



L'ESSENTIEL

> L'ADN forme quelque 10 000 boucles dans chacune de nos cellules. Nous partageons nombre de ces boucles avec d'autres organismes, dont les souris.

> Le mécanisme qui produit ces boucles s'apparente

à celui qui permet d'ajuster la longueur des sangles d'un sac à dos.

> Ces boucles interviennent dans la régulation de l'activité des gènes et sont impliquées dans certaines maladies.

L'AUTEUR



EREZ LIEBERMAN AIDEN
directeur du Centre pour l'architecture
du génome à la faculté de médecine Baylor
et à l'université Rice, à Houston, aux États-Unis

Le génome, un fouillis bien ordonné

L'ADN de nos cellules est bien plus organisé dans l'espace qu'on ne le pensait. Il présente des domaines à différentes échelles et d'étonnantes boucles mouvantes...

Multipliez la taille du génome humain par un million. À cette échelle, chaque molécule d'ADN – chaque chromosome – est aussi large qu'une nouille *rāmen*. Mis bout à bout, les 46 chromosomes du génome d'une cellule humaine s'étendraient de Paris à Helsinki, même si, repliés, ils tiennent dans une structure de la taille d'une maison – le noyau cellulaire. Ensemble, les 46 chromosomes contiennent environ 20 000 gènes en deux exemplaires. Chaque gène – aussi long qu'une voiture dans notre échelle imaginaire – porte un message codé indiquant à la cellule comment fabriquer une protéine particulière.

Si vous observez le noyau, vous verrez l'ADN s'y tortiller. Lorsque j'étais doctorant, il y a environ dix ans, je me demandais, en remuant les nouilles *rāmen* de mon dîner,

comment le génome évitait de s'emmêler en un fouillis qui empêcherait l'envoi de ses messages génétiques.

En 2014, mes collègues et moi avons apporté un élément de réponse à cette question, qui a renforcé la prise de conscience actuelle que la structure du génome dans le noyau est loin d'être aléatoire. Sous la houlette de mes étudiants Suhas Rao, Miriam Huntley et Adrian Sanborn, notre équipe a découvert que le génome humain se replie de telle façon qu'il forme plusieurs milliers de boucles.

Ces boucles semblent aider à contrôler l'activité des gènes. Toutes les cellules portent les mêmes gènes, et si l'activité de ces gènes était la même dans toutes les cellules, le corps ne pourrait pas exister : rien ne distinguerait une cellule du muscle cardiaque d'une cellule du cerveau. La façon dont ces différents profils sont orchestrés a longtemps été une énigme, mais les boucles apportent une explication. >

- > Elles seraient l'un des contrôleurs des profils d'expression des gènes, un chef d'orchestre qui influencerait sur le moment où des gènes particuliers deviennent suffisamment actifs pour modifier le fonctionnement des cellules.

UN FACEBOOK POUR LE GÉNOME

Mes réflexions sur l'enchevêtrement de l'ADN étaient liées à une question plus large : comment l'agencement tridimensionnel de l'ADN dans le noyau cellulaire influe-t-il sur l'activité des gènes ? Depuis la fin des années 1970, s'accumulaient les preuves indiquant que de petits segments d'ADN, nommés « séquences stimulatrices », ou *enhancers*, étaient nécessaires pour activer les gènes. Ces séquences se trouvaient parfois très loin de leurs gènes cibles sur la molécule d'ADN. Pour actionner l'interrupteur qui active un gène – un segment d'ADN adjacent au gène et connu sous le nom de « promoteur » –, la molécule d'ADN devait donc rapprocher du promoteur la séquence stimulatrice, vraisemblablement en se repliant sur elle-même. Mais cette conjecture était-elle correcte ? Ce problème m'a captivé et je n'ai trouvé qu'une seule façon de le résoudre : trouver toutes les boucles.

Le plan pour y parvenir était simple en théorie. Si deux personnes se fréquentent particulièrement souvent, il est logique de supposer qu'elles sont amies. De même, si deux segments d'ADN – ou « loci » – très éloignés l'un de l'autre le long du chromosome ont tendance à se fréquenter particulièrement souvent, alors l'ADN s'est probablement replié en boucle. Pour tester cette hypothèse, il nous fallait donc un moyen de mesurer la fréquence d'interaction des différentes parties du génome : une sorte de Facebook du génome humain.

Pour transformer notre idée en réalité, nous avons adapté une méthode décrite en 1993 par Katherine Cullen, alors à l'université Vanderbilt, aux États-Unis, et ses collègues. À l'époque, la structure du génome avait résisté à toutes les formes connues d'imagerie : comme un mauvais sujet de portrait, les chromosomes remuaient sans arrêt. Mais Katherine Cullen a tourné cette agitation en un avantage. Lorsque les chromosomes s'agitent, différentes parties du génome se heurtent les unes aux autres. Et les parties les plus proches dans l'espace se heurtent probablement souvent, tandis que celles éloignées se rencontrent rarement. Ainsi, en mesurant la fréquence des collisions, il devait être possible de déterminer quelles parties du génome étaient proches dans l'espace.

Pour cela, Katherine Cullen et ses collègues ont développé une technique qu'ils ont appelée le « test de ligature nucléaire ». En

substance, prenez des cellules et, sans détruire leurs noyaux, stabilisez leurs génomes. Ensuite, envoyez une enzyme couper l'ADN en petits morceaux et déployez une protéine qui fusionne les extrémités de deux fragments voisins, les rassemblant en un seul brin. Enfin, examinez la séquence des paires de bases de l'ADN (les lettres appariées du code de l'ADN) dans l'ensemble des fragments fusionnés obtenus. Si, d'une cellule à l'autre, vous observez des fusions de deux fragments d'ADN qui d'habitude sont éloignés le long d'un chromosome, vous pouvez conclure que ces deux fragments se retrouvent souvent proches l'un de l'autre dans le noyau.

Cette approche a permis à Katherine Cullen de montrer que deux morceaux d'ADN s'entrechoquaient bien plus souvent que le hasard ne l'aurait prédit. En d'autres termes, l'ADN formait une boucle qui les rapprochait.

