

Un siècle de génétique avec des souris

One century of genetics with mice

Par Jean-Louis GUÉNET⁽¹⁾

(communication présentée le 6 octobre 2005)

RÉSUMÉ

Les souris de laboratoire dérivent de spécimens appartenant à plusieurs espèces du genre *Mus*, élevés depuis les temps les plus reculés en Europe et en Asie. Elles ont joué un rôle primordial dans le développement de la génétique des mammifères depuis le début du vingtième siècle, grâce notamment à l'existence d'une grande variété de lignées consanguines. De nombreuses mutations, survenues spontanément ou induites par des agents mutagènes, ont également fourni de précieux modèles pour l'étude de la pathologie humaine. Au cours des vingt dernières années et grâce aux progrès de la technique, l'homme s'est donné les moyens de modifier, presque à volonté, le génome de la souris soit en ajoutant *in vitro* un ou plusieurs gènes d'origine exogène, soit en invalidant un gène préalablement ciblé et connu uniquement par sa séquence. Enfin, le génome de la souris, comme celui de l'homme, est aujourd'hui entièrement séquencé et l'inventaire des ARNs messagers et des protéines qui sont codés (le transcriptome et le protéome) est en train d'être réalisé. On s'aperçoit alors que seule une très petite partie de ces gènes est connue et un long travail d'annotation va commencer, dans lequel la souris jouera, à n'en pas douter, un rôle primordial. La connaissance de la séquence complète des génomes de l'homme et de la souris constitue une référence pour d'autres espèces, en particulier pour celles qui ont un intérêt zootechnique. Elle permettra aussi, au travers de comparaisons, de faire d'intéressantes observations aussi bien pour comprendre la régulation de l'expression des gènes que pour comprendre certains mécanismes mis en place par l'évolution.

Mots-clés : génétique, souris, modèle animal, génome.

SUMMARY

Laboratory mice stem from specimens belonging to different species of the *Mus* genus, bred since time immemorial both in Europe and Asia. They played a key role in the development of mammalian genetics since the early twentieth century, particularly helped by the existence of various inbred strains. Numerous mutations, whether spontaneous or induced by mutagenic agents, have also produced valuable models to study human diseases. With technical progress achieved over the past twenty years, scientists are now able to modify the mouse genome almost at will, either by adding *in vitro* one or several exogenous genes, or by knocking out a previously-targeted gene known only by its sequence. The mouse genome, like the human one, is entirely sequenced, and the inventory of messenger RNAs and their resulting proteins (transcriptome and proteome) is under way. It is now evident that only a small fraction of these genes is currently known, and that a long process of gene annotation has to be undertaken in which mice are bound to play a central role. Extensive knowledge of both human and mouse genome sequences will provide a reference for other species, especially those with zootechnical interest. Comparisons will also generate interesting data to help us understand the regulation of gene expression as well as the mechanisms at work during evolution.

Key words : mouse, genetics, animal model, genome.

(1) Chef de Service émérite à l'Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris CEDEX 15.

• LES DÉBUTS DE L'HISTOIRE

Sans qu'il y ait de preuves formelles, les historiens ont de bonnes raisons de penser que Gregor Mendel a utilisé des souris pour réaliser une partie de ses expériences sur la transmission des caractères, mais il est probable que sa hiérarchie ecclésiastique n'a pas accepté que des expériences réalisées avec un mammifère, autrement dit une créature assez proche de l'homme, soient poursuivies dans le monastère où il vivait (PAIGEN, 2003). Sans cette interdiction, les débuts de la génétique de la souris auraient tout simplement coïncidé avec ceux de la génétique tout court. C'est donc Lucien Cuénot, Professeur de Biologie à l'Université de Nancy, qui doit être considéré comme le premier généticien à avoir utilisé ce petit mammifère pour démontrer que les lois de Mendel étaient universelles et qu'elles s'appliquaient aussi bien aux plantes qu'aux animaux (CUÉNOT, 1902). Le choix de la souris pour ce genre d'expérience était assez logique car il existait, déjà à cette époque, toute une collection de lignées pures, porteuses de mutations faciles à identifier, qui permettaient de faire des croisements dont les résultats étaient relativement simples à interpréter.

Après la seconde guerre mondiale, dans les années 1950, lorsque la question s'est posée d'évaluer le risque que représentait l'utilisation de l'énergie nucléaire pour le patrimoine génétique de l'homme, ce sont encore des souris que l'on a utilisées. On en a irradié des millions, aussi bien à Oak Ridge, aux États-Unis, qu'à Harwell en Grande-Bretagne et, comme les rayonnements sont de puissants agents mutagènes, des centaines de nouvelles mutations ont ainsi été découvertes. Impliquées dans toute une variété de croisements, ces

mutations ont permis de réaliser une carte génétique très élaborée et, à la fin des années 1970, la souris était le mammifère qui possédait, et de loin, la carte génétique la plus précise et la plus dense (figure 1).

Lorsque les généticiens ont entrepris d'étudier les lois qui gouvernent l'acceptation ou du rejet des greffes, c'est toujours vers la souris que l'on s'est tourné. Dans cette espèce, il existe des lignées hautement consanguines, autrement dit des populations d'animaux chez lesquels les greffes tissulaires sont toujours acceptées (individus histocompatibles), tandis que les transplantations réalisées entre animaux provenant de deux lignées consanguines différentes sont systématiquement rejetées. On pouvait donc croiser des souris selon un protocole précis et isoler, pour les identifier individuellement, chacun des gènes impliqués dans la compatibilité tissulaire. Cet énorme travail a reçu sa reconnaissance, en 1980, au travers d'un prix Nobel, accordé aux Professeurs G. D. Snell, J. Dausset et B. Benacerraf.

Encore plus récemment, c'est la souris que l'on a choisie comme espèce pour réaliser des organismes chimères (figure 2), permettant entre autres choses de suivre le devenir des cellules au cours du développement embryonnaire et c'est aussi une souris qui fut le premier mammifère transgénique. Certes, la souris n'a pas été le premier mammifère cloné, mais très peu de temps après la brebis Dolly, des clones ont aussi été produits avec succès dans cette espèce et, si l'on en croit la littérature sur le sujet, il est même possible de cloner des souris en série sur deux ou trois "générations" ce qui, pour l'instant, n'est réalisable dans aucune autre espèce.

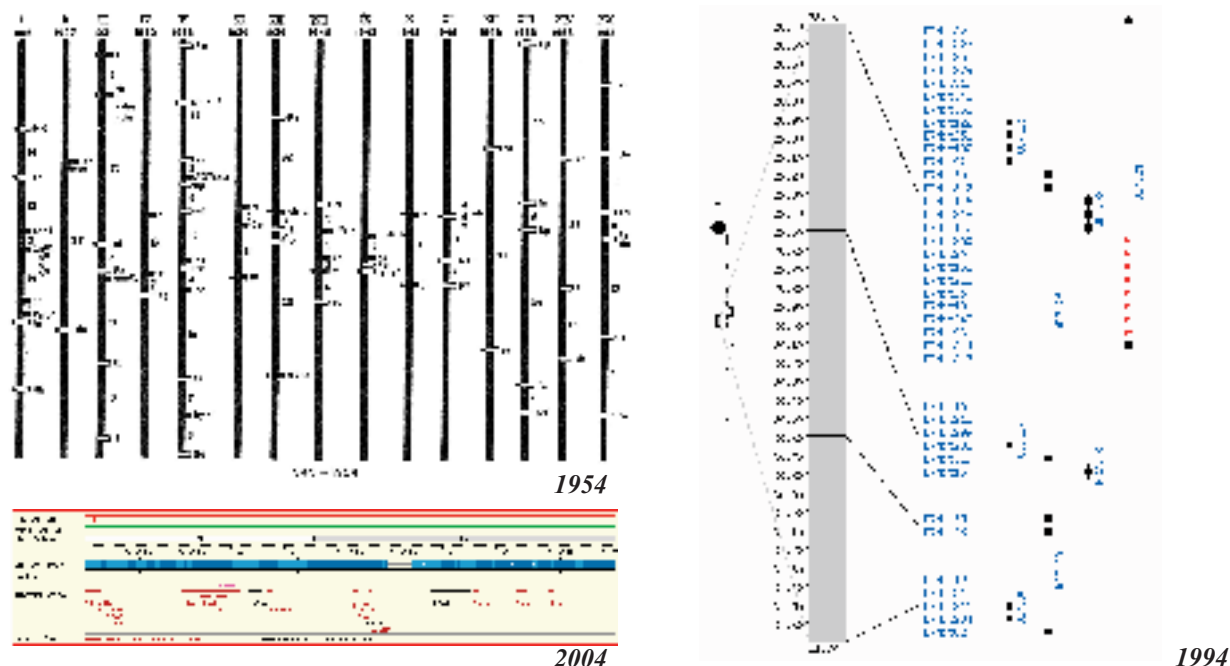


Figure 1 : La carte génétique de la souris représentée en trois étapes au cours des cinquante dernières années. En 1954, peu de gènes étaient localisés et tous correspondaient à des mutations ayant un effet phénotypique clair. A cette même date, les groupes de liaison n'avaient pas été positionnés sur un chromosome particulier et n'étaient que des représentations schématiques de la liaison génétique. La carte représentée était néanmoins la carte la plus élaborée connue chez les mammifères. En 1994, la carte génétique est devenue très dense et se compose presque uniquement de marqueurs moléculaires. Les éléments D7Mit sont des marqueurs microsatellites amplifiables par PCR. Les éléments verticaux figurant à la droite du chromosome correspondent à des ADN clonés dans des chromosomes de levure (YACs) ou dans des chromosomes bactériens (BACs). En 2004, le génome est séquencé : c'est la carte absolue !



