, green and red cells in the cell population with only one promoter expressing its fluorescent protein and Y cells with a yellow fluorescent phenotype consequence of the combination of both active promoters expressing *mCherry* and *gfp*. This strategy has a limitation: Y_0 and PDA are not able to be sorted because the promoter combination is not the same for Y, Y_0 and PDA, and, additionally, the long axon present in PDA make the cell sorting a pointless strategy. Because of that inconveniences, only Y cell profile could be determined.

Aim of the project

I started by setting up a DamID-based approach to uncover the genes actively expressed during the transdifferentiation from a rectal into a neuronal cell identity and further characterize the function of these genes in maintenance and conversion of cell fates was the main objective of this PhD work. I tested several systems to achieve tissue specific control systems and a time control system for this. Finally, I chose to used Dam-RNA Polymerase subunits fusions together with fluorescence markers establishing a fluorescent "double gate" to isolate the cells by fluorescence activated cell sorting (FACS) and carry out DamID on achieving the desired cell specificity. Additionally, we have taken advantage of the DamID system to gain more insights into the transdifferentiation event, by seeking to obtain in parallel an image of the nuclear organization of our cell of interest. This was achieved by fusing Dam with different DNA binding proteins such as LMN-1 (nuclear lamina protein interacting with heterochromatin) besides RNA Polymerase II to have an approach of how other regions of the genome are organized. This approach provided. The different steps of my project were the following ones:

1- Setting up a robust, constitutive and cell type *dam::RNAPolII* profiling active genes in *Caenorhabditis elegans*.

2- Establishing a cell sorting system in embryos using single or dual fluorescence markers to get unique cells for DamID.

3- Characterizing the genes differentially expressed in multiple tissues and in Y during the beginning of the natural transdifferentiation Y-to-PDA system.

2. Material and Methods

2.1 Worm strains.

Strain name	Meister Lab code	Genotype
BN515	PMW533	*bqSi515[p1484(unc-119(+) Phsp16.41::dam::ama-1)] II*
BN196	PMW124	*bqSi196[pBN67(unc-119(+) hsp16.41p::gfp::myc::dam)] II*
BN195	PMW123	*bqSi195[pBN65(unc-119(+) hsp16.41p::dam::myc::lmn-1)] II*
IS3675	PMW754	*fpSi2[hsp16.20p::lox::degron::Dam::rpb-6::unc54 3'UTR] II*
IS3467	PMW732	*fpIs76[egl-5p (6.5 kb) delta pes-10mcherry; Bluescript] ; fpIs92[hlh-16::pGFP::hlh-16 (p+E1 +I1 complete)::hlh-163'UTR;pFR4] ; bqSi196[pBN67(unc-119(+) hsp16.41p::gfp::myc::dam)] II*
IS3468	PMW733	*fpIs76[egl-5p (6.5 kb) delta pes-10mcherry;Bluescript] ; fpIs92[hlh-16::pGFP::hlh-16 (E1 + I1 complete)::hlh-163'UTR;pFR4] ; bqSi515[p1484(unc-119(+) hsp16.41::dam::ama-1)] II; unc-119(ed9) III*
IS3564	PMW746	*fpIs76[egl-5p (6.5 kb) delta pes-10mcherry; Bluescript] ; fpIs92[hlh-16::gfp::hlh-16 (E1 + I1 complete)::hlh-163'UTR;pFR4] ; bqSi195[pBN65 [unc-119 (+) hsp16.41p::dam::lmn-1] II*
IS3685	PMW757	*fpIs76[egl-5p (6.5 kb) delta pes-10mcherry;Bluescript] I ; fpIs92[hlh-16::gfp::hlh-16 (E1 + I1 complete)::hlh-163'UTR;pFR4] ; fpSi2[unc-119(+) hsp::lox::degron::dam::rpb-6::unc-54] II*
IS3654	PMW756	*bqSi195[pBN65 (unc-119(+) hsp16.41p::dam::myc::lmn-1)] II ; irIs25[elt-2::nls::gfp::lacz + rol-6(su1006)] V*
IS3687	PMW763	*fpSi2[unc-119 (+) hsp::lox::degron::dam::rpb-6::unc-54] II ; irIs25[elt-2::nls::gfp::lacZ + rol-6(su1006)] V*
IS3653	PMW755	*bqSi196[pBN67(unc-119(+) hsp16.41p::gfp::myc::dam)] II ; irIs25[elt-2::NLS::GFP::lacZ + rol-6(su1006)] V*
IS3684	PMW762	*ccIs4251[(pSAK2) myo-3p::gfp::Lacz::nls + (pSAK4) myo-3p::mitochondrial gfp + dpy-20(+)] I ; fpSi2[unc-119 (+) hsp::lox::degron::dam::rpb-6::unc-54] II*
IS3651	PMW760	*ccIs4251[(pSAK2) myo-3p::GFP::LacZ::NLS + (pSAK4) myo-3p::mitochondrial GFP + dpy-20(+)] I ; bqSi196[pBN67(unc-119(+) hsp16.41p::gfp::myc::dam)] II*
IS3652	PMW761	*ccIs4251[(pSAK2) myo-3p::GFP::LacZ::NLS + (pSAK4) myo-3p::mitochondrial GFP + dpy-20(+)] I ; bqSi195[pBN65(unc-119(+) hsp16.41p:: dam::myc::lmn-1)] II*
DAG958	PMW739	*domSi958[hsp16.20p::lox::mCherry::lox::degron::dam::rpb-6::unc54 3'UTR] II*
IS3568	PMW747	*heSi142[elt-2p::elg-13nls::Cre::tbb-2 3'UTR] X; domSi958[hsp16.20p::lox::mCherry::lox::degron::dam::rpb-6::unc54 3]'UTR*
PMW748	PMW748	*heSi148[myo-3p::elg-13nls::Cre::tbb-2 3'UTR] X ; domSi958[hsp16.20p::lox::mCherry::lox::degron::dam::rpb-6::unc54 3]'UTR*
PMW636	PMW636	Ex-array with plasmids PM423, PM414 and PM417 injected in N2
PMW682	PMW682	Ex-array #444 + #445 injected in wPM 641
PMW678	PMW678	*ubsSi40[unc-119(+) hsp16.41::lox:mcherry::lox::degron::gfp::dam::unc-54 3'UTR] IV* *eSi57 [eft-3p::TIR1::mRuby::unc-54 3'UTR + Cbr-unc-119(+)] II* *heSi142[elt-2p::elg-13nls::Cre::tbb-2 3'UTR] @ ttTi14042 X;*
PMW681	PMW681	*ubsSi40[unc-119(+) hsp16.41::lox:mcherry::lox::degron::gfp::dam::unc-54 3'UTR] IV* *eSi57 [eft-3p::TIR1::mRuby::unc-54 3'UTR + Cbr-unc-119(+)] II* *heSi148[myo-3p::elg-13nls::Cre::tbb-2 3'UTR] ttTi14042 X;*
PMW687	PMW687	*ubsSi40[unc-119(+) hsp16.41::lox:mcherry::lox::degron::gfp::dam::unc-54 3'UTR] IV ; ieSi57[eft-3p::TIR1::mRuby::unc-54 3'UTR unc-119(+)] II ; heSi148[hlh-8::elg-13nls::Cre::tbb-2 3'UTR]* X
IS2922	PMW530	*fpIs76[egl-5p (6.5 kb) delta pes-10mcherry; Bluescript] ; fpIs92[hlh-16::pGFP::hlh-16 (E1 + I1 complete)::hlh-163'UTR;pFR4] ; oxTi177[unc-119(+)*
MS438	PMW766	*irIs25[elt-2p::nls::gfp::lacz+rol-6(su1006)]V*
PD4251	PMW765	*ccIs4251[(pSAK2)myo-3p::gfp::lacz::nls+(pSAK4)myo-3p::mitochondrial gfp + dpy-20(+)]I*

PMW653	PMW653	*ieSi57[eft-3p::TIR1::mRuby::unc-54 3'UTR+Cbr.unc-119(+)] II ; ubsSi39[+Cbr.unc-119(+)hsp16.41::degron::gfp::dam::unc-54 3'UTR] IV*
PMW654	PMW654	*ieSi57[eft-3p::TIR1::mruby::unc-54 3'UTR unc-119(+)] II ; ubsSi40[unc-119(+) hsp16.41::lox:mcherry::lox::degron::gfp::dam::unc-54 3'UTR] IV*
PMW640	PMW640	*ubsSi39[+Cbr.unc-119(+) hsp16.41::degron::gfp::dam::unc-54 3'UTR] IV*
PMW641	PMW641	*ubsSi40[unc-119(+) hsp16.41::lox:mcherry::lox::degron::gfp::dam::unc-54 3'UTR] IV*
CA1200	PMW509	*ieSi57[eft-3p::TIR1::mRuby::unc-54 3'UTR+Cbr.unc-119(+)] II*
CA1202	PMW510	*ieSi57[eft-3p::TIR1::mRuby::unc-54 3'UTR+Cbr.unc-119(+)] II; ieSi58[eft-3p::degron::GFP::unc-54 3'UTR+Cbr.unc-119(+)] IV*
SV1438	PMW375	*heSi141[hlh-8p::elg-13nls::Cre::tbb-2 3'UTR] X*
SV1439	PMW376	*heSi142[elt-2p::elg-13nls::Cre::tbb-2 3'UTR] X*
SV1461	PMW377	*heSi148[myo-3p::elg-13nls::Cre::tbb-2 3'UTR] X*
SV1361	PMW677	*heIs105[rps-27::loxP::NLS::mCherry::let858 UTR::loxP::NLS::GFP::let-858 UTR unc-119(+)] IV*
N2	PMW037	Wild-type Bristol N2

Table 2. List of worm strains used and created during this PhD work.

2.2 Plasmid collection.

Plasmid name	Meister number	Plasmid information	Backbone
#371	#371	hsp16.41::degron::GFP::Dam::unc-54 3'UTR	pCFJ151
#372	#372	hsp16.41::degron::Dam::ama-1::unc-54 3'UTR	pCFJ151
#423	#423	egl-5::FLP-CAI::F2A::2NLS::mKate2::T2A::2NLS::mCherry::unc-54-3'UTR	pCFJ150
#417	#417	hsp 16.2::loxP::mCherry::loxP::Degron:: Dam::ama-1::unc-54 3' UTR	pCFJ151
#418	#418	hsp 16.2::loxP::mCherry::loxP::Degron:: GFP::Dam::unc-54 3' UTR	pCFJ151
#145	#145	Unc-54 3' UTR BP entry clone	pDONRP2rP3
#361	#361	egl-5 BP entry clone	pDONR4
#424	#424	FLP-CAI::F2A::2NLS::mKate::T2A::2NLS::mCherry BP entry clone	pDONR221
#434	#434	hsp16.20::LoxP::mCherry::LoxP::degron::Dam::rbp-2::unc-54	pCFJ151
#444	#444	*egl-5::FLP-CAI:: unc-54-3'UTR*	pCFJ151
#445	#445	hlh16::FRP::PHGFP::FRP::Creexon1::intron1hlh16::Creexon2::unc-54-3'UTR	pCFJ151
#414	#414	hlh16::FRP::PHGFP::FRP::Creexon1::intron1hlh16::Creexon2::T2A::2NLS::GFPw.introns::F2A::2NLS::GFPw.introns::unc-54-3'UTR	pCFJ151

pHW394	#463	(15xUAS::GFP::let-858 3'UTR)	
pHW531	#464	pHW531 (Peft-3::cGAL-C::AD::let-858 3' UTR)	
pHW533	#465	pHw533 (Peft-3::DBD::cGAL-N::let-858 3' UTR)	
#466	#466	15xUAS::Cre::let-858 3'UTR	pCFJ151
#467	#467	hlh-16p::cGAL-C::AD::let-858 3'UTR	pCFJ151
#468	#468	egl-5p::DBD::cGAL-N::let-858 3'UTR	pCFJ151
#457	#457	hsp.16.41::Degron::Dam::Rpb-2::unc-54::3'UTR	pCFJ151
pSJ4646	#503	hsp16.2::lox::mCherry::lox::degron::Dam::rpb-6::unc54 3'UTR	pCFJ151
#509	#509	hsp16.2::degron::Dam::rpb-6::unc54 3'UTR	pCFJ151
	#510	pie-1p::Cre	pCFJ355

Table 3. List of plasmids used and created during this PhD work.

2.3 Growth media and worm culture.

2.3.1 General worm culture.

Worms were grown in Petri dishes (60 mm x 15 mm) on solid Nematode Growth medium (NGM): 3 g/L of NaCl, 2.5 g/1L of peptone, 20 g/1L of agar, 1 Ml/1L of cholesterol (5 mg/ml in ethanol) 1 mL/1L of 1 M $CaCL_2$, 1mL/1L of 1M $MgSO_4$ and 25 mL/1L of 1 M (pH 6.0) KPO_4. The NGM plates were seeded with OP50 (*Escherichia Coli*) for standard worm maintenance (94). For experiments which require DNA isolation from worms expressing *dam::fusion* proteins, we used NGM plates seeded with GM48, a *dam* negative *E. Coli* strain to avoid unspecific Dam signal from bacteria. All the worms grown up at 20-22ºC to avoid the risk of heat shock. The plates were stored no longer than one month at 4ºC.

2.3.2 Auxin plates and media.

For auxin treatment in solid culture, we used auxin indole-3-acetic acid (IAA) purchased from Alfa Aesar (#A10556) (143). A 400 mM stock solution was prepared adding auxin powder into 100% ethanol. The stock solution was stored up to one month at 4ºC in darkness. For auxin plates preparation, stock solution was diluted into the Nematode Growth Medium (NGM) agar when was cooled to 50ºC just before pouring plates. In high concentration of auxin, the bacterial growth can be delayed. In those cases, we left the plates during 48 hours in darkness to produce a good bacterial growth. Worms were in a solution of M9 (3 g KH_2PO_4, 6 g Na_2HPO_4, 5 g NaCl, 1 ml 1 M $MgSO_4$, H_2O to 1 L) + the proper amount auxin added according to the experiment during 14 hours to produce a synchronized culture of embryos. During the time between bleaching and hatching, the tubes containing embryos

were on a nutator and were kept in darkness. To avoid risk of bacterial or fungal contamination on auxin plates, we pour the plates inside a laminar flow hood (143).

2.3.3 Cell sorting plates.

Worms were grown up in Petri plates (150 mm x 15 mm) on solid medium: 1.2 g/1l NaCl, 20g/1l bacto-peptone, 25g/1l bacto-agar, 1ml cholesterol 5mg/ml, 1ml/1l 1M $MgSO_4$, 25ml KH_2PO_4 pH=6 and 1ml streptomycin 50 mg/ml. The bacteria used to feed the worms was HB101 because OP50 and GM48 did not grow properly in presence of streptomycin.

2.4 Worm general traits.

2.4.1 Synchronized worm cultures.

Worms were grown in Petri plates (60 mm) on solid Nematode Growth medium (NGM): 3 g/l of NaCl, 2.5 g/l of peptone, 20 g/l of agar, 1 ml/l of cholesterol (5 mg/ml in ethanol) 1 ml/l of 1 M $CaCL_2$, 1ml/l of 1M $MgSO_4$ and 25 ml/l of 1 M (pH 6.0) KPO_4. The plates were washed with M9 (3 g KH_2PO_4, 6 g Na_2HPO_4, 5 g NaCl, 1 ml 1 M $MgSO_4$, H_2O to 1 L) and the worms collected in 15 ml Falcon tubes. After collect the worms in M9 buffer (3 g KH_2PO_4, 6 g Na_2HPO_4, 5 g NaCl, 1 ml 1 M $MgSO_4$, H_2O to 1 L), we added 3 ml of bleaching solution (5ml NaOH 1M, 5 ml Bleach and 40 ml water) and incubated the worms into the bleaching solution during 5 minutes on a nutator. Then, the sample was centrifugated to pellet down the animals and re-bleached with other 3 minutes with bleaching solution. After the 2^{nd} incubation time, we added 10 ml of M9 buffer to neutralize the bleach. We carried out up to 4 washes. Eggs were incubated overnight on a shaker and 16 hours later we got a population of starved and synchronized L1 larvae ready to be platted.

2.4.2 Freezing worms.

Two large plates (150 mm x 15 mm) with an abundant population of not starved L1 larvae were prepared and washed with M9 buffer (3 g KH_2PO_4, 6 g Na_2HPO_4, 5 g NaCl, 1 ml 1 M $MgSO_4$, H_2O to 1 L) to collect the L1 worms. After centrifugation to pellet down the L1s, 5 cryotubes were prepared. We added 6 ml of freezing solution (20 ml/200ml 1 M NaCl, 10 ml/200mL 1M KH_2PO_4 pH=6, 60 ml/200mL 100% glycerol, sterile) and mixed the solution with the L1 larvae. We added 1 ml per cryotube. One cryotube will be used to test efficiency of the freezing after two weeks at -80ºC, the other 5 are divided in 2 tubes for -80ºC stock, 2 tubes for liquid nitrogen stock and the one that remains for my personal stock in case of need.

2.4.3 Heat shock experiments.

Plates with a synchronized population of worms in the desired developmental stage of the life cycle were sealed with parafilm and introduced under the water in a water bath at 30ºC during 30 minutes. After 30 minutes, the plates were at room temperature during at least 3 hours (recovery time). After the recovery time, the expression of the genes under the control of heat shock promoter should be visible at the scope (for fluorescence genes) or detectable by PCR (for expression of recombinant proteins). Subsequent heat shock could be carried out to increase the expression of the protein.

2.4.4 Crossing worms.

For animal crossing, male induction was necessary. Males can be induced putting 20 worms in late L3/early L4 stage per plate (60 mm x 15 mm). We used to use 5 plates with 20 worms to have guarantees of male generation (100 worms in total). The plates were place into a thermocycler during 6 hours at 30ºC and later at 20ºC. After the induction, worms were placed into new plates (5 worms per plate) and three days after, a male selection was carried out.

Once the males were selected, three hermaphrodites and two males were placed into “honey moon plates” (30 mm x 15 mm plates) during 24 hours. After 24 hours the males were discarded and the hermaphrodites were separated in new plates, one worm per plate (P_0). After three days we singled the F1 population, and, three days later we singled the F2 population. The day after single the F2, after egg laying, we genotyped the F2 worms who laid eggs to check the presence and homozygosis of the desired insertion/mutation.

2.5 Plasmid generation.

2.5.1 Gibson assembly.

All the plasmid used during this PhD project were created using Gibson assembly (E2611L). The desired fragments were amplified by PCR with a tail of at least 30 nucleotides with homology between one fragment and another. For two to three fragments assembly, we used a DNA concentration from 0.02-0.5 pmols. For more than three fragments assembly, we used 0.2-1 pmols. We have always respected the proportion between concentration of each fragment being the insertion three times more concentrated than the vector if the insertions were bigger than 200 base pairs and five times more concentrated when they were smaller than 200 base pairs. The Gibson assembly mix was composed by DNA of interest + water + Gibson Assembly master mix (E2611L). The incubation time of the Gibson assembly was 1-hour incubation at 50ºC being extended to 3 hours in assemblies of more than six fragments.

2.5.2 Plasmid purification.

Purification of PCR fragments and linearized plasmid DNA were carried out using optimized 1.8X Seramag Speed Beads able to reach the efficiency of the AMPure XP beads (https://ethanomics.files.wordpress.com/2012/08/serapure_v2-2.pdf)

2.5.3 Transformation and electrotransformation.

Heat shock transformation was carried out using DH5-alfa (C2987) competent cells and DH10-beta (C3019I) competent cells. The standard protocol for both was 5-30 minutes incubation on ice of cells mixes with the plasmid of interest, 30 seconds 42ºC heat shock followed by 5 minutes on ice again. After transformation we added 500 ul of Luria Bertani solution (10g/1000ml tryptone, 10g/1000ml NaCl, 5g/1000ml yeast extract pH=7) as a recovery medium. The transformed cells were incubated during one hour at 37ºC and plated on selective antibiotic plates.

Electrocompetent cells (60242-1) were used to clone the Gibson assembly products. We diluted the electrocompetent cells with water (10:90) and spitted the volume in two tubes, 50 ul each. Before electroporation, we chilled the electroporation cuvettes on ice. After addition of the Gibson assembly mix, electroporation was carried out (10 uF, 600 Ohms, 1.8 V) using a Biorad Micropulser or an Eppendorf electroporator 2510 with the same parameters. After electroporation, a time constant between 4.5-5.5 was required. Then, we added 500 ml of Luria Bertani solution (10g/1000ml tryptone, 10g/1000ml NaCl, 5g/1000ml yeast extract pH=7) as a recovery medium. The cells were transferred into a new tube and incubated at 37ºC 200 rpm during one hour before be platted on selective antibiotic plates.

To decrease the risk of mutations on the plasmid sequence or unspecific recombination, we also carried out the incubation after the heat shock and the electroporation at 30ºC instead of 37ºC during 1 hour.

2.6 Generation of transgenic animals.

2.6.1 Generation of transgenic strains by MosSCI.

Transgenic animal strains were generated by MosSCI microinjection using high purity DNA. MosSCI is a technology largely used in *C. elegans* research field to create strains with single copy insertions in known locations in *C. elegans* genome. Microinjections in the gonad of a mix of Mos transposase, plasmid of interest (using a specific vector plasmid) and a fluorescent co-marker can be carried out in different strains of worms carrying a *ttTi5065* cassette composed by two *ttTi5065* sites flanking a Mos1 transposon. Those *ttTi5065* sites are homologous with *ttTi5065* sites presents in the plasmid of interest. Once the

microinjection is carried out the Mos transposase will excise the Mos1 transposon creating a DNA break that will be repaired by the cellular machinery incorporating the plasmid of interest into the genome because of the homology of the plasmid to the breakpoint. The vector carrying out the construct of interest to be inserted also include a positive selection marker *unc-119*, that will produce will type phenotype in the target *unc* (uncoordinated) worms correcting the defective phenotype. Also fluorescent markers can be used to identify the presence of real insertions vs extra-chromosomal arrays (Figure 15)(144)(145).

Worms were immobilized on an injection pad made by 20 ul of 2% agarose on a cover slide (25 x 40 mm). Injection needles were crafted pulling out 0.78 x 1.00 x 100mm glass capillaries using a Flaming/Brown micropipette puller model P-97.

Injection mixes contained the plasmid of interest (30-50ng/ul), a Mos transposase (50ng/ul) and the desired fluorescence markers which are necessaries for the microscopy selection.

