



**HAL**  
open science

## Comment l'approche génomique aide à comprendre le processus d'initiation de la réplication

Benoit Miotto

► **To cite this version:**

Benoit Miotto. Comment l'approche génomique aide à comprendre le processus d'initiation de la réplication. *Médecine/Sciences*, 2017, 33 (2), pp.143-150. 10.1051/medsci/20173302009 . inserm-02132556

**HAL Id: inserm-02132556**

**<https://inserm.hal.science/inserm-02132556>**

Submitted on 17 May 2019

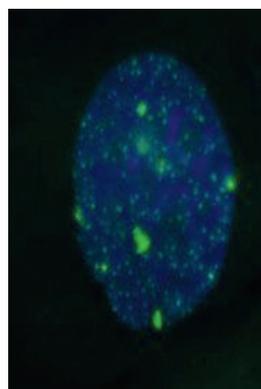
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

► La réplication de l'ADN s'initie au niveau de régions précises du génome appelées origines de réplication. Ces dernières années ont vu l'émergence de technologies sophistiquées à même d'identifier les sites d'initiation et les sites de liaison des facteurs de la réplication à l'échelle du génome. Malgré ces avancées, le mystère demeure quant aux facteurs (génétiques et chromatiniques) capables de recruter la machinerie de réplication au niveau des origines, notamment chez les mammifères. Cette revue a pour but de clarifier nos connaissances sur la nature des origines de réplication chez l'homme, les liens avec les réarrangements chromosomiques fréquemment observés dans les pathologies humaines et de mettre en lumière les défis scientifiques dans ce domaine. ◀

## Comment l'approche génomique aide à comprendre le processus d'initiation de la réplication

Benoit Miotto



Inserm U1016, Institut Cochin, Paris, France ; CNRS, UMR8104, Paris, France ; Université Paris Descartes, Sorbonne Paris Cité, 24, rue du Faubourg Saint-Jacques, 75014 Paris, France.  
[benoit.miotto@inserm.fr](mailto:benoit.miotto@inserm.fr)

La réplication de l'ADN dépend de l'assemblage, au niveau de régions particulières du génome appelées origines de réplication, de plusieurs complexes multiprotéiques essentiels au recrutement des polymérase de l'ADN. Cet assemblage se réalise dans un ordre bien précis et de manière finement régulée au cours du cycle cellulaire pour permettre l'initiation de la réplication lors de la phase S, l'activation des origines une seule fois par cycle, et de manière temporelle au cours de la phase S [1, 2]. Chez les eucaryotes, l'étape initiale consiste en l'assemblage du complexe de pré-réplication (pré-RC) au cours de la phase G1 du cycle cellulaire. Ce complexe est formé par l'assemblage successif du complexe ORC1-6 (*origin recognition complex 1-6*), des protéines CDC6 (*cell division cycle 6*) et CDT1 (*chromatin licensing and DNA replication factor 1*), et finalement de deux complexes MCM2-7 (*mini-chromosome maintenance 2-7*) [1]. À l'entrée en phase S, l'activation des kinases CDK2 (*cyclin dependent kinase 2*) et CDC7 (*cell division cycle 7-related protein kinase*) conduit au recrutement des facteurs d'initiation de la réplication et à la duplication de l'ADN [1, 2].

Le développement des technologies modernes de séquençage en masse, ainsi que l'établissement de protocoles sophistiqués, a permis de cartographier la position des sites d'initiation de la réplication et des sites de liaison des protéines de réplication dans des espèces aussi variées que les levures, la drosophile, la souris ou l'homme [3]. De nombreuses similitudes ont été retrouvées lors de l'analyse des données d'une espèce à l'autre. Ainsi, l'initiation de la réplication se produit dans des régions (ou zones) particulières et de manière bidirectionnelle. Les origines de réplication sont souvent associées à des éléments régulateurs de la transcription, comme les promoteurs et les séquences *cis*-régulatrices, suggérant que les mécanismes impliqués dans la régulation de la transcription pourraient également s'appliquer à la régulation de l'initiation de la réplication [4-10].

Une analyse plus approfondie des données démontre, toutefois, que les mécanismes de recrutement du complexe ORC et d'initiation de la réplication diffèrent d'une espèce à l'autre [3]. Chez la levure *Saccharomyces cerevisiae*, un motif ADN de 11 paires de bases permet de recruter ORC et de spécifier une origine de réplication. À l'inverse, chez l'homme ou la souris, aucune séquence spécifique n'est détectée au niveau des régions auxquelles se lie le complexe ORC et les sites d'initiation de la réplication sont plus diffus. Ces différences pourraient s'expliquer par des propriétés biochimiques nouvelles des protéines composant le pré-RC ou encore l'évolution de la structure et de la taille des génomes. De fait, l'extrapolation de nombreuses observations d'un système cellulaire (ou animal) à l'autre se révèle hasar-

deuse et de nombreux modèles historiques du contrôle de l'initiation de la réplication chez l'homme se révèlent en fait obsolètes.

Cette revue a pour but de faire un état des lieux des connaissances sur les mécanismes de détermination des origines de réplication dans le génome humain, tout en mettant en avant les limitations des méthodologies actuelles et les questions qui restent en suspens.

### Combien d'origines et de classes d'origines de réplication dans le génome humain ?

Des efforts récents ont permis de développer des protocoles expérimentaux capables d'identifier la position des origines à l'échelle du génome. Il est ainsi possible de purifier les bulles de réplication qui se forment aux origines, lors de l'initiation de la réplication, de purifier les segments d'ADN nouvellement synthétisés dans la bulle (SNS), de purifier les fragments d'Okazaki [41] (→), ou de purifier les séquences ADN liées par le complexe ORC (Figure 1). Couplées à un séquençage en masse et une analyse informatique, ces approches permettent d'identifier les zones d'initiation de la réplication et de liaison du facteur ORC dans le génome humain, dans différents types cellulaires et en réponse à différentes conditions environnementales [4-10].