Figure 2 : Chimères allophéniques ou tétraparentales. Les souris représentées sur ce cliché, dont la robe est tachetée, proviennent de la fusion in vitro de deux embryons précoces d'origine génétique différente (Cliché Charles Babinet).

Aujourd'hui, le génome de la souris est tout entier séquencé (WATERSTON *et al.*, 2002) et des projets très ambitieux sont en cours pour identifier chacun des 25 000 à 30 000 gènes de cette espèce (GUÉNET, 2005). Cet énorme travail, qui nécessitera sans doute plusieurs années pour arriver à son terme, consacrera alors la souris comme le mammifère modèle par excellence pour les recherches en génétique.

À l'heure actuelle, la génétique de la souris est devenue une spécialité à part entière. Il existe une Société Internationale comportant environ 500 membres qui se réunissent en un congrès annuel. Il existe des publications et une très grande variété de sites web spécialisés. Il existe aussi des Centres, tels que le célèbre Jackson Laboratory, à Bar Harbor aux États-Unis, ou le MRC d'Harwell en Angleterre ou bien encore le RIKEN, à Tsukuba au Japon, où des milliers de génotypes sont archivés sous forme "respirante" ou congelée. La France possède elle aussi, à Orléans la Source, un centre qui est rattaché à un réseau européen (l'*European Mouse Mutant Archive*, plus connu par l'acronyme EMMA), où de nombreuses lignées sont disponibles sous forme d'embryons ou de sperme congelés. Enfin, en association avec l'Université Paris VI et l'École Nationale Vétérinaire d'Alfort, l'Institut Pasteur dispense, chaque année, un cours spécialisé sur la "Génétique de la Souris".

Nous allons voir comment on a pu en arriver là, après un siècle de recherches et de progrès scientifique.

• DE L'ÉTAT SAUVAGE À LA DOMESTICATION

La souris appartient à l'ordre des Rongeurs et au sous-ordre des *Myomorphes*, le sous-ordre de mammifères qui comporte le plus grand nombre d'espèces sur la planète (et peut être même le plus grand nombre d'individus en valeur absolue). Les souris de laboratoire appartiennent au genre *Mus*, un genre qui comporte une douzaine d'espèces d'apparences très homogènes (figure 3) (GUÉNET et BONHOMME, 2004). D'après les données de la paléontologie, l'origine du genre *Mus* remonte à environ cinq millions d'années et se situe dans le sous-continent indien. Si l'on consi-

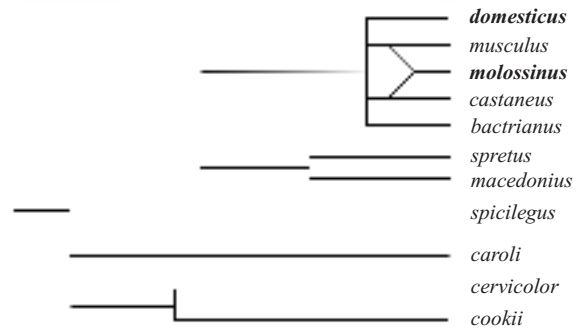


Figure 3 : Phylogénie dans le genre Mus. Ce schéma représente les relations phylogéniques dans le genre *Mus*. L'origine du genre remonte à environ 5 millions d'années. Les souris de laboratoire ne correspondent pas à une espèce naturelle mais, au contraire, sont des hybrides entre deux composantes majeures : *Mus musculus domesticus* qui vit en Europe occidentale et *Mus musculus molossinus* qui vit au Japon. Au laboratoire, on peut obtenir des hybrides entre plusieurs espèces mais dans certains cas les mâles sont stériles ou très peu viables. (D'après GUÉNET et BONHOMME, 2004)

dère les données les plus récentes sur l'émergence des mammifères, établies essentiellement à partir de l'analyse structurale des génomes nucléaires et mitochondriaux, la souris diverge de l'homme depuis 70 à 80 millions d'années et du rat, depuis 12 à 15 millions d'années.

On a la preuve qu'il y avait déjà des souris en Chine, en Grèce et en Israël, au milieu du Pléistocène. On en a retrouvé plusieurs spécimens dans le site archéologique néolithique de Çatal Hüyük en Turquie et, à l'âge du Fer, les souris étaient à coup sûr arrivées en Europe. C'est avec les conquêtes romaines, l'amélioration des communications et le changement des pratiques agricoles que les souris se sont répandues en Europe occidentale.

Bien qu'elles soient relativement rares en Afrique Centrale, on trouve des souris de l'espèce *Mus musculus domesticus* et, à un moindre degré, *Mus musculus musculus* sur pratiquement tous les continents. En Europe Occidentale, la sous-espèce la plus répandue est *Mus m. domesticus*. Elle se présente sous plusieurs aspects (les évolutionnistes parlent de morphotypes) selon qu'on la capture dans les pays du nord ou dans les contre-forts pyrénéens, mais tous ces individus appartiennent néanmoins à la même sous-espèce et l'on peut, dans tous les cas, obtenir des descendants fertiles en croisant ces animaux entre eux (GUÉNET et BONHOMME, 2003).

Certaines espèces du genre *Mus* vivent à l'extérieur et creusent des tunnels ou élèvent des tumulus pour s'abriter et procréer. D'autres vivent dans des régions boisées pour se mettre à l'abri des oiseaux prédateurs. D'autres enfin vivent en commensales de l'homme, c'est-à-dire dans sa maison ou des maisons construites par lui et se nourrissent à ses dépens. Le commensalisme n'est cependant ni une règle absolue ni une obligation pour les souris européennes et l'on connaît des îles, au nord de l'Écosse, qui ont été abandonnées par l'homme mais pas par les souris. On connaît aussi des régions dans le sud de l'Europe où les souris sortent lorsqu'arrive la période estivale et rentrent dans les habitations seulement pour y passer l'hiver. Les souris sauvages construi-

sent des nids d'architecture très variable. Dans nos régions, le nid est en général rudimentaire et transitoire. Dans d'autres pays, il est plus élaboré, parfois complètement fermé et souvent composé de matériaux isolants.

Les populations de souris du genre *Mus* présentent parfois des variations numériques d'amplitude considérables. Il peut arriver, après un hiver très rigoureux, une période de disette intense, une épizootie ou une campagne d'éradication énergique, que la population tombe à un niveau très bas sans que l'espèce soit pour autant menacée d'extinction. Il arrive aussi, à l'inverse, et pour des raisons inconnues, que le nombre d'individus atteigne des proportions si considérables qu'il est possible de parler de véritable infestation. La littérature rapporte que les habitants du village de Crystal Brook, dans le sud de l'Australie, capturèrent 126 000 souris en une seule semaine et, dans nos campagnes, où il était classique dans l'ancien temps de récompenser la destruction de cet animal dévastateur, des records similaires ont également été rapportés.

Les souris sont des animaux à activité essentiellement nocturne chez lesquels la vue est peu développée. La plupart des souris de laboratoire sont d'ailleurs aveugles ou n'ont qu'une vision crépusculaire du monde qui les entoure, sans que leur mode de vie soit (apparemment) perturbé. L'odorat et le toucher sont, par contre, les sens les plus importants. Il existe dans l'urine des mâles, et à un moindre degré dans celle des femelles, des molécules odoriférantes très puissantes (des phéromones) qui servent à marquer le territoire et à régler l'activité sexuelle ou le comportement social. Chez les souris sauvages, il existe, entre les mâles, une compétition considérable pour la reproduction. Des mâles "dominants" règnent sur un groupe de femelles, constituant ainsi une unité de reproduction relativement consanguine, que l'on appelle un dème (2 mâles et 6 femelles en moyenne), dont la composition change avec le temps. Il existe aussi des mâles "célibataires" itinérants.

• LA DOMESTICATION ET L'ÉLEVAGE POUR LES BESOINS DE L'EXPÉRIMENTATION

Les biologistes se sont très tôt intéressés à la souris pour de multiples raisons. D'abord et sans doute, parce qu'il s'agit d'un mammifère de très petit format (25 à 30 grammes à l'état adulte), qui était déjà disponible sous forme d'animal favori et qui, par conséquent, était assez facile à élever en captivité. La souris est un rongeur et la mise au point d'une formule couvrant l'ensemble des besoins alimentaires ne pose aucun problème et ne coûte pas très cher. La souris présente aussi l'énorme avantage de se reproduire très vite et sans saison "creuse". La femelle est en général pubère à l'âge de sept semaines, mère à 10-11 semaines (la gestation est de 19,5 à 20,5 jours) et peut produire une soixantaine de petits (et parfois beaucoup plus) dans sa vie génitale.

