

RÉGULATION DE L'EXPRESSION GÉNIQUE

1. ENJEUX SCIENTIFIQUES

1.1. LES AVANCÉES ET LES ENJEUX

Le champ disciplinaire de la régulation génique est extrêmement dynamique et compétitif. Un objectif majeur des recherches dans ce domaine est de progresser dans la compréhension des propriétés du vivant conférées par les modalités d'expression du génome. Les dernières années ont été marquées par la découverte de très nombreux ARNs non codants, les avancées en épigénétique et l'accès à des technologies nouvelles (séquençage massif, études de cellules uniques...). Ces développements nous amènent à reconsidérer sans cesse notre vision de ce qu'est l'expression du génome et des relations génotype/phénotype qui sont au centre de la réflexion en génétique. Par ailleurs ces découvertes permettent d'appréhender avec de nouveaux outils des phénomènes biologiques complexes (différenciation, développement, vieillissement, adaptation...) et de revisiter des notions aussi importantes que la pénétrance incomplète, la vigueur hybride, l'effet de l'environnement sur l'expression des gènes, etc. Les enjeux sociétaux sont également importants, les altérations et autres variations des mécanismes de régulation génique sous-tendant de nombreuses pathologies humaines, y compris le cancer, et étant sans aucun doute responsable d'une part considérable de la composante héritable de la plupart des traits complexes, qu'ils concernent la santé, l'agriculture ou les biotechnologies.

1.2. APPORTS DE LA DIVERSITÉ DES ORGANISMES DE RÉFÉRENCE ÉTUDIÉS

Le recours à plusieurs organismes de référence demeure toujours essentiel pour la découverte des molécules et des mécanismes impliqués dans les processus de régulation génique. Les modèles bactériens sont les premiers qui ont permis d'aborder le caractère stochastique de l'expression et le rôle du bruit dans la régulation génétique, de modéliser les boucles de régulation et étudier les phénomènes de bi-stabilité. Ils sont également critiques pour le design de modules régulateurs (biologie synthétique) et sont au premier plan dans l'étude du rôle des petites molécules dans l'expression génique (riboswitches...).

En ce qui concerne l'épigénétique, la richesse de la diversité des modèles étudiés est flagrante. L'épigénétique est définie comme l'étude des changements, héritaires au cours des divisions cellulaires, qui affectent la fonction des génomes sans altération de leur séquence ADN. Ainsi, en est-il du rôle des protéines des groupes Polycomb et Trithorax dans le maintien au travers des divisions cellulaires de l'état respectivement réprimé ou actif de gènes, qui doit beaucoup à la *Drosophile*. C'est également le cas en ce qui concerne l'identification chez les plantes et la souris d'épimutations se maintenant à travers les générations ou encore la mise à jour de nouveaux mécanismes épigénétiques impliquant les ARN non-codants, petits ou grands, à partir de travaux effectués sur la souris, des invertébrés (*C. elegans*), des microorganismes (levures et ciliés) et les plantes (*Arabidopsis*). Ces exemples soulignent l'importance qu'il y a à maintenir une diversité tant des modèles expérimentaux que des approches dans le champ prometteur de l'épigénétique.

1.3. LES FUTURES LIGNES DE FRONT

1.3.1. TRANSCRIPTION ET TRANSCRITS.

La machinerie de transcription des eucaryotes est particulièrement complexe. En effet, il existe trois ARN polymérase (Pol) possédant chacune son ensemble de facteurs généraux de la transcription. Le cas de la Pol II est de loin le plus compliqué car en plus des cinq facteurs généraux totalisant une trentaine de polypeptides, la machinerie fait appel à des complexes co-activateurs multiprotéiques (jusqu'à une trentaine de sous-unités !) modifiant ou remodelant la chromatine ou permettant la réponse aux activateurs de la transcription spécifique de l'ADN. La manière dont l'ensemble de la machinerie de transcription concourt à l'activation des gènes *in vivo* n'est pas comprise, pas plus que les règles qui gouvernent la dépendance de certains gènes vis-à-vis de facteurs ou co-activateurs particuliers. La machinerie de la transcription est par ailleurs en interaction constante avec l'environnement chromatinien, que ce soit lors de l'activation et la répression de la transcription mais aussi lors de la phase d'élongation des transcrits qui fait d'ailleurs appel à des facteurs de transcription spécifiques dont certains sont impliqués dans des leucémies. La compréhension des mécanismes de la transcription implique de comprendre comment la machinerie de transcription influence l'organisation chromatinienne des génomes et inversement. Ces relations ont des implications fortes dans les phénomènes épigénétiques.