En 1993, les expériences utilisant le test de ligature nucléaire étaient difficiles à réaliser. Heureusement, à l'époque où j'ai vu l'article de Katherine Cullen, au milieu des années 2000, il existait une référence du génome humain et le séquençage de l'ADN devenait très bon marché. Avec trois collègues de l'institut Broad, aux États-Unis, – Chad Nusbaum, Andreas Gnirke et Eric Lander – nous avons esquissé une approche pour analyser la fréquence de contact non pas d'une seule paire de fragments dans l'ADN, mais de toutes celles du génome



Les boucles s'étaient conservées pendant au moins 60 millions d'années d'évolution



en même temps. Cette approche nous permettait aussi de déterminer exactement d'où venait chaque fragment fusionné.

Nous avons décidé de fonder notre nouvelle méthode sur une variante de la procédure de Katherine Cullen développée par Job Dekker, de la faculté de médecine de l'université du Massachusetts. Au lieu de conserver les noyaux cellulaires intacts, Job Dekker les a pulvérisés et a effectué les étapes du test dans une solution très diluée. Cette variante, qu'il a nommée « capture de la conformation chromosomique »,

ou «3C», semblait fournir une estimation plus fiable de la fréquence de collisions.

Ensuite, nous avons ajouté quelques étapes à la méthode 3C. Avant de coller les fragments ensemble, nous avons attaché des étiquettes aux extrémités de l'ADN cassé afin de marquer les endroits où deux morceaux fusionneraient. Et une fois les fragments collés, nous les avons coupés en plus petits morceaux, parmi lesquels nous n'avons conservé que ceux portant les étiquettes – et donc les jonctions. En appliquant cette procédure avec Job Dekker, Nynke van Berkum, alors chercheuse en postdoctorat dans son laboratoire, et Louise Williams, de l'institut Broad, nous avons découvert que nous pouvions identifier des millions de contacts d'un seul coup. Nous avons publié cette méthode, nommée Hi-C, en 2009.

DES CHROMOSOMES DIVISÉS EN DOMAINES

Nos toutes premières cartes de génomes entiers obtenues par cette méthode ont montré que les chromosomes, malgré leur agitation, ne se repliaient pas en un amas aléatoire à l'intérieur du noyau. Au contraire, chaque chromosome était divisé en domaines: des portions d'ADN contenant des segments en contact fréquent les uns avec les autres. Les loci d'un domaine interagissaient moins fréquemment avec ceux d'autres domaines. De plus, les domaines se répartissaient en deux grandes régions dans le noyau, que nous avons nommées «compartiments A et B».

Le compartiment A était riche en marqueurs de l'activité génétique, tels que des ARN messagers – des émissaires des gènes qui donnent des instructions au reste de la cellule. Le compartiment B était plus dense et en grande partie inactif. Lorsque des domaines étaient activés ou désactivés, ils passaient d'un compartiment à l'autre. (Aujourd'hui, on sait que les noyaux des cellules contiennent plusieurs sous-compartiments A et B.)

La découverte de cette compartimentation dynamique confirmait que la structure à grande échelle du génome n'est pas aléatoire, mais plutôt intimement liée à l'activité des gènes. Toutefois, à ma grande déception, une caractéristique de repliement n'apparaissait pas dans les données de Hi-C: les boucles...

Ces données sont souvent représentées sous la forme d'une «carte thermique»: une représentation graphique montrant à quelle fréquence deux loci d'un chromosome entrent en contact l'un avec l'autre. De telles représentations ressemblent à un jeu de bataille navale où les chiffres et les lettres seraient remplacés par les loci d'un chromosome. On obtient alors un quadrillage où chaque point représente un contact possible entre deux loci et où la fréquence de contact est indiquée par la

luminosité du point. Une boucle devait donc se manifester par un point exceptionnellement lumineux correspondant aux deux loci qui entrent en contact pour la former. Or nous n'avions pas observé de tels pics de luminosité. Et sans la preuve de l'existence des boucles, impossible d'examiner si les séquences stimulatrices activaient les gènes en se rapprochant de leurs promoteurs.

CARTOGRAPHIER LES BOUCLES

Ce problème nous a paralysés pendant trois ans. Puis, en 2012, Suhas Rao et Miriam Huntley ont compris d'où il venait: un aspect de la méthode Hi-C – la destruction des noyaux des cellules avant la fusion des fragments d'ADN – perturbait les structures fines telles que les boucles. Ils ont donc mis au point une variante qui maintient les noyaux intacts.

La nouvelle approche, nommée «*in situ* Hi-C», a fait la différence. En l'utilisant sur des globules blancs, Suhas Rao et Miriam Huntley ont découvert que des pics lumineux apparaissaient désormais sur toutes nos cartes thermiques, chacun représentant une boucle potentielle. Ces points lumineux n'étaient-ils pas le fruit de notre imagination? Pour m'en assurer, j'ai rapporté les cartes à mon fils, Gabriel, qui avait alors 3 ans: «Vois-tu un point rouge?» «Oui», a-t-il répondu. «Peux-tu le montrer?» Il le pouvait.

Nous l'avions: une carte montrant quelque 10000 boucles, réparties le long du génome humain. Nous avons vérifié si les boucles reliaient les promoteurs des gènes et les séquences stimulatrices. C'était souvent le cas.

Dans un autre test, nous avons comparé nos cartes thermiques des cellules sanguines avec de nouvelles cartes obtenues cette fois pour des cellules du poumon. Un grand nombre de boucles étaient semblables, mais on y notait aussi de nouvelles connexions – potentiellement entre des séquences stimulatrices et des gènes différents. Ces modifications suggéraient que les boucles étaient impliquées dans la régulation des gènes qui donnent à une cellule son identité.

En observait-on dans d'autres organismes? Pour le savoir, nous avons cartographié les boucles dans des cellules de souris: la moitié de ces structures étaient positionnées à des endroits du génome correspondant à des boucles du génome humain. Elles existaient donc déjà chez l'ancêtre des souris et des humains et s'étaient conservées pendant au moins 60 millions d'années d'évolution.

Nos données avaient révélé une autre information intéressante: les boucles ne sont pas statiques. Elles semblaient en effet continuellement apparaître, se défaire, se reformer. Nous avons voulu savoir comment cela fonctionnait. Nous pensons que des centaines de ➤

► protéines étaient impliquées, mais les données ont raconté une autre histoire.