La souris possède, en plus, un avantage d'importance tout à fait capital que les premiers hommes de science ne connaissaient pas mais qui fut reconnu et apprécié plus tard par les généticiens : elle supporte bien la consanguinité. En accouplant de manière ininterrompue des frères avec leurs sœurs, on peut obtenir, après un nombre de générations suffisant,

une lignée pure, totalement consanguine, dans laquelle tous les animaux sont génétiquement identiques et homozygotes pour tous les gènes de leur génome. Il existe aujourd'hui des centaines de lignées consanguines (*inbred* en anglais) qui représentent un matériel très précieux pour les biologistes. Chacune d'elle, en effet, possède un certain nombre de caractéristiques qui lui sont propres et qui sont directement associées à l'assortiment de gènes présents dans le génome. Certaines lignées sont par exemple très sensibles aux infections virales ou parasitaires, d'autres au contraire y sont résistantes, et ceci bien sûr se transmet à chaque génération de sorte que l'on peut faire des expériences sur des lots d'animaux homogènes et comparables, à la fois dans le temps et dans l'espace. Une situation unique et très appréciée !

Les souris de laboratoire dérivent à la fois de la sous-espèce *Mus m. domesticus* et dans une moindre mesure de la sous-espèce *Mus m. molossinus* que l'on trouve au Japon à l'état sauvage. On retrouve en effet, dans de nombreuses lignées, des marqueurs génétiques qui ne sont jamais associés dans l'une ou l'autre espèce à l'état sauvage, ce qui prouve que les lignées de laboratoire ne dérivent pas d'une espèce unique mais qu'au contraire, elles ont une origine complexe, polyphylétique (BONHOMME *et al.*, 1987). Ces observations sont aujourd'hui amplement confirmées au niveau moléculaire où il a été récemment établi que le génome des souris appartenant à des lignées consanguines de laboratoire était analogue à une mosaïque, composée pour partie de segments chromosomiques d'origine *Mus m. domesticus* et pour une autre partie, de segments chromosomiques d'origine *Mus m. molossinus* (WADE *et al.*, 2002). Toutes ces observations trouvent d'ailleurs une explication rationnelle lorsqu'on considère l'origine des lignées de souris de laboratoire, puisqu'on a pu établir, de manière non-ambiguë, qu'elles dérivent d'ancêtres européens et japonais domestiqués depuis la très haute Antiquité. On sait en effet que des souris blanches étaient déjà élevées en haute Égypte, au temps des pharaons et en Extrême-Orient, à la période Edo. Il est donc logique de penser que ces animaux de compagnie, élevés dans le passé pour le plaisir ou pour satisfaire certaines croyances, aient été à l'origine des animaux utilisés de nos jours dans les laboratoires. Ceci se trouve également confirmé par l'observation que la même mutation albinos, relativement peu fréquente à l'état naturel, est fixée dans de nombreuses lignées consanguines. On sait aussi que le nombre de géniteurs ayant servi à fonder les lignées actuellement utilisées a sans doute été très petit car toutes les souris de laboratoire ont la même molécule d'ADN mitochondrial (ou mtDNA), ce qui suggère qu'elles ont toutes le même ancêtre côté femelle puisque les mitochondries sont transmises essentiellement par l'ovocyte (YONEKAWA *et al.*, 1982).

D'après des textes laissés par le biologiste américain L. C. Strong, l'un des fondateurs de la génétique de la souris, on sait aussi que des mâles sauvages ont été capturés et croisés avec des souris déjà domestiquées pour augmenter le polymorphisme des animaux de laboratoire. De ce point de vue, il faut donc reconnaître que, si elles constituent un matériel très commode pour la recherche, les souris de laboratoire représentent néanmoins des populations totalement artificielles et, pour cette rai-

son, elles ont parfois été qualifiées de *Mus laboratorius* par les systématiciens, pour bien souligner leur origine synthétique et artificielle (BONHOMME et GUÉNET, 1996).

À l'heure actuelle, les souris de laboratoire sont produites par millions d'exemplaires et des élevages industriels se sont développés dans plusieurs pays pour subvenir aux besoins de la recherche. Les souris y sont élevées sous le contrôle de vétérinaires spécialistes dans un état sanitaire quasi-parfait et un très haut degré de standardisation génétique. Les géniteurs sont prélevés *in utero*, par césarienne, dans des conditions d'asepsie rigoureuse, et placés dans des enceintes d'isolation, à l'abri de toute contamination bactérienne, virale ou parasitaire. La pureté génétique des stocks est régulièrement contrôlée, de même que tous les protocoles d'accouplement. Le prix de chaque animal varie de 1 à 20 € en fonction de son âge et de la lignée à laquelle il appartient. Certaines souris génétiquement manipulées font même l'objet d'une prise de brevet et ne sont fournies ou échangées entre laboratoires qu'après signature d'un accord de coopération (material transfert agreement ou MTA).

• UNE ESPÈCE DONT L'EMBRYOLOGIE EST PARFAITEMENT CONNUE

Pendant son transit dans les oviductes, l'embryon de souris peut être explanté, manipulé *in vitro* pendant plusieurs heures, puis réimplanté dans les oviductes ou les cornes utérines d'une femelle convenablement préparée du point de vue hormonal. On peut, assez facilement, congeler les embryons à des stades précoces du développement et constituer ainsi des conservatoires de sécurité ou cryo-embryothèques. De nos jours, on sait aussi congeler le sperme de souris mais la technique présente encore quelques aléas et n'est pas aussi fiable que chez l'homme ou les autres animaux domestiques. En injectant des hormones gonadotropes, on peut induire l'ovulation chez les femelles et récolter des ovocytes ou des œufs fécondés en grande quantité. On peut manipuler ces œufs à tous les stades préimplantatoires et produire des chimères, par exemple en réassociant des blastomères provenant de plusieurs embryons, à condition toutefois qu'ils soient à peu près au même stade de développement (figure 2). Ce genre d'expérience a été très utilisé pour élucider la genèse de tissus comme le muscle ou de certains éléments figurés du sang et de la lymphe. L'analyse de chimères a également permis d'élucider l'origine clonale de certains cancers (CONDAMINE *et al.*, 1971).

Des clones de souris ont également été réalisés en remplaçant le noyau haploïde d'un ovocyte par le noyau de cellules différenciées provenant de différents tissus (WAKAYAMA et YANAGIMACHI 1998; YANAGIMACHI, 2002). Il faut cependant reconnaître que le clonage n'est pas, dans cette espèce, une stratégie aussi intéressante que dans les espèces de rente et ceci pour deux raisons : d'abord parce qu'il n'y a aucune motivation de nature zootechnique, ensuite parce qu'il est facile de produire, à volonté, des animaux génétiquement identiques qui sont, au sexe près, l'analogue d'une population d'animaux clonés (des hybrides de première génération F1 entre deux lignées consanguines par exemple).

La souris est une espèce polytoque puisqu'elle porte plusieurs fœtus au cours d'une même gestation et l'avortement de l'un d'entre eux n'interrompt pas le développement des autres. Ceci représente un avantage appréciable pour l'étude du développement embryonnaire, puisqu'on peut identifier *in utero* la présence et parfois, la morphologie d'un embryon, même après sa mort.

• UN TRÈS GRAND NOMBRE DE MUTATIONS ET UNE CARTE GÉNÉTIQUE TRÈS ÉLABORÉE

L'intérêt d'une espèce pour le généticien s'apprécie souvent au nombre de mutations qui sont disponibles et, de ce point de vue, la souris a quelques records à faire valoir. Les mutations sont des événements dont les effets se matérialisent par l'observation d'un phénotype anormal, spontanément héréditaire. Certaines mutations sont dominantes et s'observent à chaque génération, la plupart sont récessives et se transmettent de manière mendélienne à une partie de la descendance seulement.

Les mutations sont observées spontanément à une fréquence très faible, de l'ordre de 5×10^{-5} à 10^{-6} par locus et par gamète, mais elles sont très importantes pour deux raisons : d'abord parce qu'elles mettent en évidence la fonction d'un gène au travers d'un phénotype anormal et aident ainsi à ce que les généticiens appellent l'annotation du génome ; ensuite, parce qu'elles représentent autant de modèles pour l'étude des mutations survenant chez l'homme, qui sont souvent génératrices d'une pathologie grave.

Il existe chez la souris une très grande collection de mutations (plus de 1 500) qui ont été accumulées au cours des années et dont les effets phénotypiques sont des plus variés. Comme chez l'homme, beaucoup de ces mutations affectent le système neuro-musculaire et se caractérisent par des paralysies, des ataxies, des pertes d'équilibre, des crises d'épilepsie, des tremblements, etc. Il existe aussi beaucoup de mutations affectant le squelette, l'appareil tégumentaire (peau et poils), l'œil, l'oreille interne et le système immunitaire. Ces mutations ont, en général, fait l'objet de recherches intenses et dans bien des cas, le défaut moléculaire est aujourd'hui caractérisé (le gène est cloné) et cela contribue à l'amélioration de notre connaissance du génome de l'espèce et, par voie de conséquence, du nôtre aussi (GUENET, 2005) (figure 4).