Chez l'homme, l'ensemble des études indique que les origines de réplication se présentent sous forme de zones, de quelques kilobases, bien définies sur le génome. En fonction des protocoles et des types cellulaires entre 50 000 et 200 000 origines sont identifiées [4-10]. Leur nombre pourrait toutefois être encore plus important car certaines origines pourraient ne pas être activées de manière suffisamment robuste ou fréquente pour être détectées sans ambiguïté expérimentalement.

Il existe cependant un paradoxe entre le nombre d'origines détecté par ces approches génomiques et le nombre d'origines réellement utilisées par la cellule. Ainsi, fondé sur l'analyse de brins d'ADN ou de radio-marquage, 30 000 origines seraient utilisées à chaque cycle cellulaire pour permettre la réplication du génome humain [3]. Il n'existe, pour l'heure, pas d'explication satisfaisante permettant de rendre compte du fait que seule une infime partie des origines prédites soit utilisée dans chaque cellule [2]. Il est proposé que la cellule active, de manière aléatoire, un nombre limité d'origines (~ 30 000) au sein d'une population plus importante d'origines potentielles (> 50 000). Ce processus permettrait une plus grande flexibilité et capacité d'adaptation à même de répondre aux changements environnementaux ou développementaux que les cellules rencontrent. Les origines potentielles non utilisées pourraient également constituer un programme de secours capable de compenser un défaut du programme de réplication. Ces origines, normalement inactives, seraient ainsi activées en cas de défaut de synthèse de l'ADN, au cours de la phase S, pour permettre une ré-initiation de la réplication en aval d'une fourche de réplication défectueuse. À l'heure actuelle, le débat reste ouvert entre ces deux interprétations qui sont, par ailleurs, mutuellement compatibles [2].

### Les sites d'initiation de la réplication sont fréquemment des régions régulatrices de la transcription : simple coïncidence ?

La cartographie de l'ensemble des sites d'initiation de la réplication et de liaison de ORC dans le génome humain a montré un enrichissement dans les régions riches en guanine et cytosine [4-10]. Pourtant aucun motif ADN ne reflète, de manière globale, la distribution de ces origines, suggérant que des facteurs de la chromatine ou de la transcription puissent être impliqués.

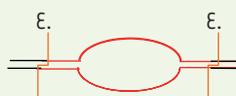
Il existe ainsi une concordance entre sites d'initiation de la réplication et séquences cis-régulatrices de la transcription, notamment les sites promoteurs des gènes [4-10]. En effet, près de 30 % des sites d'initiation de réplication et des sites ORC sont retrouvés à proximité de promoteurs (Figure 3A) ou de séquences cis-régulatrices de gènes (Figure 3B). Malgré cet enrichissement, la majeure partie des sites n'est pas située au niveau de promoteurs ou de régions cis-régulatrices, indiquant que cette localisation, au sein d'une séquence du génome, n'est pas une caractéristique essentielle de la position des origines de réplication.

La coïncidence entre la localisation des sites d'initiation de la réplication et celle des sites de transcription suggère cependant que les deux types d'activité pourraient être co-régulés. La grande majorité des promoteurs qui sont associés à une origine de réplication sont transcrits, souvent de manière constitutive dans plusieurs types cellulaires. Pourtant, des analyses sophistiquées n'ont montré aucune corrélation entre le niveau d'expression d'un gène et l'affinité du complexe ORC et/ou le niveau d'initiation de la réplication au niveau du promoteur de ce gène [5, 9, 10, 16]. Les deux événements semblent donc indépendants l'un de l'autre. Un promoteur très actif peut être une origine de réplication peu robuste et, *vice versa*, un promoteur faiblement transcrit, une origine robuste.

Une caractéristique commune aux promoteurs, aux régions cis-régulatrices et aux sites d'initiation de la réplication, est la structure chromatinienne retrouvée dans ces régions. Celle-ci est souvent ouverte, les séquences ADN sont donc accessibles et de nombreux facteurs de transcription y sont liés (Figure 3C) [4-10, 16]. Bien qu'aucun ne soit détecté au niveau de l'ensemble des origines de réplication prédites, les facteurs de transcription pourraient cependant jouer un rôle clé dans l'initiation de la réplication. En se liant à l'ADN, ils pourraient faciliter le recrutement des facteurs de réplication, le remodelage de la chromatine et *in fine* permettre l'assemblage de la machinerie de réplication. Par exemple, l'inactivation des facteurs

(→) Voir la Nouvelle de S. Marion de Procé, m/s n° 10, octobre 2015, page 821

### A Purification des bulles de réplication après digestion enzymatique



#### INFORMATIONS

- Sites ou zones d'initiation de la réplication.

#### LIMITATIONS

- Protocole complexe.
- Résolution assez faible ne permettant pas des analyses de séquences nucléotidiques.
- La méthode de détection est dépendante de la présence de sites de restriction.

### B Purification des brins d'ADN naissants



#### INFORMATIONS

- Sites ou zones d'initiation de la réplication.
- Proportionnalité entre signal et activité de l'origine.
- Très bonne résolution.

#### LIMITATIONS

- Protocole complexe.
- Possibles biais expérimentaux.
- Données faiblement reproductibles entre laboratoires.

### C Purification des fragments d'Okazaki



#### INFORMATIONS

- Sites ou zones d'initiation de la réplication.
- Direction de la fourche de réplication.
- Très bonne résolution.

#### LIMITATIONS

- Protocole complexe.
- Méthode d'analyse complexe.