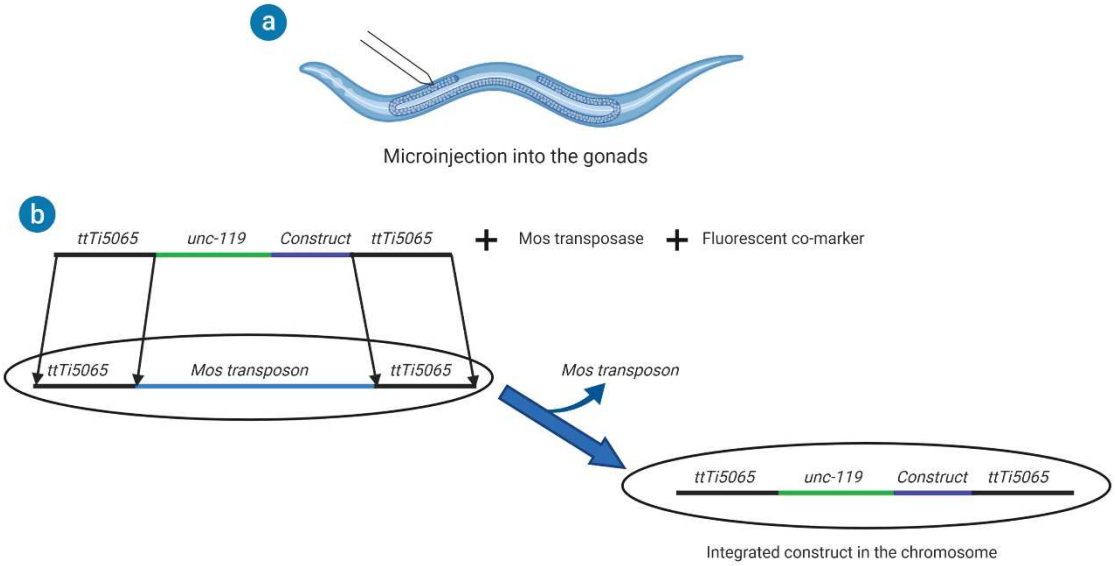


Figure 15. MosSCI microinjection. a) Microinjection in the *C. elegans* gonad. b) Schematic representation of how MosSCI is working: Injection mix containing plasmid of interest + Mos transposase + fluorescent co-marker is injected in worms carrying a Mos1 transposon in the chromosome of interest where the insertion wants to be placed. The Mos transposase will excise the transposon creating a DNA break that will be fix the cell, integrating in the chromosome the plasmid of interest during that process.

2.6.2 *In vivo* recombination of strains generated by MosSCI.

Generation of new transgenic strains from previously existing ones created by MosSCI were carried out. For strains carrying insertions with a lox cassette, an injection mix containing *pie-1p::Cre* (30ng/ul) recombining the germline by lox cassette removal.

2.7 DamID experiments using isolated DNA.

DNA was extracted from worms expressing *Dam::fusion* proteins of interest grown at 20-22ºC on NGM plates fed with GM48. Synchronized L1s larvae were lysed using a Proteinase K solution containing NTE (100mM NaCl, 50mM Tris pH7.4, 20mM EDTA), 500 ug/ml proteinase K and SDS 1% during 60 minutes, 50ºC, 750 rpm. The extraction was a Phenol-Chloroform extraction using Phase lock Gel tubes. Isolated DNA was digested with *DpnI* being the fragments ligated with specific adapters necessary to carry out the 1stDam-PCR (122) followed by amplification of ligates products by PCR as described in (126). Following the classical Illumina protocol, we sequenced our libraries by Illumina HiSeq2500. For Nanopore-Sequencing, we used the reagents and protocols provided by Oxford Nanopore Technologies.

2.8 DamID experiments using sorted cells.

Synchronized hermaphrodites gravid adults worms were grown in big Petri dishes (150 mm x 15 mm) and their eggs were collected into M9 (3 g KH_2PO_4, 6 g Na_2HPO_4, 5 g NaCl, 1 ml 1 M $MgSO_4$, H_2O to 1 L) and incubated at 25ºC during 3 hours and monitored under the dissecting scope in order to check they were in the right stage. Then, eggs were transferred to from M9 to egg buffer (25mM Hepes pH7.3, 118mM NaCl 48mM, KCl 2mM, $Cacl_2$ 2mM $MgCl_2$ in 250 ml) and 500ul of chitinase (1U/ml) was added. The eggs were incubated at room temperature during 1 additional hour. Chitinase treatment was neutralize by addition of 800ul L15 Leibovitz medium (340+ou-5mOsm) and washed one time with L15 after centrifugation 3000 rpm during 5 minutes at 4ºC. Dissociation of embryos in single cell was carried out by pipetting up and down 100 times using a P1000 pipette checking every 100 times if the eggs were dissociated or not under the scope. Additional pipetting are requested in case of undissociated embryos but no more than 150 extra times to avoid high cell mortality. Once the embryos were dissociated, we filtered the cell population to discard debris using a syringe and a 5 µm filter.

Cell sorting was carried out in a BD FACSAria Fusion sorting the cells in sterile PCR tubes containing 1ul of pick buffer (50mM Tris-HCl pH 8.3; 75mM KCl; 3mM MgCl2; 137mM NaCl).

The collected cells were frozen in liquid nitrogen and lysed by addition of 2 ul of lysis buffer (10mM TrisAc, 10 mM MgAc, 50 mM KA, 0.67% Tween20, 0.67% Igepal + 1mg/ml Proteinase K) and incubated during 2 hours at 60ºC before be the Proteinase K inactivated at 95ºC during 15 minutes.

DpnI digestion was carried out on lysates being the fragments ligated with specific adapters necessaries to carry out the Dam-PCR (122). After PCR we used the reagents provided by Oxford Nanopore and the protocols on Nanopore Community Platform (https://nanoporetech.com/community).

2.9 Next generation sequencing methods.

2.9.1 Illumina Sequencing.

We used the statistical package QuasR for our statistical analysis. We used a DamID pipeline developed in Meister laboratory (126) . The DamID parameters were: Bin.len: 100000, errors=1, mapping=T, qc =T, species= "Bsgenome.Celegans.UCSC.ce11", restr.seq= "GATC", adapter.seq= "CGCGGCCGAG". All the reads without the adapter CGCGGCCGAG and the GATC motif were discarded during our analysis. The rest of reads were mapping vs the *C. elegans* genome using RBowtie. After mapping the number of reads per motif were counted followed by a binning of the *C. elegans* genome divided in region of 100 kb. The total number of GATC reads was calculated withing those regions of 100 kb. This total number of reads per 100 kb was normalized using the total number of DamID reads calculating a ratio between Dam::POL/GFP::Dam for each replica and being the mean ration calculated using a $\log_2$fold scale.

2.9.2 Oxford Nanopore Sequencing.

NanoDamID were analyzed following the pipeline developed in Andrea Brand laboratory (146) using a MinION Mk1C device provided by Oxford Nanopore Technologies (ONT) (https://nanoporetech.com/products/minion-mk1c). The pipeline is divided in several steps (Figure 16):

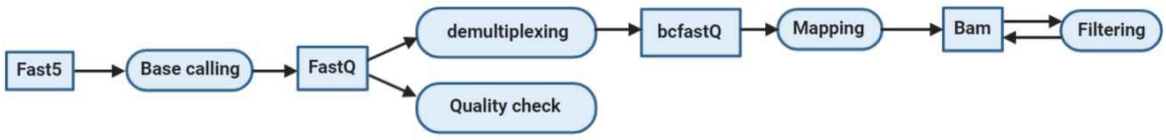


Figure 16. NanoDamID pipeline flowchart.

Basecalling: We obtained the RAW data files in Fast5 format containing only the RAW signal that need to be converted in FastQ format. For this step, we used high accuracy Guppy basecalling (provided by ONT). The FastQ file can be used to carry out a quality check, using software such as PicoQC.

Demultiplexing: In this step we identified the barcodes matching the sequence to the sample it comes from and removing the barcodes from the sequences. For that purpose, two software can be used: Deepbinner and Guppy. This step produces bcfastQ files.

Mapping: Reads are then mapped on the reference genome using Minimap2 producing BAM files. Additional filtering is necessary before analyzing the data. So as to all the sequences begin with TC and end with GA (remaining after *DpnI* GATC digestion). The sequences without that pattern on both ends, or only on one are removed from the data. In order to reduce the risk of errors, we used 16 bp window (8 bp before and 8 bp after the end of the read).

Normalization: In DamID experiments the data are usually normalized against the control condition *gfp::dam* calculating the $\log_2$fold *dam::protein-of-interest*/*gfp::dam*. However, because in our case we are comparing two fusion proteins (PolII and GFP) that methylate euchromatin, the risk of biases and a risk of significant loss of information is possible. In contrast, the use of other *dam::protein-of-interest* such as *dam::lmn-1* does not present this problem: This protein binds heterochromatin in contrast with euchromatin (*gfp::dam* or *dam::rpb-6*) producing a good normalization data. To solve those biases issues, we used the damidseq_pipeline kernel method developed by Andrea Brand laboratory (146), pipeline that was used for our NanoDamID analysis since the very beginning using BAM files. The resultants Bedgraph files produced by this analysis can be loaded in a genome browser. To check genomic datasets, we used IGV.

3. Results.

3.1 Studies to determine the right *C. elegans Dam::rnapolII* subunit for the profiling of active genes.

The first aim we had to achieve was to setup a robust, worm wide and cell specific pipeline to carry out *dam::RNApolII*-mediated active gene profiling experiments in *C. elegans*. In order to reach that aim, two different versions of *dam::RNApolII* were required: one constitutive version, expressing *dam::RNApolII* in all the cells of the body and another version carrying a *lox::mCherry::lox* cassette to get tissue specificity by *Cre/lox* recombination (Figure 17).

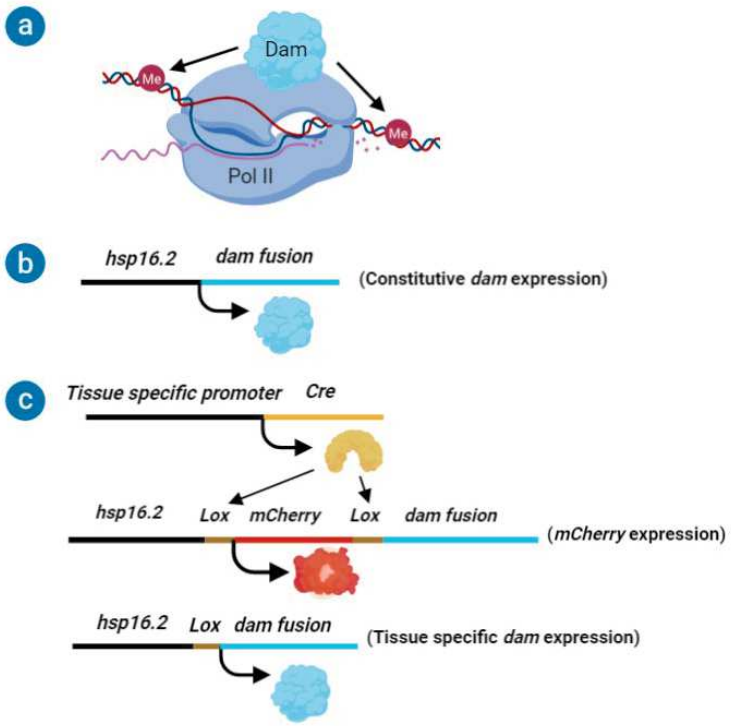


Figure 17. Dam RNA polymerase II approaches to study transcription. a) Dam RNA polymerase II profiling active genes during transcription. b) Constitutive *dam::fusion* version expressing Dam in all cells of the body, taking advantage of the leakiness of heat shock promoters non heat shocked. c) Tissue specific *dam::fusion* protein driven by *Cre/lox* recombination.

The RNA polymerase II of *Caenorhabditis elegans* is composed of 12 subunits (Chapter 1.9; Table 1), some of them associated with RNA polymerase II activity only and other additionally with RNA polymerase I and III activity (Wormbase). Initially we considered that the best subunit to be fused with Dam was AMA-1. The reason why we decided to use the AMA-1 subunit, as previously mentioned in Chapter 1.9, was the availability of ChIP-Seq data at worm wide level (139).

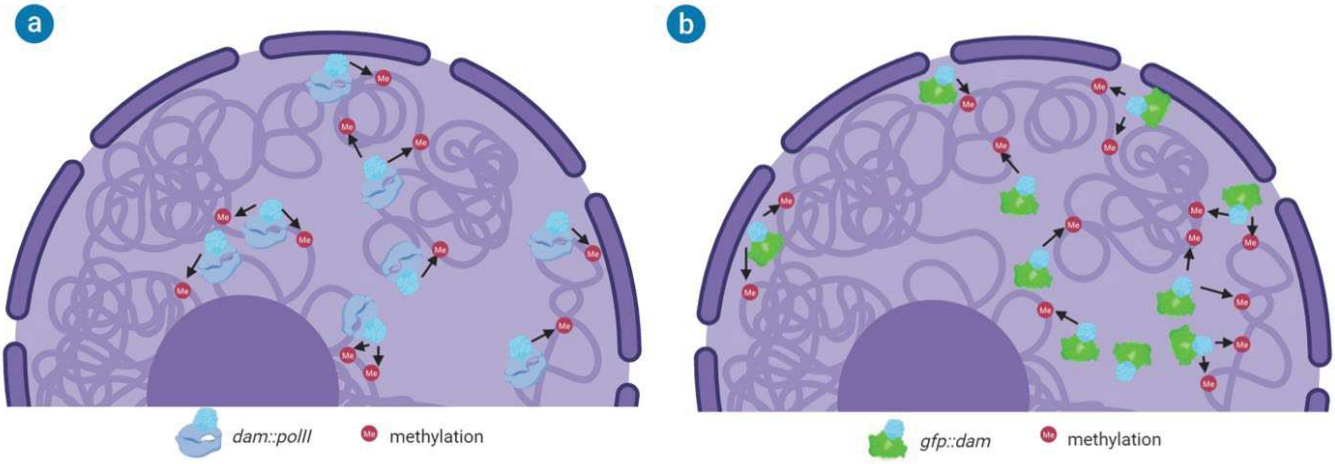


Figure 18. Different *dam::fusion* proteins profiling chromatin. a) Dam::AMA-1 is profiling active transcribing genes. b) GFP::Dam is profiling open-accessible chromatin.

We started cloning a *hsp::dam::ama-1* version to test the efficiency and the viability of a RNA polymerase II profiling *in vivo* at worm wide level. We created a transgenic worm strain expressing *dam::ama-1* in all tissues of the body integrated as a single copy insertion in chromosome II by MosSCI (Chapter 2.6.1). Additionally, another strain, already available in the Meister laboratory, expressing *gfp::dam* in all the cells of the worm body was used. GFP::Dam is a green fluorescent protein fused to the Dam methyltransferase that diffuses in the nucleus and methylates open accessible chromatin. We used GFP::Dam values to normalize the protein of interest (Dam:AMA-1) calculating the log_2fold as is described in Chapter 2.9.1 (126). Using these two *dam::fusion* proteins we expected to localize genes in the nucleus that were transcriptionally active (Figure 18).

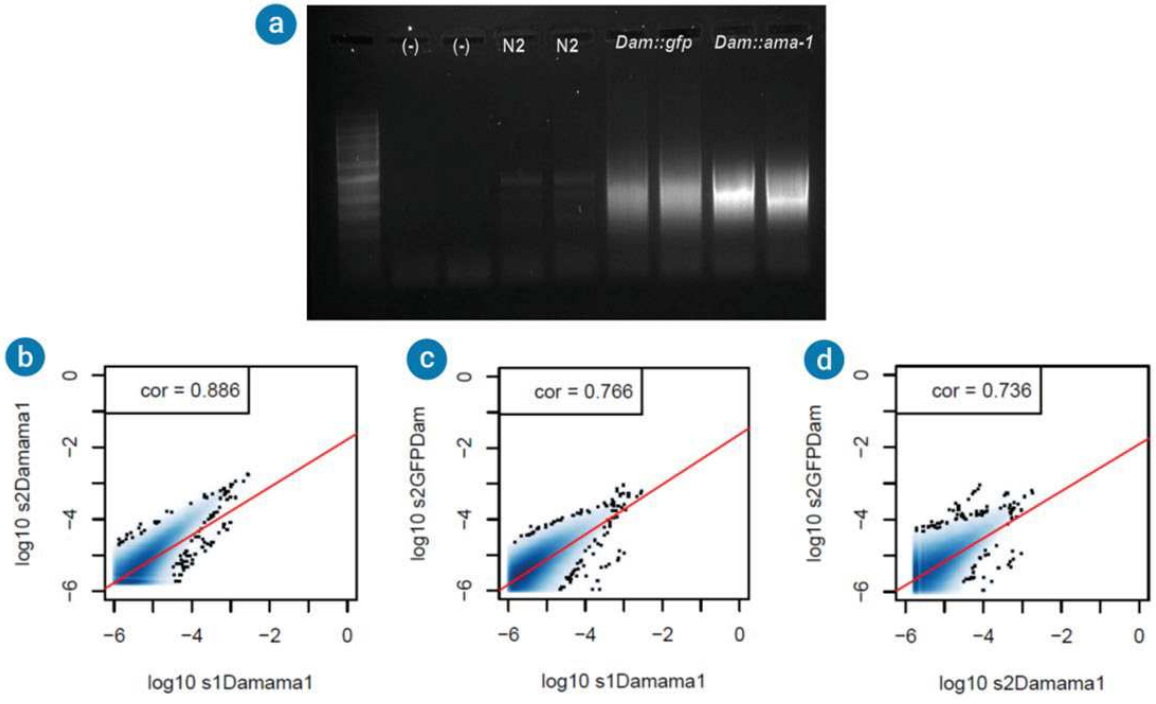


Figure 19. Correlation between sample reads (*Dam::ama-1*) and the control reads (*gfp::dam*) on DNA from whole worms. a) DamID PCR on an agarose gel (RedSafe: Catnº:21141): (-) no DNA, N2: wild type non methylated DNA, *dam::gfp*: Signal of the Dam methylated GFP, *Dam::ama-1*: Signal of the methylated subunit of the PolII: AMA-1. Correlation plots of GATC read numbers for 24 PCR cycles after adapter ligation. b) Correlation between the reads of two *dam::ama-1* samples c) Correlation between the reads of the control *gfp::Dam* and the sample number 1 *Dam::ama-1*. d) Correlation between the reads of the control *gfp::dam* and a biological replica number 2 of *Dam::ama-1*.

The correlations values obtained showed different values between *dam::ama-1* and *gfp::dam* experiments and a better correlation when we compared *dam::ama-1* vs *dam::ama-1.* This first approach allowed us to use this subunit of the RNA polymerase II to profile active transcribing genes in worms to uncover the transcriptional basis of the Y-to-PDA. (Figure 19). We did not go depth analyzing data at this level because we had to test before the possibility to reproduce the same result with tissue specificity, so, an analysis at gene level in that moment was meaningless.

Expression in Y and PDA cells *in vivo* conditions required cell specific Dam expression. The next step was the creation of a new strain carrying a *lox::mCherry::lox* cassette which was necessary to generate tissue specificity by *Cre/lox* (141). Unfortunately, because of the size of the construct (14 Kb) the efforts to generate a *dam::ama-1* transgenic strain carrying a *lox::mCherry::lox* cassette were unsuccessful.

We decided to clone another RNA polymerase II subunit, that time smaller, fused with *dam* and then try to integrate the new construct in the worm genome by MosSCI as we did with Dam::AMA-1. We tried to integrate a *dam::rpb-2* construct (Table 1). Unfortunately, we did not have any success either integrating the *lox::mCherry::lox* version of that *dam::fusion* protein as happened before with AMA-1. The reason why remains unknown, but it is possible that the fusion between Dam and RPB-2 is toxic or simply too big to be integrated (8933 bp in that case).

The difficulties to generate a transgenic worm expressing a *dam::polII* with a *lox::mCherry::lox* cassette, which is part of the molecular machinery to generate tissue specificity by *Cre/lox*, made necessary a comparative study between the sizes of the different RNA Polymerase II subunits in the worm to choose the right one (Chapter 1.9; Table 1).

Finally, I decided to create a new *dam::fusion* using one of the most little subunits of the RNA polymerase II. Based on a conversation with Dr M. Barkoulas about the best subunit to be used, we chose the subunit F of the RNA polymerase II called *rpb-6* (Chapter 1.9; Table 1). We re-started as previously, cloning a *floxed mCherry dam::rpb-6* and later injecting the plasmid to create a single copy strain with our construct integrated in the

genome of the worm, again by MosSCI. That time the microinjection was successful, and the strain was created. Using that strain and CRE drivers we could expressed *dam::rpb-6* in several cell types.

3.2 Creation of a constitutive *dam::rpb-6* profiling active genes.

In addition to the *lox::mCherry::lox dam::rpb-6* version, another strain carrying a constitutive version of *dam::rpb-6* was necessary to test the efficiency of that *dam::polymerase* subunit profiling active transcribing genes at worm wide level. This strain will be also useful to be crossed with worms expressing fluorescent markers in the Y cell (IS2922) in order to carry out prospective FACS experiments (Chapter 1.11.3).

For that purpose, instead of creating a new plasmid again without the *lox::mCherry::lox* cassette and then create a new transgenic strain by MosSCI, I chose to carry out an *in vivo* recombination to eliminate the *lox::mCherry::lox* cassette by microinjection of a plasmid expressing CRE recombinase producing *in vivo* recombination in the germline (Figure 20). Following this method, we generated a new strain (IS3675) from the pre-existing one (DAG958). I did different trials as described below:

1st trial: I injected a *hsp::Cre* (pSR33) plasmid carrying CRE recombinase under the control of a heat shock promoter, in the gonad of young adult worms. After the injection, in young adult stage, I heat shocked the worms to overexpress the CRE and recombine the *lox::mCherry::lox* cassette. Heat shock was repeated in the next generation carrying the array during the young adult stage. The result was negative: germline recombination was not successful.

2nd trial: I injected the commercial CRE recombinase from New England Biolabs (M0298) in the gonad of young adult worms. Serial dilutions with supplied buffer or in MQ water were necessary as I found that injection of pure CRE recombinase is killing the worm. Finally, I got the proper concentration with MQ water (1:100) that did not kill the injected animals; however, the result was again negative: I could not achieve germline recombination.