Les mutations représentent un matériel tellement important pour les généticiens que des stratégies ont été développées au cours du temps pour en augmenter la fréquence : ils utilisent, soit les radiations de toutes natures (surtout les rayonnements X et γ), soit une variété de molécules qui interfèrent avec l'ADN au moment de la réplication et font faire des fautes à la polymérase et aux autres molécules impliquées dans la réplication. Parmi toutes ces substances, l'éthyl-nitroso-urée (ENU) est celle qui, de très loin, possède le plus important pouvoir mutagène. Depuis la découverte de ses effets, en 1979, l'ENU est devenu l'agent mutagène par excellence chez la souris, à cause de son efficacité et de sa simplicité d'emploi. On considère aujourd'hui qu'après trois injections de 100 à 120 mg d'ENU par kilogramme de poids corporel, pratiquées à une semaine d'inter-



Figure 4 : Les mutations chez la souris. Il existe chez la souris une très grande collection de mutations et l'on peut en induire de nouvelles quasiment à volonté. Toutes ces mutations permettent d'annoter les différents gènes du génome en mettant en parallèle un phénotype anormal et une altération au niveau de la séquence du gène. Sur la figure, la souris présentée n'a plus de récepteur pour la leptine, une cytokine impliquée dans la régulation du métabolisme des lipides : elle en devient obèse et diabétique. Les mutations de la souris servent aussi de modèles pour la pathologie humaine et des obésités (rares il est vrai) sont connues, qui procèdent de la même altération génétique chez l'homme.

valle, ou une injection unique de 250 mg/kg, le taux des mutations est multiplié par un facteur 100 à 120. Exprimé différemment, cela signifie qu'en moyenne, après traitement, un gamète sur 670 produits par la souris traitée porte une mutation à un locus donné (GUÉNET, 2004).

Les phénotypes mutants, d'abord collectionnés comme des curiosités, ont aussi été utilisés pour établir la carte génétique de l'espèce. On s'est rapidement rendu compte, en effet, en croisant des souris porteuses de différentes mutations, que certains phénotypes ne pouvaient pas être facilement associés, alors que d'autres, inversement, avaient tendance à rester associés dans leurs configurations parentales. Ce phénomène de liaison (ou de linkage) est à la base de l'établissement des cartes génétiques car il permet de recenser les gènes qui sont portés par la même structure physique, autrement dit, par le même chromosome. En 1975, grâce au patient travail des généticiens, la souris était le mammifère dont la carte était la plus élaborée (SILVER, 1995). Un légitime motif de fierté ! Hélas, cet avantage ne devait pas perdurer plus de cinq ans car, à partir de 1981, la génétique moléculaire, en offrant une richesse incommensurable en marqueurs "moléculaires" (fragments d'ADN de longueur variable engendrés par les enzymes de restriction, puis microsatellites amplifiés par PCR, puis polymorphisme nucléotidiques ou SNPs observés par séquençage), donnait à la carte génétique de l'homme une avance désormais irrattrapable, qui devait aboutir *in fine* au séquençage complet du génome.

Aujourd'hui, les mutations ne servent plus de marqueurs puisque le génome de la souris est lui aussi presque complètement séquencé, mais elles continuent de servir d'outils pour annoter le génome et des projets trans-nationaux gigantesques sont en cours pour induire, par mutagenèse à l'ENU, au moins une mutation dans chaque gène du génome. En se fondant sur les résultats déjà publiés, on peut faire confiance

aux équipes qui sont impliquées dans ces projets pour parvenir à ce but. Mais le difficile restera de "phénotyper" les animaux porteurs de mutations, autrement dit, d'identifier précisément, et surtout complètement, les modifications induites par le mutagène. En effet, si on reconnaît très facilement une souris blanche dans une nichée de souris grises, il est beaucoup plus difficile de reconnaître une surdité progressive, une maladie provoquée par l'accumulation lente d'un produit mal catabolisé ou certaines maladies du système immunitaire. Enfin, à l'extrême, il est quasi impossible de reconnaître, sans un protocole précis et parfois complexe, qu'une souris est devenue brutalement sensible à un virus quelconque, parce qu'une mutation survenue dans un gène important a modifié ou tout simplement annihilé les réactions non spécifiques de son immunité innée. La route qui passe par la mutagenèse pour conduire à l'annotation de tous les gènes du génome est fiable, mais elle sera sans doute longue et hérissée de difficultés. Heureusement, il en existe d'autres qui sont complémentaires comme nous le verrons.

• LA TRANSGÉNÈSE SOUS TOUTES SES FORMES

Les vingt dernières années ont été particulièrement riches et importantes en matière d'avancées techniques dans le domaine de la génétique de la souris, avec la mise au point de la transgénèse par addition d'ADN *in ovo* d'une part et la découverte des extraordinaires propriétés des cellules souches embryonnaires d'autre part.

L'idée d'injecter directement dans l'œuf de la souris une molécule d'ADN linéarisée, en espérant qu'elle se recombinera avec l'ADN de l'œuf pour finalement s'ajouter au patrimoine héréditaire de l'embryon, n'est pas très originale et il est vraisemblable que plusieurs chercheurs y ont pensé très tôt après la découverte des enzymes de restriction, permettant de cloner les gènes des mammifères dans des vecteurs bactériens. Mais une importante étape technique restait à franchir, celle de l'injection proprement dite, qui impliquait de pouvoir manipuler l'œuf au stade d'une cellule, juste après la pénétration du spermatozoïde au travers de la zone pellucide. Une difficulté importante était en particulier de trouver des milieux de culture permettant de cultiver l'embryon *in vitro*, depuis ce stade une cellule jusqu'au stade morula mature ou jeune blastocyte, où il pouvait alors être réimplanté dans l'utérus d'une mère porteuse convenablement préparée du point de vue hormonal.

C'est d'abord à Jon Gordon, dans le laboratoire de Frank Ruddle, à Harvard (GORDON *et al.*, 1980), puis et surtout à deux Vétérinaires de l'Université de Pennsylvanie, Ralph Brinster et Richard Palmiter (BRINSTER et PALMITER 1984), que l'on doit les premiers animaux transgéniques. La publication en 1982, en couverture de la revue *Nature*, d'une photographie représentant une souris transgénique devenue géante en raison de la présence dans son génome de quelques copies du gène de l'hormone de croissance du rat, activées par un promoteur constitutif, a eu un impact considérable sur la communauté scientifique. La photographie en question a été reproduite à un très grand nombre d'exemplaires, comme pour souligner l'extraordinaire pouvoir de cette technique. En effet, il devenait tout d'un coup possible d'ajouter *in vitro*, ou

plus exactement *in ovo*, directement et mécaniquement avec une pipette de verre effilée, un nouveau gène au patrimoine génétique d'une espèce, sans que celui-ci n'ait jamais transité par le canal de l'évolution (figure 5).

Depuis cette date, la technique s'est raffinée et la transgénèse par injection d'ADN *in ovo* fait désormais partie du quotidien dans de nombreux laboratoires. On peut même obtenir ce service de certaines sociétés travaillant à façon, évitant ainsi aux laboratoires d'acheter un équipement lourd et onéreux. Cette technique a aussi été transposée aux principales espèces domestiques chez lesquelles elle est susceptible d'applications très nombreuses dans des domaines variés de la zootechnie (voir article de L. M. Houdebine dans ce numéro du Bulletin). Ce qui est très important avec la transgénèse, c'est que l'ADN linéaire injecté *in ovo*, peut tout aussi bien provenir de la même espèce que d'une espèce très différente et phylogénétiquement très éloignée de la souris. Les souris transgéniques pour un gène humain sont innombrables, ainsi que les souris transgéniques pour un gène bactérien ou viral. On peut, par exemple, créer une souris transgénique pour le gène humain qui code le récepteur du virus de la poliomyélite et la souris peut alors être infectée par ce virus, ce qu'elle n'est pas spontanément (YONEKAWA *et al.*, 1993). On peut aussi fabriquer, par génie génétique, un transgène artificiel en plaçant la séquence d'un oncogène cellulaire sous le contrôle d'un promoteur qui régule l'expression de l'oncogène en question dans la glande mammaire. On obtient alors une souris transgénique (modèle) qui développe des tumeurs mammaires avec une très haute fréquence. Une souris de ce type a été fabriquée par DuPont Technology et brevetée sous le nom commercial d'OncoMouse™, déclenchant du même coup une bataille juridique sans précédent à propos de la brevetabilité du "vivant". Le débat est loin d'être clos ! Si le transgène est constitué d'un oncogène et d'un promoteur exprimé spécifiquement dans les lymphocytes, on obtient une souris qui développe une leucémie lymphoïde. On peut aussi "fabriquer" une souris transgénique pour un gène humain codant le collagène de type I, prélevé chez un malade atteint d'ostéogenèse imparfaite et la souris développe alors le même syndrome que l'homme. C'est un



Figure 5 : Transgénèse par injection d'ADN *in ovo*. Une pipette de verre très effilée permet d'injecter des molécules d'ADN directement dans le pronucleus mâle. Dans un pourcentage de cas assez élevé, l'ADN injecté s'intègre à l'ADN de l'hôte et se transmet comme un nouveau caractère mendélien dominant (Cliché Charles Babinet).