### D Purification des sites d'incorporation du BrdU



#### INFORMATIONS

- Sites ou zones d'initiation de la réplication.
- Temps d'initiation en phase S.
- Très bonne résolution.

#### LIMITATIONS

- Protocole complexe.
- Méthode et résolution dépendante de la durée de marquage au BrdU.

### E Purification des sites de liaison à la chromatine du complexe ORC



#### INFORMATIONS

- Sites ou zones de liaison de ORC.
- Protocole standard.
- Très bonne résolution.

#### LIMITATIONS

- Protocole dépendant de la qualité de l'anticorps.
- Aucune information sur l'initiation de la réplication.

**Figure 1. Description des méthodologies mises au point pour l'identification des origines de réplication à l'échelle du génome humain entier.**

**A.** Purification des bulles de réplication (d'après [7]). Après digestion enzymatique avec l'enzyme EcoRI (une enzyme de restriction qui coupe les sites G<sup>^</sup>AATTC, notée ε.) de l'ADN génomique, les fragments d'ADN sont déposés sur une colonne d'agarose. Du fait de leur nature circulaire, les bulles de réplication sont retenues sur la colonne et éluées plus tardivement que le reste des brins d'ADN. Les ADN circulaires sont ensuite préparés et analysés par séquençage haut débit.

**B.** Purification des brins d'ADN naissants (d'après [4, 5, 8]). Après purification, l'ADN génomique est dénaturé et analysé sur une colonne de fractionnement par la taille. Les fractions correspondant aux tailles 1 à 1,5 kilobases, et contenant les brins naissants, sont purifiées. Les échantillons sont alors traités à la lambda-endonucléase pour sélectionner les brins naissants et éliminer les fragments d'ADN (provenant de cassures lors de la préparation des échantillons). Ils sont ensuite préparés et analysés par séquençage haut-débit.

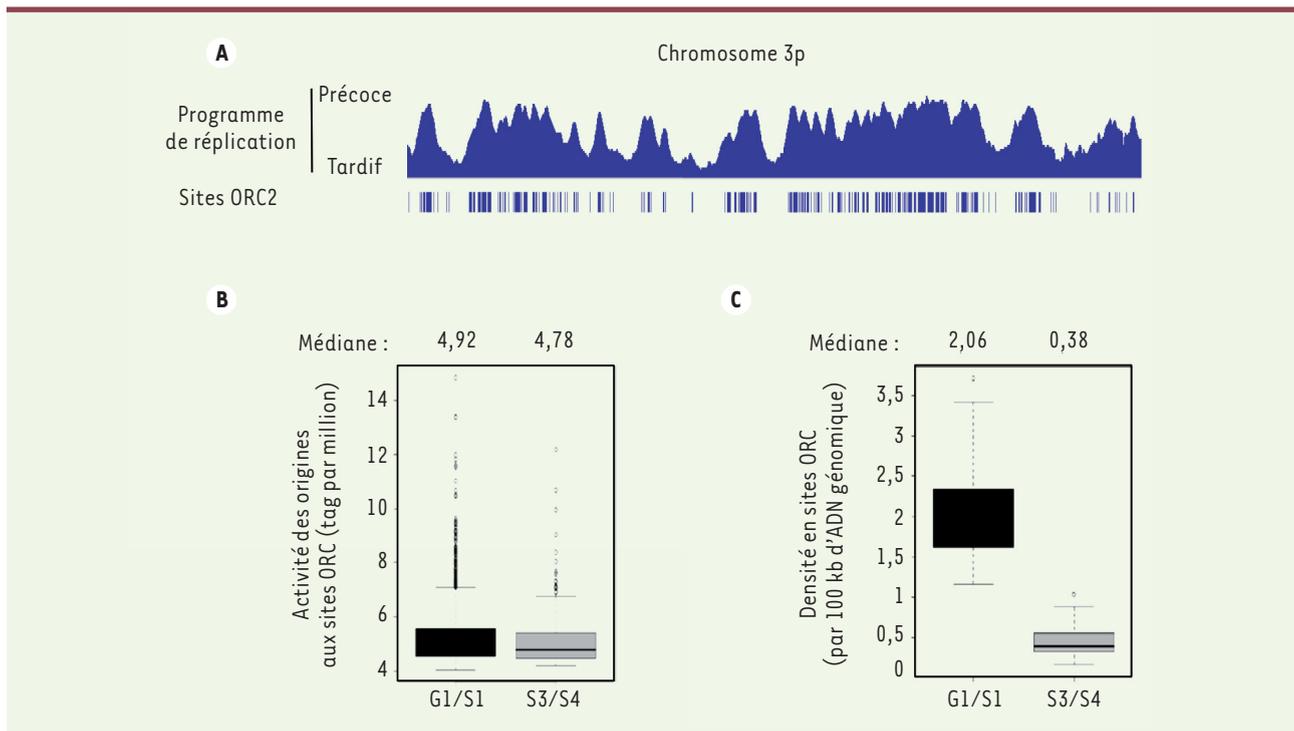
**C.** Purification des fragments d'Okazaki (d'après [9]). Après un marquage à la 5-éthynyl-2'-désoxyuridine (EdU), un analogue de la thymidine, l'ADN est isolé des cellules puis dénaturé. Après passage sur une colonne de fractionnement, les fragments d'Okazaki sont purifiés grâce à l'EdU, l'ARN hydrolysé à la soude, et les ADN simples brins obtenus sont analysés par séquençage haut-débit.

