

## B. Les CNV chez la Souris

La Souris (*Mus musculus*) est un bon modèle pour les études génétiques, car c'est un petit mammifère qui est facile à élever, prolifique et qui supporte bien la consanguinité, ce qui permet de créer des lignées consanguines où les individus sont homozygotes à tous les locus. De nombreuses lignées consanguines existent et sont très facilement accessibles. Le génome de la Souris est bien connu et annoté (*Mouse Genome Sequencing Consortium*, 2002). Connaître les variations génétiques au sein de cette espèce est nécessaire à la recherche bio-médicale, car la Souris est le mammifère le plus utilisé pour modéliser les maladies complexes humaines.

Les CNV se sont imposés comme une nouvelle source de variation génétique chez la Souris suite aux découvertes faites chez l'Homme. Plusieurs études ont recensé les CNV chez la Souris et ont montré que ces variants contribuaient à la diversité génétique, aux modifications d'expression génique, aux différences phénotypiques, et étaient susceptibles de participer à la vulnérabilité à certaines maladies (Li *et al.*, 2004 ; Snijders *et al.*, 2005 ; Cutler *et al.*, 2007 ; Egan *et al.*, 2007 ; Graubert *et al.*, 2007 ; Watkins-Chow et Pavan, 2007 ; She *et al.*, 2008 ; Cahan *et al.*, 2009 ; Henrichsen *et al.*, 2009 B).

La Souris est, à ce jour, l'animal dont les CNV ont été le plus étudiés. Graubert *et al.*, Cutler *et al.*, Egan *et al.*, et She *et al.* ont recensé les CNV dans différentes lignées de Souris. L'étude de Watkins-Chow et Pavan est, à ce jour, la seule qui ait étudié les CNV au sein d'une même lignée consanguine de Souris (C57BL/6J). Cahan *et al.* et Henrichsen *et al.*, ont étudié le transcriptome de différents tissus de Souris, afin d'étudier l'effet des CNV sur l'expression des gènes.

### 1. Les CNV identifiés entre les différentes lignées consanguines

Graubert *et al.*, en 2007, ont recensé 80 CNV dans le génome de 21 lignées de Souris de laboratoire caractérisées dans la *Mouse Phenome Database* (<http://www.jax.org/phenome> ; base de données publique où sont répertoriées les données concernant le génotype et le phénotype des lignées de Souris les plus utilisées). Ces CNV ont été détectés par hybridation comparative à haute résolution. Les génomes des vingt lignées testées ont été comparés au génome de la lignée C57BL/6J. Les CNV répertoriés avaient une taille comprise entre 21,4kb et 2Mb, la taille moyenne

était de 271,5kb. Vingt-deux CNV ont été recensés en moyenne par lignée, avec des écarts allant de 2 CNV recensés pour la lignée C57BL/6Tac à 38 pour la lignée NOD/LtJ. Graubert *et al.* soutenaient que le contenu en CNV de chaque lignée permettait de retracer la généalogie des lignées de Souris. Ils ont remarqué que le nombre de CNV recensés (80) paraissait faible (Li *et al.* en 2004 avaient recensés 278 CNV chez la Souris). Deux raisons pouvaient être à l'origine du nombre réduit de CNV détectés : une trop forte stringence des critères de sélection des CNV et un nombre réduit de sondes dans les zones répétées du génome qui sont très riches en CNV (Sharp *et al.*, 2005; Perry *et al.*, 2006 ; Graubert *et al.*, 2007).

Graubert *et al.* ont classé ces 80 CNV en trois catégories : les CNV **mono-alléliques** présents dans une seule des 21 lignées ; les CNV mono-alléliques présents dans plusieurs lignées et les CNV **multi-alléliques** présents dans une ou plusieurs lignées. Graubert *et al.* ont suggéré que les CNV mono-alléliques étaient apparus tôt dans les lignées et étaient fixés dans les populations ; alors que les CNV multi-alléliques semblaient être le résultat de plusieurs remaniements qui étaient survenus indépendamment dans différentes lignées. Graubert *et al.* ont supposé que les régions du génome de la Souris qui contenaient des CNV multi-alléliques étaient plus favorables aux réarrangements que le reste du génome. Ils ont également suggéré que ces régions dynamiques du génome, susceptibles de varier fréquemment dans leur nombre de copies, entraînaient des variations génétiques entre les lignées mais aussi au sein des lignées.