Une des surprises récentes, due au développement de méthodes combinées au séquençage à très haut débit, est l'observation chez les eucaryotes et les procaryotes d'une transcription cryptique importante du génome, souvent divergente par rapport aux gènes codant les protéines. Le rôle de cette transcription n'est pas connu mais il y a de fortes chances que ce phénomène puisse réguler la transcription d'un nombre important de gènes soit en contrôlant l'environnement chromatinien des promoteurs, soit par la production d'ARN non-codant régulateur. La compréhension des implications à l'échelle des génomes de ce phénomène fait l'objet de travaux intenses. Enfin, il a été observé très récemment qu'entre un quart et la moitié des gènes eucaryotes portent dans la première centaine de nucléotides transcrits une ARN polymérase arrêtée. La question est de savoir si la signification physiologique en est bien de permettre une activation rapide de la transcription ou au contraire s'il s'agit d'un mécanisme de répression. Il faut également comprendre quels sont les mécanismes moléculaires qui provoquent la pause de l'ARN polymérase et ceux qui lui permettent d'en sortir.

La machinerie de transcription est intimement liée aux machineries de maturation, modification, dégradation des ARNs et il est maintenant établi que l'ensemble du métabolisme des ARNs est très organisé spatialement et temporellement. Ce nouveau champ d'investigation est extrêmement actif.

1.3.2. EPIGÉNÉTIQUE ET DESTIN CELLULAIRE.

Grâce aux études fondamentales réalisées à partir d'espèces modèles, il est maintenant clairement établi que l'épigénétique est au cœur de la stabilité des profils d'expression des gènes au cours du développement et de la différenciation cellulaire. De plus, il semble que la plupart des phénomènes épigénétiques sont sous-tendus chez les eucaryotes par des changements d'états de la chromatine. L'empreinte parentale et l'inactivation du chromosome X fournissent les exemples les plus frappants chez les mammifères.

Des défauts de contrôle épigénétique sont très certainement impliqués chez l'homme dans le phénomène de vieillissement ainsi que dans la survenue de nombreuses maladies, dont le cancer, et dans les complications associées à l'assistance médicale à la procréation. Alors que

les mutations de la séquence de l'ADN sont irréversibles à l'échelle de l'individu, les épimutations peuvent être effacées de façon relativement aisée à l'aide d'inhibiteurs de modifications chromatinienne spécifiques (epidrugs), ce qui ouvre de nouvelles perspectives thérapeutiques, déjà explorées pour le cancer.

En agriculture, les modifications épigénétiques induites par des perturbations de l'environnement sur les plantes et les animaux commencent à être identifiées comme source importante de variation phénotypique héritable, avec des conséquences pour la sélection génétique, la prévention de pathologies émergentes, la domestication et l'adaptation aux changements climatiques.

Le renouvellement de notre représentation du vivant auquel contribue l'épigénétique devient également un enjeu pour l'appréhension des avancées de la recherche par la société. La question essentielle de la part de l'inné et de l'acquis dans le développement l'adaptation des organismes à leur environnement, trouve avec l'épigénétique des éléments de réponse au niveau moléculaire. Dans le même temps, la complexité des phénomènes se révèle bien plus grande qu'on ne le soupçonnait.

2. ETATS DES LIEUX

L'étude de la régulation de l'expression génétique chez les bactéries est un domaine historiquement très représenté dans les laboratoires français de Microbiologie. Depuis les années 80, les laboratoires français ont contribué significativement à notre compréhension des mécanismes fondamentaux de l'expression (transcription et traduction) et de leurs régulations (systèmes à deux composants, anti-terminaison...). Les compétences couvrent un vaste champ depuis les études génétiques jusqu'à l'analyse structurale des complexes. La France a notamment été pionnière dans la découverte de la transcription cryptique chez la levure. De plus, la recherche sur le métabolisme des ARNs, extrêmement active et bien implantée sur l'ensemble du territoire a récemment permis d'irriguer le nouveau champ d'investigation de la régulation par les petits ARN non codants. De ce fait, la France est clairement bien positionnée dans les études structurales et fonctionnelles du métabolisme des ARNs des bactéries à l'homme en passant par la levure. Les équipes de cette large communauté ont de fortes interactions (ANR, congrès bisannuel SifARN...) mais la communauté gagnerait à être plus structurée.

La recherche française en épigénétique est particulièrement riche et solide. De fait, la France peut s'enorgueillir de compter parmi les meilleures équipes mondiales dans l'étude de la chromatine et de l'organisation nucléaire. Elle est également en pointe dans l'étude des processus exemplaires que sont l'empreinte parentale et l'inactivation du chromosome X chez les mammifères, ou encore le maintien des patrons d'expression des gènes par les protéines du groupe Polycomb et Trithorax chez la drosophile, la répression stable des éléments transposables chez la souris, la drosophile et les plantes, etc. Plusieurs équipes de très haut niveau travaillent dans le domaine de l'interférence ARN et des ARN non-codants, en relation avec l'épigénétique et la régulation de l'expression génique en général. En santé humaine, les liens entre épigénétique et environnement sont abordés par plusieurs équipes de premier plan, certaines s'intéressant à l'alimentation et d'autres à la reproduction *in vitro*. De même, l'étude des processus de reprogrammation intéresse des équipes leaders travaillant directement chez l'homme ou d'autres mammifères. Cependant, les recherches portant sur la composante épigénétique du cancer sont encore peu développées en France, comme d'ailleurs en Europe. La France est enfin pionnière en agronomie pour l'étude de la composante épigénétique de l'héritabilité des caractères. Malgré l'importance des forces en présence, la recherche française dans le domaine de l'épigénétique n'est pas formellement structurée. Néanmoins, les

interactions et collaborations sont nombreuses, notamment via des projets ANR. De plus, une dizaine d'équipes françaises ont participé au réseau européen d'excellence « Epigenome » (2004-2010), et de nouveaux projets européens en épigénétique impliquant la France devraient voir le jour dans le cadre du FP7.