**D.** Purification des sites d'incorporation de la bromodésoxyuridine (BrdU)/EdU (d'après [13]). Après un marquage à la BrdU/EdU, l'ADN est isolé des cellules

puis dénaturé. Les fragments d'ADN contenant l'EdU (ou le BrdU) sont purifiés et analysés par séquençage haut-débit. **E.** Purification des sites de liaison du complexe ORC (*origin recognition complex*) (d'après [6, 10]). Après fixation au formaldéhyde et capture des complexes ADN/protéine, la chromatine est extraite des cellules et fractionnée par sonication. Les brins de chromatine liés par ORC sont ensuite purifiés sur une colonne de billes d'anticorps dirigés contre la protéine ORC2. Après purification, les ADN sont analysés par séquençage haut-débit.

généraux de la transcription, comme NFYA (*nuclear transcription factor y subunit alpha*), HMGAI (*high mobility group at-hook 1*), c-JUN, c-FOS ou MYC, altère le programme de réplication et la progression en

phase S [17-20]. Ces facteurs seraient donc importants localement et plusieurs d'entre eux pourraient fonctionner de manière équivalente au niveau d'origines



**Figure 2. Corrélation entre la distribution des sites ORC2 et des sites d'initiation de la réplication avec le programme temporel de réplication.** **A.** Illustration sur une portion du chromosome 3 de la correspondance entre la densité en sites ORC2 (*origin recognition complex 2*) et le programme temporel de réplication au cours de la phase S. **B.** Illustration de l'activité d'initiation des origines localisées dans des régions se répliquant en début de phase S (G1/S1) et en fin de phase S (S3/S4). **C.** Distance relative entre 2 sites ORC2 dans les régions du génome se répliquant à différents stades de la phase S : début de phase S (G1/S1) et fin de phase S (S3/S4).

différentes pour permettre le contrôle de l'initiation de la réplication. Par exemple, les facteurs de transcription c-JUN et c-FOS, tout comme le facteur de réplication CDT1, se lient à l'histone acétylase HBO1 [20-22] qui est essentielle pour l'assemblage du complexe pré-RC au cours de la phase G1 et l'initiation de la réplication en phase S [20, 22]. Il est donc envisageable que CDT1, en coopération avec certains facteurs de transcription, puisse définir un site d'ancrage, et de régulation, de HBO1 au niveau des origines de réplication [17-22].

Il a aussi été démontré que les voies de signalisation qui régulent la transcription agissent également sur les constituants du complexe pré-RC pour contrôler la réplication [23]. La juxtaposition des machineries de réplication et de transcription dans certaines régions du génome pourrait ainsi permettre de réguler simultanément les deux processus en réponse à des stimulus environnementaux ou développementaux communs. Il apparaît toutefois que ces effets auraient dû être examinés au niveau d'une origine en particulier, et non à l'échelle du génome. Différents laboratoires ont ainsi montré que, localement, l'activité transcriptionnelle pouvait influencer sur les zones d'initiation de la réplication [3, 24-27]. L'activité d'une origine de réplication serait en effet réduite si la région est transcrite et la position du site d'initiation affectée par le passage de la polymérase de l'ARN [24-28]. Remus et ses collaborateurs (Institut Sloan Kettering, New York, États-Unis) ont élégamment confirmé ces observations en montrant, dans un système *in vitro*, que le passage de la polymérase de l'ARN mobilise le

complexe MCM sur un brin de l'ADN, de sorte que l'initiation de la réplication se fait alors non plus au niveau du site ORC, mais dans la région de terminaison de la transcription [28]. Un tel phénomène pourrait expliquer la flexibilité du processus d'initiation de la réplication avec des sites ORC situés à distance des zones d'initiation de la réplication.

### Comment se spécifient les origines de réplication ?

L'accessibilité de l'ADN serait l'un des déterminants majeurs de la position des origines de réplication et du contrôle de leur activité. Les histones, autour desquelles s'enroule l'ADN pour former la chromatine, sont la cible de nombreuses modifications post-traductionnelles (méthylation, phosphorylation, etc.) qui résultent en une combinatoire complexe le long de la fibre chromatinienne [29-31] (→).

Certaines modifications de ces histones jouent un rôle clé à différentes étapes de

(→) Voir la Synthèse de D. Ray-Gallet *et al.*, *m/s* n° 4, avril 2005, page 384, la Synthèse de B. Miotto et K. Struhl, *m/s* n° 8-9, août-septembre 2007, page 735, et la Nouvelle de L.C. Pradel *et al.*, *m/s* n° 3, mars 2015, page 236

l'initiation de la réplication. Ainsi, Julien et ses collaborateurs (Institut de recherche en cancérologie de Montpellier) ont montré que la méthyltransférase PR-SET7, qui est responsable de la méthylation de la lysine 20 de l'histone H4 (H4K20me1), est essentielle à l'initiation de la réplication [32]. D'autre part, le domaine BAH (*bromo-adjacent homology*) de ORC1, une des 6 sous-unités du complexe ORC, reconnaît préférentiellement la diméthylation de la lysine 20 [33] et des mutations ponctuelles de ce domaine bloquent *in vitro* la liaison de ORC1 à cette marque épigénétique, et son recrutement sur la chromatine *in vivo* [33]. La méthylation de la lysine 20 semble donc jouer un rôle important dans le positionnement du complexe ORC. Malheureusement il n'existe, pour l'heure, aucune cartographie de cette marque dans les cellules humaines qui puisse être comparée à la position des sites ORC. À noter que les mutations affectant le domaine BAH sont retrouvées chez des patients atteints du syndrome Meier-Gorlin, un syndrome rare qui se caractérise par une petite taille, de petites oreilles et des rotules sous-développées [33, 34]. À n'en pas douter la cartographie des différentes formes méthylées de la lysine 20 de l'histone H4 sera une étape importante pour notre compréhension du mécanisme de sélection des origines de réplication.