Graubert *et al.*, ont comparés la localisation des 80 CNV qu'ils ont trouvés, aux duplications segmentaires cartographiées chez la Souris (Cheung *et al.*, 2003 ; Bailey *et al.*, 2004). Les duplications segmentaires sont des régions du génome de plus d'un kilobase ayant une séquence similaire à plus de 90% entre elles et retrouvées plusieurs fois dans le génome. De nombreuses études ont montré un surplus significatif en CNV (4 à 10 fois) dans les régions du génome présentant des duplications segmentaires chez l'Homme (Sebat *et al.*, 2004 ; Iafrate *et al.*, 2004 ; Sharp *et al.*, 2005) et chez la Souris (Li *et al.*, 2004; She *et al.*, 2008). Graubert *et al.* ont montré que 38 des 80 CNV (47,5%) co-localisaient ou étaient voisins de duplications segmentaires. Ils ont montré que l'association entre les CNV et les duplications segmentaires restait significative jusqu'à 2Mb de distance. Graubert *et al.* ont suggéré que les duplications segmentaires étaient propices à la formation des CNV chez la Souris, comme chez les autres mammifères, car elles favoriseraient le phénomène de recombinaison homologue non-allélique, qui entraîne des duplications, des délétions

et des inversions au sein du génome (Guryev *et al.*, 2007 ; Perry *et al.*, 2006, McCaroll *et al.*, 2008 ; Nicholas *et al.*, 2009).

Graubert *et al.* ont également étudié la position des CNV par rapport aux gènes de la Souris. Ils ont trouvé que 62 des 80 CNV (77,5%) touchaient au moins un gène ou une partie d'un gène. Ce résultat n'était pas significativement différent d'une répartition au hasard des CNV dans le génome de la Souris. Néanmoins, certaines catégories de gènes, (comme les gènes impliqués dans l'activité des récepteurs couplés à la protéine G, dans l'activité du récepteur *Rhodopsine-like*, ou encore dans l'activité des récepteurs olfactifs), étaient significativement plus touchées par les CNV que d'autres catégories de gènes. Par exemple, Graubert *et al.* ont détecté une délétion emportant le gène *Klra8*, codant pour une lectine transmembranaire de type C (analogue à un récepteur de *natural killer* humain), chez 5 lignées de Souris. Il avait déjà été montré que ce gène était un facteur de résistance aux infections par des cytomégalovirus (Lee *et al.*, 2001 ; Graubert *et al.*, 2007). Graubert *et al.* ont soutenu que les 5 lignées délétées avaient plus de risque d'être sensibles aux cytomégalovirus que les lignées non délétées. Graubert *et al.* ont également identifié, chez plusieurs lignées dont BALB/cByJ, une augmentation du nombre de copies du gène *Intlna*, codant pour une intellectine, protéine de surface des cellules épithéliales intestinales impliquée dans la réponse immunitaire innée anti-parasitaire. Cette observation confirmait la duplication de ce gène dans la lignée BALB/c, qu'une étude antérieure avait mis en évidence (Pemberton *et al.*, 2004). Une augmentation du nombre de copies de ce gène avait été associé à une résistance accrue des Souris BALB/c à *Trichinella spiralis*, par rapport aux Souris C57BL/6J qui n'étaient pas porteuses de cette duplication (Pemberton *et al.*, 2004). Graubert *et al.* ont postulé que les autres lignées porteuses de cette duplication avaient plus de chances d'être résistantes à *Trichinella spiralis*. L'étude de Graubert *et al.* a suggéré que les CNV avaient des conséquences phénotypiques, notamment concernant la différence de sensibilité aux infections entre les lignées de Souris de laboratoire. Graubert *et al.* ont également identifié 5 gènes affectés par des CNV, chez la Souris, qui l'étaient également chez l'Homme. La présence de CNV récurrents dans le génome de plusieurs espèces suggérait qu'une pression sélective avait pu encourager la conservation des CNV qui conféraient un avantage sélectif aux individus.