3rd trial: Finally, I injected the plasmid *pie-1p::Cre* in the gonad of young adult worms. *pie-1* encodes a nuclear protein PIE-1, localized to the germline blastomeres from the beginning to the end of embryo development (147). Using *pie-1* promoter, I could express CRE in the germline. The worms survived after the injection, generating F1 animals expressing the array and later, producing F2 progeny without the cassette, thereby demonstrating germline recombination. Hence, a transgenic strain expressing *dam::rpb-6* in all cell types of the worm was created.

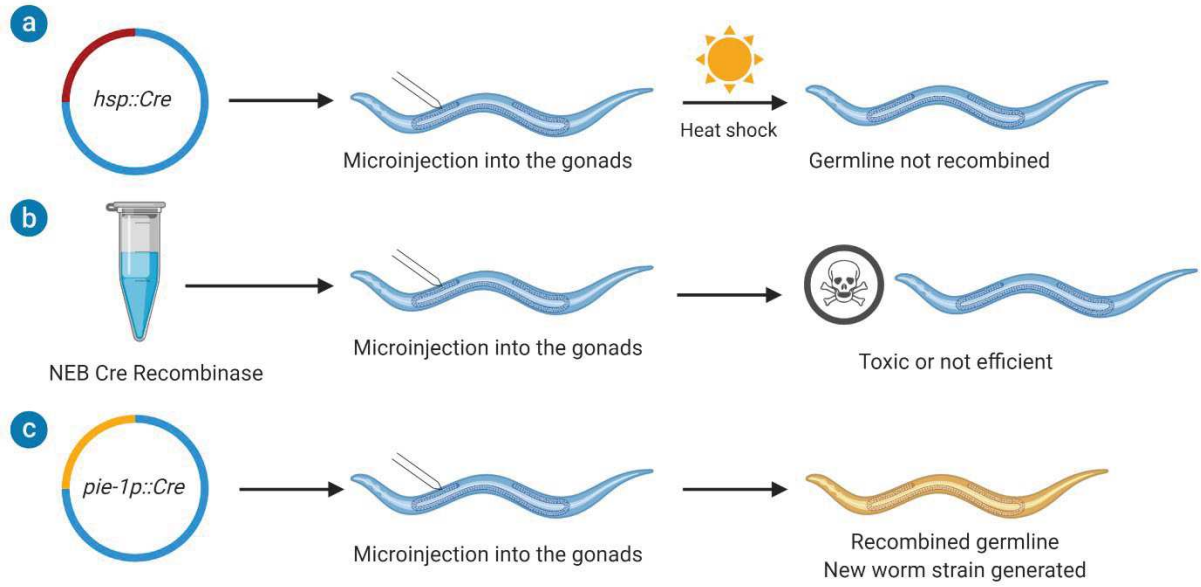


Figure 20. Different strategies to modify the germline for successful lox cassette removal in a single copy strain by microinjection. a) Microinjection of a *hsp::Cre* plasmid to overexpress CRE after heat shock and recombine the germline. b) Microinjection of a commercial pure CRE recombinase into the gonad to generate germline recombination. c) Microinjection of a *pie-1p::Cre* construct into the gonad to generate worms carrying the array in the F1 generation. Germline recombination after F2 generation.

Once the two versions *dam::rpb-6* (with and without *lox::mCherry::lox* cassette) were ready, I carried out worm wide and tissue specific DamID using recombination for cell-type specific Dam fusion expression. Additionally, I decided to use another *dam::fusion* proteins in our experiments: constitutive and *lox::mCherry::lox* cassette version of *dam::lmn-1*, already available in the Meister laboratory, which could be used in the two proposed approaches of this PhD work (cell sorting and recombination). LMN-1 is the sole nuclear lamin in *C. elegans* and could provide us information on the perinuclear-located chromatin in each cell type. Additionally, LMN-1 is a good control to contrast our experiment with *rpb-6*::*lmn-1* expression, as its expression starts in the germline (148), its binding pattern to the genome is known from ChIP-seq and DamID experiments and due to its long half-life, it was expected to lead to higher methylation levels than *rpb-6*. These two qualities were helpful to understand the characteristics of our DamID system, based on the sensitivity of the technique according to a minimal amount of interactions and the timing needed to make the system work. The third *dam::fusion* protein used in our experiments, as mentioned in Chapter 3.1, was *gfp::dam.* This fluorescent protein was co-expressed with *dam* in order to mark all accessible chromatin and was used to normalize thee data.

3.3 NanoDamID, a new tool to sequence DamID libraries.

In addition to finding the right subunit of the RNA polymerase II for polymerase footprinting, I faced an important problem for the accomplishment of the project: our access to the

Illumina HiSeq2500 device the Meister laboratory had used in the past was not possible anymore and at the IGBMC only an Illumina HiSeq4000 was available. We could not use this technology for our DamID-Seq because in the new Illumina HiSeq4000 chips, the clusters for polony (polymerase colony) amplification are not distant enough as they were in the HiSeq2500. The consequence of that new design is the impossibility to sequence fragments longer than 500 bp, which is too small for most DamID amplicons. For this reason, the development of an alternative sequencing method became a priority.

In order to solve that problem, we decided to develop a new DamID-seq method to sequence DamID libraries, changing our original Illumina method for Oxford Nanopore Technologies (ONT) (Figure 21). This technology offers several advantages, the most important for our purpose was the possibility to sequence any length of DNA, short or ultra-long and the possibility of carrying out direct sequencing of DamID amplicons. Additionally, we could sequence in real-time, directly in the laboratory. Also, using that technology, we do not need a sequencing platform anymore, carrying out the sequencing in our own bench in the lab.

The new methodology has in common with the classical protocol (122) the *DpnI* digestion, the first adapter ligation and the 1st PCR but, instead of carry out a second PCR after the end-repair, A-tailing and barcoding steps, we directly attached the Nanopore motor protein in our amplicons to create the library. The sequencing was carried out using a MinION Mk1C device.

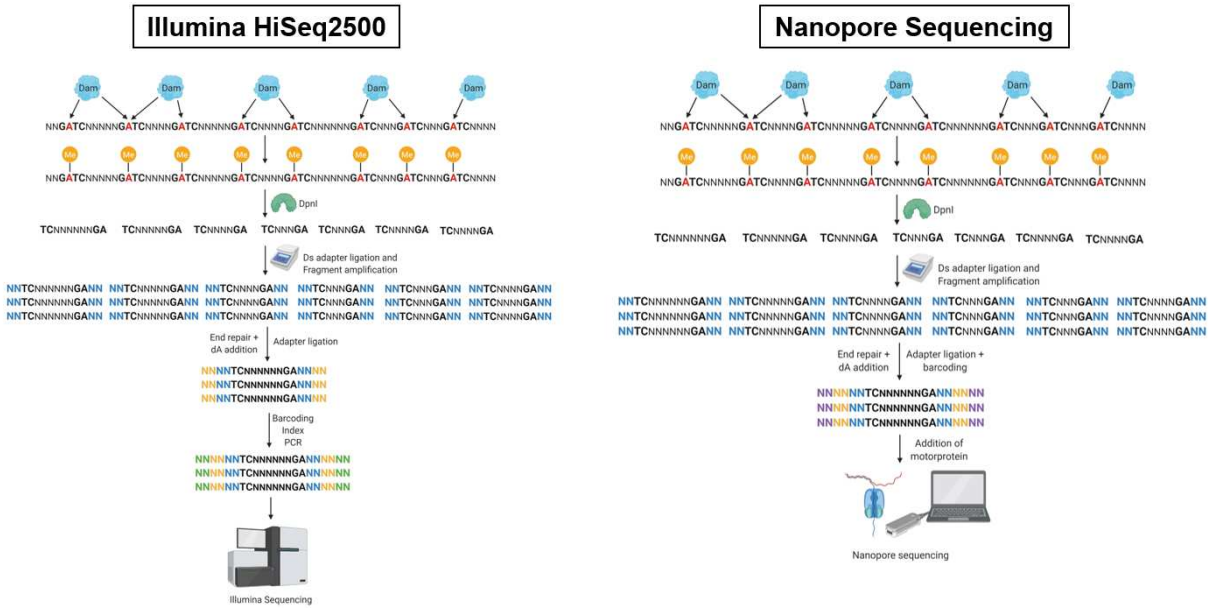
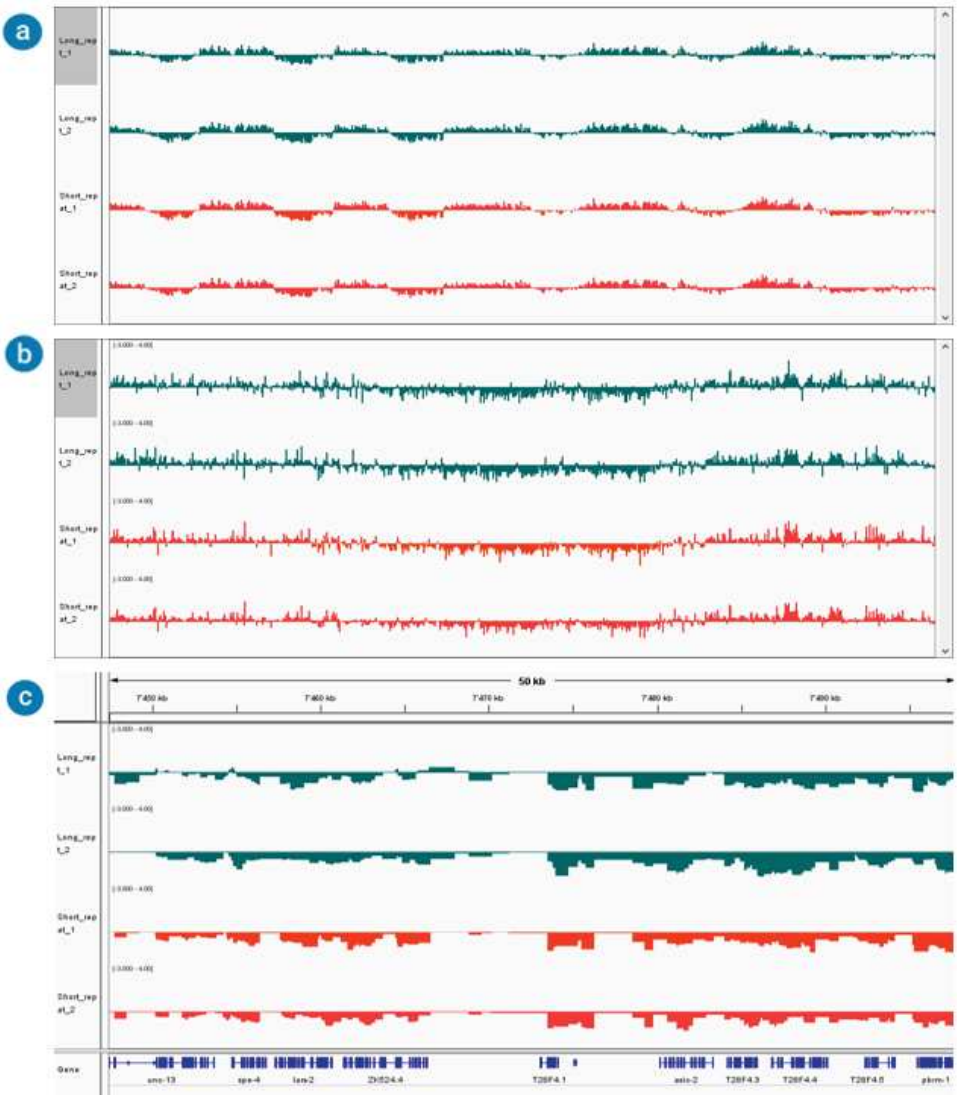


Figure 21. Comparison between the Illumina DamID-seq and the NanoDamID-seq protocols. Left: Classical Illumina DamID sequencing. After *DpnI* digestion and first PCR is carried out and the amplified fragments are end-repaired and adenine tailed, the barcoding step include a second PCR. Right: After the *DpnI* digestion and adapter ligation we directly sequence the amplicons using Oxford Nanopore technologies (ONT) adding barcodes and the motor protein.

In order to test the efficiency of the new DamID-Sequencing methodology developed in our laboratory, we carried out DamID using a *C. elegans* strain expressing *dam::lmn-1.* Our aim was compare *dam::lmn-1* profiles already available by Illumina HiSeq2500 vs *dam::lmn-1* profiles sequenced by us using Oxford Nanopore Technologies. The results of this experiment showed how similar the two sequencing methods are; having both similar correlation values, the correlation values of the samples sequenced by Illumina were however slightly better (Figure 22).



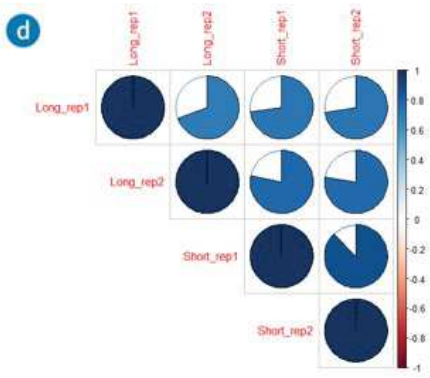


Figure 22. Correlation between Illumina (short)/Nanopore (long) DamID checked in IGV browser after calculation of the value Log2fold *Dam::lmn-1*/*GFP::Dam*. a) Comparison at genome wide level analysis between NanoDamID (blue) vs IlluminaDamID (red). b) Comparison between NanoDamID (blue) vs IlluminaDamID (red) in chromosome I. c) Comparison between NanoDamID (blue) vs IlluminaDamID (red) in a random region of chromosome I. d) Correlation values of the reads of the normalized values *Dam::lmn-1/GFP::Dam.* Nanopore (long) vs Nanopore (long); Nanopore (long) vs Illumina (short); Illumina (short) vs Illlumina (short).

3.4 Worm wide and tissue specific NanoDamID experiments.

Starting directly by expressing *dam::fusions* in the Y cell is a complex process as I showed in my introduction (Chapter 1.11). For that reason, we planned a “simple” test using *Cre/lox* recombination system, expressing CRE in different tissues in order to setup a proof of principle. For that purpose, we created a number of strains, in collaboration with the Glauser laboratory at the University of Fribourg. The chosen tissues were intestine (20 cells/worm), body wall muscle (95 cells/worm) and XXX neurons (2 cells/worm) using *elt-2*, *myo-3* and *sdf-9* promoters respectively driving expression of the CRE recombinase. We carried out DamID experiments using young adult worms. The reason why we used young adult worms was to have a large amount of methylation in the DNA, accumulated since embryogenesis. The successful results validating our tissue specific DamID using RPB-6 by *Cre/lox* are detailed in Annex 1.

The next step after this successful proof-of-principle on animals expressing worm-wide and tissue specific Dam was to setup the spatial and time-controlled recombination systems in order to have a tissue-specific time-controlled Dam profiling active gene at different stages of the development in Y-to-PDA. Unfortunately, after several trials and experimental re-designs, we have concluded that for the moment, with the available molecular technology, it is not possible to have an induced temporal controlled and tissue specific Dam in Y cell. All results and trials related to the efforts carried out in my PhD work to reach that aim are described in detail in Annex 2.

3.5. Experiments to setup NanoDamID in sorted cells.

3.5.1 Building an alternative method: FACS-NanoDamID.

As we failed to express Dam only in Y cell (Annex 2), I had to adopt a new strategy in order to obtain the transcriptional profile in the Y cell during early embryogenesis. One good alternative was to sort the Y cell by Fluorescence Activated Cell Sorting (FACS) using a double gating system designed with the same promoters proposed for the recombination cascade and the cGAL system (Annex 2). A strain carrying those fluorophores combination had been created in the Jarriault laboratory and the sorting method improved (N. Fischer and S. Jarriault, unpublished). In addition, the protocols to i) grow fairly synchronized large population of worms, ii) extract cells from *C. elegans* embryos and then iii) purify them based on their expression of 2 fluorescent proteins, had been set up and optimized in the Jarriault laboratory (N. Fischer and S. Jarriault, unpublished). The used strain, IS2922, expressed *mCherry* and GFP using *egl-5* and *hlh-16* promoters, respectively (Chapter 1.11.3). Both promoters lead to the expression of the fluorescent proteins in a number of cells, so that populations of non-fluorescent cells, red cells (*egl-5p*), green cells (*hlh-16p*) can be sorted and, to finish, the Y cell, the only cell in which these promoters are both active, this cell will be both red and green.

Using the strain IS2922 in combination with our collection of constitutive *dam::fusion* proteins (*dam::rpb-6*, *dam::lmn-1* and *gfp::dam*) we could generate new strains to carry out the desired FACS-NanoDamID. In addition to the Y cell experiments, we proposed as a control and proof of principle to carry out the same FACS-NanoDamID experiments sorting more abundant homogenous cells populations than Y. Again, as we did in our proof-of-principle (Annex 1), we choose intestinal (20 cells per embryo) and body wall muscle (81 cells per embryo) cells to have this control/proof-of-principle of our NanoDamID in sorted cells. Using two different strains, *elt-2p::gfp* and *myo-3p::gfp* expressing GFP in intestine and muscle respectively, we generated new strains expressing the same *dam::fusions* that we wanted to express in Y by worm crossing (Table 4) (Figure 23).

Strain	Fluorophores	Cell type	*Dam::fusion* protein
IS3653	GFP	Intestine	GFP::Dam
IS3654	GFP	Intestine	Dam::LMN-1
IS3687	GFP	Intestine	Dam::RPB-6
IS3651	GFP	Body wall muscle	GFP::Dam
IS3652	GFP	Body wall muscle	Dam::LMN-1
IS3684	GFP	Body wall muscle	Dam::RPB-6
IS3467	Double/GFP/mCherry	Y/hlh-16+/egl-5+	GFP::Dam
IS3468	Double/GFP/mCherry	Y/hlh-16+/egl-5+	Dam::AMA-1
IS3564	Double/GFP/mCherry	Y/hlh-16+/egl-5+	Dam::LMN-1
IS3685	Double/GFP/mCherry	Y/hlh-16+/egl-5+	Dam::RPB-6

Table 4. Strains used for FACS-NanoDamID experiments in different populations of sorted cells.

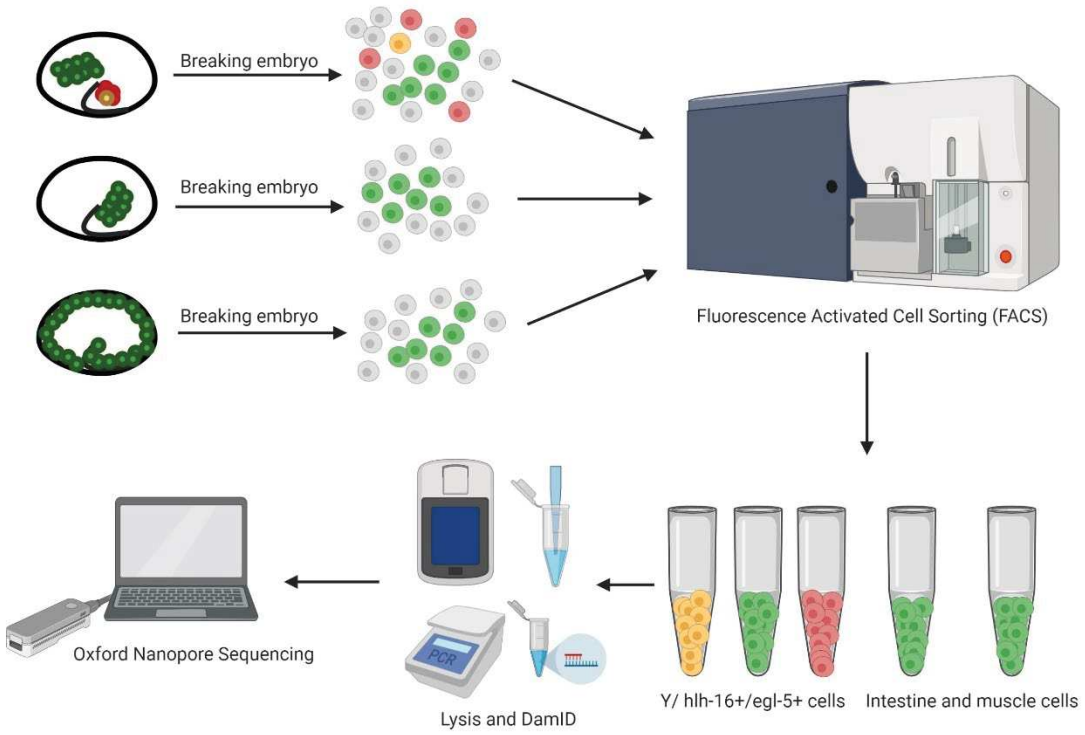


Figure 23. Tissue specificity expression using Fluorescence Activated Cell Sorting (FACS) – DamID procedure in *C. elegans*. Embryos and larvae can be dissociated in single cells. Cells sorted by FACS according to their emitted fluorescence. The cells are lysed and treated to carry out DamID and subsequently analyzed.

3.5.2 Timing assays to get tightly synchronous embryo populations.

Prior to sorting the cells of interest, I needed to establish worm synchronization and staging conditions for each strain to obtain a population of embryos as synchronized as possible. Initially, I wanted to have a 100% population of young adult worms carrying 2 eggs in order to be able to isolate homogenous (tightly synchronous) embryo populations at the 2-cells/4-

cells state of the embryogenesis. Oocytes mature and get fertilized every 23 min on average in gravid adults (149)(150), and two oocytes can be fertilized at the same time from the two gonadal arms. Hence a young adult with 4 eggs in its uterus will yield an egg population of 4 eggs all within a ~23 min time window. While isolating a small population of such embryos at that stage of the development is possible, scaling it up to the ten million of embryos necessary to perform Y extraction and sorting, turned out to be very challenging or impossible. In order to establish an approach to have the highest number of embryos in the proper stage, I therefore tested different adult development time points.

To determine the exact timing to get a tightly synchronized adult population with a given number of eggs in the uterus, we have conducted several experiments. We started each time with a population of gravid adult worms synchronized over two generations. First, an adult worms preparation was bleached to retrieve their eggs. These eggs were incubated in M9 buffer for a number of hours to obtain a synchronized and arrested population of L1 larvae, which was then plated on food for a given amount of time until early adulthood. Determination of the proportion of adults with a specific number of eggs in the uterus was then achieved by checking individual animals under the dissecting scope using at least 50 worms per condition (Table 5).