modèle parfait de la maladie humaine mais l'ennuyeux, dans ce dernier cas, est que la souris est tellement affectée par la maladie, qu'elle meurt en général avant de s'être reproduite (les spécialistes disent, dans ce cas, que le transgène a des effets dominants négatifs). Il faut alors recommencer l'expérience d'injection *in ovo* à chaque fois, pour créer des animaux transgéniques nouveaux. Enfin, on peut construire un transgène complètement artificiel avec un promoteur provenant d'un gène de souris dont on ne connaît pas très bien le "territoire" d'expression et d'une séquence codante, par exemple d'origine bactérienne ou végétale, provenant d'un gène dont le produit est facile à visualiser sous le microscope. Dans ce cas, le gène bactérien, viral ou végétal porte le nom de gène rapporteur, car il permet de mettre en évidence de manière très simple et souvent très claire, les tissus dans lesquels le promoteur étudié est activé et le moment de son activation (Figure 6). Le monde entier s'est émerveillé, il y a quelques années, de voir apparaître à la télévision une souris transgénique qui devenait verte en lumière ultraviolette. La souris verte de la chanson enfantine devenait brutalement une réalité ! Cette "green mouse" porte un transgène réalisé à partir d'un promoteur exprimé de manière ubiquitaire ou seulement dans certains tissus et de la séquence codant la *Green Fluorescent Protein* (ou GFP) de la méduse *Aequora victoria* qui rend l'animal fluorescent en U.V. Le lecteur pourra, s'il le souhaite, contempler cette souris sur le site :

<http://kumikae01.gen-info.osaka-u.ac.jp/tg/tg-ad.cfm>



Figure 6 : Embryon de souris transgénique pour un gène rapporteur. L'embryon présenté est hétérozygote pour un transgène composé d'un gène rapporteur (*lacZ*) sous le contrôle du promoteur de la Desmine, une protéine spécifique du muscle. Une réaction enzymatique fait apparaître tous les territoires où le gène de la Desmine est exprimé au cours du développement embryonnaire. Le gène de la Desmine sur l'autre chromosome est normal (Cliché Charles Babinet).

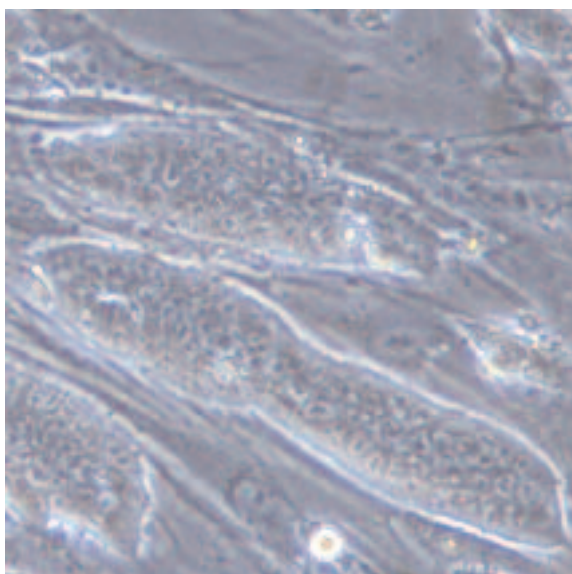


Figure 7: Cellules embryonnaires précoces en culture. Ces cellules, plus connues sous leur nom anglais d'*Embryonal Stem cells* ou *ES cells*, peuvent être cultivées *in vitro* pendant plusieurs générations et manipulées génétiquement. Replacées dans la cavité d'un blastocyte, elles peuvent ensuite être recyclées dans la lignée germinale et participer par conséquent à la formation d'un embryon (Cliché Michel Cohen Tannoudji).

Trente à 50% des transgènes implantés *in ovo* chez la souris, ou même chez d'autres espèces, sont compatibles avec une vie quasi-normale à condition que le produit du gène ne soit pas lui-même générateur d'une pathologie. Ces mêmes transgènes persistent en général au cours des générations successives, de sorte que plusieurs transgènes d'origine indépendante peuvent, s'il le faut, être réunis dans un même génome par la voie de la reproduction sexuée.

La transgénèse s'est très vite révélée une technique très utile pour étudier la fonction des gènes ou pour créer des modèles de pathologie. L'OncoMouse™, mentionnée ci-dessus, s'est très bien vendue, mais d'autres souris transgéniques, créées à partir d'une séquence codant une enzyme du bactériophage P1, la Cre-recombinase, placée sous le contrôle d'une grande variété de promoteurs, est aujourd'hui encore bien plus utilisée : nous verrons pourquoi un peu plus loin dans ce texte.

À vrai dire, il n'y a aucune limite, autre que l'imagination des généticiens (et ce n'est pas peu dire !), à la panoplie des souris transgéniques qui puissent potentiellement être créées à partir d'une séquence codante et d'un élément promoteur de régulation et, à n'en pas douter, l'avenir est encore plein de surprises. Pourtant, et pour aussi utile que soit la technique, la transgénèse a, malgré tout, une limitation intrinsèque due au fait que la manipulation expérimentale permet d'ajouter une information génétique exogène, disons un gène, dans un génome de souris, mais ne permet pas de soustraire ou d'éliminer quelque chose existant dans de ce même génome. Heureusement, les mêmes vingt dernières années ont enrichi l'arsenal technique des généticiens d'une manière incroyablement puissante, en permettant de faire ce qu'il est désormais

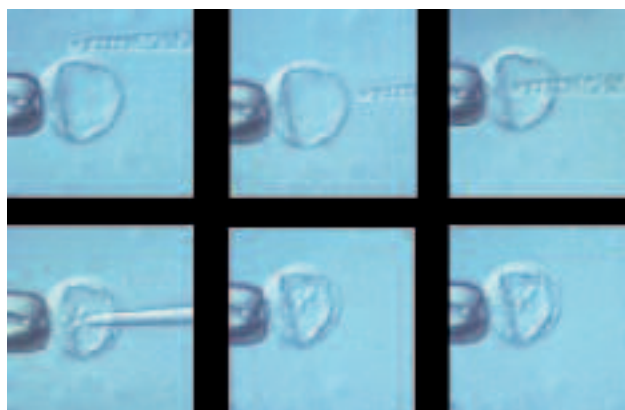


Figure 8: Recyclage des cellules ES dans la lignée germinale. Une fois manipulées *in vitro*, les cellules ES peuvent être ré-injectées dans la cavité d'un blastocyte et participer à la formation d'un embryon chimère. Lorsque ces mêmes cellules ES ont conservé la compétence de participer à la lignée germinale la mutation peut passer dans les gamètes (Cliché Charles Babinet).

convenu d'appeler de la recombinaison homologue *in vitro*, dans les cellules embryonnaires. Pour l'instant, malheureusement, cette technique n'est utilisable que chez la souris, mais dans cette espèce, elle est extrêmement puissante et fait désormais l'objet d'une abondante littérature. La place manque pour aller dans les détails, mais nous allons malgré tout essayer d'expliquer le principe de cette stratégie. Retenons simplement, dès à présent, que grâce à la recombinaison homologue dans les cellules embryonnaires, il est désormais possible de modifier n'importe quelle séquence du génome de la souris, à volonté, de manière très précise et très ciblée, de manière conditionnelle, c'est-à-dire lorsqu'on le décide, et parfois même de manière réversible !

L'outil de base pour cette stratégie est constitué par les cellules souches embryonnaires ou cellules ES (*Embryonal Stem cells* ou *ES cells*). Ces cellules, explantées de la masse cellulaire interne d'un embryon au stade blastocyte, sont maintenues *in vitro* dans un milieu de culture particulier et transplantées en série comme des cellules somatiques ordinaires (ROBERTSON *et al.*, 1986) (figure 7). Elles peuvent ainsi être manipulées génétiquement, sélectionnées, clonées et éventuellement congelées. Un grand nombre de ces cellules conservent toutes les propriétés des cellules embryonnaires et en particulier, une fois replacées dans la cavité d'un blastocyte receveur, elles peuvent participer à la formation de la lignée germinale et sous la forme de gamète, à la formation d'un nouvel individu. Une fois manipulées *in vitro*, ces cellules peuvent être analysées au niveau génétique (au niveau de l'ADN) et, lorsque la manipulation est jugée convenable, on peut alors décider de les ré-implanter dans la cavité d'un blastocyte receveur et récupérer, quelques jours plus tard, une petite souris chimère qui transmettra le résultat de la manipulation génétique dans ses gamètes (figure 8).