Cutler *et al.* en 2007, ont détectés par hybridation comparative (génome de référence C57BL/6J) les CNV chez 41 lignées de Souris. Ils ont identifié 2094 CNV, dont 793 duplications d'une taille moyenne de 183 kb et 1303 délétions d'une taille moyenne de 207kb. Ils ont trouvé en

moyenne 51 CNV par lignée, qui couvraient en moyenne 10 Mb, soit 0,38% du génome de chaque lignée. Cutler *et al.*, ont identifié 26 fois plus de CNV que Graubert *et al.* ; cette différence importante pourrait être due à l'emploi de puces d'hybridation génomique comparative différentes, mais surtout, à l'utilisation d'algorithmes d'analyse des CNV différents.

Cutler *et al.* ont observé que les lignées les plus éloignées génétiquement de la lignée de référence C57BL/6J étaient celles qui avaient le plus grand nombre de CNV. Par exemple, la lignée SPRET/EiJ présentait plus de 36Mb d'ADN qui variaient par rapport à C57BL/6J. Cutler *et al.* affirmaient qu'un arbre phylogénique construit à partir des CNV recensés dans les différentes lignées représenterait l'histoire évolutive de ces lignées. Cette observation était concordante avec plusieurs autres études des CNV chez la Souris (Graubert *et al.*, 2007 ; Henrichsen, Vinckenbosch *et al.*, 2009).

Cutler *et al.* ont étudié le contenu génique des CNV qu'ils ont identifié. Ils ont trouvé que les délétions étaient plus souvent hors des gènes que les duplications. Par contre, les délétions contenaient deux fois plus de **pseudogènes** que les autres régions du génome. Cette étude montrait également que les gènes et les **îlots CPG** (courtes séquences méthylées présentes au niveau de la région régulatrice de certains gènes) étaient moins touchés par les CNV. Ces résultats suggéraient qu'une pression de sélection négative pourrait jouer sur la conservation des CNV qui touchent des gènes ou leurs éléments de régulation, et que l'ADN fonctionnel aurait tendance à être exclu des CNV.

Cutler *et al.* ont remarqué que les CNV touchaient significativement plus de gènes appartenant à de grandes familles multigéniques, que de gènes ayant pas ou peu de **paralogues**. Ils suggéraient que le nombre de copies des gènes redondants dans le génome était plus flexible que pour les autres gènes. Cutler *et al.* ont montré que les gènes impliqués dans la liaison aux phéromones et dans la réponse immunitaire étaient plus souvent touchés par des CNV que d'autres catégories fonctionnelles de gènes. Ces gènes appartiennent à de grandes familles qui évoluent rapidement. Au contraire, les gènes impliqués dans des mécanismes cellulaires de base, comme par exemple la régulation du cycle cellulaire, étaient significativement moins touchés par les CNV (Cutler *et al.*, 2007). La pression de sélection a certainement joué un rôle majeur dans l'élimination ou la conservation de certains CNV au cours de l'évolution. Les CNV et les duplications

segmentaires leur ont semblé une force importante pour l'évolution, car ils apportaient la matière nécessaire à la formation de nouveaux gènes (Cutler *et al.*, 2007).

Cutler *et al.* ont réalisé une étude d'association entre les CNV qu'ils avaient identifié et un caractère multigénique complexe : la prise alimentaire. Le comportement alimentaire de chaque lignée avait été étudié par Seburn *et al.* (Seburn *et al.*, 2001). Cutler *et al.* ont identifié 3 CNV associés à la prise alimentaire, dont un qui touchait le gène *Glp1r* (*glucagon-like peptide 1 receptor*) gène qui joue un rôle dans la satiété et l'homéostasie du poids chez l'Homme ainsi que dans le besoin énergétique chez la Souris (Cutler *et al.*, 2007). Ainsi, Cutler *et al.* soutenaient, comme Graubert *et al.*, que les CNV étaient impliqués dans les différences phénotypiques entre les lignées consanguines de Souris.