Boucle après boucle, deux facteurs protéiques se sont distingués. L'un, appelé CTCF, a été découvert en 1990 par Victor Lobanenko, alors au Royal Cancer Hospital, à Londres, et ses collègues. Il contient onze structures dites «à doigt de zinc» qui lui permettent de se lier étroitement à certains endroits de l'ADN. Le second facteur, la cohésine, décrit en 1997 par Kim Nasmyth, aujourd'hui à l'université d'Oxford, est un complexe en forme d'anneau constitué de plusieurs protéines. À l'époque de nos travaux, les biologistes pensaient que deux anneaux de cohésine étaient capables de se lier et de fonctionner ensemble, chacun encerclant l'ADN et glissant librement sur lui comme un anneau sur un collier.

Retrouver ces protéines n'était pas si surprenant : de nombreuses études antérieures avaient suggéré qu'elles intervenaient dans le repliement du génome. Mais leur implication dans un mécanisme aussi omniprésent que la formation des boucles d'ADN était inattendue.

DES MOTIFS À BOUCLE

C'est alors que nous avons fait une découverte très étrange. Suhas Rao, Miriam Huntley et moi avons demandé à Ido Machol, un nouvel informaticien au laboratoire, d'étudier la distribution des histones – des protéines qui aident à compacter l'ADN dans le noyau – à proximité des facteurs CTCF. Ido Machol a remarqué que les histones étaient plus nombreuses juste à l'extérieur des boucles qu'à l'intérieur, comme si ces protéines savaient où une boucle était positionnée par rapport aux molécules CTCF. Ce résultat ressemblait à une erreur, mais au fil des semaines, Ido Machol n'a détecté aucun bug dans son programme.

Nous avons donc recherché une explication biologique. Victor Lobanenko avait montré que CTCF ne s'attache pas à des positions arbitraires sur l'ADN. Au contraire, le facteur se lie toujours à une séquence spécifique d'environ 20 bases, nommée «motif». L'ADN étant une double hélice, il comporte deux brins orientés en opposition. Chaque brin peut porter des motifs, par conséquent orientés dans un sens ou l'autre sur la longue molécule d'ADN. En général, l'orientation relative des motifs le long de l'ADN est aléatoire, comme un tirage au sort : il y a 50% de chances qu'un motif donné pointe dans un sens et 50% de chances qu'il pointe dans l'autre. Nous pensions donc qu'il en serait de même pour les motifs de liaison à CTCF.

Or, à notre grand étonnement, les deux minuscules motifs de liaison à CTCF qui intervenaient dans la formation d'une boucle – même s'ils étaient séparés par des millions de bases dans l'ADN déplié – pointaient ►

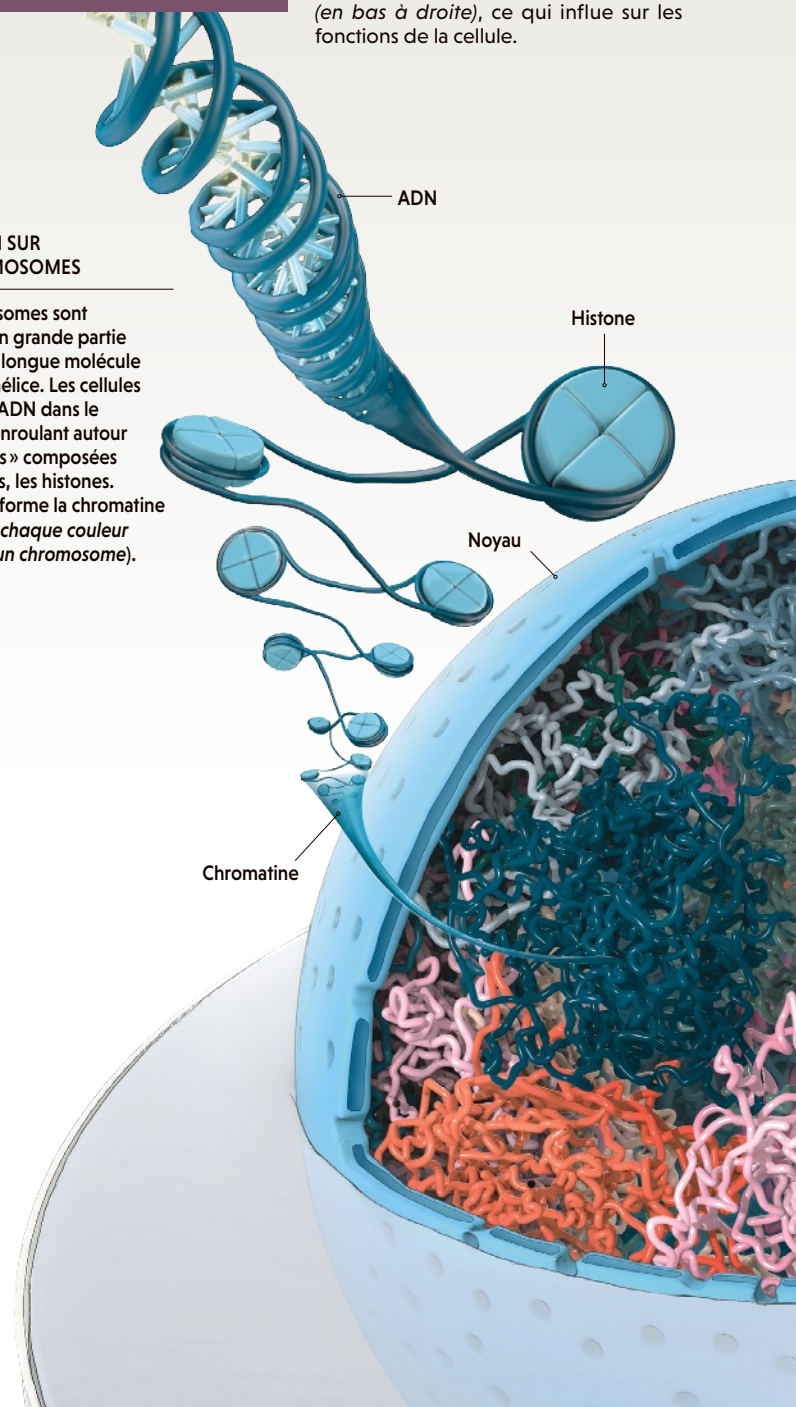
UN GÉNOME BOUCLÉ

Au microscope, le génome d'une cellule – l'ensemble de ses chromosomes – ressemble à un enchevêtrement de nouilles. Pourtant, la disposition est loin d'être aléatoire. Par exemple, avec une meilleure résolution que celle utilisée ici, on verrait que le génome se plie en environ 10 000 boucles qui ne s'entremêlent pas. Des travaux récents ont révélé un processus clé de formation des boucles, nommé «extrusion» (en haut à droite). Ces boucles favorisent l'expression ou l'activation de certains gènes dans différentes cellules (en bas à droite), ce qui influe sur les fonctions de la cellule.