L'acétylation des histones et l'acétylase HBO1 sont essentielles à la formation du pré-RC et l'initiation de la réplication [20, 22, 35, 36]. Notre analyse des sites de liaison de ORC sur le génome humain confirme ces observations [10]. La présence d'une région chromatinienne « ouverte » et enrichie en acétylation des queues d'histones est très prédictive de la liaison de ORC [10]. Le modèle le plus probable est que dans un environnement chromatinienn ouvert, le complexe ORC reconnaît aspécifiquement une séquence ADN. Cette interprétation est cohérente avec les données biochimiques qui ont été accumulées depuis de nombreuses années indiquant que le complexe ORC lie l'ADN indépendamment de sa séquence [1, 2, 37]. Ces observations suggèrent donc que le complexe ORC se lierait à la chromatine de manière opportuniste au niveau de régions de chromatine ouverte [10, 38]. Ceci pourrait expliquer l'enrichissement des promoteurs, et autres régions associées à la transcription, au niveau des sites ORC et des origines puisque ces régions sont intrinsèquement accessibles aux facteurs nucléaires.

Les séquences ADN retrouvées au niveau des origines de réplication sont plus riches en dinucléotides CpG (cytosine-phosphate-guanine) que le reste du génome [4-10]. D'ailleurs le niveau d'activation d'une origine et l'ouverture de la chromatine sont corrélés à la richesse en CpG [5, 8, 16]. Les séquences riches en CpG sont de plus susceptibles de former des structures secondaires comme des « quadruplexes » de guanines (ou G4) [5, 8, 9, 39, 40] et, pour plusieurs origines modèles, une séquence susceptible de former un G4 contribue à l'initiation de la réplication [39, 40]. Une information de structure de l'ADN, plutôt que de séquence, permettrait donc le recrutement du complexe ORC et l'établissement des origines. Toutefois, des approches de biochimie n'ont pas permis de mettre en évidence une affinité de ORC avec des structures de ce type *in vitro* ([37] et données personnelles). De plus, le fait que ces structures G4 soient absentes au niveau de nombreux sites d'initiation suggère qu'elles ne sont pas un déterminant de la position des origines, mais plus probablement requises pour l'activité de certaines d'entre elles [4, 5, 8, 9, 39, 40].

## La distribution des origines de réplication le long de bras des chromosomes est-elle aléatoire ?

Les approches sur brin d'ADN indiquent que la distance entre deux origines de réplication dans une cellule est, en moyenne, comprise entre 40 et 200 kilobases. Les approches génomiques confirment ces observations. La distance moyenne entre deux sites ORC est d'environ 94 kilobases et, entre deux sites d'initiation de la réplication, de 30 à 60 kilobases [4-10].

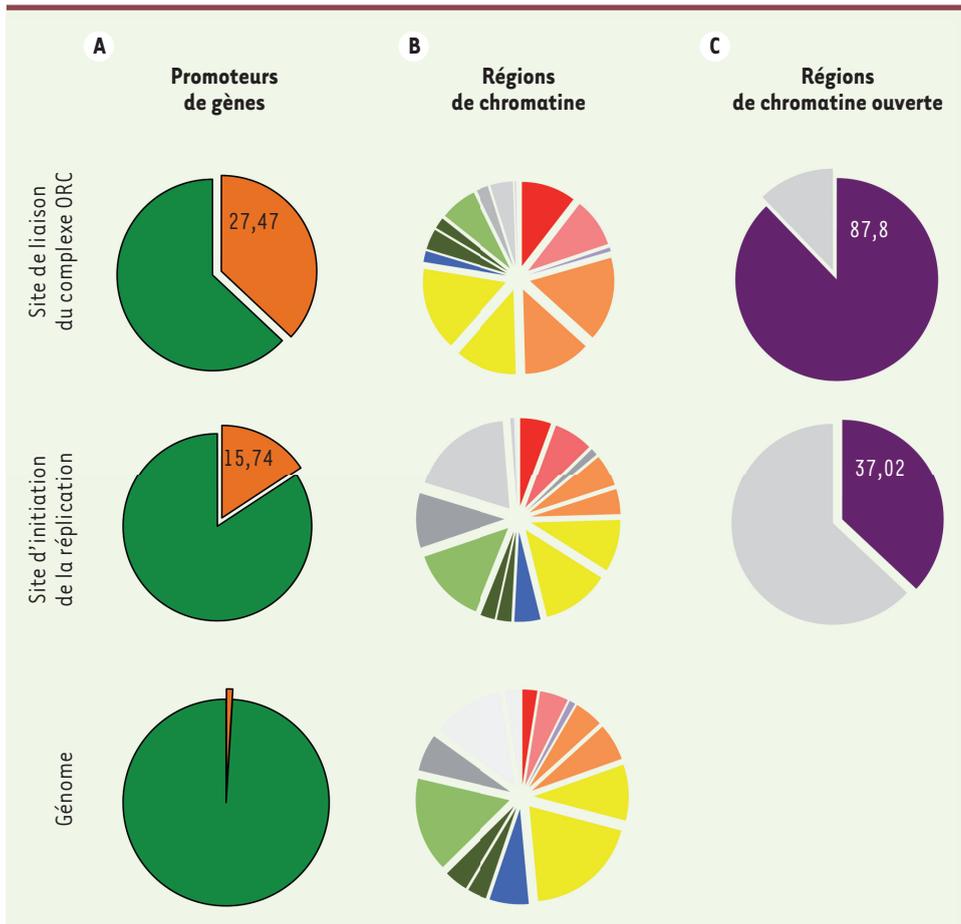
La distribution des sites d'initiation de la réplication et des sites ORC n'est toutefois pas homogène sur le génome. Il existe des régions « riches » en origines, où la distance entre deux origines est plus faible, et des régions « pauvres » en origines. Cette distribution serait en lien direct avec le programme temporel de réplication. La densité des sites d'initiation et des sites ORC est en effet plus élevée dans les régions se répliquant en début de phase S que dans les régions se répliquant en fin de phase S (Figure 2A) [5, 6, 8, 10]. Cette observation fait écho à de nombreux travaux démontrant que les régions répliquées au début de phase S sont généralement riches en gènes transcrits et en zones de chromatine ouverte [2, 3].