Egan *et al.* en 2007 ont recherché les CNV récurrents chez la Souris, c'est-à-dire ceux qui sont retrouvés dans plusieurs lignées. L'objectif de l'étude d'Egan *et al.* était de déterminer la fréquence d'apparition des CNV, dans différentes régions du génome. Ils ont étudié 14 sous-lignées dérivées récemment de la lignée C57BL/6. La lignée C57BL/6 est très utilisée en recherche biomédicale et de nombreuses sous-lignées sont élevées dans différents centres de recherche. Les génomes de ces sous-lignées ont divergé les uns des autres suite à l'accumulation de mutations spontanées (Egan *et al.*, 2007).

Egan *et al.* ont trouvé 38 CNV parmi les 14 sous-lignées dérivées de la lignée C57BL/6. Les 38 CNV touchaient 60 gènes et leur taille variait de 4kb à 4Mb.

Egan *et al.* ont recherché ces 38 CNV chez 12 autres lignées de Souris (non apparentées à C57BL/6). Sur les 38 CNV, Egan *et al.* ont observés 20 CNV qui n'apparaissaient que dans une seule des 14 sous-lignées C57BL/6. Ils ont suggéré que ces CNV mono-alléliques étaient le fruit d'un événement mutationnel unique dont ils ont estimé la fréquence d'apparition à  $1,2 \times 10^{-6}$ . Les 18 autres CNV étaient récurrents. Ils ont estimé que la fréquence d'apparition de ces CNV récurrents était comprise entre  $3,6 \times 10^{-3}$  et  $1,1 \times 10^{-2}$ . D'après Egan *et al.* les CNV récurrents seraient le résultat de plusieurs événements mutationnels indépendants, au même locus, dans différentes lignées. Cette étude suggérait que la fréquence d'apparition des CNV au sein du génome de la Souris pourrait varier d'un facteur de  $10^4$ . Egan *et al.* ont supposé que certaines régions du génome (surtout celles comprenant des éléments répétés) étaient susceptibles de muter plusieurs fois sur quelques générations seulement.

Des CNV récurrents ont également été observés entre différentes races de Chiens (Nicholas *et al.*, 2009).

She *et al.*, en 2008, ont étudié la localisation des duplications segmentaires dans le génome de la Souris et le contenu en CNV de ces régions. Ils ont tout d'abord identifié les duplications segmentaires dans la lignée C57BL/6J, puis ont recherché les CNV associés à ces duplications segmentaires chez 15 autres lignées de Souris. She *et al.* ont observé que 4,94% du génome de la Souris était couvert par des duplications segmentaires. Ce pourcentage était similaire à celui trouvé chez l'Homme (5,5%) (Sharp *et al.*, 2005). En revanche, la répartition des duplications segmentaires entre ces deux espèces était très différente. Les duplications segmentaires de la Souris étaient le plus souvent intrachromosomiques, en tandem, et regroupées dans certaines parties du génome, alors que les duplications segmentaires chez l'Homme étaient presque toutes interchromosomiques et dispersées dans le génome. She *et al.* ont montré que 60% de la séquence des duplications segmentaires (soit 20Mb d'ADN en moyenne) étaient polymorphes entre les lignées de Souris. Ils ont trouvé que les CNV ayant la plus grande fréquence de mutation ( $10^{-2}$ ) étaient ceux situés entièrement ou presque entièrement (de 80 à 90%) dans les duplications segmentaires. Ils ont suggéré que ces régions d'homologies, alignées en tandem dans le génome de la Souris, favorisaient le phénomène de recombinaison homologue non allélique et donc l'apparition des CNV.