GROS PLAN SUR LES CHROMOSOMES

Les chromosomes sont constitués en grande partie d'ADN, une longue molécule en double hélice. Les cellules emballent l'ADN dans le noyau en l'enroulant autour de «bobines» composées de protéines, les histones. L'ensemble forme la chromatine (ci-dessous, chaque couleur représente un chromosome).

© Falconieri Visuals

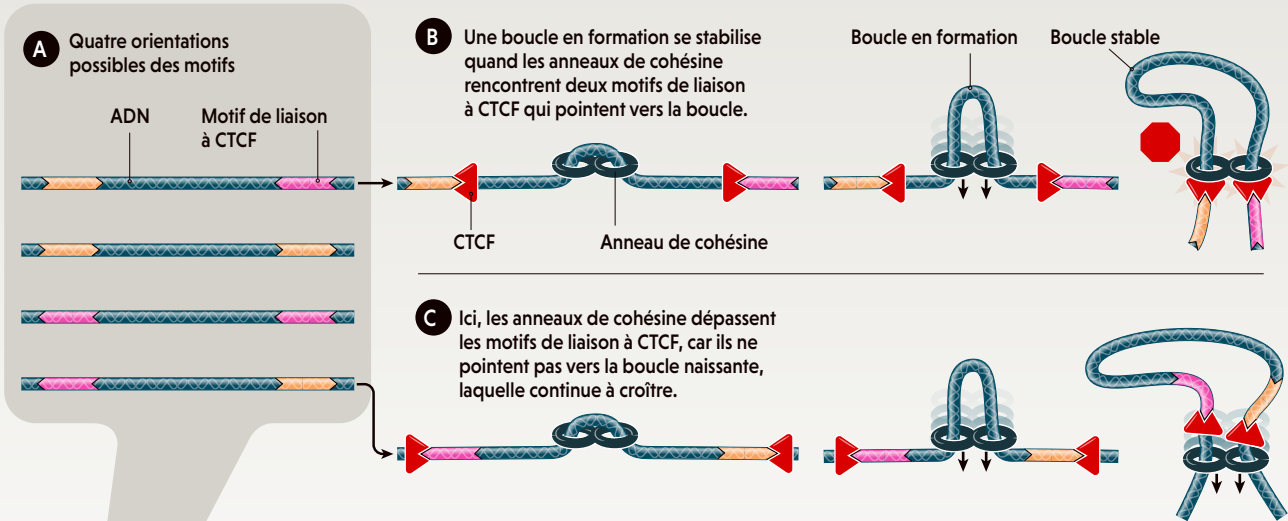


COMMENT LES BOUCLES SE FORMENT

L'extrusion d'une boucle commence quand un « complexe d'extrusion » rencontre l'ADN et s'y arrime. La boucle se forme lorsque deux sous-unités du complexe, comportant chacune une structure en anneau nommée « cohésine », se mettent à glisser dans des directions opposées

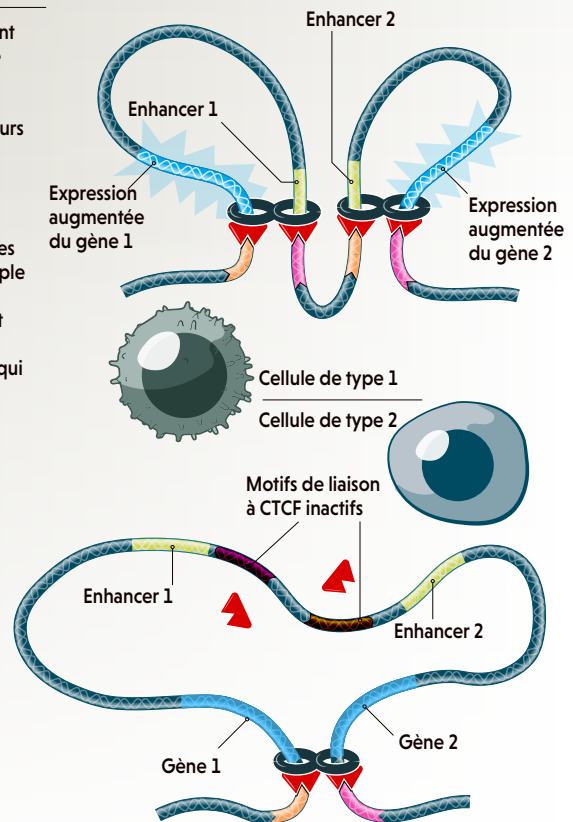
le long de l'ADN tout en restant attachées ensemble. Une protéine, CTCF, est capable d'arrêter la croissance de la boucle dans certaines conditions. Elle se lie à une séquence particulière de l'ADN, laquelle pointe dans un sens ou l'autre de la molécule (A). Lorsqu'un anneau de cohésine rencontre une protéine CTCF liée à un motif qui

