

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 ➤