Les cellules ES ont aussi une autre caractéristique : elles sont perméables à la transfection par des molécules d'ADN exogènes. On peut alors fabriquer *in vitro* une molécule hybride⁽²⁾ dont la séquence est rigoureusement homologue de la séquence

(2) La réalisation d'une molécule d'ADN hybride relève de techniques désormais utilisées en routine dans les laboratoires de biologie moléculaire. Dans un premier temps, on coupe l'ADN des différentes espèces avec des enzymes de restriction, puis on mélange les fragments et on "recolle" avec d'autres enzymes, des ligases. Pour amplifier les molécules recombinantes, on les clone dans un vecteur bactériens ou viral, en général un plasmide.

d'un gène connu mais porte, en substitution à un segment quelconque, par exemple un exon, un petit gène d'origine bactérienne codant par exemple l'enzyme *neomycine phosphotransférase* (en abrégé *neo^r*) (figure 9). Lorsque la molécule pénètre dans la cellule et se recombine avec l'ADN de la souris, les cellules ES deviennent alors résistantes à une substance chimique analogue de la néomycine, le G418, et peuvent alors être séparées des cellules non-recombinantes qui sont tuées par cette substance chimique. Parmi les cellules recombinantes, la majorité ne sont pas des recombinantes "homologues" car la molécule d'ADN transfectée s'est intégrée n'importe où dans le génome, rendant certes les cellules résistantes au G418, mais pas très intéressantes pour le but poursuivi. Celles dans lesquelles la molécule transfectée a remplacé exactement la séquence de l'ADN hôte, sont les seules intéressantes. Elles sont recherchées par une autre technique et utilisées pour produire des souris chimères. Chez les descendants de ces souris, un gène sur les deux, celui qui résulte de la manipulation génétique, est invalidé et non-fonctionnel : il manque dans l'ARN transcrit de ce gène la séquence codée par le segment éliminé (un exon dans notre cas) : c'est un "knockout".

La stratégie que nous venons de décrire a volontairement été simplifiée pour raison de clarté mais un catalogue détaillé de toutes les stratégies utilisables pour manipuler génétiquement les cellules ES *in vitro* a été récemment publiée (GUÉNET, 2004), où il est montré que les raffinements techniques sont innombrables. En utilisant différentes molécules pour sélectionner les cellules manipulées *in vitro*, on peut facilement retrouver les cellules recombinantes homologues (les *knockout*) parmi toutes les cellules transfectées ; on peut aussi procéder à deux recombinaisons en cascade, l'une après l'autre, dans les mêmes cellules, et remplacer la séquence

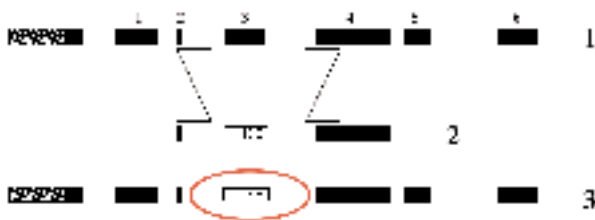


Figure 9 : Recombinaison homologue dans des cellules ES *in vitro*. La recombinaison homologue dans les cellules ES s'effectue en trois étapes. 1.- Un segment d'ADN ciblé pour être modifié est préalablement cloné et modifié génétiquement *in vitro*. Au terme de la modification, il doit continuer de présenter une grande homologie de séquence avec le segment *in situ* normal sur une grande partie de ses extrémités, mais il porte en son milieu une modification susceptible d'invalider le gène. Dans le cas représenté sur la figure, l'exon 3 du gène est remplacé par un segment d'ADN exogène n'ayant rien à voir avec la séquence naturelle. Il s'agit d'un mini-gène d'origine bactérienne qui code pour l'enzyme neomycin phosphotransférase. 2. Les cellules sont ensuite transfectées, en général par électroporation, avec les molécules d'ADN modifiées linéarisées. Les cellules transfectées deviennent alors résistantes à la néomycine et peuvent donc être sélectionnées. 3. – Dans un petit nombre de cas intéressants, la recombinaison est strictement homologue et l'ADN modifié remplace l'ADN natif. Les cellules continuent d'être *neo*-résistantes mais n'ont plus d'exon 3 dans une copie du gène ciblé. C'est un *knockout*. Il suffit ensuite de récupérer le gène recombiné à l'état hétérozygote dans une souris, puis de le mettre à l'état homozygote par la reproduction sexuée pour étudier les effets de l'invalidation.

d'un gène donné par une séquence qui n'en diffère que par une seule et unique paire de base... autrement dit, une mutation strictement ponctuelle (figure 10).

La technique que nous venons de décrire permet donc de faire, strictement à volonté, des mutations aboutissant à des allèles non-fonctionnels (des *knockout*) pour n'importe quel gène, dès lors qu'on connaît sa séquence. Malheureusement, la technique a ses faiblesses. En effet, il se peut qu'une mutation dans un tissu particulier ait des effets sévères au point de faire mourir l'embryon à un stade précoce et, dans ce cas, on ne peut rien savoir sur la fonction de ce gène dans d'autres tissus de l'adulte, par exemple. Les généticiens ont tourné la difficulté en mettant au point une stratégie connue sous le nom de *Cre-loxP*, qui permet de faire des *knockout* conditionnels et ciblés dans certains tissus.

L'invalidation de gènes par la technique *Cre-loxP* se fait en plusieurs étapes. On commence d'abord par définir très précisément la séquence que l'on souhaite modifier. On peut, par exemple, décider comme précédemment de supprimer un exon d'un gène donné pour le rendre non-fonctionnel. Lorsque la décision est prise, on procède alors à une recombinaison homologue *in vitro*, dans des cellules ES, qui aboutira à placer en amont (les généticiens disent en 5') et en aval (en 3') de la séquence ciblée (l'exon dans notre cas), deux séquences particulières, qui sont des sites reconnus par l'enzyme *Cre*-recombinase du bactériophage P1 dont nous avons parlé plus haut. Ces sites, que nous appelons dans notre jargon des sites *loxP*, n'existent pas dans le génome des mammifères et, s'ils sont convenablement placés dans la séquence du gène (par exemple dans les introns), ils ne gênent pas la transcription. Lorsque la souris porteuse de la recombinaison avec deux sites *loxP* a été réalisée, on dit que le gène ciblé est désormais "floxed". La mutation est en quelque sorte "préméditée", mais elle n'est pas effective. Elle le deviendra lorsque la souris à gène *floxed* sera croisée avec une souris transgénique qui synthétise la *Cre*-recombinase car, dans les

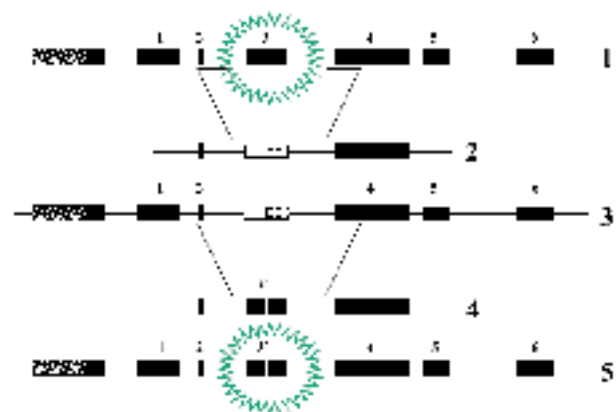


Figure 10 : Induction d'une mutation ponctuelle par recombinaison homologue. Après avoir réalisé une première mutation par recombinaison homologue, comme indiqué précédemment, on peut procéder à une seconde recombinaison dans les mêmes cellules et remplacer le segment porteur de la "cassette" *neo^r* par un segment absolument homologue au segment normal, à une paire de base près. Il faut pour cela utiliser d'autres drogues pour sélectionner les cellules ayant perdu *neo^r*.

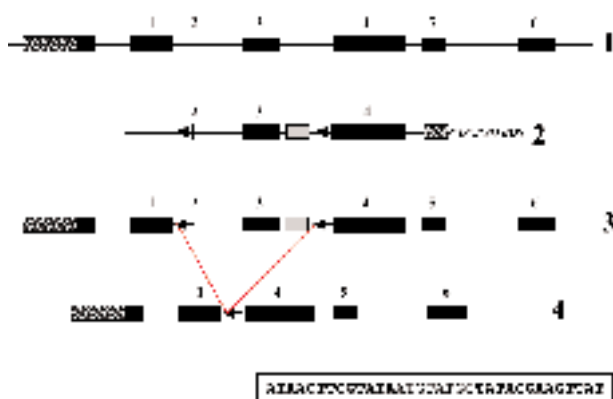


Figure 11 : Induction de mutations ciblées, conditionnelles par la technique Cre-loxP. Sur le schéma, la flèche pleine orientée à gauche matérialise un site loxP dont la séquence apparaît en cartouche sur la figure. Cette séquence, qui est reconnue par la recombinaise Cre, n'existe pas dans le génome des mammifères. Au terme d'une recombinaison homologue, les sites loxP sont intercalés à des positions bien étudiées pour ne pas gêner la transcription du gène. Le gène est "floxxé". La mutation est alors réalisée de manière cryptique. Elle deviendra effective lorsque l'enzyme Cre sera synthétisée dans les cellules ES en raison de la délétion du segment floxxé. L'utilisation d'une variété de promoteurs, inductibles ou non, permet de faire exprimer la mutation au moment et dans le tissu choisis.

cellules d'une partie des descendants de ces deux souris, l'enzyme qui coupe aux sites loxP et le gène floxxé seront mis en présence. La délétion se produira et l'exon sera éliminé (figure 11). L'avantage de cette technique est évident : on peut induire la délétion quand on veut et dans le tissu que l'on aura choisi. Il suffit par exemple d'utiliser pour construire le transgène Cre, un promoteur inductible par une substance exogène (la tétracycline par exemple), et le transgène Cre ne sera activé que si la souris boit une solution de tétracycline ou la reçoit par inoculation (figure 12). On peut aussi choisir un promoteur très spécifique d'un tissu et la délétion ne se produira que dans le tissu en question. D'autres raffinements sont encore disponibles dans l'arsenal des généticiens de la souris. Avec des transgènes plus élaborés et des substances inductibles, comme la tétracycline ou la doxycycline, on peut désormais faire fonctionner les transgènes de manière réversible.