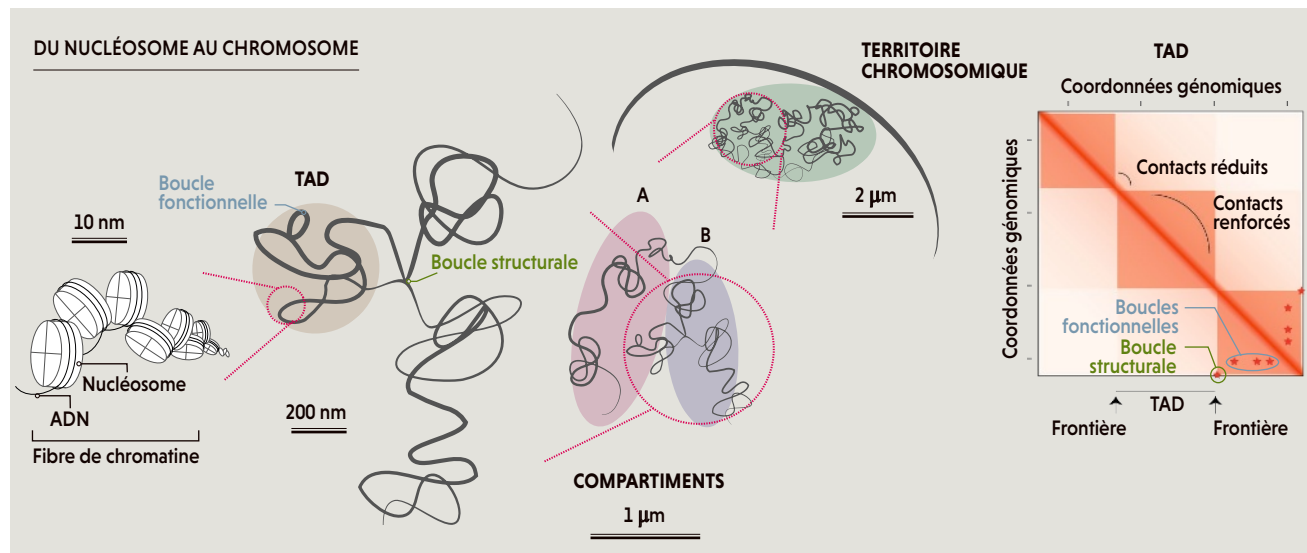
pointe vers la boucle, l'anneau s'arrête. Si le motif est orienté dans l'autre sens, l'anneau continue d'avancer. Quand les deux anneaux rencontrent des protéines CTCF liées à des motifs pointant vers l'intérieur, ils s'arrêtent donc tous les deux, ancrant la boucle (B). Dans toutes les autres configurations, la boucle continue de croître (C).



L'INTÉRÊT DES BOUCLES

Pour fonctionner correctement, certains gènes doivent entrer en contact avec un segment d'ADN nommé « enhancer ». Mais les enhancers se trouvent souvent loin des gènes qu'ils activent. Les boucles rapprochent les enhancers et leurs gènes (*en haut*). Toutes les cellules présentent les mêmes gènes et les mêmes motifs de liaison aux protéines CTCF, mais elles ne forment pas toutes les mêmes boucles. Il arrive par exemple que les motifs de liaison à la protéine CTCF soient désactivés à certains endroits de l'ADN, ce qui empêche la protéine de s'y lier (*en bas*).





UNE ORGANISATION À TOUTES LES ÉCHELLES

L'organisation spatiale du génome est loin d'être aléatoire au sein des noyaux de nos cellules. On connaissait le premier échelon de cette organisation, le nucléosome, dont la structure cristallographique avait été révélée en 1997. Le nucléosome permet d'enrouler la molécule d'ADN et représente donc un premier niveau de compaction de l'ADN pour former la fibre de chromatine (voir la figure ci-dessus à gauche). À l'autre bout de l'échelle, on savait, grâce à des approches cytologiques et de microscopie photonique, que les chromosomes s'organisent en territoires chromosomiques individuels, occupant des positions préférentielles dans les noyaux suivant leur activité génique, leur densité en gènes ou encore leur taille. En revanche, au milieu de l'échelle, les choses paraissaient beaucoup plus floues. On avait notamment observé des repliements à large échelle entre des régions génomiques éloignées, ainsi que des regroupements de gènes même lorsqu'ils étaient situés sur

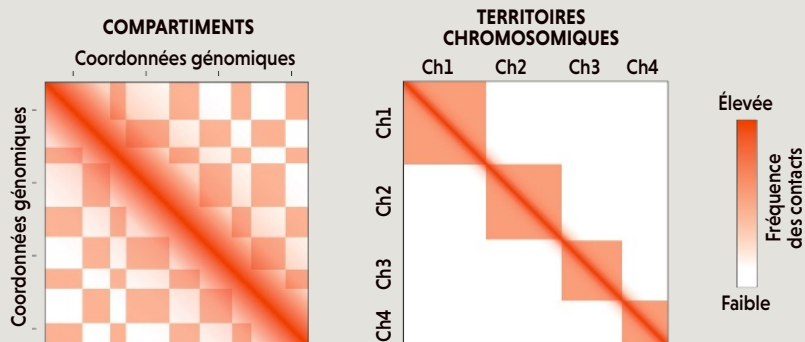
des chromosomes différents. Bien sûr, on soupçonnait aussi l'existence de repliements de la fibre de chromatine et, par là même, de rapprochements physiques entre les promoteurs des gènes et leurs séquences stimulatrices ou « enhancers » – les boucles qui font l'objet de cet article. Les enhancers sont de véritables boutons d'activation, parfois situés à plusieurs dizaines de milliers de paires de bases des gènes qu'ils régulent. Des avancées majeures dans la compréhension de l'organisation spatiale du génome, de la centaine de kilobases (100 000 paires de bases) jusqu'à la mégabase (1 million de paires de bases), ont vu récemment le jour grâce à l'application de nouvelles techniques de biologie moléculaire, telles que la méthode Hi-C.

En 2009, on a vu apparaître les premières cartes de Hi-C, qui saisissent l'ensemble des contacts entre régions génomiques dans l'espace tridimensionnel du noyau. Plus précisément, ces représentations cartographient la fréquence de chacun des contacts possibles sous la forme de points plus ou moins intenses (voir ci-dessus à droite). Les premières cartes avaient une résolution relativement faible, mais on y distinguait déjà des organisations préférentielles de larges régions génomiques et, notamment, l'organisation en compartiments actif et réprimé, appelés respectivement compartiments A et B. En 2012, l'application de la méthode Hi-C ou de dérivées avec une résolution accrue a permis à trois équipes – la nôtre, celle d'Edith Heard, à l'institut Curie, et celle de Bing Ren, à l'institut Ludwig pour la recherche sur le cancer à La Jolla,

aux États-Unis – de découvrir de nouveaux domaines d'organisation de la chromatine, qui sont désormais appelés TAD (pour *topologically associating domains*).