Time window	Time between bleaching and plating	Growing time after plating (20°C)	Number of worms (n)	Adults without eggs (n/%)	Adults with 2-4 eggs (n/%)	Adults with 8-12 eggs (n/%)	Adults with >12 eggs (n/%)
Time window 1	18	54	50	14/28	8/16	10/20	18/36
Time window 2	18	55	50	10/20	9/18	11/22	20/40
Time window 3	24*	58	50	23/46	15/30	7/14	5/10
Time window 4	24*	60	50	21/42	17/34	5/10	7/14
Time window 5	24*	62	50	12/24	30/60	2/4	6/12
Time window 6	24*	65	50	7/14	22/44	10/20	11/22

Table 5. Time windows to determine the proper timing to generate tightly synchronized populations of young adults carrying 2-4 eggs. *Extended time between bleaching and plating in order to produce high starvation and force the worm population to prospectively re-enter into the life cycle at the same time.

We concluded that the synchronization of a huge number of embryos in the same stage is impossible for the different environmental conditions of the worm culture; likely, small variations in humidity, temperature or bacterial lawn density would impact the worm synchronization, thereby producing different population of eggs. Additionally, we could not use a time window in which our desired number of worms was less than 50% because even

if the rest of the population is not producing eggs (hence no desynchronization problems), we could not have isolated enough eggs to perform cell sorting. The only alternative was therefore to increase the length of our time window, thereby generating a slightly heterogenous population of embryos with the higher abundancy of them in the 1.5-fold stage and 3.0-fold stage (Figure 24).

IS2922 (25°C)	0 min	30 min	1h	1h 30 min	2h	2h 30 min	3h		4h
2 to 8 cell stage (50-100min)	22,9%	3,5%	0,3%	0,8%	0,0%	0,0%	0,0%		0,0%
Eggs laid outside (150 min)	60,0%	54,5%	29,4%	37,6%	18,3%	14,5%	10,0%		2,8%
Ventral cleft closed (330 min)	13,3%	31,5%	41,3%	13,5%	24,1%	22,7%	27,9%		24,2%
Bean stage (360 min)	0,0%	1,4%	3,5%	7,2%	8,5%	9,0%	5,5%		5,5%
Comma stage (430 min)	1,0%	0,0%	2,4%	5,9%	8,0%	6,6%	10,1%		10,4%
1,5-fold (460 min)	0,0%	1,4%	2,4%	6,8%	13,4%	11,3%	14,8%	1h in chitinase	15,2%
1,7-fold (480 min)	0,0%	0,0%	0,7%	3,4%	2,2%	3,5%	7,1%		13,1%
2-fold (490 min)	1,9%	1,4%	4,2%	7,2%	7,1%	7,0%	7,5%		6,6%
3-fold (550 min)	0,0%	1,4%	11,2%	13,5%	14,7%	19,5%	13,0%		17,0%
4-fold (600-800 min)	1,0%	4,9%	4,5%	4,2%	3,6%	5,9%	4,1%		5,2%
Total number of embryos scored	105	143	286	237	224	256	253,5		289
L1s	-	-	-	~15 L1s in 15µL drop	~20 L1s in 15µL drop	~20 L1s in 15µL drop	~48 L1s in 15µL drop		~90 L1s in 15µL drop

Figure 24. Final time window used to generate a heterogeneous population of embryos able to be dissociated after chitinase digestion and subsequently sorted by FACS. Source: Isaia D, Jarriault J/Unpublished data.

Using this enlarged time window, we could determine that the embryos used in our sorting experiments will be expressing GFP after 200 minutes (ventral cleft closed stage) in intestine, counting from the first embryonic cell division, after 300 minutes (bean stage) in body wall muscle and ≈400 minutes (entering in comma stage) in Y, Green (HLH-16+) and Red (EGL-5+) cells (Wormbase).

3.5.3 FACS-NanoDamID sensitivity tests.

The next step was to carry out a DamID sensitivity test. We wanted to carry out bulk DamID-Seq to compare the genes detected by DamID vs RNA-Seq performed by other members of the Jarriault laboratory in Y/hlh-16+/egl-5+ cells. For RNA-Seq experiments the optimal minimal amount were 200 cells for each biological replica (N. Fischer, S. Jarriault, unpublished data). We started sorting 50, 100 and 250 cells per tube. However, these amounts were not enough. Finally, we determined that 500 cells per tube were required to see an appropriate methylation signal on gel after the DamID-PCR. DamID worked

successfully in 500 cells bulks from worms expressing *gfp::dam* and *dam::lmn-1*, but, unfortunately the DamID never worked in cells expressing *dam::ama-1*. One possible reason could be the huge size of the subunit A of the RNA polymerase II that requires a long time to generate a functional protein, which is then able to methylate the DNA. However, this issue does not happened when we sorted cells expressing *dam::rpb-6*, subunit F of the RNA polymerase II, smaller than subunit A (Figure 25).

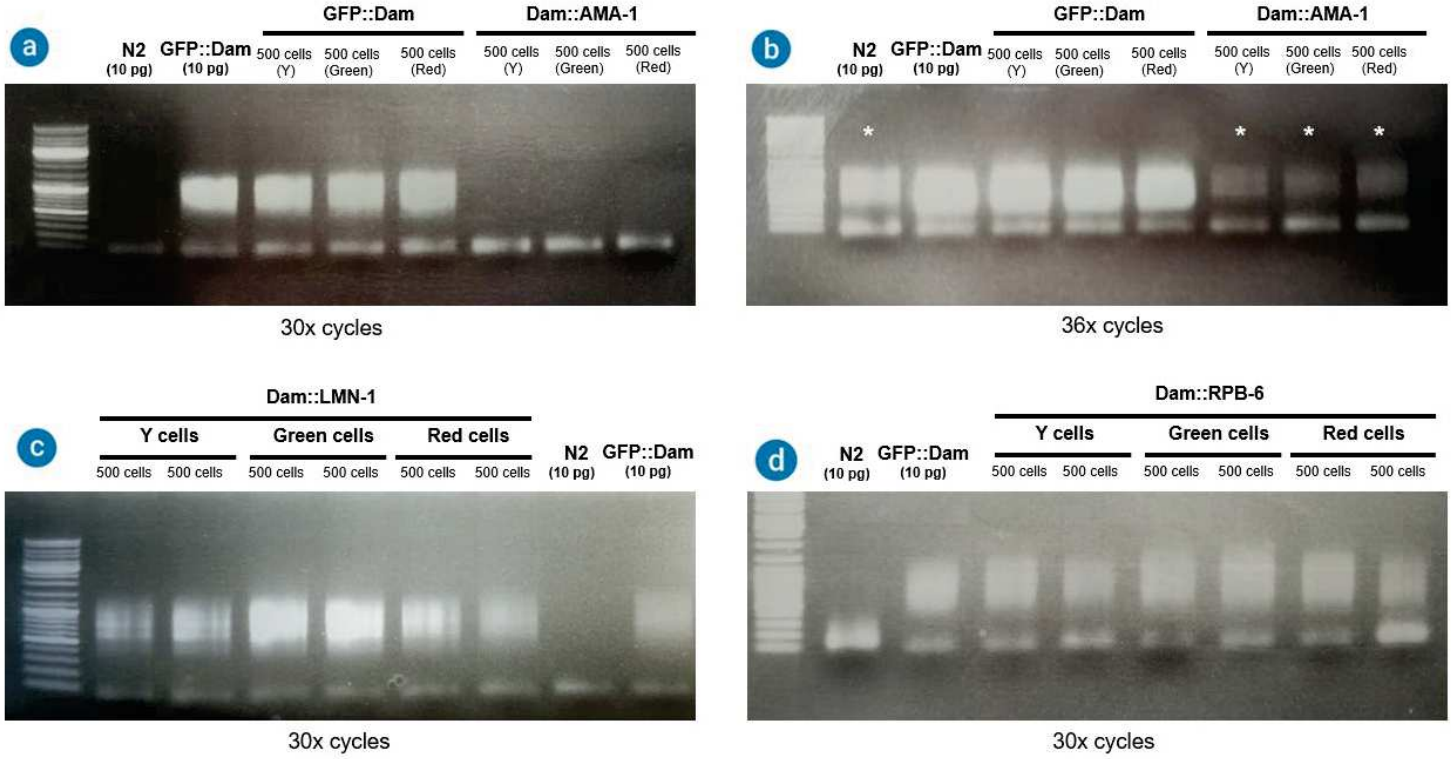


Figure 25. Sensitivity FACS-NanoDamID test. Proof of principle checking FACS-DamID sensitivity in sorted cells expressing *GFP::Dam* and *Dam::ama-1* after x30 (a) and x36 (b) DamID-PCR cycles. *: background produced by excess of PCR cycles. c) Complete FACS-NanoDamID experiment using a strain (IS3564) expressing mCherry, GFP fluorophores and *dam::lmn-1*. Two technical replicas for each cell population, 1000 cells in total each biological replica. d) Complete FACS-NanoDamID experiment using a strain (IS3685) expressing mCherry, GFP fluorophores and *dam::rpb-6*. Two technical replicas for each cell population, 1000 cells in total each biological replica.

Using these amounts of cells to perform our experiments we could carried out NanoDamID on the mentioned cell populations: intestine, body wall muscle and Y, Green (HLH-16+) and Red (EGL-5+) cells.

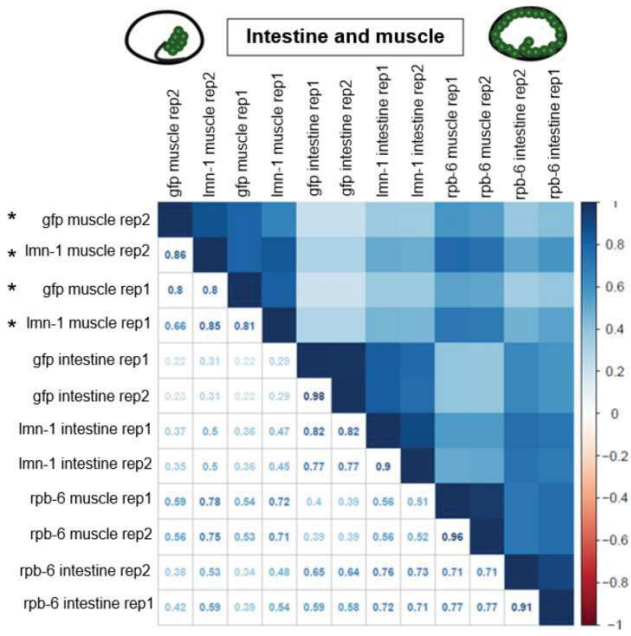
3.6 FACS-NanoDamID results.

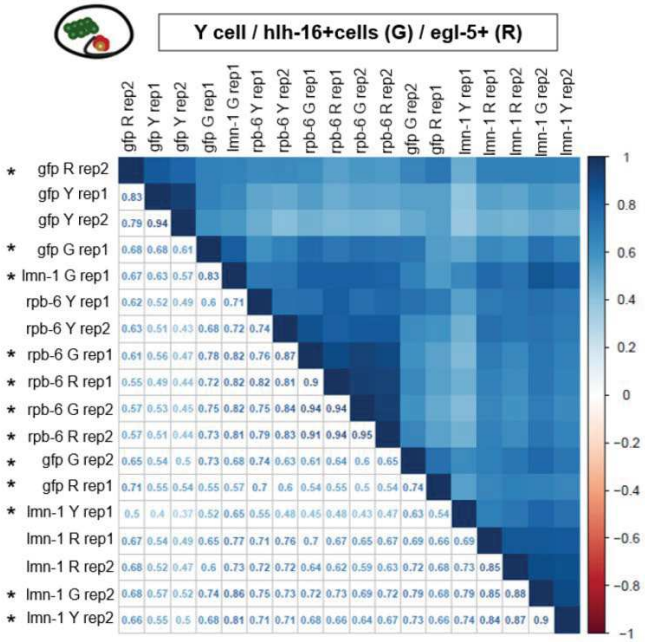
3.6.1 FACS-NanoDamID statistical analysis.

The results produced by FACS-NanoDamID were promising. For a quality control, we clustered the different libraries based on the values for the different *Dpn*I restriction fragments amplified by the DamID PCR. For Intestine and Muscle embryonic cells, the two repeats of the different libraries were clustering together, thereby demonstrating that the variability between different Dam fusions was higher than between replicates of the same Dam fusion, even at the single *DpnI* restriction fragment scale, the smallest which can be analyzed (Figure 26a). However, for cells expressing the markers identifying uniquely Y (R for *hlh-16* and G for *egl-5*), the duplicates of the individual Dam fusions did not cluster together. This indicates that the variability between replicates is higher than the variability between the different Dam fusions. Several explanations could be invoked for this lack of reproducibility. First, for the R and G samples, sorting does produce highly heterogenous populations of cells, as both these markers are expressed in a variety of different cells types. This leads to sampling differences between the different cell types in replicas, ultimately producing variable DamID libraries. The fact that Y libraries replicates for *rpb-6* and *gfp dam* fusions cluster close together would be a further argument for this explanation. Alternatively, Dam methylation is a stochastic event (123) and, starting from a tiny population of cells, this would lead to different methylation patterns between cells and between replicas. This would translate into noisy amplification at the individual restriction fragment level. If this would be the case, one would expect that analyzing methylation in larger regions would lead to better reproducibility. We performed such analysis by first analyzing average methylation on larger domains. This did not improve repeat clustering, even for very large domains up to 100’000 bp (Figure 26b). Alternatively, we smoothed the data using a sliding window of 500 bp shifted by 250 bp. By combining different fragments, we hoped to increase correlations between replicates. As the mean distance between the *DpnI* restriction sites GATC in the genome is 370 bp and the median 206 bp (126), this method would average the amplification values for several restriction fragments. This indeed led to a better clustering of the different fusion proteins yet did not improve the correlation between individual replicates nor the clustering of the R and G fusions together (Figure 26c). We concluded that for muscle and intestinal cells, and likely for Y cells too, the reproducibility between experiments was high enough (Figure 26d) to move further and analyze the enrichment values for *rpb-6*. Yet for R and G samples, which we FACSed as controls when purifying the Y cell, the variability between replicas was too high to confidently analyze polymerase footprinting or perinuclear localization of specific regions.

a

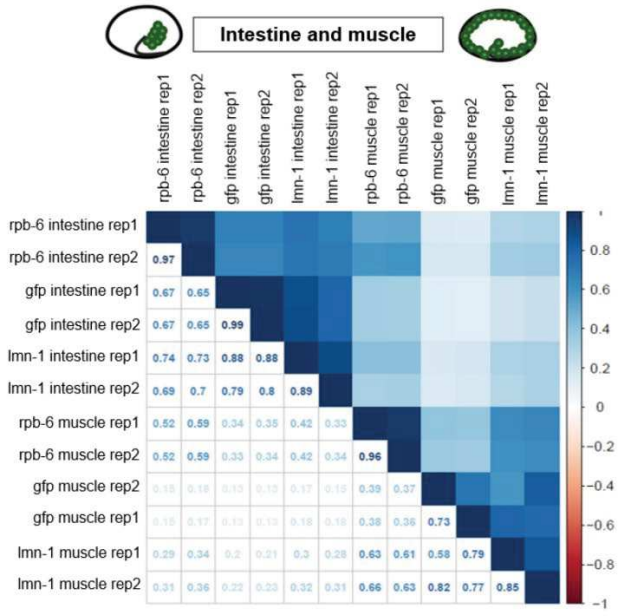
Correlations per GATC fragment

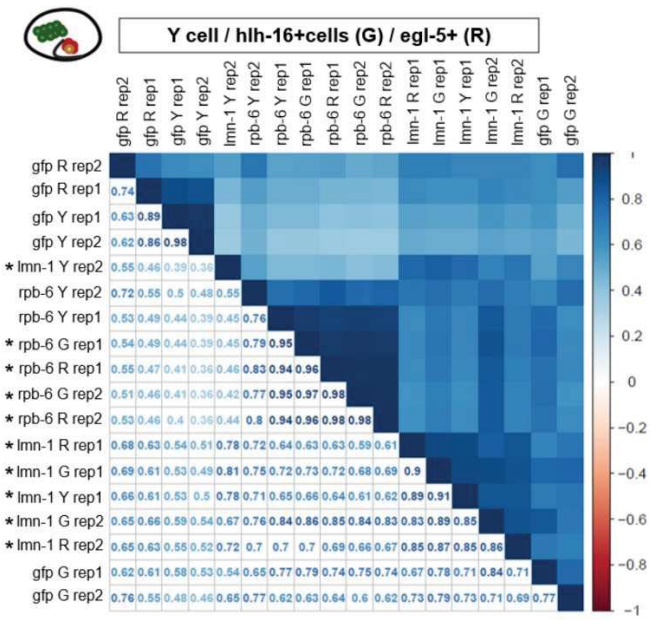




b

Sliding window 500bp with a 250shift





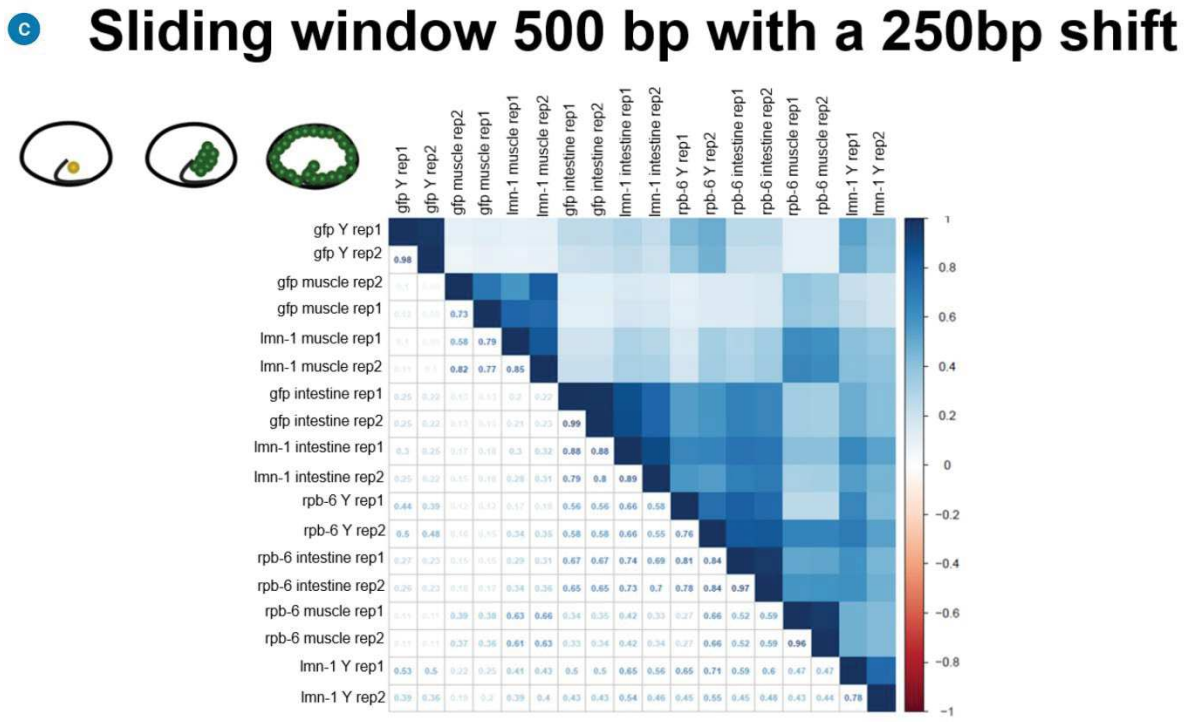


Figure 26. Correlation values at GATC level and sliding windows. a) Pearson's correlation at GATC level in intestine, body wall muscle and Y/G/R cells. b) Pearson's correlation with sliding window of 500 bp shifted by 250 bp in intestine, body wall muscle and Y/G/R cells. c)Pearson's correlation with sliding window of 500 bp shifted by 250 bp in intestine, body wall muscle and Y cell. *: replicas that are not clustering properly. Source: Osuna-Luque J, Beaumann E, Jarriault S, Meister P.

3.6.2 FACS-NanoDamID study of genomic datasets.

Smoothing the data using a sliding window of 500 bp shifted by 250 bp, checking normalized DamID tracks on IGV browser, we could see that the results were promising:

-Experiments using tissue specific *dam::lmn-1*: We expected methylation abundancy in the periphery and less methylation in the center of the chromosome. Chromatin is more open in the middle of the chromosome because of the absence of centromere in *C. elegans* (151). Our data (Figure 27) showed expected *dam::lmn-1* methylation (152) with abundance of methylation signal at the ends of the chromosomes and less signal in the middle, as observed before (153)(154).