She *et al.* ont identifié 2424 CNV chez les 15 lignées de Souris, dont 1259 duplications et 1958 délétions. Les résultats de She *et al.* suggéraient que la Souris possédait autant, voire un peu plus, de CNV que l'Homme. She *et al.* ont identifié 353 gènes qui présentaient un nombre de copies augmenté ou diminué. Trente et un de ces gènes étaient touchés par différents allèles d'un même CNV, dans différentes lignées. She *et al.* ont trouvé significativement plus de CNV dans des familles multi-géniques associées à la spermatogénèse, à la gestation, à la viviparité, à la sensibilité aux phéromones et à la réponse immunitaire. Cutler *et al.* en 2007 avaient déjà mis en évidence que ces deux dernières catégories de gènes étaient enrichis en CNV. Chez l'Homme également, les CNV semblaient surreprésentés dans les familles de gènes impliqués dans la réponse immunitaire (Redon *et al.*, 2006).

She *et al.* ont, par exemple, montré que les gènes *Defcr21*, *22*, *23* et *Def5b1*, codant pour des *defensin*, et la famille génique des *Klra* (*killer cell lectin-like receptor*) faisaient partie des blocs dupliqués riches en CNV. Ces deux familles de gènes avaient déjà été associées à la variation de

sensibilité aux infections entre les lignées de Souris (Lee *et al.*, 2001 ; Graubert *et al.*, 2007 ; She *et al.*, 2008).

## 2. Les CNV au sein des lignées consanguines

Watkins-Chow et Pavan en 2008 ont étudiés les CNV au sein de la lignée C57BL/6J. Jusque là seuls les CNV entre différentes lignées de Souris avaient été recherchés (Li *et al.*, 2004; Snijders *et al.*, 2005; Cutler *et al.*, 2007 ; Egan *et al.*, 2007 ; Graubert *et al.*, 2007). La lignée C57BL/6J est la plus utilisée et la mieux connue dans la recherche biomédicale. Cette lignée, dérivée de la lignée C57BL en 1921 par Dr C.C. Little, est maintenue au *Jackson Laboratory* (Etats-Unis) depuis plus de 70 ans par croisements frère-sœur, afin de conserver une lignée consanguine où les individus sont homozygotes à tous les locus.

Watkins-Chow et Pavan ont détectés deux duplications, l'une incluant le gène *Ide* (*Insulin degrading enzyme*) codant pour une enzyme de dégradation de l'insuline et l'autre incluant le gène *Fgfbp3* (*Fibroblast growth factor binding protein 3*) codant pour une protéine se liant au facteur de croissance fibroblastique. Une recherche de la duplication touchant le gène *Ide* chez 39 reproducteurs C57BL/6J a montré que 64% de ces animaux étaient hétérozygotes pour cette duplication (ils possédaient une copie dupliquée et une non dupliquée), 23% étaient homozygotes non dupliqués et 13% étaient homozygotes dupliqués. Watkins-Chow et Pavan ont supposé que cette duplication, dont les fréquences alléliques respectaient l'**équilibre de Hardy-Weinberg**, était apparue rapidement après la création de cette lignée.

Watkins-Chow et Pavan ont également étudié la différence d'expression des gènes *Ide* et *Fgfbp3* entre les Souris dupliquées et non dupliquées. Les animaux porteurs de la duplication qui touchait le gène *Ide* présentaient une expression significativement plus élevée du gène dans la rate et le cerveau (qui étaient les deux tissus testés). Les animaux porteurs de la duplication qui touchait *Fgfbp3* présentaient une expression plus élevée de ce gène dans la rate mais pas dans le cerveau. Ces résultats suggéraient que les éléments de régulation spécifiques de la rate du gène *Fgfbp3* étaient altérés par la duplication, ce qui n'était pas le cas pour les éléments de régulation spécifiques du cerveau ou bien que des mécanismes de contrôle agissaient pour corriger les modifications d'expression dans le cerveau.

Cette étude suggérait que les CNV pouvaient avoir des conséquences sur l'expression des gènes chez des individus d'une même lignée.