En d'autres termes, et pour conclure sur le sujet, on peut dire que de nos jours, n'importe quel élément de la séquence du génome de la souris, qu'il soit codant ou non, peut être délété ou modifié extrêmement finement. En combinant les délétions par Cre-loxP et l'addition d'information par transgénèse classique, on peut aussi étudier les effets du dosage génique. En effet, on peut déléter la copie d'un gène qui est sur un chromosome (haploïdie fonctionnelle), ou bien les deux copies, mais on peut aussi ajouter une copie, voire deux ou trois ou plus. Ainsi, un modèle de mongolisme, très fidèle, est en train d'être mis au point par des chercheurs de plusieurs pays, qui ajoutent au génome de la souris des copies surnuméraires de gènes provenant d'une région critique du chromosome 21 de l'homme, connue pour être impliquée dans le syndrome du mongolisme. D'énormes progrès sont à attendre de ces développements.

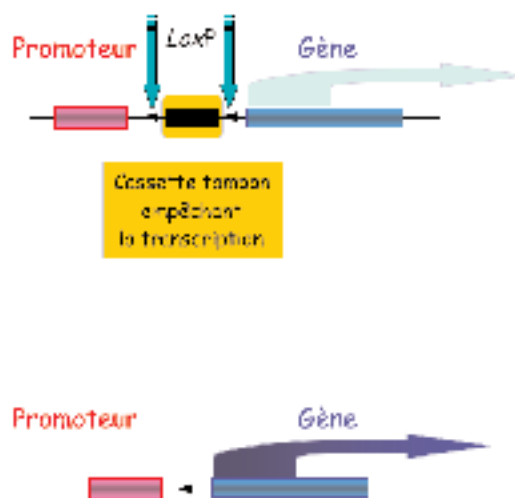


Figure 12 : Induction de l'expression d'un transgène par le système Cre-loxP – A. Une cassette spéciale, préalablement floxxée, est introduite entre le promoteur d'un transgène et le gène proprement dit. Cette cassette empêche la transcription du gène aussi longtemps qu'elle est en place. **B.** Sa délétion par la Cre recombinase permet l'expression du gène.

• LE SÉQUENÇAGE DU GÉNOME ET SES ENSEIGNEMENTS

Le génome de la souris est aujourd'hui presque complètement séquencé (WATERSTON *et al.*, 2002). Ceci veut dire que les généticiens connaissent avec précision l'agencement linéaire des 2,8 milliards de paires de bases qui constituent le génome nucléaire de cette espèce. Certes, quelques ambiguïtés subsistent et quelques trous (ou *gaps*) restent à combler, mais le travail avance bon train et il est prévu que tout soit achevé en 2007. Le génome de la souris sera alors le deuxième génome de mammifère à être séquencé en totalité, après celui de l'homme. Le plus difficile et le plus long sera sans doute d'établir la séquence des régions dupliquées car les spécialistes du séquençage sont relativement désarmés dans ces cas très précis. Cet énorme travail a été réalisé en moins de quatre ans par un consortium de plusieurs laboratoires anglo-saxons (il n'y eut, hélas, aucune contribution française dans ce projet, contrairement au génome humain), avec une incroyable précision, puisqu'on sait que le taux d'erreur dans les séquences terminées est inférieur à 10^{-5} . Pour en arriver là, il aura fallu séquencer sept fois le génome et retenir comme référence la séquence consensus (la plus fréquente) parmi les sept séquences primaires.

La connaissance de cette séquence est intéressante à un double titre :

- d'abord, elle permet de faire l'inventaire complet des séquences codantes, que les informaticiens savent maintenant repérer assez précisément grâce à des programmes informatiques très élaborés. Elle permet aussi de préparer très finement les manipulations génétiques que nous avons décrites ci-dessus ;
- ensuite et surtout, elle permet de faire des comparaisons entre espèces, proches ou lointaines, ce qui conduit à interpréter ces différences et ces ressemblances en termes de fonction.

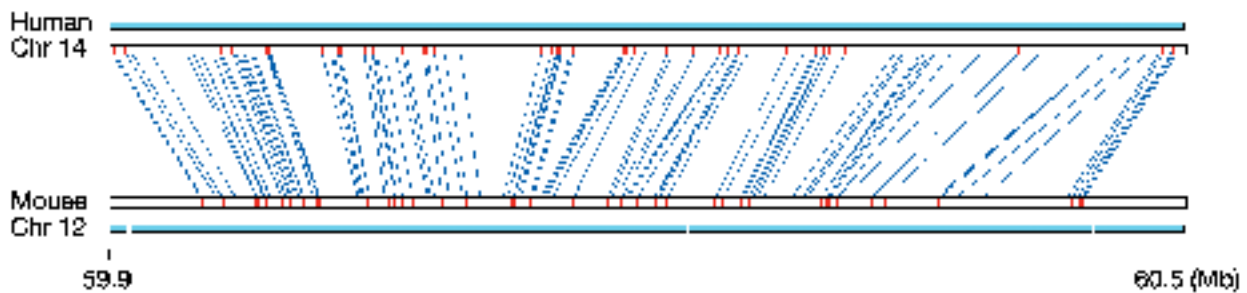


Figure 13 : Le génome de la souris est entièrement séquencé. Le nombre de gènes est pratiquement le même chez l'homme, le rat et la souris : il est voisin de 28 000. La plupart (>90 %) des gènes humains ont un orthologue direct chez la souris et le rat et vice-versa. L'alignement des gènes sur les chromosomes est très conservé sur des courtes distances. Cela permet de faire des prédictions dans une espèce en se fondant sur les données solidement établies dans d'autres. De ce point de vue, la souris sert de modèle pour l'identification des gènes dans les espèces à génome non encore séquencé (D'après WATERSTON et al., 2002).

Comme on pouvait s'y attendre, le génome de la souris ressemble beaucoup à celui de l'homme. Lorsqu'on aligne les séquences, on s'aperçoit que 95 % au moins des gènes d'une espèce ont un orthologue (c'est-à-dire un équivalent parfait avec une séquence très voisine et la même fonction) dans l'autre espèce et que les gènes en question présentent souvent le même arrangement linéaire dans les deux espèces, témoignant ainsi de l'existence d'un ancêtre commun (figure 13). Le génome de l'ancêtre commun a évolué indépendamment, pendant 70 à 80 millions d'années, vers les deux espèces et de nombreux réarrangements se sont produits pour défaire et refaire de manière différente, l'arrangement linéaire des gènes sur les chromosomes. Chaque génome pourrait en fait être réduit à 350 segments parfaitement homologues mesurant entre 65 Mégabases (Mb) pour le plus long et seulement 300 kilobases (kb) pour le plus court (moyenne 7 Mb). Moins attendu et très intéressant, est le fait que 5 % des séquences du génome de la souris se montrent beaucoup plus conservées que le reste du génome, lorsqu'on fait des comparaisons avec plusieurs espèces plus ou moins apparentées, dont le génome est également séquencé (par exemple l'homme, le rat, le porc, le chien, le chimpanzé, le poulet etc...). Cela veut dire que ces séquences, à la différence du reste du génome, sont soumises à une pression de sélection pour ne pas s'écarter trop d'une séquence canonique et que, vraisemblablement, elles doivent avoir une fonction. Un et demi pour cent de ces séquences sont codantes et la conservation de la séquence s'explique assez naturellement si on considère qu'elles codent une protéine dont la fonction, déterminée par la séquence, est essentielle. Le reste (3,5 %) correspond sans doute aux séquences qui sont impliquées dans la régulation de l'expression des gènes et à propos desquelles on connaît encore bien peu de choses. Leur identification sera la prochaine étape dans l'analyse structurale du génome et, ici encore, d'importants progrès sont attendus.

La compilation des séquences codantes aboutit à dénombrer environ 250 000 exons chez la souris et 25 à 28 000 gènes au total, c'est-à-dire un peu plus que chez l'homme. Mais l'inventaire des gènes a peu de sens pour un généticien car certains gènes sont dupliqués dans une espèce et pas dans l'autre, et rien ne prouve que tous les gènes sont fonctionnels et indispensables. Ce qui est important, c'est le nombre de transcrits potentiellement présents dans un génome mais cette donnée,

malheureusement, est très difficile à obtenir avec précision parce que certains gènes sont exprimés (donc transcrits) dans de très petits territoires et/ou pendant très peu de temps. À l'heure actuelle, les Japonais de l'Institut RIKEN de Yokohama ont entrepris de faire l'inventaire complet de tous les messagers transcrits dans le génome de la souris (CARNINCI *et al.*, 2005). Ils ont réalisé une base de données (FANTOM pour Functional Annotation of the Mouse genome) et disposent d'une librairie qui comporte 103 000 clones de cDNA complets. Si l'on tient compte des redondances (faciles à estimer) et des lacunes (inappréciables sur des bases rationnelles), on peut considérer qu'un génome de souris peut se traduire en 65 000 protéines environ. Mais toutes ces protéines ont-elles une fonction essentielle ? C'est, pour l'instant, impossible à savoir et il faudra attendre que tout le génome soit annoté et notamment que tous les gènes aient été invalidés l'un après l'autre pour se prononcer. Un joli programme pour la (pour les) décennie(s) à venir !