3. ANALYSE STRATÉGIQUE

3.1 FORCES

En ce qui concerne les mécanismes de la transcription, la principale force est le rassemblement des moyens humains dans des grands centres même si quelques équipes isolées contribuent également de façon déterminante au domaine à l'échelle nationale. Pour le métabolisme des ARNs, la principale force est la taille et la qualité de la communauté ainsi que sa large implantation géographique.

Avec le Royaume Uni, la France occupe la première place en Europe pour les recherches en épigénétique. Cela se traduit notamment par une forte proportion de participants français dans les projets européens et par une implication importante dans l'organisation de conférences internationales tant en Europe qu'aux Etats-Unis. De plus, de nombreuses équipes/unités ont été créées ces dernières années dans le domaine de l'épigénétique. Enfin, la France s'est positionnée très tôt dans la mise en œuvre des approches épigénomiques pour l'étude de systèmes modèles (Drosophile, plantes, levure).

3.2. FAIBLESSES

Les avancées récentes dans l'étude de l'expression génique ont des répercussions profondes sur l'appréhension de la relation génotype – phénotype. La réversibilité et l'instabilité des épimutations nous obligent notamment à reconsidérer la base de la pénétrance incomplète de certains phénotypes et sous-tendent vraisemblablement une part importante de l'hérédité dite « non-mendélienne ». Plus généralement, il nous faut dorénavant considérer des sources additionnelles de variation intra- et inter-individus et revoir nos méthodes d'analyse de la composante génétique des traits quantitatifs. Ceci a provoqué l'émergence d'une discipline qui pour appréhender cette complexité se veut non réductionniste : la biologie des systèmes.

La biologie des systèmes peine à émerger en France malgré plusieurs actions incitatives ciblées. La faible taille de la communauté et les contours flous de la discipline en rendent l'évaluation difficile.

Par ailleurs, les approches épigénomiques sont à consolider chez les organismes modèles où elles ont été mises en œuvre initialement et restent en grande partie à développer chez les mammifères. C'est notamment vrai pour l'étude de la relation épigénétique et cancer, encore peu avancée en France.

3.3. GOULOTS D'ÉTRANGLEMENT

Le principal goulot d'étranglement actuellement est la mise en œuvre des approches de séquençage à très haut débit et peut-être de manière encore plus critique le développement des analyses bioinformatiques et biostatistiques associées.

4. PRIORITÉS SCIENTIFIQUES

Les bouleversements récents de notre appréhension de l'expression génique engendrent des questions très variées, qui ont pour objet final de comprendre comment les modalités d'expression du génome contribuent à conférer aux organismes vivants les propriétés qui leur sont spécifiques.

Quelles sont les propriétés fondamentales de la machinerie de transcription ?

Comment cela impacte-t-il les propriétés des systèmes biologiques ?

Quel est le rôle du bruit dans l'expression génétique ?

Quel est le degré d'hétérogénéité de l'expression génique au sein des populations microbiennes ? Quelles en sont les causes et conséquences ?

Quelle est la contribution – amplitude et types de processus – des petits RNA non-codants à l'expression du génome ?

Comment intégrer les différents niveaux de régulation – de la transcription jusqu'à l'activité protéique – dans le fonctionnement cellulaire, sa compréhension et sa maîtrise ?

Comment s'organisent au sein du noyau les machineries d'expression du génome ?

Quelle est l'implication de l'épigénétique dans le développement, la reproduction, l'apparition de maladies et le cancer ?

Quels sont les mécanismes épigénétiques sous-jacents et comment identifier les marques épigénomiques associées ?

Comment ces marques sont-elles apposées sur la chromatine ? Comment sont-elles effacées ?

Quel est l'impact du vieillissement, de l'alimentation, de l'environnement et des stress sur les mécanismes épigénétiques ?

Jusqu'à quel point les modifications épigénétiques contribuent-elles à la variabilité des caractères complexes au sein d'une population ?

5. PROPOSITIONS D'ACTIONS

Encourager l'émergence de la biologie systémique. Nous préconisons un soutien fort à la prise de risque au sein des programmes blancs (notamment jeunes équipes) accompagné d'une structuration rapide des équipes du domaine (GDR ou autre).

Faciliter l'accès aux approches de séquençage à très haut débit et favoriser le développement des moyens en bioinformatique et biostatistique (les propositions dans ce sens sont développées dans un autre document).