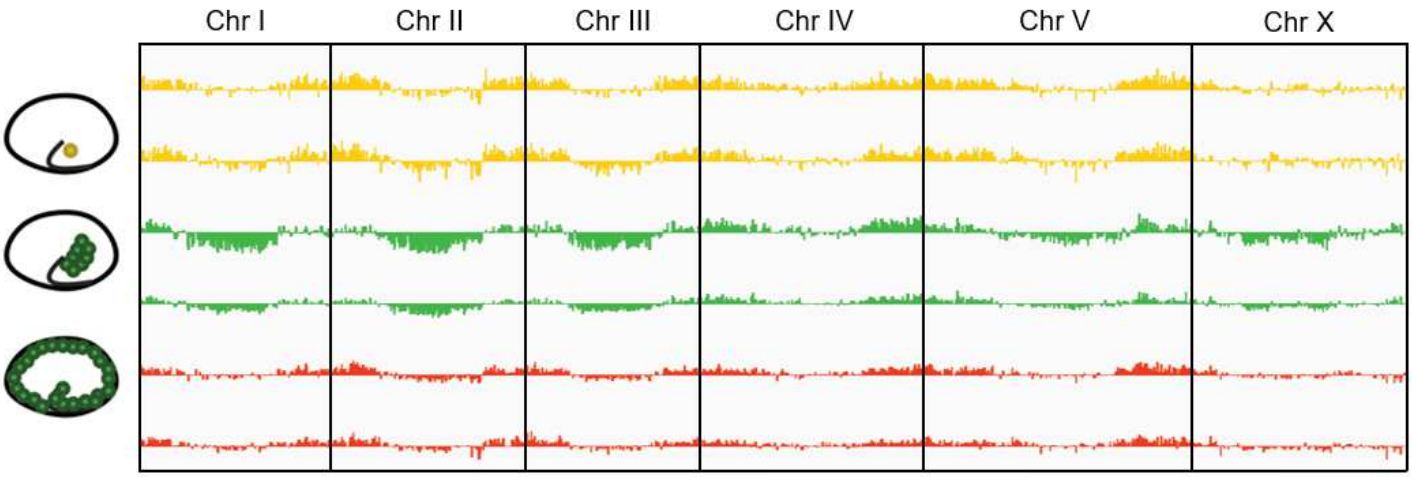
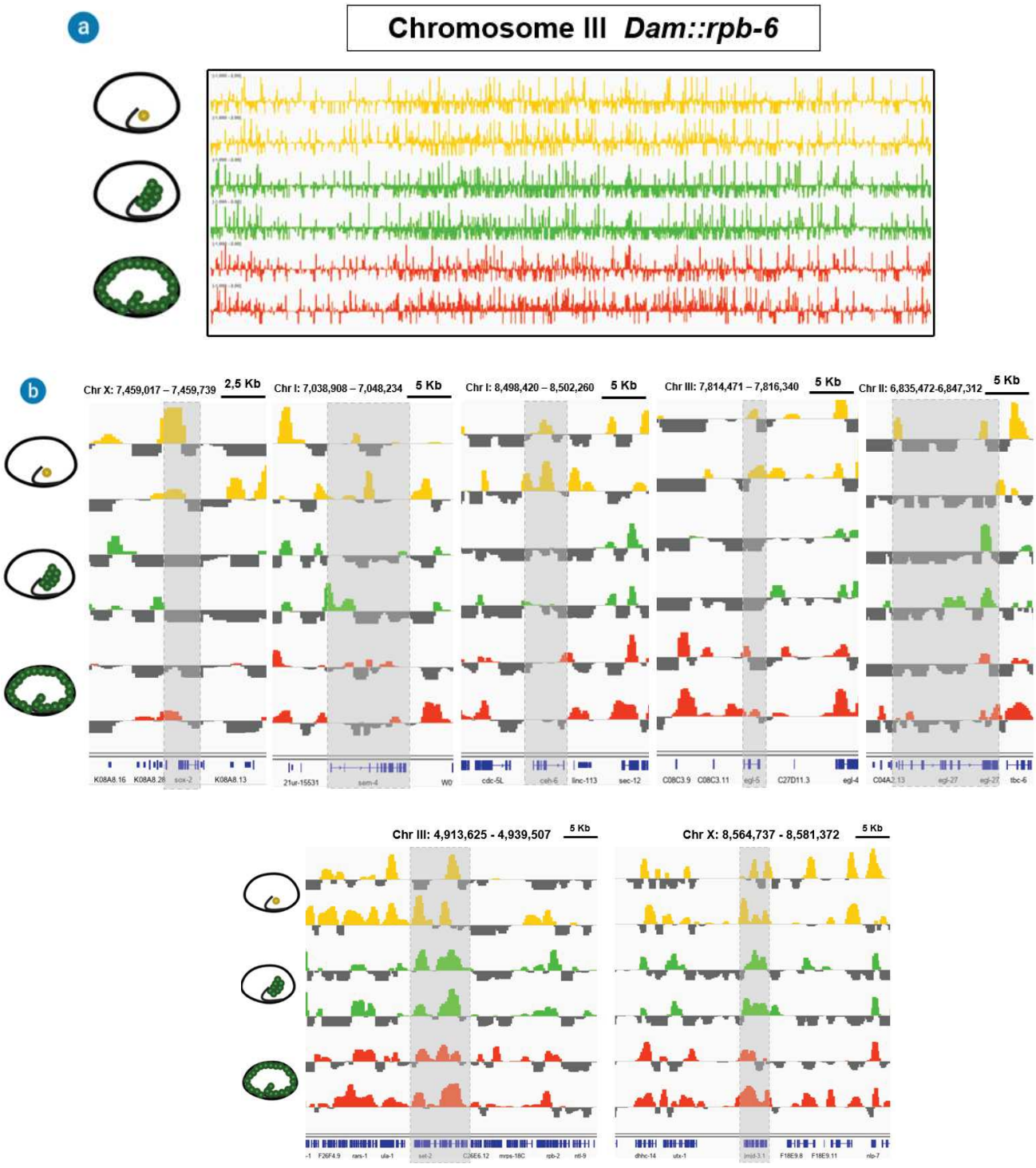


Figure 27. Cell specific *dam::lmn-1* using IGV browser. Y (yellow), Intestine (green) and body wall muscle (red) *dam::lmn-1* worm wide levels profiles. Higher methylation signal corresponds with higher compacted chromatin in the flanks of the chromosomes. Scale: -2 to +2.

-Experiments using tissue specific *dam::rpb-6*: Comparison between Y, intestine and body wall muscle cells showed cell-type specific methylation patterns (Figure 28). Expected methylation abundancy in the center of the chromosomes was observed at chromosome level (Figure 28a), in agreement with the fact that many housekeeping genes are located in the center of the chromosomes (155). Genes such as *sox-2, ceh-6, sem-4, egl-27* and *egl-5* are encoding transcription factors which activity is critical for the Y-to-PDA transdifferentiation initiation (110), providing therefore valuable positive controls to assess the polymerase footprinting. *sox*-2, *ceh-6* and *egl-5* showed high Dam::RBP-6 methylation pattern, in contrast to *sem-4* and *egl-27,* which showed a lower methylation profile in comparison with the first three genes; however, *egl-5* showed high methylation. It has been described that EGL-27 is modulating the activity of *egl-5* gene (110), for that reason, *egl-27* gene cannot be inactive. We hypothesized that the low methylation peaks in *sem-4* and *egl-27* could either be caused by the stochasticity of the DamID (141) or because of a low expression level of those transcription factors during embryogenesis in Y. Indeed, *sem-4* expression has been shown in the laboratory to start fairly late during embryogenesis (Daniele T, Jarriault S, unpublished data). It is important to remember that Y-to-PDA initiation starts in L1 (106) and it is possible the higher expression levels of those transcription factors occurs in that step of the life cycle. To finish, *set-2,* and *jmjd-3.1,* two genes involved in robustness of Y-to-PDA showed high methylation levels. *set-2* is encoding a catalytical subunit of the SET-1 histone methyltransferase and its absence disrupted invariant transdifferentiation (113), for that reason high methylation was expected. *jmjd-3.1* encode a histone demethylase which plays an important role in PDA redifferentiation. During *C. elegans* life cycle, this gene is expressed in Y from embryogenesis to L3 stage (113). Additionally, typical *C. elegans* housekeeping genes *act-1, spt-5, egl-45* and *gpd-4* (156)

were checked and as expected, high Dam::RBP-6 methylation levels were detected (Figure 28c).



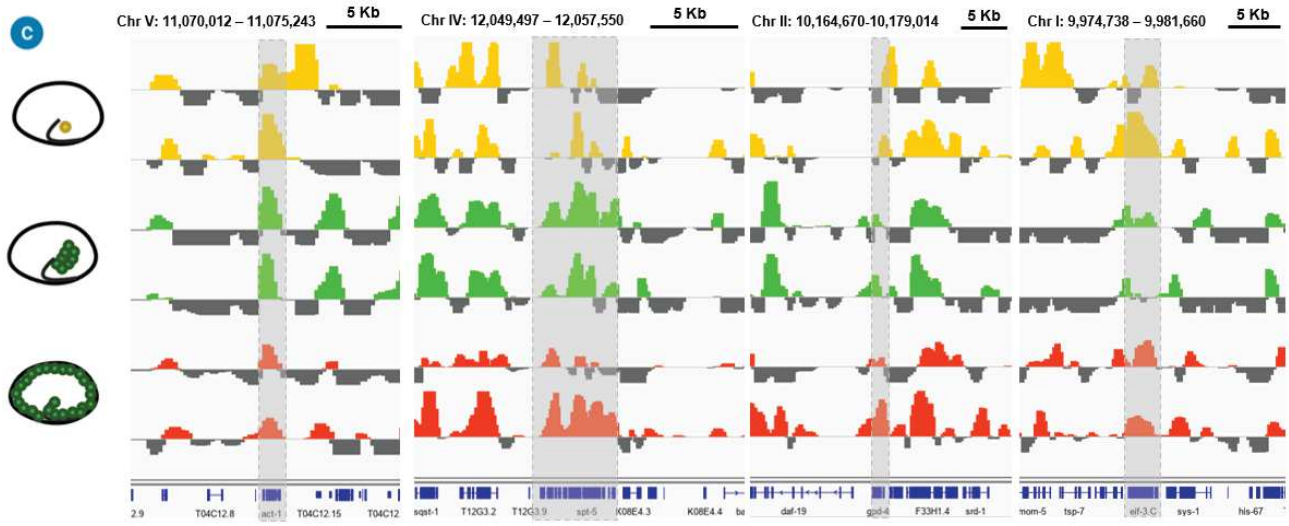
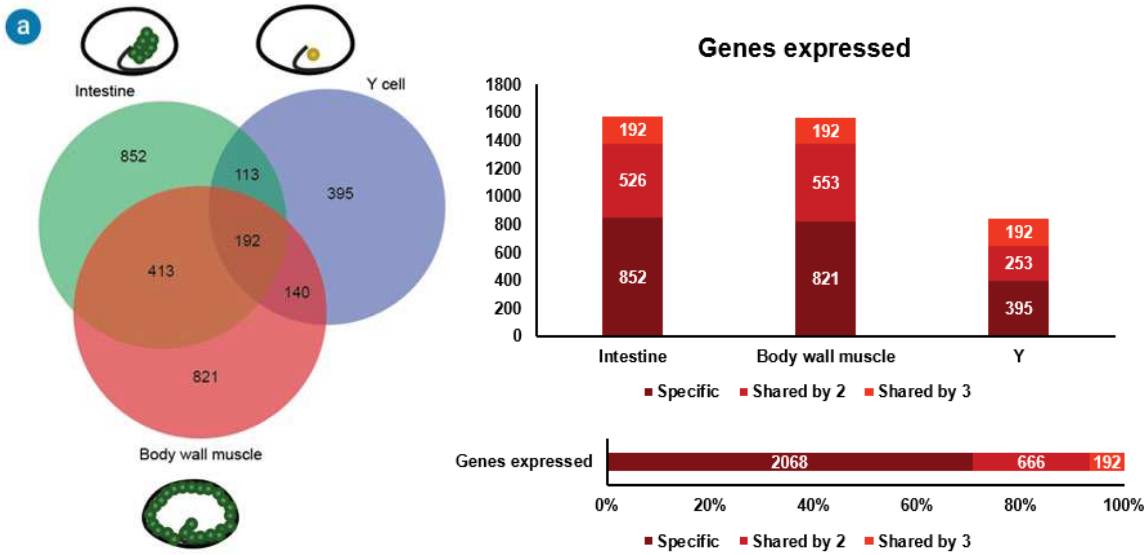


Figure 28. Cell specific *dam::rpb-6* using IGV browser. *dam::rpb-6* showed high methylation in genes involved in Y-to-PDA. Y (yellow), intestine (green) and body wall muscle (red). a) Chromosome level. b) Y-to-PDA initiation specific gene level expression patterns. c) *C. elegans* housekeeping genes. Scales (a) -1 to +2 (b) and (c): -0.1 to +2.

3.6.3 Gene Ontology (GO) enrichment analysis.

Comparisons between intestinal, body wall muscle and Y cells showed 852 specific genes in intestine, 821 specific genes in body wall muscle and 395 specific genes in Y cells. In total, 2068 genes are specific of one tissue, 666 are shared by 2 tissues and 192 are shared by three tissues (Figure 29).



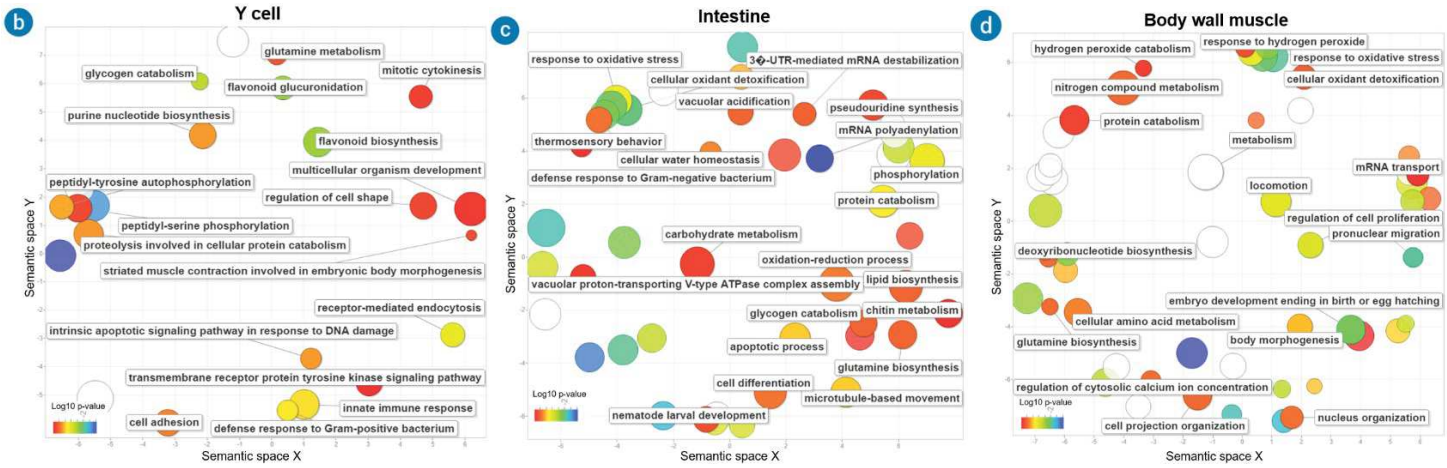


Figure 29. Gene expression distribution per tissue and Gene Ontology analysis. Venn diagram and gene expression distribution between cell types (a). Gene Ontology enrichment plots in Y cell (b). Gene Ontology enrichment plots in intestine (c). Gene Ontology enrichment plots in body wall muscle (d).

Gene Ontology (GO) enrichment analysis on our gene sets showed expected functions for intestinal cells such as vacuolar acidification (157), protein and glycogen catabolism, lipid biosynthesis and carbohydrate metabolism. Gene Ontology enrichment analyses in body wall muscle showed activities such as locomotion, metabolism, glutamine biosynthesis, amino acid biosynthesis and oxidation-reduction processes. To finish, GO enrichments in Y cell showed cell adhesion, metabolism, regulation of cell shape and other common cellular processes as expected for a rectal cell like Y cell during embryo development and also expected during L1 stage. This fact can explain the high difficulty to find specific genes presents only in Y cell and the abundancy of shared genes with other cells of the rectum (Jarriault laboratory, unpublished data).

4. Discussion and outlook.

4.1 NanoDamID, a new methodology to study DNA-protein interactions.

This PhD work provided a new powerful methodology, NanoDamID, to study DNA-protein interactions at different tissue level using ONT sequencing technologies instead of the classical Illumina sequencing. Furthermore, during the development of this project, a new method to carry out DamID in embryonic cell populations was optimized down to the last detail. FACS-NanoDamID combination is currently the most accurate method to isolate and sort a minimal amount of *C. elegans* cells to carry out DamID experiments. Additionally, our Dam RNA polymerase II version (*Dam::rpb-6*) profiling *in vivo* of actively transcribing genes by DamID was a technology never developed in differentiated tissues using *Caenorhabditis elegans* as a model organism. The only precedent using ONT was the full length transcriptome of *C. elegans* using direct RNA sequencing in embryonic cells (158).

Identification of all identified key players during the initiation of Y-to-PDA in Y cell and not in muscle and intestinal cells is a strong evidence to support the accuracy of our methodology. Additionally, body wall muscle and intestinal genes were not expressed in Y cell. Y-to-PDA transdifferentiation initiation occurs at the end of the embryogenesis during the L1 stage (106). Thus, high Dam::RPB-6 methylation levels in Y cell profiles, indicating the expression of those molecular players since Y cell birth, correlate well with cytological observations using fluorescent reporters *in vivo* (Daniele T, Jarriault S), thereby validating our method. Transcription factors (SOX-2, CEH-6, SEM-4, EGL-27 and EGL-5) (110) are involved in the initiation of the transdifferentiation. On the other hand, histone modifiers (SET-1 and JMJD-3.1) are responsible of the robustness of the process (113). The previous biological validation of those genes (110)(113) can be used as a solid argument to consider our findings as a *bona fide* RNA polymerase footprints as all known Y specific markers we identified were previously biologically validated in the Jarriault laboratory (106)(110)(113). The next step to get a complete understanding of how Y-to-PDA initiation is regulated will be characterize the 395 specific Y cell genes identified in this PhD work. We found 15 specific genes encoding transcription factors in Y cell (Table 6). The next step of this research should be then characterizing the importance of the identified genes during natural transdifferentiation initiation and its role in that process. One promising candidate gene could be *odd-1* which is the homologous of *osr-1* in mammals involved in direct transcriptional reprogramming of adult cells to embryonic nephron progenitors (159), however, proper biological validation is required to address that affirmation and additional experiments are needed. Genes of interest such as *odd-1* will be knock-downed using RNAi by either feeding or dsRNA injection. We expect that knocking-down these transcripts will lead to a block in

the epithelial or the transition cell fate, which will be identified using specific fluorescent reporters allowing us to setup a gene regulatory network in combination with genes already identified by Jarriault laboratory. Gain-of-function experiments and detailed analysis of the available mutants will depend greatly of the type of regulator involves to uncover target genes and/or pathways supporting the transdifferentiation process. Those experiments will shed light on natural transdifferentiation initiation.

Y specific genes encoding TF	
nhr-9	*odd-1*
nhr-21	*unc-37*
nhr-31	*F19B2.6*
nhr-47	*nhr-247*
nhr-90	*nhr-30*
nhr-109	*nhr-273*
nhr-115	*nhr-137*
nhr-120	

Table 6. Y specific genes encoding transcription factors.

Additionally, experiments expressing *dam::fusions* in intestine, body wall muscle and XXX adult cells using the *Cre/lox* recombination system to address cell specificity, showed as FACS-DamID described above, clear tissue-specific transcription patterns. In addition, specific genes of each cell type were identified successfully (annex 1). The identified transcription patterns could be considered cleaner or better than using FACS-DamID in embryos for several reasons: first, FACS is not a 100% efficient method, normally we never get more than 90% of purity of our desired cell population. Similar unspecific expression percentages could be expected using recombination systems such as FLP/FRT (140) or *Cre/lox* (141) in some tissues. It was described that FLP/FRT or *Cre/lox* tissue specific expression in vulva (*hlh-8* promoter) or seam cells (*nhr-82* promoter) can be lower than 70%, however for the tissues we studied, intestine and body wall muscle, the specific expression was close to 100% (140)(141). Second, the stochasticity could be higher in the embryonic cells because the DamID experiment started with small cell populations (500 cells) and this fact would lead us to slightly different methylation patterns between each cell type and its replica. To finish, *Cre/lox* experiments were carried out in young adults instead of embryos,

therefore, Dam had more time to methylate a higher number of GATC motifs over the course of complete development, increasing the methylation levels (but then reflecting an accumulation of different methylation patterns over time).

4.2 Temporal control system, current perspectives.

The development of an inducible temporal control system is currently one of the aims of the *Caenorhabditis elegans* community. Currently, auxin inducible degradation system is not fully optimized. Recently, a study based on the degradation speed of different auxin analogs has been published. This study showed differences between degradation of degron-tagged proteins using synthetic auxin 1-naphthaleneacetic acid (NAA) and natural auxin indole-3-acetic acid (IAA) (160) in worms expressing the molecular components of the AID system. In my experiments, I used IAA as was described in the original AID work described by Dernburg laboratory (143). Faster degradations occurs using NAA instead of IAA (160). However, all the results using both auxins highlighted the same issue: a small percentage of the degron-tagged protein is not depleted in the presence of the auxin or its analog. As DamID is highly sensitive to protein levels, the small amounts of remaining Dam fusions are sufficient to methylate a significant number of GATC motifs over the course of animal development, blurring the methylation footprint. This AID limitation using DamID was tested in our studies: we could not see any difference between worms grown on auxin versus worms grown in the absence of auxin. This was the conclusive evidence that the AID system could be useful if the experimental aim is the depletion of a fluorescent protein such as GFP. However, for technique in which the protein needs to be completely absent such as DamID due to the amplifying effect of the enzymatic reaction, this method is pointless: even if the depletion is 95% efficient, trace amounts of Dam::RPB-6 will methylate GATC motifs in active transcribing genes (Annex 2).

In addition to the AID system, another temporal control system is currently in development by the *C. elegans* community. The Greiss laboratory is trying to expand the genetic code of *Caenorhabditis elegans* developing different photo-caged amino acid to create light activatable protein variants. Initial experiments have shown that the addition of photo-caged artificial amino acids included in the CRE recombinase (replacing an amber stop codon) can be reverted by illuminating the worms, thereby inducing the expression of the CRE recombinase and subsequently the desired *dam::fusion*. Once CRE is expressed, *Cre/lox* recombination should work allowing the researcher to express tissue specific *dam::fusions* in a time-controlled manner. However, this technology is still in development and has not been fully tested yet.

4.3 Spatial control system improvements.

Based on our successful double fluorophore gate combining GFP and mCherry fluorophores (Chapter 1.11.3) in Y cell with simultaneous co-expression of those fluorescent proteins under the control of *hlh-16* and *egl-5* promoters respectively, we could think that creating our initial double gate combining recombination systems is feasible. However, there were some potential factors that could make the double gate recombination cascade (Chapter 1.11.1) and the double gate split cGAL cascade inefficient (Chapter 1.11.2). The first and most obvious problem is the positioning of the *hlh-16* intron, which the Jarriault laboratory has found necessary for the expression in the Y cell (MC. Morin, S. Jarriault, unpublished data). This intron was positioned between two CRE exons in the Double recombination system or after CRE exons in the cGAL recombination system. For our double fluorophore version, the GFP fluorophore was co-expressed with the *hlh-16* transcriptional construct (including its intron and its 3'UTR); however, in our cascade we removed *hlh-16* first exon and only used the intron to regulate the expression of CRE under transcriptional control of the *hlh-16* promoter. Maybe this fact and not only the positioning of the intron had been playing an important role in the regulation of CRE expression by the *hlh-16* promoter. One good option to be tested could be the co-expression of CRE with the *hlh-16* gene and its intron to have a successful CRE expression under the control of the *hlh-16* promoter, using endogenous trans-splicing or 2A cleavage. al. In conclusion, there was no reason to think that the variation on *hsp::cre* template sequence could be the cause of our problem to setup the double gate recombination cascade and double gate split cGAL-CRE. The future efforts to get an *in vivo* recombination system expressing *dam::fusions* in Y-to-PDA should be focus on the study of *hlh-16* expression in order to express the second piece of the recombination cascade properly.