Par une approche de modélisation mathématique, nous avons montré que la distribution hétérogène des sites ORC2 sur le génome et une activation purement aléatoire de l'initiation à certains de ces sites ORC2 au cours de la phase S, permettent de rendre compte du programme temporel de réplication que l'on observe expérimentalement [10]. En d'autres termes, la densité en sites ORC2 établis au cours de la phase G1 (indépendamment de leur activation en phase S) serait un déterminant essentiel du programme temporel de réplication, et fortement suggestif d'un processus aléatoire du choix des origines activées au cours de la phase S. Ce modèle expliquerait que les sites d'initiation de la réplication soient à des positions variables, au sein de zones, entre les cellules d'une même population.

De manière surprenante, il existe de nombreuses portions du génome, de plusieurs mégabases, pour lesquelles peu de sites ORC et d'origines sont détectables expérimentalement [5, 8, 10]. Ces régions coïncident fréquemment avec les sites fragiles communs (SFC) du génome<sup>1</sup> [10-12] (→).

(→) Voir la Nouvelle de A. Letessier et al., m/s n° 8-9, août-septembre 2011, page 707

<sup>1</sup> Les sites fragiles communs (SFC) sont des locus instables en cas de stress réplicatif et des lieux préférentiels de réarrangements dans les tumeurs. Les SFC sont souvent associés aux grands gènes du génome. Il a été proposé que la transcription de ces gènes et des défauts de réplication dans ces régions soient à l'origine de cassures de l'ADN.



**Figure 3. Distribution des sites ORC (origin recognition complex) et des sites d'initiation de la réplication sur le génome : lien avec les sites d'initiation de la transcription.** **A.** Pourcentage de sites ORC2 et de sites d'initiation localisés au niveau des promoteurs de gènes ( $\pm 1$  kilobase du site d'initiation de la transcription). **B.** Répartition des sites ORC2 et des sites d'initiation de la réplication en fonction de leur statut chromatinien et transcriptionnel à l'échelle du génome. En intégrant les données de statut chromatinien, de transcription et de liaison de nombreux facteurs de transcription, un modèle probabiliste de l'alternance des différents types d'éléments le long des séquences génomiques identifie 15 états de la chromatine. Ces états sont : les promoteurs actifs (rouge), les promoteurs faibles (rose), le promoteur inactif (gris foncé), les régions *cis*-régulatrices actives (orange), les régions *cis*-régulatrices faiblement actives (jaune), les régions insulatrices (bleu), les régions transcrites (vert foncé), les régions faiblement transcrites (vert clair), les régions contrôlées par les protéines de la famille Polycomb (gris foncé) et les régions hétérochromatiniennes (gris clair). **C.** Pourcentage de sites ORC2 et de sites d'initiation coexistant avant des zones de chromatine ouverte (définie par une accessibilité accrue à la nucléase DNase I).

trices actives (orange), les régions *cis*-régulatrices faiblement actives (jaune), les régions insulatrices (bleu), les régions transcrites (vert foncé), les régions faiblement transcrites (vert clair), les régions contrôlées par les protéines de la famille Polycomb (gris foncé) et les régions hétérochromatiniennes (gris clair). **C.** Pourcentage de sites ORC2 et de sites d'initiation coexistant avant des zones de chromatine ouverte (définie par une accessibilité accrue à la nucléase DNase I).

Les SFC sont répliqués très tardivement au cours du cycle cellulaire. Ils sont donc fréquemment sujets à des cassures et des réarrangements lorsque le programme de réplication se prolonge [11, 12]. L'absence de ces localisations particulières dans les données obtenues par génomique est probablement associée au fait que les origines de réplication sont distribuées de manière diffuse et variable dans ces portions du génome, ce que les programmes informatiques actuels ne permettent pas d'identifier de manière robuste. L'absence d'origines bien localisées serait ainsi une des causes de la fragilité de ces régions du génome [10-12].

Les délétions récurrentes et les régions où s'accumulent les mutations ponctuelles dans les cancers sont, pour la plupart, localisées dans des régions répliquées en fin de phase S [10, 13-15]. Inversement, les régions répliquées en début de phase S contiennent des régions fréquemment amplifiées [13-15]. Il existerait donc une relation, peut-être causale, entre la densité en origines de réplication d'une région du génome et la nature des réarrangements chromosomiques que l'on observe dans les cancers.

### Perspectives

La carte des sites de liaison du complexe ORC et des sites d'initiation de la réplication a ouvert la porte à de nombreuses études sur les mécanismes de régulation de l'initiation de la réplication, l'organisation du génome, et l'importance de la structure chromatinienne dans ce processus. Les mécanismes de l'initiation de la réplication étaient précédemment fondés sur l'analyse d'un nombre limité d'origines de réplication chez l'homme, ou extrapolés à partir d'autres organismes modèles, comme la levure ou la drosophile. Les données collectées sont donc essentielles et, de toute évidence, elles permettent de revisiter certains modèles historiques des mécanismes d'initiation, chez l'homme. Un effort important sera cependant nécessaire pour normaliser les protocoles et les systèmes cellulaires étudiés. En effet, les données accumulées n'ont pas permis une

description et une caractérisation satisfaisantes des origines de réplication. La portée des conclusions de chaque étude et les observations sont soumises à d'interminables débats sur les biais (ou artéfacts) liés au protocole expérimental choisi. Il sera donc essentiel de confirmer et reproduire les observations de chacun.