Le long des chromosomes, les TAD représentent des zones du génome caractérisées par des interactions internes bien plus fortes que les interactions de TAD adjacents. Entre les TAD, on trouve de véritables frontières, qui conduisent à la partition du génome en une série de TAD. L'obtention de cartes de Hi-C à très haute résolution a permis de mettre en évidence de nombreuses boucles d'ADN, décrites dans cet article. Certaines, dites « structurales », séparent les TAD les uns des autres. D'autres, situées à l'intérieur des TAD, joueraient divers rôles fonctionnels. Certaines rapprocheraient des promoteurs et des enhancers, d'autres les deux extrémités d'un gène (ce qui renforcerait la direction selon laquelle le gène est transcrit), d'autres encore se formeraient entre des sites reconnus par des protéines impliquées dans la répression de nombreux gènes – les protéines du groupe Polycomb –, renforçant leur action. En d'autres termes, les TAD permettent de mieux confiner certaines régions du génome, ce qui facilite des repliements de la fibre chromatinienne et la régulation de l'expression des gènes.

La découverte de ces différents échelons de l'organisation des chromosomes a ouvert de nouvelles voies d'investigation sur la compréhension du fonctionnement du génome, ainsi que sur celle de maladies dont l'origine restait mystérieuse. Notamment, des réarrangements chromosomiques, du type



L'organisation de la chromatine aux différentes échelles (à gauche) apparaît lorsque l'on cartographie la fréquence des contacts entre les différentes régions du génome (à droite).

inversion ou délétion, surviennent au niveau des régions frontières des TAD. Cela a pour effet d'entraîner des défauts d'isolement entre les gènes. Ainsi, des séquences stimulatrices localisées dans le TAD voisin peuvent communiquer et activer de manière erronée des gènes dans des tissus où ils sont normalement éteints. C'est par exemple le cas de gènes du développement directement impliqués dans des malformations des membres (syndrome F, polydactylie...), des défauts neurologiques ou des dysplasies. De même, des réarrangements chromosomiques peuvent entraîner l'activation de gènes nommés « proto-oncogènes », normalement éteints dans les cellules et susceptibles de conduire à la transformation des cellules et à l'apparition de tumeurs. Il est donc primordial de prendre en compte l'organisation 3D du génome pour expliquer les nombreux défauts cellulaires liés à la dérégulation des gènes.

De façon plus générale, grâce à tous ces travaux, on comprend bien mieux l'organisation 3D des génomes. Néanmoins, des questions primordiales demeurent : quels mécanismes précis définissent la taille et les bordures des TAD ? Quel est leur rôle dans la régulation des fonctions du génome ? Enfin, sera-t-il possible de prédire, dans un type cellulaire donné, la régulation de son expression génique en connaissant la séquence de l'ADN et son organisation tridimensionnelle ? Les recherches dans ce domaine révéleront sans doute de nouvelles surprises...

**FRÉDÉRIC BANTIGNIES
ET GIACOMO CAVALLI**

directeurs de recherche au CNRS,
Institut de génétique humaine
(CNRS, université de Montpellier)

➤ toujours l'un vers l'autre et vers la boucle. Cette «règle de convergence» expliquait comment les histones évitaient les boucles: il leur suffisait de déterminer dans quelle direction le motif de liaison à CTCF pointait. En revanche, le fait que l'orientation des motifs ne soit pas aléatoire défiait notre entendement.

Pour mettre les choses en perspective, grossissons à nouveau le génome un million de fois. Les motifs mesurent à présent chacun 5 millimètres de long et sont séparés par 1 kilomètre de nouille génomique. Et pourtant, comme guidés par une boussole magique, ceux situés aux extrémités d'une boucle pointent toujours l'un vers l'autre. Comme tout bon tour de magie, la règle de convergence semblait impossible. De plus, elle contredisait l'idée que l'on avait alors de la façon dont les boucles se forment.

À l'époque, presque tout le monde – y compris nous-mêmes – pensait qu'elles apparaissaient par diffusion. Dans cette hypothèse, une protéine nécessaire à la formation d'une boucle se lie à une extrémité d'un segment d'ADN. Une autre fait de même à l'autre extrémité. Comme à son habitude, l'ADN se tortille. Si finalement sa danse réunit les deux protéines, elles forment un lien physique, créant ainsi une



Presque tout le monde pensait que les boucles apparaissaient par diffusion

boucle. Le problème est que la chaîne d'ADN a tellement de place pour s'agiter que, si le modèle de diffusion est correct, l'orientation relative des motifs de liaison à CTCF n'a plus d'importance. Or, pourtant, on assistait à une convergence. Et de fait, au cours de l'année, deux équipes, l'une dirigée par Suzana Hadjur, de l'University College de Londres, et l'autre par Yijun Ruan, du laboratoire Jackson, aux États-Unis, avaient confirmé la règle de convergence avec leurs propres données. La règle était là pour rester, et les boucles ne se formaient donc pas par diffusion. Comment apparaissaient-elles alors ? Et quels étaient les rôles de CTCF et de la cohésine ? Décontenancés, nous avons commencé à jouer avec les câbles de nos écouteurs.