4.4 SwitchDamID: Double gate cascade combining heat-shock/cell-specific promoters for spatial and temporal control expression of *dam::fusion* protein.

In order to solve the above-mentioned issues with the inefficiency of the double gate recombination cascade, I proposed at the end of my PhD an additional experiment to setup another double recombination gate system to create at the same time, tissue specificity and time control of our *dam::fusions*, without using the auxin inducible degradation system. The only weakness of that system will be its unavailability to express *dam* on tissues or cells which require more than one tissue specific promoter acting at the same time such as Y-to-PDA. However, it should be a suitable method to express *dam* in many tissues that require only one promoter for tissue-specific expression. This system was thought specially to be used in experiments related to oscillatory gene expression or chronobiology, to study tissue-

specific gene expression at specific moments of the development without accumulating previous Dam methylations, as we wanted to initially achieve using the auxin inducible degradation system (Chapter 2.3.2).

The new system uses the previous double gate recombination cascade as a base to generate tissue specificity and time control expression of *dam::fusions*. This system is divided in two different steps: First, after heat shock, the FLP recombinase will be expressed under the control of a heat shock promoter acting as a switch. Then, after FLP/FRT recombination, the second construct which includes a tissue specific constitutive promoter expressing CRE recombinase will express CRE to recombine the *lox::mCherry::lox* cassette. After the second recombination, Dam will be expressed in the desired tissue. In absence of heat shock, the first part of the cascade will be broken because there will not be expression. Only a heat shock in the desired moment of the development that we would like to study will activate the cascade producing spatial and temporal Dam expression (Figure 30).

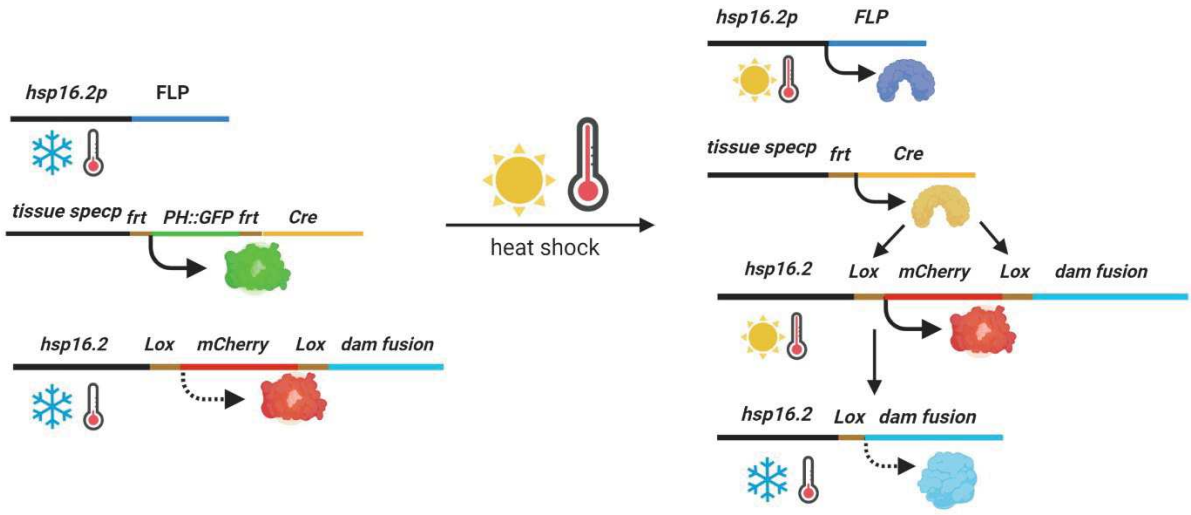


Figure 30. Switch DamID system. After heat shock FLP recombinase is expressed and the cascade switched ON. FLP/FRT recombination will allow the system to express CRE recombining the *lox::mCherry::lox cassette*. The result will be a Dam expressed in the proper moment of the development only in the tissue where CRE is expressed.

This new approach to setup a spatial and temporal control system could have two potential limitations. First, timing between heat shock and proper FLP recombinase expression level to carry out FLP/FRT recombination needs to be adjusted, therefore, timing should be calibrated in order to know when is the best moment to carry out the heat shock according to the developmental stage that one would like to study. Second: After the heat shock, we expected mCherry fluorescent protein overexpression, which is the direct consequence of the heat shock but not *dam::fusion* protein of interest expression. Knowing that a heat shock of 10 minutes at 30ºC could produce expression after 3 hours at room temperature in

hsp::lox::mcherry::lox::gfp::dam strains (MeisterLab/Unpublished data) we must be really careful choosing when is the best moment to heat shock our worms. However, there are reasons to be optimistic: First the heat shock will not produce immediate *dam::fusion* expression because first, the FLP recombinase has to be expressed, then the FRT cassette have to be recombined and once the FRT cassette recombines, the molecular machinery of the *Cre/lox* system has to be active. The time required for that could be more than enough to produce again leaky Dam signal driven by the heat shock promoter (Figure 30). Additionally, the versatility of the *C. elegans* allows us to grow up worms at 15ºC delaying the development as much as possible to slow down the life cycle.

5. Annexes

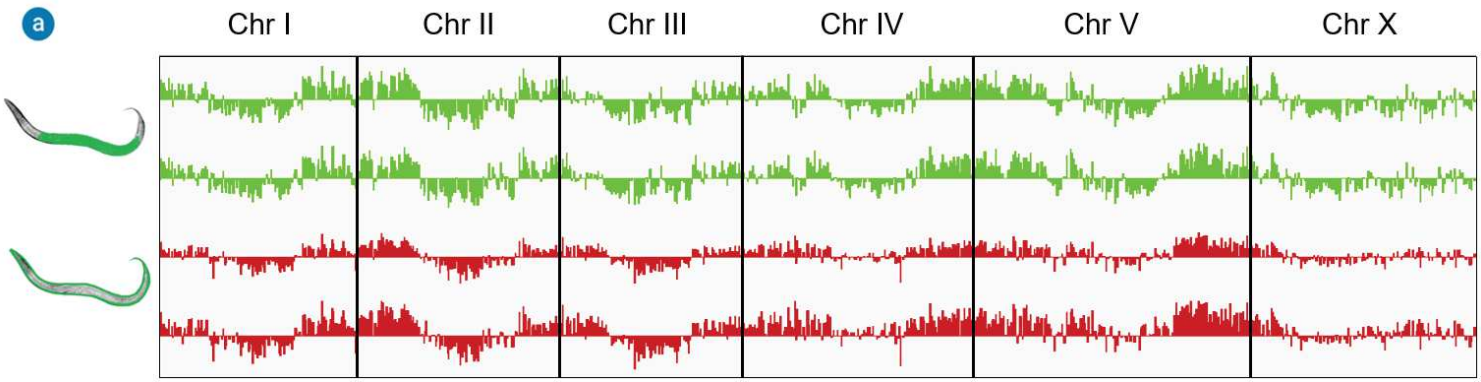
5.1 Annex 1 – Proof of principle using *Cre/lox* recombination system.

We performed several worm crosses using strains carrying single copy insertions (Table 2). Our aim was express cell specific *dam::fusions* by *Cre/lox* recombination. We chose intestine (*elt-2p::cre*), body wall muscle (*myo-3p::cre*) and XXX neurons (*sdf-9p::cre*) as a target tissues for *dam::fusion* expression. For our worm crosses, we used strains expressing *cre* under the control of the mentioned promoters and our *lox::mCherry::lox::dam::fusion* protein-of-interest collection (*dam::rpb-6*, *gfp::dam* and *dam::lmn-1*).

The study of normalized DamID tracks was really promising:

-Experiments using tissue specific *dam::lmn-1*: Expected abundancy of methylation in the periphery of the chromosome and less methylation in the center of the chromosome where the chromatin is more open because of the absence of centromere. Those results showed an expected *dam::lmn-1* methylation (152) being a good positive control of the accuracy of our NanoDamID experiments (Figure 31a).

-Experiments using tissue specific *dam::rpb-6*: Prospective Dam target genes exhibit abundancy of methylation according to their tissue specificity. High methylation was detected in intestine specific genes such as *elt-2*, *asp-1* and *spp-2* (Figure 31b). Specific body wall muscle genes such as *myo-3*, *mlc-3* and *pat-10* also showed high methylation levels (Figure 31b). To finish, XXX neuron specific genes sdf-*9*, *daf-9* and *eak-4* also showed high methylation levels (Figure 31b). According to those tracks, we can consider that our results as a *bona fide* information: methylation present in each specific gene matched with the specific cell type without be present in the other two selected tissues for that study. Furthermore, correlation values showed high degree of similarity when each tissue is compared with its biological replica and lower similarity when different cell types are compared (Figure 31c).



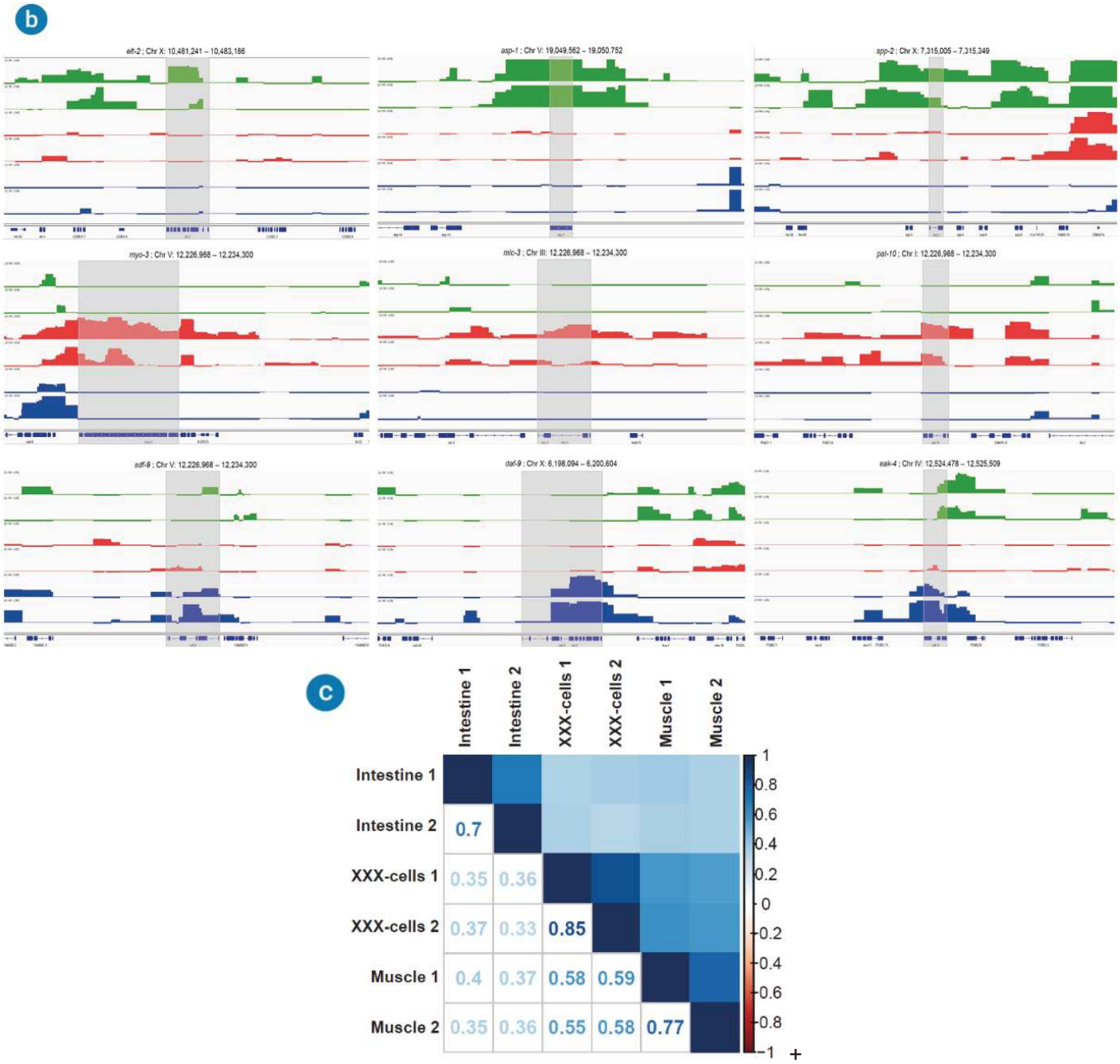


Figure 31. DamID using *dam::lmn-1* and *dam::rpb-6* as a *dam::fusion* proteins expressed in specific tissues by *Cre/lox*. a) Methylation *dam::lmn-1* tracks values expressed in intestine and body wall muscle showing higher interactions in the periphery of the chromosome and less in the center of the chromosome. b) *dam::rpb-6* methylation tracks values expressed in intestine (Green), body wall muscle (Red) and XXX cells (Blue), showing specific methylation in each cell type. c) Pearson/s correlation plot at GATC level of methylated genes by *dam::rpb-6* in muscle intestine and XXX expressed by *Cre/lox* recombination. Scale (a): -2 to +2. Scale (b-c): -0.1 to +2

Comparisons between intestinal, body wall muscle and XXX cells showed 2620 specific genes for one tissue, 1241 genes shared in two tissues and 536 shared in three. On the other hand, Comparisons between the three cell type populations vs worm wide experiments

showed 2509 genes shared in only one tissue, 1273 shared in two tissues, 799 shared in three tissues and 536 shared in four (Figure 32).

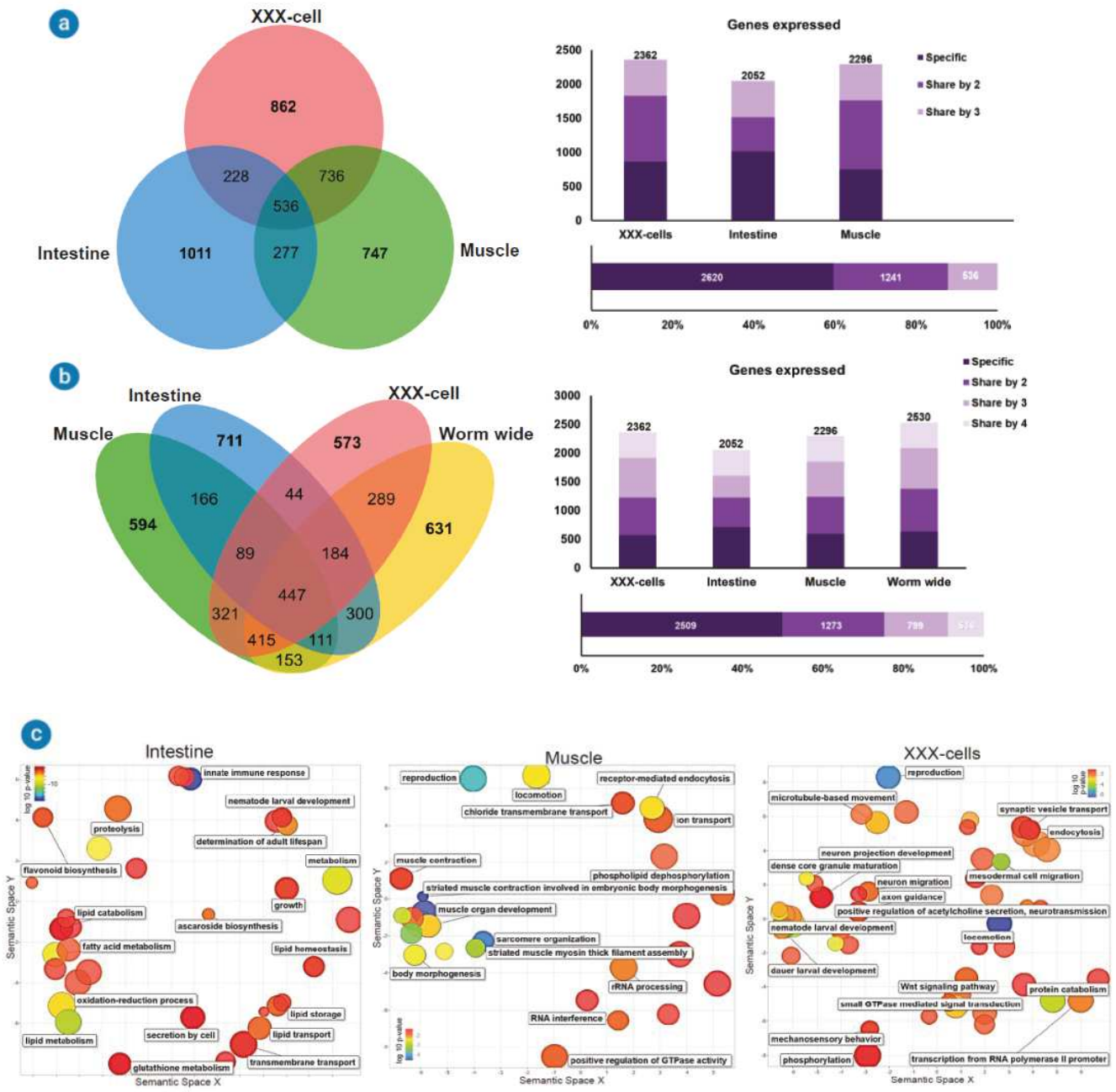


Figure 32. Venn diagrams and Gene Ontology plots.

GO enrichment plots showed specific functions for each cell type according to the expected function (Figure 32): intestine GO plot showed expected activities such as metabolism, oxidation-reduction processes, lipid metabolism, glutathione metabolism, lipid transport, storage and lipid homeostasis. Body wall muscle GO plot in showed locomotion, phospholipid dephosphorylation, striated muscle contraction involved in body

morphogenesis, muscle development and sarcomere organization. To finish, XXX cells GO plots showed neuron projection development, axon guidance and neuron migration.

Altogether, these evidences showed that tissue specific NanoDamID by *Cre/lox* is a suitable method to study transcription in different tissues of the *C. elegans* using a RNA polymerase II profiling active genes during the development.

5.2 Annex 2 – Preliminary experiments to determine the best method to carry out NanoDamID to study Y-to-PDA transdifferentiation.

5.2.1 Temporal control system to express *Dam* in different time points of the development.

Gene mapping since Y cell birth until PDA formation from L1 to L3 stage distinguishing every step of the development was one of our main initial aims. We expressed separately the F-box TIR1 under the control of an ubiquitous *eft-3* promoter expressing a red fluorescent marker and the other part of the system, the auxin induced degron, co-expressed with a *dam::fusion* protein under the control of a heat shock promoter non-heat shock to produce the minimal *Dam* signal. The chosen conditions for our experiments were absence of auxin, 250 nM auxin and 1 mM auxin.

Worms were grown in auxin plates seeded with GM48 bacteria as a source of food (Chapter 2.3.2). In order to avoid a possible *degron::gfp* expression, worms and eggs populations were exposed to auxin all the time, even after bleaching, eggs were incubated in a M9+auxins solution (Chapter 2.3.2) to ensure the protein depletion by AID system. In our experiments we grown up to three generations of nematodes in auxin conditions, F2 progeny was the population used for our experiments. After heat shock worms grown in every condition to overexpress *degron::gfp::dam* activating our heath shock promoter we could observed a clear depletion of the GFP in worms grown in both auxin conditions and non-depletion of the GFP protein in worm grown in absence of auxin. This result showed the efficiency of the system to deplete fluorescent proteins (Figure 33).

For DamID experiments, we started testing the efficiency of the system using isolated DNA from synchronized L1 worms expressing *degron::dam::gfp*. We compared two different conditions: 0 nM and 1mM auxin. Experiments at worm wide level and at tissue specific level (intestine and body wall muscle) were carried out expressing *dam*. After PCR-DamID step, we could easily see the same signal in worms grown without and with auxin exposure. Apparently, 7-8% of the expression of the proteins remains non-depleted (Figure 33). These results showed that the protein depletion is not 100% efficient. Low *dam* expression was enough to achieve DNA methylation even in presence of auxin, not allowing us to carry out time control experiments.

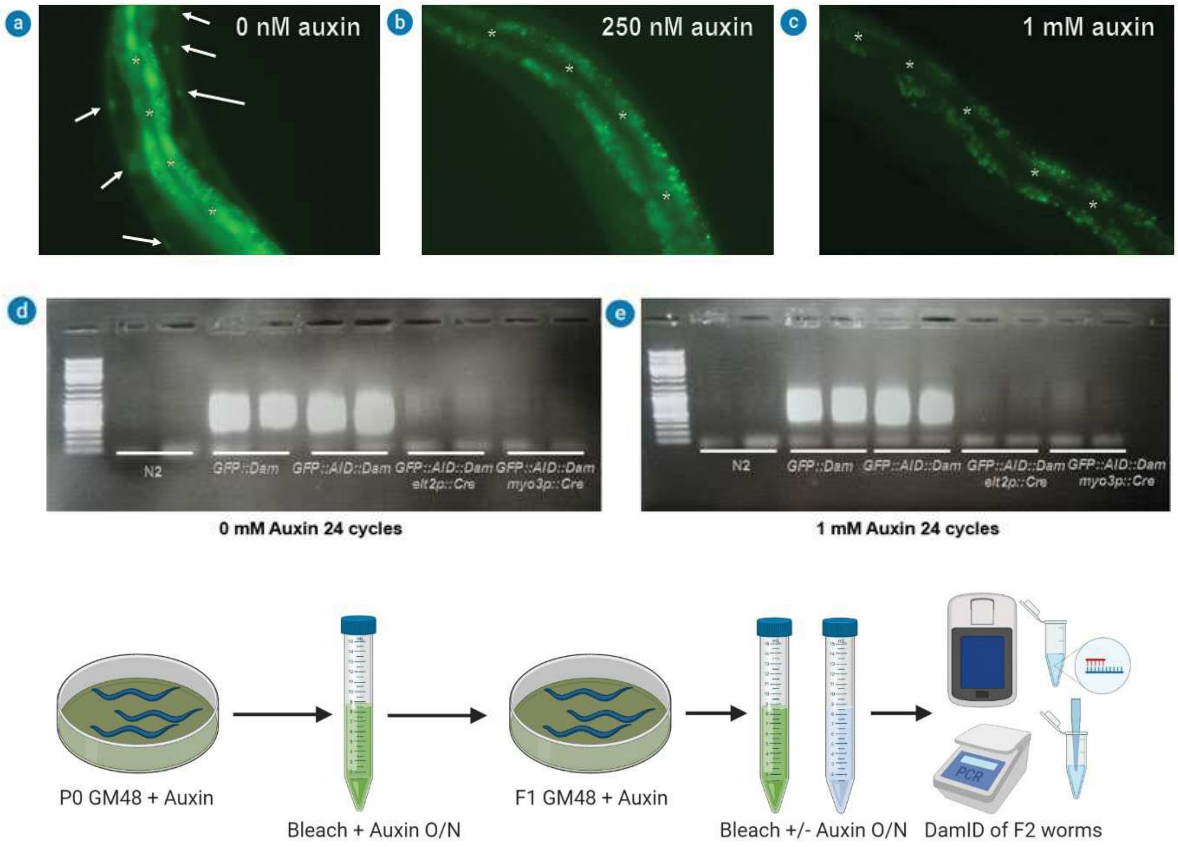


Figure 33. Auxin degradation system (AID) as a temporal control system. 0 nM auxin (a) *degron::gfp::dam* is expressed in the nuclei. 250nM (b) and 1000nM (c) of auxin: GFP fluorescence is not visible because tagged GFP will be ubiquitinated and degraded in the proteasome. Arrows: GFP::Dam fluorescence in the nucleus *: Autofluorescence of the intestine. d) and e) PCR-DamID signal between worm incubated without auxin (0 nM) and auxin (1mM).