L'identification de CNV au sein de lignées de Souris de laboratoire pourrait augmenter la diversité des lignées disponibles pour la recherche biomédicale. Etant donné l'importance attribuée aux CNV dans la prédisposition à certaines maladies humaines, un recensement précis des CNV chez la Souris semble crucial pour renforcer le potentiel de la Souris comme modèle des variations génétiques humaines. Par exemple, des variants proches du locus *Ide* ont été associés au risque de développer un diabète de type 2 chez l'Homme (Scott *et al.*, 2007) et la maladie d'Alzheimer (Mueller *et al.* 2007). Des Souris porteuses d'une duplication spontanée à ce locus et qui présentent des variations d'expression de ce gène, pourraient être utile pour comprendre le rôle du gène *Ide* dans ces deux maladies.

La recherche des CNV dans l'étude de Watkins-Chow et Pavan ne couvrait qu'une partie très limitée du génome. Ceci suggérait que d'autres CNV pourraient ségréger dans la lignée C57BL/6J, bien que cette lignée soit maintenue « pure » avec précaution. Ces résultats remettaient en question l'homozygotie à tous les locus des individus de cette lignée. La présence de ces variants pourrait influencer l'interprétation d'expériences réalisées sur cette lignée. C'est pourquoi cartographier les CNV de cette lignée de manière exhaustive semble indispensable. Des CNV pourraient être présents au sein d'autres lignées également. Une prévalence élevée des CNV au sein des lignées consanguines compliquerait l'étude des maladies complexes et l'identification de locus contrôlant des caractères complexes, qui s'appuient sur des **lignées recombinantes** ou **congéniques**. En effet, des polymorphismes à certains locus, hors des régions sélectionnées, pourraient être présents. Watkins-Chow et Pavan soutenaient que les CNV, avec l'environnement, les phénomènes épigénétiques, les événements stochastiques et sûrement d'autres phénomènes, participaient à la variation phénotypique des individus au sein d'une lignée.

### 3. Les effets des CNV sur l'expression des gènes

Cahan *et al.* en 2009 ont recensé, par hybridation comparative, les régions comportant des CNV de 20 lignées de Souris et ont étudié le profil d'expression des gènes touchés par ces CNV *in vivo*. Ils ont identifié 1333 régions distinctes touchées par des CNV. L'ensemble de ces régions couvrait 85Mb, soit 3,2%, du génome de la Souris. La taille des CNV détectés variait de 1,87 kb à 3,84 Mb, avec une moyenne de 64 kb. La majorité des CNV ne dépassait pas 10 kb, mais ces petits

CNV ne contribuaient qu'à faire varier 0,13% du génome. Cahan *et al.* ont observé que la fréquence des CNV diminuait quand leur taille augmentait. Le nombre de bases couvert par des CNV était assez homogène entre les lignées (entre 26 Mb et 48 Mb, soit en moyenne 39 Mb). Soixante-sept pour cent des régions présentant des CNV étaient retrouvées chez plusieurs lignées. Cahan *et al.* ont comparé la localisation des CNV qu'ils avaient identifiés à la répartition des régions répétées dans le génome de la Souris (Cheung *et al.*, 2003 ; Bailey *et al.*, 2004 ; She *et al.*, 2008). Comme She *et al.*, ils ont mis en évidence que les éléments répétés du génome, tels que les duplications segmentaires, les **LINE** et les **LTR**, étaient enrichis en CNV de plus de 10 kb, par rapport au reste du génome. Néanmoins, Cahan *et al.* ont observé que les petits CNV (<10 kb) n'étaient pas plus fréquents dans les duplications segmentaires et les LINE. Ils ont également trouvé que les **SINE** étaient significativement moins riches en CNV de plus de 10 kb que le reste du génome.

Cahan *et al.* ont trouvé que 679 gènes étaient touchés par les 1333 régions présentant des CNV. Les CNV affectaient moins souvent des gènes que ne le voudrait le hasard. Ces résultats étaient en accord avec ce qu'avait trouvé She *et al.* en 2008, mais étaient inversés par rapport à ce qui semblait se produire chez l'Homme (Redon *et al.*, 2006) ou chez le Rat (Guryev *et al.*, 2008), espèces chez lesquelles les CNV semblaient être enrichis en gènes.