➤

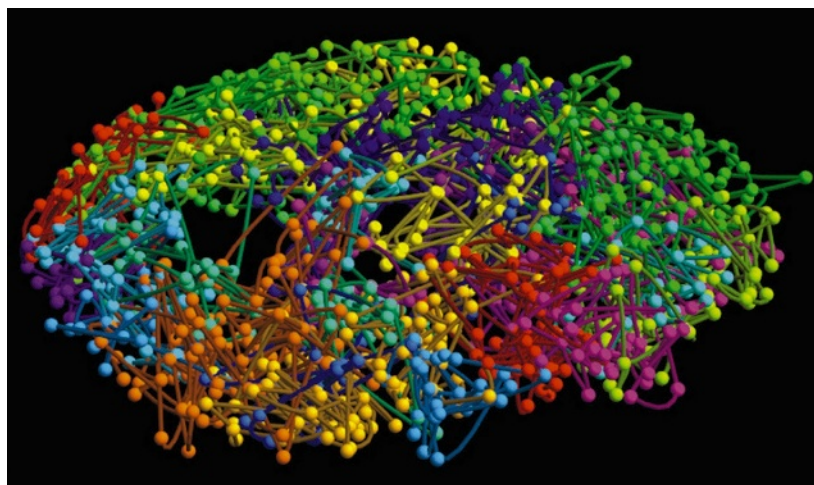
> La plupart des personnes qui travaillent sur le repliement du génome gardent à portée de main un objet long et flexible: un morceau de fil, un tube en plastique. Lorsque vous êtes bloqué sur un problème difficile, vous sortez cet objet et jouez avec. Un jour, Suhas Rao et moi faisons passer les écouteurs d'un côté à l'autre pour explorer les modèles possibles de formation de boucles, quand soudain, nous avons compris que la réponse était non pas dans nos écouteurs, mais sur nos sacs à dos.

UNE BOUCLE DE SANGLE QUI FAIT DES BOUCLES

Prenez la boucle de sangle qui en ajuste la longueur des bretelles. Elle est plus ou moins constituée de deux anneaux attachés l'un à l'autre. La sangle entre dans le premier et sort dans le second. Pour régler sa longueur, vous en tirez une partie à travers l'un des anneaux et commencez à former une boucle. Vous pouvez l'agrandir tant que vous ne butez pas sur l'extrémité de la sangle.

Et si les paires d'anneaux de cohésine fonctionnaient comme des boucles de sangle? Au début, elles s'attacheraient n'importe où sur le génome, l'ADN entrant par un anneau et sortant par l'autre. Mais ensuite, les deux anneaux glisseraient dans des directions opposées, extrudant au fur et à mesure une boucle. Mais ils ne glisseraient pas indéfiniment. L'un d'eux finirait par s'approcher d'un site où une molécule de CTCF est liée. Avec un motif de liaison à CTCF pointant vers l'anneau, celui-ci s'arrêterait au contact. Mais avec un motif en sens inverse, la cohésine l'ignorerait et continuerait de glisser. Le deuxième anneau glisserait de la même manière jusqu'à un motif de liaison à CTCF orienté vers l'intérieur, ce qui terminerait la boucle.

Si les anneaux de cohésine fonctionnaient réellement de cette manière, alors les boucles ne se formeraient qu'entre des paires de motifs de liaison à CTCF obéissant à la règle de convergence. Nous avons vite compris que ce processus d'extrusion apporterait un avantage crucial aux cellules. Si les boucles apparaissaient par diffusion, elles s'entremêlèrent facilement, conduisant les chromosomes à former des nœuds et à



Cartographier à la fois l'organisation du génome et l'expression des gènes dans l'espace 3D du noyau d'une même cellule humaine, c'est la nouvelle étape que vient de franchir l'équipe de Xiaowei Zhuang, à l'université Harvard. Grâce à une technique d'imagerie à haut débit et haute résolution, les chercheurs ont détecté en même temps la position de quelque 1000 régions du génome (à gauche, où chaque couleur se rapporte à un chromosome, et à droite en gris) et la trace de l'expression de quelque 1100 gènes localisés dans ces régions (à droite, où chaque sphère repère l'expression d'un gène, et sa taille, l'intensité de cette expression).

s'enchevêtrer. Cela compliquerait le fonctionnement des gènes et empêcherait les chromosomes de se séparer lors de la division cellulaire. En revanche, les boucles produites par extrusion ne forment pas de nœuds ou d'enchevêtrements – à l'instar des bretelles de votre sac à dos.

Le modèle était de la pure spéculation. Il reposait sur de nombreuses hypothèses pour lesquelles nous n'avions pas la moindre preuve directe, comme l'idée que la cohésine glissait le long de l'ADN. Nous avions peur d'être partis dans un délire. Mais en nous plongeant dans la littérature sur la cohésine, nous avons appris que Kim Nasmyth lui-même avait proposé en 2001 que la cohésine puisse extruder l'ADN.

Adrian Sanborn a alors effectué des simulations détaillées qui rendaient compte avec précision des données de nos cartes. Et lorsque Suhas Rao a effectué des expériences sur de l'ADN réel, la position des boucles changeait exactement de la manière prédite par le modèle d'Adrian Sanborn: la suppression d'un motif de liaison à CTCF impliquait dans la formation d'une boucle éliminait celle-ci; l'inversion de l'orientation d'un motif faisait de même et entraînait la formation d'une boucle de l'autre côté; l'ajout d'un motif de liaison à CTCF produisait aussi une nouvelle boucle – tant qu'il pointait dans la bonne direction; enfin, il était possible d'ajouter et de supprimer des boucles à volonté.

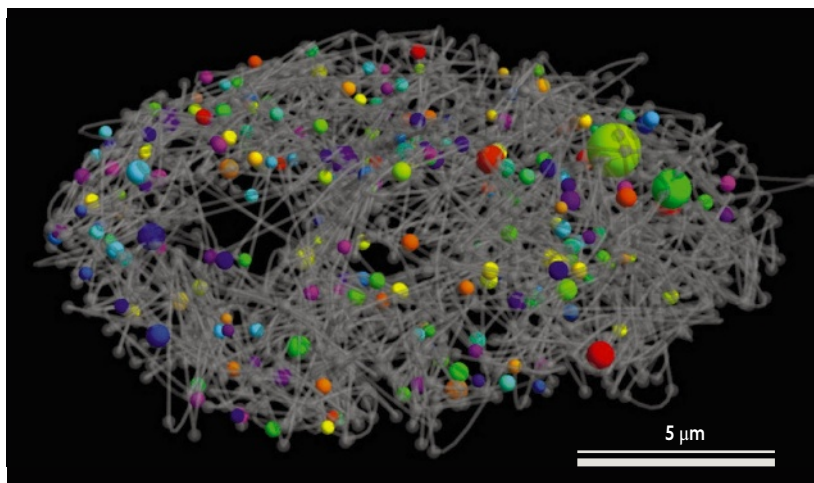
Le domaine était en ébullition. Fin 2015, à quelques semaines d'intervalle, trois équipes dont la nôtre ont publié des articles montrant que ce type de chirurgie 3D du génome fonctionne. De même, trois équipes – celle de Leonid Mirny, à l'institut de technologie du Massachusetts, la nôtre et deux chercheurs de l'université Emory, Michael Nichols et Victor Corces – ont argumenté que la règle de convergence favorise un modèle dans lequel les boucles se forment par extrusion. La communauté scientifique commençait enfin à démêler l'écheveau.