Une autre question essentiellement sémantique s'offre au domaine. Que signifie, ou recouvre, le terme « origine de réplication » ? En effet, initialement utilisé chez la bactérie et la levure, ce terme reflète l'existence de séquences génomiques précises associées à l'assemblage du complexe de réplication et à son initiation. Cette notion ne s'applique probablement pas aussi clairement au génome humain. Les sites ORC et les sites d'initiation ne sont pas nécessairement superposables, et les sites d'initiation sont souvent des zones de plusieurs kilobases. Ces derniers points peuvent prêter à confusion dans la littérature actuelle où certaines observations à partir de la cartographie des sites d'initiation de la réplication sont extrapolées et considérées comme pertinentes pour l'étude des mécanismes de recrutement du complexe ORC (et *vice versa*). Finalement, les approches génomiques, encore balbutiantes, dans le domaine de la réplication n'ont pas permis d'offrir une vision claire et définitive des déterminants de la position des origines de réplication et de leurs mécanismes d'activation dans le génome humain. De manière surprenante, les études mettent en avant le caractère vraisemblablement aspécifique et stochastique du processus, aussi bien de recrutement de ORC que d'activation des origines. L'identification des déterminants moléculaires régulant ces processus stochastiques est le défi qui s'offre aux scientifiques étudiant les mécanismes d'initiation de la réplication. ♦

## SUMMARY

### How genomic approaches help the understanding of the initiation of DNA replication

Eukaryotic DNA replication starts at multiple sites distributed throughout the genome called origins of replication. The identification of the position of these origins in the genome, as well as the description of their sequence features and chromatin structure, have been the center of many studies over the last years. A handful of sophisticated methods has been developed to capture replication initiation events, to detect replication factors binding sites onto the chromatin and to computationally analyze these data at the genome scale. Yet, we remain far from fully understanding how these origins of replication are specified in the genome and how DNA replication initiation is regulated. The aim of this review is to provide the reader with a detailed and curated list of the latest findings regarding the nature of origins of replication in the human genome. ♦

## LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

## RÉFÉRENCES

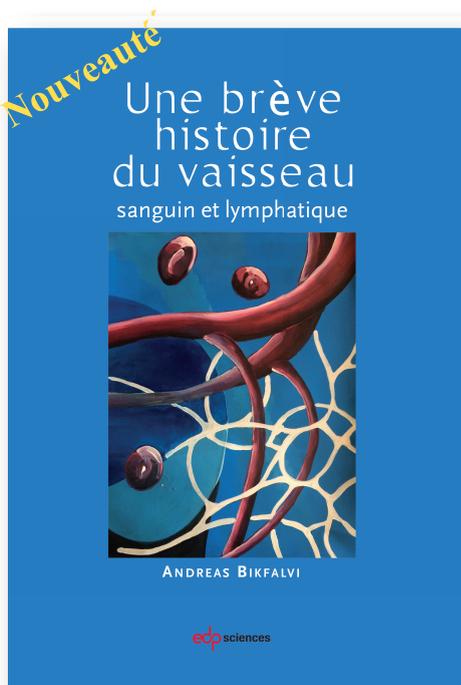
- Bell SP, Dutta A. DNA replication in eukaryotic cells. *Annu Rev Biochem* 2002 ; 71 : 333-74.
- Fragkos M, Ganier O, Coulombe P, Méchali M. DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol* 2015 ; 16 : 360-74.
- Hyrien O. Peaks cloaked in the mist : the landscape of mammalian replication origins. *J Cell Biol* 2015 ; 208 : 147-60.
- Cadoret JC, Meisch F, Hassan-Zadeh V, et al. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci USA* 2008 ; 105 : 15837-42.
- Besnard E, Babled A, Lapasset L, et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol* 2012 ; 19 : 837-44.
- Dellino GI, Cittaro D, Piccioni R, et al. Genome-wide mapping of human DNA-replication origins : levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res* 2013 ; 23 : 1-11.
- Mesner LD, Valsakumar V, Cieslik M, et al. Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Res* 2013 ; 23 : 1774-88.
- Picard F, Cadoret JC, Audit B, et al. The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet* 2014 ; 10 : e1004282.
- Petryk N, Kahli M, d'Aubenton-Carafa Y, et al. Replication landscape of the human genome. *Nat Commun* 2016 ; 7 : 10208.
- Miotto B, Ji Z, Struhl K. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proc Natl Acad Sci USA* 2016 ; 113 : E4810-9.
- Letessier A, Birnbaum D, Debatisse M, Chaffanet M. La pauvreté en sites d'initiation de la réplication rend-elle fragile certaines régions du génome ? *Med Sci (Paris)* 2011 ; 27 : 707-9.
- Letessier A, Millot GA, Koundrioukoff S, et al. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* 2011 ; 470 : 120-3.
- Ryba T, Battaglia D, Chang BH, et al. Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome Res* 2012 ; 22 : 1833-44.
- Smith L, Plug A, Thayer M. Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proc Natl Acad Sci USA* 2001 ; 98 : 13300-5.
- Polak P, Karličić R, Koren A, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* 2015 ; 518 : 360-4.
- Lombrana R, Almeida R, Revuelta I, et al. High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *EMBO J* 2013 ; 32 : 2631-44.
- Thomae AW, Pich D, Brocher J, et al. Interaction between HMG1a and the origin recognition complex creates site-specific replication origins. *Proc Natl Acad Sci USA* 2008 ; 105 : 1692-7.
- Benatti P, Belluti S, Miotto B, et al. Direct non transcriptional role of NF- $\kappa$ B in DNA replication. *Biochim Biophys Acta* 2016 ; 1863 : 673-85.
- Dominguez-Sola D, Ying CY, Grandori C, et al. Non-transcriptional control of DNA replication by c-Myc. *Nature* 2007 ; 448 : 445-51.
- Miotto B, Struhl K. HBO1 histone acetylase is a coactivator of the replication licensing factor Cdt1. *Genes Dev* 2008 ; 22 : 2633-8.
- Miotto B, Struhl K. Differential gene regulation by selective association of transcriptional coactivators and bZIP DNA-binding domains. *Mol Cell Biol* 2006 ; 26 : 5969-82.
- Feng Y, Vlassis A, Roques C, et al. BRPF3-HBO1 regulates replication origin activation and histone H3K14 acetylation. *EMBO J* 2016 ; 35 : 176-92.
- Miotto B, Struhl K. JNK1 phosphorylation of Cdt1 inhibits recruitment of HBO1 histone acetylase and blocks replication licensing in response to stress. *Mol Cell* 2011 ; 44 : 62-71.
- Grégoire D, Brodolin K, Méchali M. HoxB domain induction silences DNA replication origins in the locus and specifies a single origin at its boundary. *EMBO Rep* 2006 ; 7 : 812-6.
- Hyrien O, Maric C, Méchali M. Transition in specification of embryonic metazoan DNA replication origins. *Science* 1995 ; 270 : 994-7.
- Powell SK, MacAlpine HK, Prinz JA, et al. Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO J* 2015 ; 34 : 531-43.
- Norio P, Kosiyatrakul S, Yang Q, et al. Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Mol Cell* 2005 ; 20 : 575-87.
- Gros J, Kumar C, Lynch G, et al. Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Mol Cell* 2015 ; 60 : 797-807.
- Ray-Gallet D, Gérard A, Polo S, Almouzni G. Variations sur le thème du code histone. *Med Sci (Paris)* 2005 ; 21 : 384-9.