Suite à ce recensement des CNV, Cahan *et al.* ont étudié l'impact des CNV sur l'expression des gènes, dans des cellules souches hématopoïétiques et dans des cellules adipeuses et hypothalamiques de Souris. Cette étude a montré que 28% de l'expression génique spécifique de lignée, dans les cellules souches hématopoïétiques, était due aux CNV. Ces résultats étaient similaires à ceux obtenus par Stranger *et al.*, qui montraient que les CNV contribuaient à hauteur de 18% aux variations d'expression génique chez l'Homme (Stranger *et al.*, 2007).

Par exemple, Cahan *et al.* ont montré, chez plusieurs lignées, qu'une surexpression de la protéine Glo1 (*Glyoxalase 1*) dans les trois types cellulaires testés était due à une duplication qui touchait le gène *Glo1* en entier. La surexpression de cette protéine avait été associée au développement d'un comportement anxieux chez la Souris, dans une étude antérieure (Cahan *et al.*, 2009 ; Williams *et al.*, 2009). Cet exemple montrait l'impact possible d'une duplication sur l'expression d'un gène et son effet sur le phénotype.

D'après Cahan *et al.*, les CNV pourraient influencer l'expression des gènes par différents mécanismes, dont l'altération du dosage génique (nombre de copies fonctionnelles d'un gène), la suppression ou relocation de séquences régulatrices, ou encore par une modification de la structure chromatinienne locale. Cahan *et al.* ont suggéré que les CNV qui altéraient la structure chromatinienne d'une région pouvaient affecter le profil d'expression d'un plus grand nombre de gènes, y compris les gènes avoisinant le CNV. La plupart des spéculations récentes concernant l'impact des CNV sur la variabilité phénotypique était centrée sur l'effet de dosage génique. Or il est probable que ce mécanisme soit minoritaire, et que les phénomènes d'altération des séquences régulatrices ou d'action sur la structure chromatinienne locale soient plus importants que ce qui est suspecté aujourd'hui. L'étude de Cahan *et al.* montrait que seuls 7,7% des CNV agissaient directement sur le dosage génique.

Deux autres études ont tenté d'évaluer les effets des CNV sur le transcriptome de tissus de Rats (Guryev *et al.*, 2008) et de Souris (Henrichsen *et al.*, 2009 B). Ces deux études rapportaient que seulement 5 à 18% des gènes avaient une expression fortement corrélée au nombre de copies du gène. Dans 2/3 des cas, le nombre de copies n'avait aucun effet sur le niveau d'expression des gènes dans les tissus, ce qui suggérait que des mécanismes de régulation compensateurs intervenaient, ou bien que le variant ne touchait pas ou pas complètement les éléments de régulation du gène. De plus, des phénomènes épigénétiques pourraient moduler les effets des CNV, voire masquer leur présence (Henrichsen *et al.*, 2009 A ; Wain *et al.*, 2009).

Ces deux études montraient aussi que les CNV n'affectaient pas seulement l'expression des gènes qu'ils touchaient, mais aussi celle des gènes voisins, jusqu'à 500 kb de distance (Guryev *et al.*, 2008 ; Henrichsen *et al.*, 2009 B). Henrichsen *et al.* ont proposé que certains CNV pourraient provoquer une séparation physique des gènes voisins et de leurs éléments de régulation, ou une altération de la transcription par modification de la structure chromatinienne locale. Cette idée était aussi partagée par Cahan *et al.*. De tels mécanismes pourraient expliquer comment certaines voies métaboliques sont altérées sans être directement touchées par un variant (Cahan *et al.*, 2009 ; Henrichsen *et al.*, 2009 A). Henrichsen *et al.* ont observé que les gènes touchés par des CNV présentaient globalement une plus faible expression que ceux localisés en dehors des variants. Les modifications de dosage des gènes, exprimés dans le cerveau, étaient moins fréquentes et avaient tendance à être compensées par des mécanismes de régulation de la transcription plus étroits que pour les gènes exprimés dans d'autres tissus (Henrichsen *et al.*, 2009 B).