5.2.2. Spatial control systems to drive expression of a tissue specific *Dam*.

5.2.2.1 Double recombination cascade expressing FLP/FRT and *Cre/lox*.

We created extrachromosomal-array strains with the three constructs that composed our recombination cascade in order to carry out a quick recombination test to check double recombination efficiency (Chapter 1.11.1). We got a successful strain with the three constructs expressed as an ex-array but the characterization of the system was not really promising. After testing by PCR several populations of worms we did not have a clear result that showed that the recombination of FLP/FRT and *Cre/lox* in cascade is working only in Y cell. The results showed that only the FLP/FRT recombination was successful but something was blocking *cre* expression. We discovered that the problem was a frameshift caused by a wrong *hlh-16* intron sequence in the second construct. Because of that mistake the expression of the CRE was interrupted and the cascade was broken. The issue was fixed but the result remains the same: there was not *cre* expression after FLP/FRT recombination

(Figure 34). We hypothesized that the problem could be the positioning of the *hlh-16* intron responsible to regulate the promoter activity. The same experiment was repeated injecting construct *egl-5* and 2 into a transgenic strain carrying *rps-27p::lox::mcherry::lox::GFP* in order to have a validation by fluorescence in Y cell. The result was also negative. Previously to these assays, the Jarriault laboratory created a worm strain expressing two different fluorophores, mCherry and GFP under the control of two different promoters, *egl-5* and *hlh-16* respectively to create a fluorescent double gate in Y generating yellow fluorescence combining both fluorophores (Chapter 1.11.3). In this version, GFP fluorophore and *hlh-16* gene with its corresponding intron are co-expressed together. A similar approach however, expressing *cre* and *hlh-16* gene + intron was not tested in this PhD work.

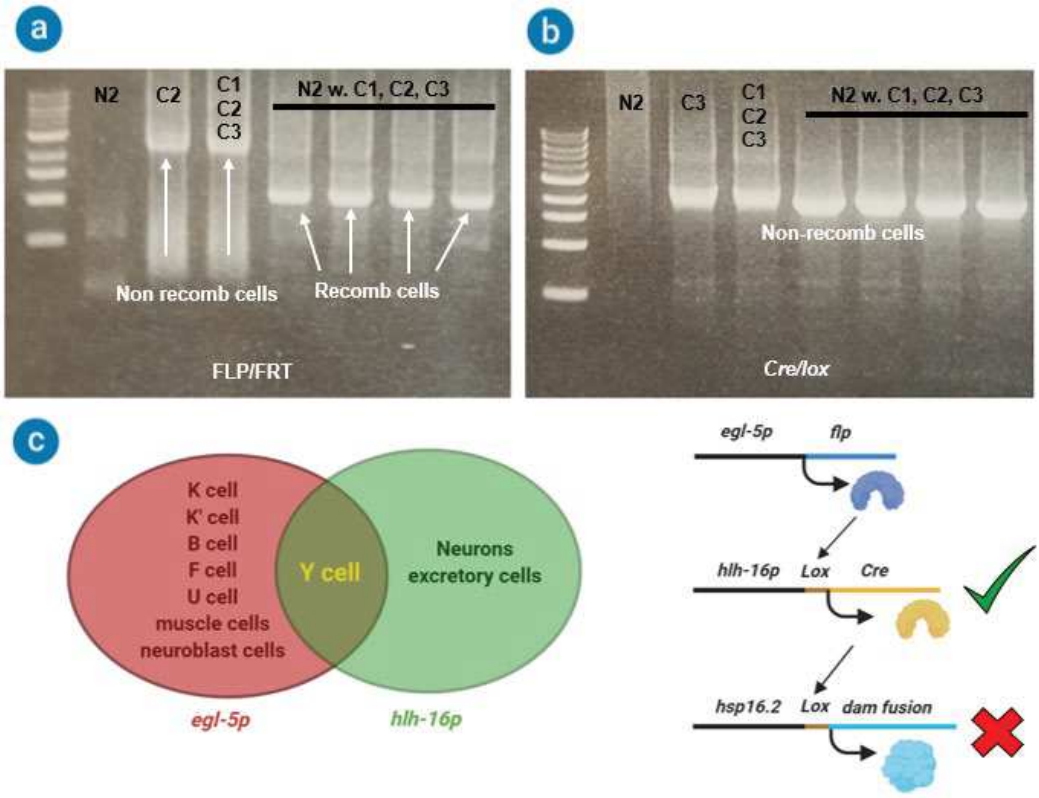


Figure 34. Recombination cascade system. N2: wild type worm; C2 is construct two DNA, C1, C2, C3 are the three constructs; N2 w.C1, C2, C3 are wild type worms injected with the three constructs. The FLP/FRT recombination was successful (a) but not the *Cre/lox* (b). c) Representation of the double gate and the constructs to express both recombinant proteins only in Y.

5.2.2.2 Double gate split cGAL-CRE recombination system.

Because of the initial roadblock with the original cascade design, we tested a second tissue-specific approach using the cGAL system from *Drosophila melanogaster* recently adapted in *C. elegans* (142). Expressing each halve of the GAL4 transcriptional activator under the control of *egl-5* and *hlh-16* promoters we expected to express CRE recombinase only in Y cell under the transcriptional control of the upstream activating sequence. Therefore,

lox::mCherry::lox cassette is recombined and Dam fusions proteins of interest are uniquely activated in Y cell. We followed the same procedure that with the double gate recombination cascade creating extrachromosomal-array strains injecting the three constructs required to make the system works. We generated the arrays injecting a strain carrying *rps-27p::lox::mCherry::lox::GFP* as a single copy insertion. If recombination occurs, the GFP will be expressed only in Y cell and will be visible at the fluorescent scope. Two different tests were carried out: a fluorescent test that was negative because we could not see any fluorescence in Y cell, and a PCR test to check the presence or absence of *lox::mCherry::lox* cassette. Testing recombination by PCR we could observe one lower band that will correspond with the cells without the cassette. Based on that result, we can consider that event as a putative *Cre/lox* recombination, however we expected also an upper band corresponding with cells that does not recombine the cassette. Because of all reasons described in this annex we decided to set aside too sophisticated strategies expressing *in vivo dam::fusions* only in Y cell.

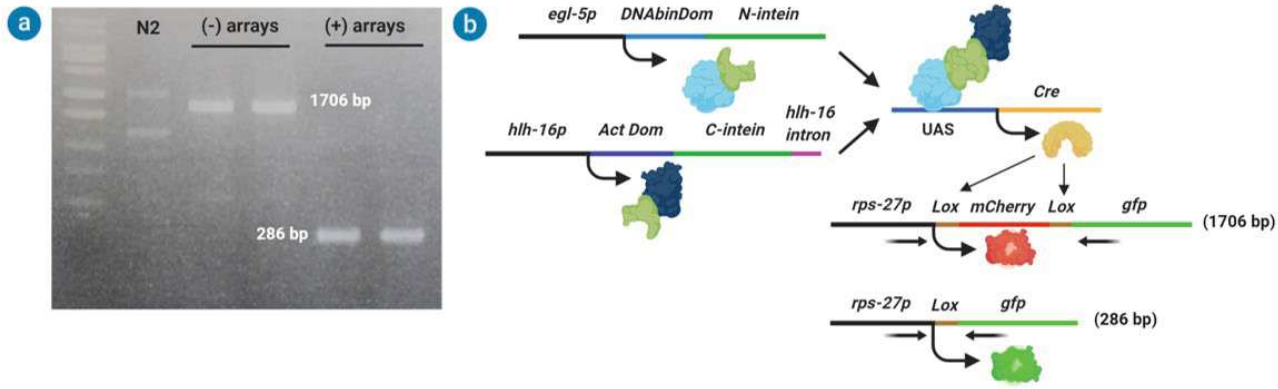


Figure 35. Split cGAL system using a split intein to create a refined spatiotemporal control system in Y/PDA. a) Worm without the array, (-) array did not showed any recombination, in worm with the array the recombination worked successfully. Unfortunately, we do not see any upper band, so all the cells are recombined. b) cGAL/Cre system arrays working on the transgenic strain carrying *rps-27p::lox::mCherry::lox::GFP* as a single copy array.

6. Bibliography

1. J. D. Watson, Crick. F. H. C., Molecular Structures of Nucleic Acids. *Nature* **171**, 737–738 (1953).

2. E. S. Lander, The Heroes of CRISPR. *Cell* (2016) https:/doi.org/10.1016/j.cell.2015.12.041.

3. P. E. Mc Govern, D. L. Glusker, L. J. Exner, P. E. Mc Govern, Neolithic resinated wine. *Nature* (1996) https:/doi.org/10.1038/381480a0.

4. , *Scientific Frontiers in Developmental Toxicology and Risk Assessment* (2000) https:/doi.org/10.17226/9871.

5. E. Abouheif, Developmental genetics and homology: A hierarchical approach. *Trends Ecol. Evol.* (1997) https:/doi.org/10.1016/S0169-5347(97)01125-7.

6. T. A. Eyre, M. W. Wright, M. J. Lush, E. A. Bruford, HCOP: A searchable database of human orthology predictions. *Brief. Bioinform.* (2007) https:/doi.org/10.1093/bib/bbl030.

7. K. Howe, *et al.*, The zebrafish reference genome sequence and its relationship to the human genome. *Nature* (2013) https:/doi.org/10.1038/nature12111.

8. D. D. Shaye, I. Greenwald, Ortholist: A compendium of C. elegans genes with human orthologs. *PLoS One* (2011) https:/doi.org/10.1371/journal.pone.0020085.

9. A. Sánchez Alvarado, S. Yamanaka, Rethinking differentiation: Stem cells, regeneration, and plasticity. *Cell* (2014) https:/doi.org/10.1016/j.cell.2014.02.041.

10. S. Mitalipov, D. Wolf, Totipotency, pluripotency and nuclear reprogramming. *Adv. Biochem. Eng. Biotechnol.* (2009) https:/doi.org/10.1007/10_2008_45.

11. V. E. Prince, R. M. Anderson, G. Dalgin, *Zebrafish Pancreas Development and Regeneration: Fishing for Diabetes Therapies (Zebrafish at the Interface of Development and Disease Research - Chapter Seven)* (2017) https:/doi.org/10.1016/bs.ctdb.2016.10.005.

12. J. Rothman, S. Jarriault, Developmental plasticity and cellular reprogramming in caenorhabditis elegans. *Genetics* (2019) https:/doi.org/10.1534/genetics.119.302333.

13. A. Weismann, W. N. Parker, H. Ronnfeldt, The Germ-Plasm: A Theory of Heredity. *J. Anthropol. Inst. Gt. Britain Irel.* (1894) https:/doi.org/10.2307/2842101.

14. E. H. P. Haeckel, *Zur Entwicklungsgeschichte der Siphonophoren* (1869).

15. E. N. Browne, The production of new hydranths in Hydra by the insertion of small grafts. *J. Exp. Zool.* (1909) https:/doi.org/10.1002/jez.1400070102.

16. H. Spemann, H. Mangold, über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. *Arch. für Mikroskopische Anat. und Entwicklungsmechanik* (1924) https:/doi.org/10.1007/BF02108133.

17. Colucci, No Title. *Sulla rigenerazione parziale dell'occhio nei Tritoni- Istogenesi e Svilupp. Stud. Sper.*

18. Wolff, No Title. *Entwicklungsphysiologische Stud. I. Die Regen. der Urodelenlinse G. Wolff*.

19. J. C. R. Mario Ivankovic, Radmila Haneckova, Albert Thommen, Markus A. Grohme, Miquel Vila-Farré, Steffen Werner, Model systems for regeneration: planarians. *Development* (2019).

20. K. A. Gurley, *et al.*, Expression of secreted Wnt pathway components reveals unexpected complexity of the planarian amputation response. *Dev. Biol.* (2010) https:/doi.org/10.1016/j.ydbio.2010.08.007.

21. P. W. Reddien, A. L. Bermange, A. M. Kicza, A. Sánchez Alvarado, BMP signaling regulates the dorsal planarian midline and is needed for asymmetric regeneration. *Development* (2007) https:/doi.org/10.1242/dev.007138.

22. S. Piraino, D. De Vito, J. Schmich, J. Bouillon, F. Boero, Reverse development in Cnidaria. *Can. J. Zool.* (2004) https:/doi.org/10.1139/Z04-174.

23. S. Piraino, F. Boero, B. Aeschbach, V. Schmid, Reversing the life cycle: Medusae transforming into polyps and cell transdifferentiation in Turritopsis nutricula (Cnidaria, Hydrozoa). *Biol. Bull.* (1996) https:/doi.org/10.2307/1543022.

24. J. Schmich, *et al.*, Induction of reverse development in two marine Hydrozoans. *Int. J. Dev. Biol.* (2007) https:/doi.org/10.1387/ijdb.062152js.

25. E. C. Carla', P. Pagliara, S. Piraino, F. Boero, L. Dini, Morphological and ultrastructural analysis of Turritopsis nutricula during life cycle reversal. *Tissue Cell* (2003) https:/doi.org/10.1016/S0040-8166(03)00028-4.

26. R. D. Ruppert, Edward E.; Fox, Richard, S.; Barnes, *Invertebrate Zoology, 7th edition. Cengage Learning* (2004).

27. E. M. Tanaka, Regenerating tissues. *Science (80-.).* (2018) https:/doi.org/10.1126/science.aat4588.

28. J. Holtfreter, A study of the mechanics of gastrulation. *J. Exp. Zool.* (1944) https:/doi.org/10.1002/jez.1400950203.

29. J. Holtfreter, Neural induction in explants which have passed through a sublethal cytolysis. *J. Exp. Zool.* (1947) https:/doi.org/10.1002/jez.1401060205.

30. R. Briggs, T. J. King, Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proc. Natl. Acad. Sci.* (1952) https:/doi.org/10.1073/pnas.38.5.455.

31. J. B. Gurdon, T. R. Elsdale, M. Fischberg, Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. *Nature* (1958) https:/doi.org/10.1038/182064a0.

32. K. H. S. Campbell, J. McWhir, W. A. Ritchie, I. Wilmut, Sheep cloned by nuclear transfer from a cultured cell line. *Nature* (1996) https:/doi.org/10.1038/380064a0.

33. I. A. Polejaeva, *et al.*, Cloned pigs produced by nuclear transfer from adult somatic cells. *Nature* (2000) https:/doi.org/10.1038/35024082.

34. K. Okita, T. Ichisaka, S. Yamanaka, Generation of germline-competent induced pluripotent stem cells. *Nature* (2007) https:/doi.org/10.1038/nature05934.

35. K. Takahashi, S. Yamanaka, Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* (2006) https:/doi.org/10.1016/j.cell.2006.07.024.

36. P. Hou, *et al.*, Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science (80-.).* (2013) https:/doi.org/10.1126/science.1239278.

37. W. Li, *et al.*, Generation of Rat and Human Induced Pluripotent Stem Cells by Combining Genetic Reprogramming and Chemical Inhibitors. *Cell Stem Cell* (2009)

https:/doi.org/10.1016/j.stem.2008.11.014.

38. S. Sekiya, A. Suzuki, Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* (2011) https:/doi.org/10.1038/nature10263.

39. T. Vierbuchen, *et al.*, Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* (2010) https:/doi.org/10.1038/nature08797.

40. Q. Zhou, J. Brown, A. Kanarek, J. Rajagopal, D. A. Melton, In vivo reprogramming of adult pancreatic exocrine cells to β-cells. *Nature* (2008) https:/doi.org/10.1038/nature07314.

41. C. Jopling, S. Boue, J. C. I. Belmonte, Dedifferentiation, transdifferentiation and reprogramming: Three routes to regeneration. *Nat. Rev. Mol. Cell Biol.* (2011) https:/doi.org/10.1038/nrm3043.

42. K. Kikuchi, *et al.*, Primary contribution to zebrafish heart regeneration by gata4+ cardiomyocytes. *Nature* (2010) https:/doi.org/10.1038/nature08804.

43. A. Raya, *et al.*, Activation of Notch signaling pathway precedes heart regeneration in zebrafish. *Proc. Natl. Acad. Sci. U. S. A.* (2003) https:/doi.org/10.1073/pnas.1834204100.

44. K. D. Poss, L. G. Wilson, M. T. Keating, Heart regeneration in zebrafish. *Science (80-.).* (2002) https:/doi.org/10.1126/science.1077857.

45. T. Gerber, *et al.*, Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. *Science (80-.).* (2018) https:/doi.org/10.1126/science.aaq0681.

46. D. L. Stocum, Mechanisms of urodele limb regeneration. *Regeneration* (2017) https:/doi.org/10.1002/reg2.92.

47. R. Mirsky, K. Jessen, "Early events in Schwann cell development" in *The Biology of Schwann Cells: Development, Differentiation and Immunomodulation*, (2007) https:/doi.org/10.1017/CBO9780511541605.003.

48. R. Mirsky, *et al.*, Novel signals controlling embryonic Schwann cell development, myelination and dedifferentiation in *Journal of the Peripheral Nervous System*, (2008) https:/doi.org/10.1111/j.1529-8027.2008.00168.x.

49. K. R. Jessen, W. D. Richardson, S. S. Scherer, J. L. Salzer, *Axon–Schwann cell interactions during peripheral nerve degeneration and regeneration* (2012).

50. P. A. Tsonis, M. Madhavan, E. E. Tancous, K. Del Rio-Tsonis, A newt's eye view of lens regeneration. *Int. J. Dev. Biol.* (2004) https:/doi.org/10.1387/ijdb.041867pt.

51. J. B. L. Bard, Waddington's Legacy to Developmental and Theoretical Biology. *Biol. Theory* (2008) https:/doi.org/10.1162/biot.2008.3.3.188.

52. S. Yamanaka, H. M. Blau, Nuclear reprogramming to a pluripotent state by three approaches. *Nature* (2010) https:/doi.org/10.1038/nature09229.

53. V. Kashyap, *et al.*, Regulation of Stem cell pluripotency and differentiation involves a mutual regulatory circuit of the Nanog, OCT4, and SOX2 pluripotency transcription factors with polycomb Repressive Complexes and Stem Cell microRNAs. *Stem Cells Dev.* (2009) https:/doi.org/10.1089/scd.2009.0113.

54. M. Pardo, *et al.*, An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. *Cell Stem Cell* (2010) https:/doi.org/10.1016/j.stem.2010.03.004.

55. D. L. C. van den Berg, *et al.*, An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. *Cell Stem Cell* (2010) https:/doi.org/10.1016/j.stem.2010.02.014.

56. I. Chambers, *et al.*, Nanog safeguards pluripotency and mediates germline development. *Nature* (2007) https:/doi.org/10.1038/nature06403.

57. J. Silva, *et al.*, Nanog Is the Gateway to the Pluripotent Ground State. *Cell* (2009) https:/doi.org/10.1016/j.cell.2009.07.039.

58. P. Ellis, *et al.*, SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Dev. Neurosci.* (2004) https:/doi.org/10.1159/000082134.

59. H. Kondoh, Y. Kamachi, SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int. J. Biochem. Cell Biol.* (2010) https:/doi.org/10.1016/j.biocel.2009.09.003.

60. Y. H. Loh, *et al.*, The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* (2006) https:/doi.org/10.1038/ng1760.

61. T. Kuroda, *et al.*, Octamer and Sox Elements Are Required for Transcriptional cis Regulation of Nanog Gene Expression. *Mol. Cell. Biol.* (2005) https:/doi.org/10.1128/mcb.25.6.2475-2485.2005.

62. M. Nakagawa, *et al.*, Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* (2008) https:/doi.org/10.1038/nbt1374.

63. D. Huangfu, *et al.*, Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat. Biotechnol.* (2008) https:/doi.org/10.1038/nbt.1502.

64. A. Haase, *et al.*, Generation of Induced Pluripotent Stem Cells from Human Cord Blood. *Cell Stem Cell* (2009) https:/doi.org/10.1016/j.stem.2009.08.021.

65. J. Kim, *et al.*, Direct reprogramming of mouse fibroblasts to neural progenitors. *Proc. Natl. Acad. Sci. U. S. A.* (2011) https:/doi.org/10.1073/pnas.1103113108.

66. J. C. D. Heng, *et al.*, The Nuclear Receptor Nr5a2 Can Replace Oct4 in the Reprogramming of Murine Somatic Cells to Pluripotent Cells. *Cell Stem Cell* (2010) https:/doi.org/10.1016/j.stem.2009.12.009.

67. H. Xie, M. Ye, R. Feng, T. Graf, Stepwise reprogramming of B cells into macrophages. *Cell* (2004) https:/doi.org/10.1016/S0092-8674(04)00419-2.

68. J. Nichols, *et al.*, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* (1998) https:/doi.org/10.1016/S0092-8674(00)81769-9.