## RÉFÉRENCES

30. Miotto B, Struhl K. De la régulation du génome à la progression tumorale. *Med Sci (Paris)* 2007 ; 23 : 735-40.
31. Pradel LC, Vanhille L, Spicuglia S. Projet européen *Blueprint*. *Med Sci (Paris)* 2015 ; 31 : 236-8.
32. Tardat M, Brustel J, Kirsh O, et al. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat Cell Biol* 2010 ; 12 : 1086-93.
33. Kuo AJ, Song J, Cheung P, et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 2012 ; 484 : 115-9.
34. Bicknell LS, Walker S, Klingseisen A, et al. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nat Genet* 2011 ; 43 : 350-5.
35. Suchyta M, Miotto B, McGarry TJ. An inactive geminin mutant that binds cdt1. *Genes (Basel)* 2015 ; 6 : 252-66.
36. Miotto B, Struhl K. HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Mol Cell* 2010 ; 37 : 57-66.
37. Hoshina S, Yura K, Teranishi H, et al. Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J Biol Chem* 2013 ; 288 : 30161-71.
38. Sequeira-Mendes J, Gómez M. On the opportunistic nature of transcription and replication initiation in the metazoan genome. *Bioessays* 2012 ; 34 : 119-25.
39. Bartholdy B, Mukhopadhyay R, Lajugie J, et al. Allele-specific analysis of DNA replication origins in mammalian cells. *Nat Commun* 2015 ; 6 : 7051.
40. Valton AL, Hassan-Zadeh V, Lema I, et al. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J* 2014 ; 33 : 732-46.
41. Marion de Procé S. Des fragments d'ADN synthétisés par l'ADN polymérase  $\alpha$  modifient notre génome. *Med Sci (Paris)* 2015 ; 31 : 821-3.

### TIRÉS À PART

B. Miotto



ISBN : 978-2-7598-1863-1

202 pages

25 €

Ce livre, intéressant et lisible à la fois pour le spécialiste et le grand public, apporte des observations originales et nouvelles concernant l'angiogenèse, et notamment l'histoire des différentes découvertes, et discute les aspects et les concepts plus généraux en les plaçant dans le contexte de la philosophie des sciences.

Facile à lire, bien illustré, cet ouvrage cherche à comprendre et à faire comprendre les enjeux de la recherche sur l'arbre vasculaire en développement et en pathologie. Il intéressera non seulement les étudiants et post-doctorants en biologie, mais aussi les chercheurs actifs dans ce domaine de recherche ainsi que toute personne intéressée par la biologie et la médecine et par l'histoire des sciences. Un voyage passionnant à travers l'histoire et les concepts les plus actuels concernant les recherches sur le vaisseau sanguin.

*Andreas Bikfalvi est Professeur à l'université de Bordeaux et Directeur d'une unité de recherche Inserm sur le cancer et la biologie vasculaire. Il est, par ailleurs, membre senior de l'Institut Universitaire de France (IUF) et reconnu internationalement pour ses recherches dans le domaine de l'angiogenèse tumorale.*

**BON DE COMMANDE**

À retourner à EDP Sciences, 109, avenue Aristide Briand, 92541 Montrouge Cedex, France  
Tél. : 01 41 17 74 05 - Fax : 01 49 85 03 45 - E-mail : [françois.flori@edpsciences.org](mailto:françois.flori@edpsciences.org)

NOM : ..... Prénom : .....

Adresse : .....

Code postal : ..... Ville : .....

Pays : .....

Fonction : .....

Je souhaite recevoir

**Une brève histoire de vaisseau : 25 € + 3 € de port = 28 € TTC**

en ..... exemplaire, soit un total de ..... €

Par chèque, à l'ordre de EDP Sciences

Par carte bancaire :  Visa  Eurocard/Mastercard

Carte n° | | | | | | | | | | | | | | | | | | | | | |

Date d'expiration : | | | | | | | | N° de contrôle au dos de la carte : | | | | | | | | Signature :