## 4. Conclusion

L'ensemble des publications qui ont étudiées les CNV chez la Souris soutenaient que ces variants contribuaient à la variabilité phénotypique entre les lignées (Cutler *et al.*, 2007 ; Egan *et al.*, 2007 ; Graubert *et al.*, 2007 ; She *et al.*, 2008 ; Cahan *et al.*, 2009 ; Henrichsen *et al.*, 2009 B). Les résultats de ces différentes études étaient difficilement comparables, car ces travaux de recherche n'étudiaient pas les mêmes lignées de Souris, et n'utilisaient pas les mêmes puces à ADN ni les mêmes algorithmes pour identifier les CNV. De ce fait, le nombre de CNV détectés variait énormément entre les études (Tableau 1).

Une cartographie exhaustive des CNV au sein des lignées de Souris fait encore défaut. Or les CNV qui ségrègent au sein des lignées sont importants à prendre en compte pour la modélisation des maladies et pour l'identification de caractères complexes (Watkins-Chow et Pavan, 2008). Des études complémentaires sur les mécanismes d'action moléculaires des CNV seraient également nécessaires pour comprendre comment ils perturbent la structure génomique et comment ils influencent l'expression des gènes (Henrichsen *et al.*, 2009 B). Une connaissance plus pointue des CNV chez la Souris permettra sans doute aussi de mieux comprendre l'effet des CNV chez l'Homme.

**Tableau 1. Tableau récapitulatif des résultats des différentes études de recensement des CNV publiées chez la Souris (jusqu'en juillet 2009).**

Etudes	Méthode	Nb de lignées étudiées	Nb de CNV recensés	Taille min des CNV	Taille max des CNV	Taille moyenne des CNV	Nb moyen de CNV par lignée (nb de bases couvert par les CNV)	Nb de gènes touchés
<b>Li <i>et al.</i>, 2004</b>	CGH	15	238	-	-	-	-	-
<b>Snijders <i>et al.</i>, 2005</b>	CGH	9	74	-	-	-	-	-
<b>Cutler <i>et al.</i>, 2007</b>	CGH	42	2094	-	-	-	51 (10Mb)	6027
<b>Egan <i>et al.</i>, 2007*</b>	CGH	14	38	4kb	4Mb	-	-	60
<b>Graubert <i>et al.</i>, 2007</b>	CGH	21	80	21kb	2Mb	271kb	22	-
<b>Watkins-Chow et Pavan, 2008**</b>	SNP	1	2	-	-	-	-	4
<b>She <i>et al.</i>, 2008***</b>	CGH	16	2424	-	-	-	-	353
<b>Cahan <i>et al.</i>, 2009</b>	CGH	20	1333 <sup>oo</sup>	1,9kb	3,8Mb	64kb	331 (39Mb)	679
<b>Henrichsen <i>et al.</i>, 2009</b>	CGH	13+21 <sup>o</sup>	3786 <sup>oo</sup>	43kb	3,35Mb	-	-	6246

\*L'étude d'Egan *et al.* recensait les CNV de 14 sous-lignées dérivées de la lignée C57BL/6, puis recherchait ces CNV chez 12 autres lignées non apparentées à C57BL/6.

\*\*Watkins-Chow et Pavan ont recherché, sur une partie du génome réduite, les CNV au sein de la lignée C57BL/6J.

\*\*\*She *et al.* n'ont recensé que les CNV présents dans ou en périphérie des duplications segmentaires, et non sur l'ensemble du génome.

•L'étude de Snijders *et al.* était la seule à utiliser une technique d'hybridation comparative (CGH) en référence au génome de la lignée FVB/N. Les autres études utilisaient le génome de la lignée C57BL/6J comme référence.

°Henrichsen *et al.* ont recensé les CNV de 3 à 4 individus de 12 lignées consanguines différentes, de 5 individus C57BL/6J et de 21 Souris sauvages.

°° Cutler *et al.* et Henrichsen *et al.* ont identifié respectivement 1333 et 3786 régions distinctes du génomes comportant des CNV. Ils ont considérés les CNV qui se chevauchaient ou qui étaient redondants entre les lignées comme une unité.

Nb : nombre

Min : minimale

Max : maximale