Les progrès se sont poursuivis à un rythme effréné. Aux instituts Gladstone, Benoit



La fonction principale des boucles reste peut-être à découvrir





Bruneau et ses collègues ont montré qu'en interférant avec CTCF, on affaiblissait considérablement les boucles. Au Laboratoire européen de biologie moléculaire (EMBL), François Spitz et ses collègues ont obtenu un résultat similaire en éliminant une protéine connue pour déposer la cohésine sur l'ADN. À l'Institut néerlandais du cancer, l'équipe de Benjamin Rowland a observé que la suppression d'un facteur qui retire la cohésine de l'ADN conduisait à des boucles plus grandes, sans doute parce que la cohésine avait alors la possibilité de glisser plus longtemps. Et, dans notre laboratoire, Suhas Rao a montré qu'en dégradant la cohésine elle-même, on éliminait en quelques minutes toutes les boucles associées.

Mais nous rêvions tous d'une confirmation directe : voir l'extrusion en action. C'est ce qu'ont fait, en avril 2018, Cees Dekker, de l'université de technologie de Delft, aux Pays-Bas, et ses collègues. En utilisant la condensine de la levure – un complexe protéique très proche de la cohésine –, ils ont réalisé un film que nombre de chercheurs du domaine n'oublieront jamais. La condensine atterrit sur un ruban d'ADN, formant un petit nodule d'ADN qui grandit et se développe jusqu'à ce que le spectateur comprenne ce qu'il est vraiment : une boucle extrudée.

DES BOUCLES ET DES MALADIES

Aujourd'hui, à mesure que les mécanismes de formation des boucles et leurs règles se précisent, l'importance de ces structures pour la santé et la compréhension des maladies devient de plus en plus claire. Par exemple, Frederick Alt, de l'université Harvard, et ses collègues ont commencé à expliciter leur rôle dans la production d'anticorps. Votre organisme produit des anticorps contre des agents pathogènes qu'il n'a jamais rencontrés auparavant en coupant et collant des fragments de gènes d'anticorps. L'équipe de Frederick Alt a montré que ce processus s'appuie sur l'extrusion de boucles : de

multiples boucles délimitées par CTCF se forment, puis sont éliminées.

L'équipe de Stefan Mundlos, de l'institut Max-Planck de génétique moléculaire à Berlin, a montré que la modification d'un seul motif de liaison à CTCF chez la souris entraîne des malformations des pattes. De même, les mains d'humains présentant une mutation similaire ont des malformations semblables. Et Rafael Casellas, des Instituts américains de la santé (NIH), a montré que la perturbation des motifs de liaison à CTCF dans un plasmocytome de souris – une forme de cancer – pouvait ralentir de 40% la croissance de la tumeur.

Cependant, à mesure que la notion d'extrusion de boucles s'est imposée, des théories plus fondamentales sur le rôle des boucles dans la régulation des gènes se sont effondrées. Pendant des décennies, les biologistes avaient vu ces structures comme des interrupteurs : lorsqu'une boucle rapprochait une séquence stimulatrice et un promoteur, le gène correspondant s'activait. On pensait donc que supprimer la cohésine des cellules déréglerait l'expression des gènes en modifiant le niveau d'activité de milliers d'entre eux. Comme prévu, l'expression de nombreux gènes a changé. Mais les changements sont restés assez faibles. Les boucles – du moins celles formées par extrusion – ne sont donc finalement pas des interrupteurs binaires. Elles semblent plutôt fonctionner comme des boutons qui augmentent ou diminuent un peu l'activité des gènes, ajustant avec précision l'approvisionnement d'une cellule en différentes protéines.

En d'autres termes, notre vision était trop simpliste. Peut-être même que la régulation des gènes n'est qu'un accessoire des boucles et que leur fonction principale reste à découvrir... Comme tout explorateur en territoire inconnu, nous avons besoin de cartes plus précises. Dans le cadre du projet *Encode* (Encyclopédie des éléments de l'ADN) des NIH, Yijun Ruan et moi-même travaillons avec nos collègues à la création du premier atlas des boucles du génome humain en les cartographiant dans les tissus humains. Nos équipes et nombre d'autres se sont aussi réunies au sein du consortium 4D Nucleome, qui développe de nouvelles méthodes pour s'attaquer à ces problèmes. Et Olga Dudchenko, chercheuse en postdoctorat au laboratoire, a créé le DNA Zoo – un consortium de laboratoires universitaires, de zoos et d'aquariums du monde entier qui rassemble les génomes de centaines d'espèces et retrace l'évolution des boucles dans l'arbre du vivant.

Pour les chercheurs, la fin d'une histoire scientifique est toujours le début d'une autre. Il y a deux milliards d'années, avant l'émergence du noyau cellulaire, le processus d'extrusion de l'ADN est apparu. Pourquoi ? Qui sait, répondre à cette question permettra peut-être de boucler la boucle. ■

BIBLIOGRAPHIE

J.-H. Su et al., **Genome-scale imaging of the 3D organization and transcriptional activity of chromatin**, *Cell*, vol. 182, pp. 1641-1659, 2020.

M. Ganji et al., **Real-time imaging of DNA loop extrusion by condensin**, *Science*, vol. 360, pp. 102-105, 2018.

A. L. Sanborn et al., **Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes**, *PNAS*, vol. 112(47), pp. E6456-E6465, 2015.

S. S. P. Rao et al., **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping**, *Cell*, vol. 159(7), pp. 1665-1680, 2014.

T. Misteli, **La vie agitée du génome**, *Pour la Science*, n°403, mai 2011.

K. E. Cullen et al., **Interaction between transcription regulatory regions of prolactin chromatin**, *Science*, vol. 261, pp. 2003-2006, 1993.