69. C. H. Waddington, Towards a theoretical biology. *Nature* (1968) https:/doi.org/10.1038/218525a0.

70. C. T. Wu, J. R. Morris, Genes, genetics, and epigenetics: A correspondence. *Science (80-.).* (2001) https:/doi.org/10.1126/science.293.5532.1103.

71. P. Meister, B. D. Towbin, B. L. Pike, A. Ponti, S. M. Gasser, The spatial dynamics of tissue-specific promoters during C. elegans development. *Genes Dev.* (2010) https:/doi.org/10.1101/gad.559610.

72. A. Gonzalez-Sandoval, *et al.*, Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in C. elegans Embryos. *Cell* (2015)

https:/doi.org/10.1016/j.cell.2015.10.066.

73. J. Chen, *et al.*, H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* (2013) https:/doi.org/10.1038/ng.2491.

74. R. R. E. Williams, *et al.*, Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J. Cell Sci.* (2006) https:/doi.org/10.1242/jcs.02727.

75. R. Lister, *et al.*, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* (2009) https:/doi.org/10.1038/nature08514.

76. E. L. Greer, *et al.*, DNA methylation on N6-adenine in C. elegans. *Cell* (2015) https:/doi.org/10.1016/j.cell.2015.04.005.

77. T. Kouzarides, Chromatin Modifications and Their Function. *Cell* (2007) https:/doi.org/10.1016/j.cell.2007.02.005.

78. F. Chen, H. Kan, V. Castranova, "Methylation of lysine 9 of histone H3: Role of heterochromatin modulation and tumorigenesis††disclaimer:" in *Handbook of Epigenetics*, (2011) https:/doi.org/10.1016/B978-0-12-375709-8.00010-1.

79. A. Barski, *et al.*, High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* (2007) https:/doi.org/10.1016/j.cell.2007.05.009.

80. Y. Liu, *et al.*, RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes Dev.* (2007) https:/doi.org/10.1101/gad.1544207.

81. L. A. Boyer, *et al.*, Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* (2006) https:/doi.org/10.1038/nature04733.

82. S. D. Fouse, *et al.*, Promoter CpG Methylation Contributes to ES Cell Gene Regulation in Parallel with Oct4/Nanog, PcG Complex, and Histone H3 K4/K27 Trimethylation. *Cell Stem Cell* (2008) https:/doi.org/10.1016/j.stem.2007.12.011.

83. S. Dal-Pra, C. P. Hodgkinson, M. Mirotsou, I. Kirste, V. J. Dzau, Demethylation of H3K27 Is Essential for the Induction of Direct Cardiac Reprogramming by MIR Combo. *Circ. Res.* (2017) https:/doi.org/10.1161/CIRCRESAHA.116.308741.

84. C. Sawan, Z. Herceg, *Histone Modifications and Cancer* (2010) https:/doi.org/10.1016/B978-0-12-380866-0.60003-4.

85. B. E. Bernstein, *et al.*, A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* (2006) https:/doi.org/10.1016/j.cell.2006.02.041.

86. R. J. Klose, E. M. Kallin, Y. Zhang, JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* (2006) https:/doi.org/10.1038/nrg1945.

87. Y. H. Loh, W. Zhang, X. Chen, J. George, H. H. Ng, Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* (2007) https:/doi.org/10.1101/gad.1588207.

88. M. K. Seymour, K. A. Wright, C. C. Doncaster, The action of the anterior feeding apparatus of Caenorhabditis elegans (Nematoda: Rhabditida). *J. Zool.* (1983) https:/doi.org/10.1111/j.1469-7998.1983.tb05074.x.

89. H. Schulenburg, M. A. Félix, The natural biotic environment of Caenorhabditis elegans. *Genetics* (2017) https:/doi.org/10.1534/genetics.116.195511.

90. J. R. Chasnov, K. L. Chow, Why are there males in the hermaphroditic species Caenorhabditis elegans? *Genetics* (2002).

91. J. Hodgkin, Two types of sex determination in a nematode. *Nature* (1983) https:/doi.org/10.1038/304267a0.

92. J. E. Sulston, H. R. Horvitz, Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev. Biol.* (1977) https:/doi.org/10.1016/0012-1606(77)90158-0.

93. J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* (1983) https:/doi.org/10.1016/0012-1606(83)90201-4.

94. S. Brenner, The genetics of Caenorhabditis elegans. *Genetics* (1974).

95. L. Byerly, R. C. Cassada, R. L. Russell, The life cycle of the nematode Caenorhabditis elegans. I. Wild-type growth and reproduction. *Dev. Biol.* (1976) https:/doi.org/10.1016/0012-1606(76)90119-6.

96. , Genome sequence of the nematode C. elegans: A platform for investigating biology. *Science (80-.).* (1998) https:/doi.org/10.1126/science.282.5396.2012.

97. C. H. Lai, C. Y. Chou, L. Y. Ch'ang, C. S. Liu, W. C. Lin, Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics. *Genome Res.* (2000) https:/doi.org/10.1101/gr.10.5.703.

98. A. K. Corsi, B. Wightman, M. Chalfie, A transparent window into biology: A primer on Caenorhabditis elegans. *Genetics* (2015) https:/doi.org/10.1534/genetics.115.176099.

99. T. Fukushige, M. Krause, The myogenic potency of HLH-1 reveals wide-spread developmental plasticity in early C. elegans embryos. *Development* (2005) https:/doi.org/10.1242/dev.01774.

100. M. Krause, A. Fire, S. W. Harrison, J. Priess, H. Weintraub, CeMyoD accumulation defines the body wall muscle cell fate during C. elegans embryogenesis. *Cell* (1990) https:/doi.org/10.1016/0092-8674(90)90494-Y.

101. J. Zhu, T. Fukushige, J. D. McGhee, J. H. Rothman, Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. *Genes Dev.* (1998) https:/doi.org/10.1101/gad.12.24.3809.

102. S. Quintin, G. Michaux, L. McMahon, A. Gansmuller, M. Labouesse, The Caenorhabditis elegans gene lin-26 can trigger epithelial differentiation without conferring tissue specificity. *Dev. Biol.* (2001) https:/doi.org/10.1006/dbio.2001.0294.

103. J. S. Gilleard, J. D. McGhee, Activation of Hypodermal Differentiation in the Caenorhabditis elegans Embryo by GATA Transcription Factors ELT-1 and ELT-3. *Mol. Cell. Biol.* (2001) https:/doi.org/10.1128/mcb.21.7.2533-2544.2001.

104. M. A. Horner, *et al.*, pha-4, an HNF-3 homolog, specifies pharyngeal organ identity in Caenorhabditis elegans. *Genes Dev.* (1998) https:/doi.org/10.1101/gad.12.13.1947.

105. S. Zuryn, T. Daniele, S. Jarriault, Direct cellular reprogramming in Caenorhabditis elegans: Facts, models, and promises for regenerative medicine. *Wiley Interdiscip. Rev. Dev. Biol.* (2012) https:/doi.org/10.1002/wdev.7.

106. S. Jarriault, Y. Schwab, I. Greenwald, A Caenorhabditis elegans model for epithelial-neuronal transdifferentiation. *Proc. Natl. Acad. Sci. U. S. A.* (2008)

https:/doi.org/10.1073/pnas.0712159105.

107. J. P. Richard, *et al.*, Direct in vivo cellular reprogramming involves transition through discrete, non-pluripotent steps. *Development* (2011) https:/doi.org/10.1242/dev.063115.

108. M. Sammut, *et al.*, Glia-derived neurons are required for sex-specific learning in C. Elegans. *Nature* (2015) https:/doi.org/10.1038/nature15700.

109. J. G. White, E. Southgate, J. N. Thomson, S. Brenner, The Structure of the Nervous System of the Nematode Caenorhabditis elegans. *Philos. Trans. R. Soc. B Biol. Sci.* (1986) https:/doi.org/10.1098/rstb.1986.0056.

110. K. Kagias, A. Ahier, N. Fischer, S. Jarriault, Members of the NODE (Nanog and Oct4-associated deacetylase) complex and SOX-2 promote the initiation of a natural cellular reprogramming event in vivo. *Proc. Natl. Acad. Sci. U. S. A.* (2012) https:/doi.org/10.1073/pnas.1117031109.

111. S. Zuryn, S. Le Gras, K. Jamet, S. Jarriault, A strategy for direct mapping and identification of mutations by whole-genome sequencing. *Genetics* (2010) https:/doi.org/10.1534/genetics.110.119230.

112. S. Marro, *et al.*, Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* (2011) https:/doi.org/10.1016/j.stem.2011.09.002.

113. S. Zuryn, *et al.*, Sequential histone-modifying activities determine the robustness of transdifferentiation. *Science (80-.).* (2014) https:/doi.org/10.1126/science.1255885.

114. A. M, G. B, C. MA, K. JM, M. JD, A fork head/HNF-3 homolog expressed in the pharynx and intestine of the Caenorhabditis elegans embryo. *Dev. Biol.* (1996).

115. D. H. Park, *et al.*, Activation of Neuronal Gene Expression by the JMJD3 Demethylase Is Required for Postnatal and Adult Brain Neurogenesis. *Cell Rep.* (2014) https:/doi.org/10.1016/j.celrep.2014.07.060.

116. O. Hobert, Neurogenesis in the nematode Caenorhabditis elegans. *WormBook* (2010) https:/doi.org/10.1895/wormbook.1.12.2.

117. S. Shaham, Glial development and function in the nervous system of Caenorhabditis elegans. *Cold Spring Harb. Perspect. Biol.* (2015) https:/doi.org/10.1101/cshperspect.a020578.

118. I. Abdus-Saboor, *et al.*, Notch and Ras promote sequential steps of excretory tube development in C. elegans. *Development* (2011) https:/doi.org/10.1242/dev.068148.

119. M. V. Sundaram, M. Buechner, The caenorhabditis elegans excretory system: A model for tubulogenesis, cell fate specification, and plasticity. *Genetics* (2016) https:/doi.org/10.1534/genetics.116.189357.

120. A. Hecht, M. Grunstein, Mapping DNA interaction sites of chromosomal proteins using immunoprecipitation and polymerase chain reaction. *Methods Enzymol.* (1999) https:/doi.org/10.1016/S0076-6879(99)04024-0.

121. M. H. Kuo, C. D. Allis, In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods A Companion to Methods Enzymol.* (1999) https:/doi.org/10.1006/meth.1999.0879.

122. B. Van Steensel, S. Henikoff, Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* (2000)

https:/doi.org/10.1038/74487.

123. J. Redolfi, *et al.*, DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat. Struct. Mol. Biol.* (2019) https:/doi.org/10.1038/s41594-019-0231-0.

124. V. J. Simpson, T. E. Johnson, R. F. Hammen, Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res.* (1986) https:/doi.org/10.1093/nar/14.16.6711.

125. A. Bird, DNA methylation patterns and epigenetic memory. *Genes Dev.* (2002) https:/doi.org/10.1101/gad.947102.

126. R. Sharma, D. Ritler, P. Meister, Tools for DNA adenine methyltransferase identification analysis of nuclear organization during C. elegans development. *Genesis* (2016) https:/doi.org/10.1002/dvg.22925.

127. J. Kind, *et al.*, Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. *Cell* (2015) https:/doi.org/10.1016/j.cell.2015.08.040.

128. G. N. Aughey, T. D. Southall, Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdiscip. Rev. Dev. Biol.* (2016) https:/doi.org/10.1002/wdev.205.

129. T. I. Lee, R. A. Young, Transcription of Eukaryotic Protein-Coding Genes. *Annu. Rev. Genet.* (2000) https:/doi.org/10.1146/annurev.genet.34.1.77.

130. R. D. Kornberg, The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U. S. A.* (2007) https:/doi.org/10.1073/pnas.0704138104.

131. K. Gupta, D. Sari-Ak, M. Haffke, S. Trowitzsch, I. Berger, Zooming in on Transcription Preinitiation. *J. Mol. Biol.* (2016) https:/doi.org/10.1016/j.jmb.2016.04.003.

132. S. Sainsbury, C. Bernecky, P. Cramer, Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* (2015) https:/doi.org/10.1038/nrm3952.

133. P. Cramer, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: RNA polymerase II at 2.8 ångstrom resolution. *Science (80-.).* (2001) https:/doi.org/10.1126/science.1059493.

134. M. Cojocaru, *et al.*, Genomic location of the human RNA polymerase II general machinery: Evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription. *Biochem. J.* (2008) https:/doi.org/10.1042/BJ20070751.

135. D. B. Nikolov, S. K. Burley, RNA polymerase II transcription initiation: A structural view. *Proc. Natl. Acad. Sci. U. S. A.* (1997) https:/doi.org/10.1073/pnas.94.1.15.

136. V. Grishkevich, T. Hashimshony, I. Yanai, Core promoter T-blocks correlate with gene expression levels in C. elegans. *Genome Res.* (2011) https:/doi.org/10.1101/gr.113381.110.

137. J. A. Powell-Coffman, J. Knight, W. B. Wood, Onset of C. elegans gastrulation is blocked by inhibition of embryonic transcription with an RNA polymerase antisense RNA. *Dev. Biol.* (1996) https:/doi.org/10.1006/dbio.1996.0232.

138. G. Seydoux, M. A. Dunn, Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. *Development* (1997).

139. M. Zhong, *et al.*, Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. *PLoS Genet.* (2010) https:/doi.org/10.1371/journal.pgen.1000848.

140. C. Muñoz-Jiménez, *et al.*, An efficient flp-based toolkit for spatiotemporal control of gene expression in caenorhabditis elegans. *Genetics* (2017) https:/doi.org/10.1534/genetics.117.201012.

141. S. Ruijtenberg, S. Van Den Heuvel, G1/S Inhibitors and the SWI/SNF Complex Control Cell-Cycle Exit during Muscle Differentiation. *Cell* (2015) https:/doi.org/10.1016/j.cell.2015.06.013.

142. H. Wang, J. Liu, K. P. Yuet, A. J. Hill, P. W. Sternberg, Split cGAL, an intersectional strategy using a split intein for refined spatiotemporal transgene control in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* (2018) https:/doi.org/10.1073/pnas.1720063115.

143. L. Zhang, J. D. Ward, Z. Cheng, A. F. Dernburg, The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. *Dev.* (2015) https:/doi.org/10.1242/dev.129635.

144. C. Frøkjær-Jensen, *et al.*, Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. *Nat. Methods* (2014) https:/doi.org/10.1038/nmeth.2889.

145. C. Frøkjær-Jensen, *et al.*, Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat. Genet.* (2008) https:/doi.org/10.1038/ng.248.

146. O. J. Marshall, A. H. Brand, Damidseq-pipeline: An automated pipeline for processing DamID sequencing datasets. *Bioinformatics* (2015) https:/doi.org/10.1093/bioinformatics/btv386.

147. C. C. Mello, *et al.*, The PIE-1 protein and germline specification in C. elegans embryos. *Nature* (1996) https:/doi.org/10.1038/382710a0.

148. J. Liu, *et al.*, Essential roles for Caenorhabditis elegans lamin gene in nuclear organization, cell cycle progression, and spatial organization of nuclear pore complexes. *Mol. Biol. Cell* (2000) https:/doi.org/10.1091/mbc.11.11.3937.

149. K. L. McNally, F. J. McNally, Fertilization initiates the transition from anaphase I to metaphase II during female meiosis in C. elegans. *Dev. Biol.* (2005) https:/doi.org/10.1016/j.ydbio.2005.03.009.

150. A. Singson, Every sperm is sacred: Fertilization in Caenorhabditis elegans. *Dev. Biol.* (2001) https:/doi.org/10.1006/dbio.2000.0118.

151. F. A. Steiner, S. Henikoff, Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *Elife* (2014) https:/doi.org/10.7554/eLife.02025.

152. R. Sharma, *et al.*, Differential spatial and structural organization of the X chromosome underlies dosage compensation in C. elegans. *Genes Dev.* (2014) https:/doi.org/10.1101/gad.248864.114.

153. B. D. Towbin, *et al.*, Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* (2012) https:/doi.org/10.1016/j.cell.2012.06.051.

154. K. Ikegami, T. A. Egelhofer, S. Strome, J. D. Lieb, Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biol.* (2010) https:/doi.org/10.1186/gb-2010-11-12-r120.

155. R. S. Kamath, *et al.*, Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* (2003) https:/doi.org/10.1038/nature01278.

156. Y. Zhang, D. Chen, M. A. Smith, B. Zhang, X. Pan, Selection of reliable reference genes in caenorhabditis elegans for analysis of nanotoxicity. *PLoS One* (2012) https:/doi.org/10.1371/journal.pone.0031849.

157. K. Baintner, Demonstration of acidity in intestinal vacuoles of the suckling rat and pig. *J. Histochem. Cytochem.* (1994) https:/doi.org/10.1177/42.2.7507141.

158. N. P. Roach, *et al.*, The full-length transcriptome of C. elegans using direct RNA sequencing. *Genome Res.* (2020) https:/doi.org/10.1101/gr.251314.119.

159. C. E. Hendry, *et al.*, Direct transcriptional reprogramming of adult cells to embryonic nephron progenitors. *J. Am. Soc. Nephrol.* (2013) https:/doi.org/10.1681/ASN.2012121143.

160. M. A. Q. Martinez, *et al.*, Rapid degradation of Caenorhabditis elegans proteins at single-cell resolution with a synthetic auxin. *G3 Genes, Genomes, Genet.* (2020) https:/doi.org/10.1534/g3.119.400781.

7. Acknowledgements

I would like to dedicate this PhD work to my parents, Francisca Luque and Manuel Osuna, to all my family, friends and also to all the people I met during those four fascinating years. I felt all of you as a part of a great family. I will never thank you enough all the experiences I lived during those years with all of you.

Special thanks to my two supervisors Peter Meister and Sophie Jarriault for giving me the opportunity to work in their labs during those years. Peter, thank you so much to transfer me your passion for the most cutting-edge science and technology and guide me in all the troubleshooting I had during my PhD, all I learnt with you about molecular biology is going to be useful the rest of my career as a scientist wherever I go, thank to you, I made omics my work but also my passion. Sophie, thank you to transmit me your fascination for the science and the academia, your calm and your serenity even in the most uncertain moments of this PhD project and specially for believe in all the ideas I had during those four years and give me free rein to carry out them. I cannot quantify how much I learnt from both of you.

I would like to thank to my external supervisors, Peter Askjaer and Francesca Palladino to revise my PhD, also to my internal supervisor in Strasbourg, László Tora and my mentor in Bern Benoît Zuber.

I would like to thank all the past and present members of Meister and Jarriault laboratories. Here, I would like to mention Julie Campos y Sansano and Alisha Marti, two people very important for me. They really made me feel like at home since the beginning of my project. Special thanks to Ringo Püschel and Francesca Coraggio to help me to start with my project and rescue me when I was lost with my experiments. Also, I would like to thanks the current members Bolaji Isiaka, Moushumi Das and specially Jennifer Semple, a person who was always kind and helpful since the beginning of the project and even before when I met her in Barcelona. My appreciation also goes to all the people I met at the IGBMC from different team, people like Claudia Riva, Pietro Berico and Giovanni Gambi among others who made my life in Strasbourg funnier during my PhD. Additionally, I would like to thank all administration people at IZB and IGBMC standing out a person among the rest: Karin Heimberg. Karin's help and assistance were a constant during my PhD, without her, I had been lost with paperwork and bureaucracy.

I would like to specially thank all the team members of the Ochsenreiter laboratory that shared their space with me during all those four years. I would like to thank Hélène Baudouin and Simona Amodeo for their friendship and support, with science and paperwork and specially thank Anneliese Hoffmann for being there for me every time I need help.

To finish and most important, I would like to express my gratitude to my fiancé Charline David. Since I met her in 2017 during the second year of PhD she was always a faithful companion, strong support and the person who faced with me all the problems and difficulties of this PhD. She always encouraged me to never give up and believe in my work and capacities to finish this PhD project.

Curriculum Vitae

Personal Information

Name, Surname:	Jaime Osuna Luque
Date of birth:	03.08.1990
Nationality:	Spanish (European Union); B EU/EFTA permit
Current Address:	Bolligenstrasse 91 Bolligen, Kanton Bern, 3065, Switzerland
Languages:	English (Fluent), Spanish (Native), French (Elementary)
E-mail:	josuna1412@gmail.com

Education

March 2016 – 13th May 2020

PhD candidate Biochemistry and molecular biology (**University of Bern**)

Prof. Dr Peter Meister, Cell Fate and Nuclear Organization. Institut für Zellbiologie, Bern, Switzerland (Swiss host lab)

PhD candidate Sciences de la vie et de la santé (**University of Strasbourg**)

Prof. Dr. Sophie Jarriault, Development and Stem Cells, IGBMC Illkirch-Graffenstaden, France. (French host lab)

October 2013 - October 2014

Master in Translational Biomedical Research, University of Cordoba.

September 2008 - September 2013

Bachelor's degree in Biochemistry. University of Cordoba.

Professional Experience

November 2014-November 2015

Research associate

Project: Analysis of the action mechanism of Risperidone and aripiprazole on the nervous system of *C.elegans*

Prof. Dr. Manuel Ruiz Rubio, University of Cordoba

Publications

Osuna-Luque, J, Rodriguez-Ramos, A, Del Estal-Gámez, MM, Ruiz-Rubio, M. Behavioral mechanisms that depend on dopamine and serotonin in *Caenorhabditis elegans* interact with the antipsychotics risperidone and aripiprazole. **Neuroscience Insights.** 2018 https:/doi.org/10.1177/1179069518798628

Courses

Name of the course: 13th Course of epigenetics "Chromatin dynamics and epigenetics: Evolution"
Dates (from – to): 8th to 15th March 2017
Location: Institute Curie. Paris

Name of the course: 1st Course of Introduction to *C. elegans*.
Dates (from – to): 13th to 20th September 2015
Location: Centre for Genomic Regulation. Barcelona.

Fellowships and awards

2016-2020	SBFI-COST PhD Grant	180.000 CHF
2018	Germaine de Stael Grant 2019	4.600 CHF
2017	SwissUniversities Cotutelle de these Grant	8.000 CHF
2017	Germaine de Stael Grant 2018	4.600 CHF
2017	International mobility Grant (UniStra)	750€

References

Sophie Jarriault	sophie.jarriault@igbmc.fr	+33 (0)3 88 65 33 92
Peter Meister	peter.meister@izb.unibe.ch	+41 31 631 46 09
Manuel Ruiz Rubio	g1rurum@uco.es	+34 957212422