Si les génomes de l'homme et de la souris présentent de très grandes similitudes, ils présentent aussi quelques différences intéressantes. Ainsi, on a déjà repéré que certains gènes, présents dans une espèce n'étaient pas retrouvés chez l'autre. La souris, par exemple, n'a pas le gène *Il8* qui code l'Interleukine 8 qui existe chez l'homme (*IL8*). Mais elle a encore le gène (devenu inutile) qui code le récepteur de cette même cytokine (*Ilr8*). Le gène qui code la protéine acide du lait dénommée *whey acidic protein* (*Wap*) de la souris n'existe pas chez l'homme. À l'inverse, les gènes qui codent les *major urinary proteins* (*Mup*), des protéines du genre lipocalines qui sont chez la souris (et le rat) support des phéromones excrétée dans l'urine, n'existent pas chez l'homme.

Au-delà de ces différences ponctuelles, il en existe d'autres qui portent sur les variations du nombre de copies de certains gènes dans un génome et pas dans l'autre. Le gène *OAS1*, qui code une enzyme dénommée Oligo-Adenylate synthétase, dont le rôle est fondamental pour la mise en œuvre des défenses innées de l'organisme vis-à-vis de certaines infections virales (virus à ARN positifs tels que dengue, fièvre jaune, etc.), est présent dans le génome de l'homme sous forme d'une unique copie; tandis que chez la souris, il en existe rien moins que huit copies arrangées en tandem (MASHIMO *et al.*, 2003). Pour l'heure, on ne connaît pas très bien les raisons de cette ampli-

fication génique dans une espèce et pas dans l'autre. Est-ce fortuit, comme certains le pensent, et dans ce cas, les sept copies supplémentaires de la souris n'auraient pas de fonction indispensable et seraient de simples pseudo-gènes, ou bien au contraire est-ce une adaptation à un environnement plus agressif ? La vérité doit être entre ces deux extrêmes, mais pour le savoir, il faut expérimenter, c'est-à-dire, dans le cas présent, invalider une à une chacune des huit copies et voir les effets produits. Ce travail est en cours, notamment dans notre laboratoire. Il sera très riche en enseignements, mais il est long et coûteux. D'après les premiers résultats, on sait qu'il y a plusieurs pseudo-gènes parmi ces huit copies mais on sait aussi que d'autres copies sont fonctionnelles et indispensables à la mise en œuvre de défenses innées non-spécifiques contre les infections à flavivirus. On pense que cette duplication de gènes dans certains génomes et pas dans d'autres, procède d'une stratégie assez souvent utilisée par l'évolution, stratégie que les auteurs anglo-saxons appellent le *genome shaping*. Ce serait un des moyens pour les espèces d'adapter leur génétique aux variations du milieu ambiant.

• VERS OÙ ALLONS NOUS ?

Il est clair que le séquençage du génome de la souris a eu les effets d'une véritable révolution dans le microcosme des "génomiciens". D'un côté, on peut faire l'inventaire de tous les gènes, comparer les séquences en les alignant avec celles d'autres espèces et utiliser la plus ou moins grande conservation des homologies observées comme critère pour

en tirer des conclusions (ou simplement pour formuler des hypothèses) sur la fonction de la séquence étudiée. De l'autre, on peut préparer, avec une précision d'horloger, un changement dans la séquence et observer ses effets chez un animal transgénique. Aucune autre espèce ne permet de faire ce travail expérimental qui conduira à une connaissance véritablement fondamentale du génome en associant, directement, la structure (la séquence) à la fonction des gènes. De ce point de vue, il n'est pas exagéré de dire que la souris a joué (et continue de jouer) pour les généticiens de mammifère le rôle d'*Arabidopsis thaliana* pour les généticiens des végétaux : aucune de ces deux espèces n'a d'intérêt en elle-même mais l'outil qu'elles représentent est irremplaçable.

Au-delà des progrès attendus dans la connaissance du génome des mammifères, on peut prévoir de substantiels progrès dans notre connaissance de la génétique quantitative, cette branche de la génétique qui s'intéresse à l'hérédité des caractères à déterminisme multigénique, tels que la stature, la répartition des graisses dans l'organisme, la résistance aux maladies infectieuses, les maladies autoimmunes comme le diabète, les hypercholestérolémies, etc. L'hérédité des caractères à déterminisme multigénique est difficile à étudier car, en plus des séquences codantes qui sont responsables du polymorphisme observé au niveau des protéines, elle implique les séquences qui régulent l'activité transcriptionnelle des gènes ; elle dépend aussi du nombre de copies de ces mêmes gènes et de leurs interactions de toute nature. La souris est presque irremplaçable en la matière...

BIBLIOGRAPHIE

- BONHOMME F, GUÉNET JL (1996) The laboratory mouse and its wild relatives. In: RASTAN S, LYON MF, BROWN SDM editors : "Genetic variants and strains of the laboratory mouse". Oxford University Press, Oxford, 1577-1596.
- BONHOMME F, GUÉNET JL, DOD B, MORIWAKI K, BULFIELD G. (1987) The polyphyletic origin of laboratory inbred mice and their rate of evolution. *Biol. J. of the Linnean Society*, **30**, 51-58.
- BRINSTER RL, PALMITER RD (1984) Introduction of genes into the germ line of animals. *Harvey Lect.*, **80**, 1-38.
- CARNINCI P, KASUKAWA T, KATAYAMA S, GOUGH J, FRITH MC, MAEDA N *et al.*, (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559-1563.
- CONDAMINE H, CUSTER RP, MINTZ B (1971) Pure-strain and genetically mosaic liver tumors histochemically identified with the -glucuronidase marker in allophenic mice. *Proc. Natl. Acad. Sci. USA*, **68**, 2032-2036.
- CUÉNOT L (1902) La loi de Mendel et l'hérédité de la pigmentation chez la souris. *Arch. Zool. Exp. Gén.*, 3^e série **3**, 27-30.
- GORDON JW, SCANGOS GA, PLOTKIN DJ, BARBOSA JA, RUDDLE FH (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. USA*, **77**, 7380-7384.
- GUÉNET JL (2004) Chemical mutagenesis of the mouse genome: an overview. *Genetica*, **122**, 9-24.
- GUÉNET JL (2005) *The Mouse Genome. Genome Res.* (sous presse)
- GUÉNET JL, BONHOMME F (2003) Wild mice : an ever-increasing contribution to a popular mammalian model. *Trends Genet.*, **19**, 24-31.
- GUÉNET JL, BONHOMME F (2004) Origin of the Laboratory Mouse and related subspecies. In: HEDRICH H editor. *The Laboratory Mouse (Handbook of Experimental Animals)* ISBN 0-1233-6425-6, pp 3-12.
- KOIKE S, HORIE H, SATO Y, ISE I, TAYA C, NOMURA T, YOSHIOKA I, YONEKAWA H, NOMOTO A (1993) Poliovirus-sensitive transgenic mice as a new animal model. *Dev. Biol. Stand.*, **78**, 101-107.
- MASHIMO T, GLASER P, LUCAS M, SIMON-CHAZOTTES D, CECCALDI PE, MONTAGUTELLI X, DESPRES P, GUENET JL. (2003) Structural and functional genomics and evolutionary relationships in the cluster of genes encoding murine 2',5'-oligoadenylate synthetases. *Genomics*, **82**, 537-552.
- PAIGEN K (2003) One hundred years of mouse genetics : an intellectual history. I. The classical period (1902-1980). *Genetics*, **163**, 1-7.
- PALMITER RD, BRINSTER RL, HAMMER RE, TRUMBAUER ME, ROSENFELD MG, BIRNBERG NC, EVANS RM (1982) Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature*, **300**, 611-615.

- ROBERTSON E, BRADLEY A, KUEHN M, EVANS M. (1986) Germ-line transmission of genes introduced into cultured pluripotent cells by retroviral vector. *Nature*, **323**, 445-448.

- SILVER LM (1995) *Mouse Genetics: Concepts and applications*. Oxford University Press Oxford and New York. 362p, pp 133-283.

- WADE CM, KULBOKAS EJ 3RD, KIRBY AW, ZODY MC, MULLIKIN JC, LANDER ES, LINDBLAD-TOH K, DALY MJ (2002) The mosaic structure of variation in the laboratory mouse genome. *Nature*, **420**, 574-578.

- WAKAYAMA T, YANAGIMACHI R (1998) Fertilisability and developmental ability of mouse oocytes with reduced amounts of cytoplasm. *Zygote*, **6**, 341-346.

- WATERSTON RH, LINDBLAD-TOH K, BIRNEY E et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520-562.

- YANAGIMACHI R (2002) Cloning: experience from the mouse and other animals. *Mol Cell Endocrinol.*, **187**, 241-248.

- YONEKAWA H, MORIWAKI K, GOTOH O, MIYASHITA N, MIGITA S, BONHOMME F, HJORTH JP, PETRAS ML, TAGASHIRA Y (1982) Origins of laboratory mice deduced from restriction patterns of mitochondrial DNA. *Differentiation*, **22**, 222-226.

