



**HAL**  
open science

# Rôles relatifs de la transcription et de la dynamique de réplication dans l'instabilité des Sites Fragiles Communs humains

Dana Azar

► **To cite this version:**

Dana Azar. Rôles relatifs de la transcription et de la dynamique de réplication dans l'instabilité des Sites Fragiles Communs humains. Génétique. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT: 2015PA066158 . tel-01206005

**HAL Id: tel-01206005**

**<https://theses.hal.science/tel-01206005>**

Submitted on 28 Sep 2015

**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.



*Thèse de doctorat de l'université Pierre et Marie Curie  
Ecole Doctorale : Complexité du vivant*

*Rôles relatifs de la transcription et de la  
dynamique de réplication dans l'instabilité des  
Sites Fragiles Communs humains*

*Présentée par  
Dana Fouad AZAR*

*Préparée sous la direction du Dr. Olivier Brison  
Dans le laboratoire du Pr. Michelle Debatisse,  
UMR3244 à l'Institut Curie (Paris)*

Membres du jury

Pr. Frédéric Devaux	Président
Dr. Philippe Pasero	Rapporteur
Dr. Angelos Constantinou	Rapporteur
Dr. Marie-Noëlle Prioleau	Examinatrice
Dr. Filippo Rosselli	Examineur
Pr. Michelle Debatisse	Examinatrice
Dr. Olivier Brison	Directeur de thèse



« La vérité, c'est de chercher toujours la vérité. »

« La vérité est une illusion et l'illusion est une vérité. »

« Vérité dans un temps, erreur dans un autre. »

« La vérité peut attendre, elle y est habituée. »

« Ce sont rarement les réponses qui apportent la vérité, mais l'enchaînement des questions. »

« Ce qu'il y a de terrible quand on cherche la vérité, c'est qu'on la trouve. »

« Le moyen d'être cru est de rendre la vérité incroyable. »

« Le doute est un hommage que l'on rend à la vérité. »

# Remerciements

---

Je profite de ces quelques lignes pour remercier tous ceux qui m'ont accompagnée ces cinq dernières années et sans qui ce manuscrit n'aura jamais pu être réalisé.

Je tiens d'abord à remercier les différents membres de mon jury, merci M. Pasero et M. Constantinou d'avoir accepté d'être mes rapporteurs malgré vos emplois de temps chargés. Merci à mes deux examinateurs, Mme. Prioleau et M. Rosselli qui ont suivi mon travail de près pendant toutes ces années et l'ont nourri avec leurs remarques et commentaires. Je remercie enfin Pr. Devaux pour avoir accepté de présider cette thèse.

Je tiens à remercier M. Olivier Brison, j'aurai jamais pu accomplir ce travail sans vous, merci de m'avoir fait confiance, merci pour votre disponibilité et votre soutien quotidien, pour votre patience, je sais que cela n'a pas été toujours facile et que je suis loin d'être l'étudiante parfaite, maintenant, vous allez avoir votre retraite pour vous en remettre 😊 merci encore pour l'aide pour la rédaction de ce manuscrit.

Je tiens à remercier Mme. Michelle Debatisse, merci de m'avoir permis d'effectuer ma thèse dans un environnement scientifique de très grande qualité, merci pour votre disponibilité et pour l'intérêt assidu que vous avez manifesté pour mon travail. Je me souviens encore du premier cours que j'ai eu avec vous à l'UPMC, le soir même je vous ai *pubmedé* et *googlé* et j'ai espéré pouvoir travailler un jour dans votre laboratoire.

Je tiens à exprimer ma reconnaissance à Michelle et Olivier qui m'ont soutenue dans mes différentes demandes de bourses pour que je puisse rester au laboratoire pour ma thèse ainsi que pour la quatrième année.

Merci à Mme. Anne-Marie Lachagès pour tout le travail fourni et dont les fruits sont visibles dans ce manuscrit, merci pour tout Anna-Maria! Merci d'être toujours à l'écoute, de savoir toujours trouver les mots qui réconfortent et qui motivent, merci d'avoir toujours rit à mes blagues pourries! Et répondre à mes questions bizarres, je continuerai à t'en envoyer par mail.

Merci à Mme. Sandra Carignon (Sandrou) quoi te dire... merci d'avoir toujours été là pour moi, pour ton soutien, je savais que je pouvais toujours compter sur toi. Sandrouuu je veux des JEFF !!

Merci à Mme. Anne Letessier, merci pour toutes nos discussions, nos fou-rires, et les moments agréables que nous avons partagés ensemble.

Merci à Mlle. Marion Blin, merci de m'avoir supporté dans mes moments les plus extrêmes ;) c'est toi la prochaine. Bon courage pour la dernière ligne droite.

Merci à Benoit Le Tallec, je penserai à toi à chaque fois que j'écouterai *let it be* et *elle est d'ailleurs* et plein d'autres. Et non j'ai plus besoin de la hotte !!

Merci à Stéphane Koundrioukoff, que j'ai connu comme professeur avant de le connaître comme collègue de travail.

Merci à Mme. Marie-France Lavigne. Merci pour tout ce que tu as fait pour moi, merci d'avoir été là pendant les moments difficiles pour me faire sourire comme tu sais très bien le faire.

Merci à toute l'unité UMR3244, travailler parmi vous c'était un plaisir. Merci Karina, Marina, Anne G, Sara, Eléonore, Julien, Rachida, Mylène, Win-Yan, Nicolas, Raphaëlle...

Merci à Thérèse Wilhelm (Rèse), tu sais bien que tu étais beaucoup plus qu'une simple collègue. Merci pour tous ces moments partagés au laboratoire et en dehors du laboratoire, pour ton amitié, tes conseils, ton soutien... inutile de te dire à quel point tu m'as manquée.

Merci à Hervé Técher (Jijou), c'était un plaisir de partager le bureau avec toi, merci pour tous les moments agréables et les discussions scientifiques et non scientifiques que nous avons eu.. Shou !! 😊

Merci à Pr. Claude Thermes et M. Chunlog Chen ainsi que toute l'équipe pour notre collaboration.

Merci à Dr. Bernard Dutrillaux de m'avoir consacré un peu de son temps et de son expertise pour analyser mes expériences.

Merci aux membres de la plateforme de cytométrie, merci Zosia, Annick et Sophie.

Merci Papa, Merci Maman, Merci Ahlam, Aline, Amal, mon très cher petit frère Milad. Merci à mes neveux et nièces, à mes deux petits Christina et Nicolas qui ont toujours su me faire rire dans les moments les plus durs. Merci à mon beau-frère Elie. Merci pour votre amour et votre présence.

Merci à Yasmine Rattel, un grand merci pour ton amitié, ton soutien et ta présence. Merci à Elie Abdo (Lalous) et Paul Sawan (Boulos). On est riche que de ses amis 😊

Merci à Denise Abboud (Dandoun) loin de mes yeux mais proche de mon cœur, merci d'être toi-même. Une source d'enthousiasme et d'énergie. J'attends de voir ton manuscrit bientôt.

Merci à Sarah Ayad avec qui j'ai partagé les meilleurs moments parisiens. A bientôt dans le pays des cèdres.

Merci à Gérard Chalhoub, merci de faire partie de ma vie.

# Résumé

---

Les Sites Fragiles Communs (SFC) sont des loci incluant des grands gènes où des cassures chromosomiques apparaissent chez des cellules soumises à un stress réplicatif. Il est couramment admis que ces sites sont des régions dont la réplication n'est pas terminée lorsque les cellules entrent en mitose mais le mécanisme sous-jacent restait mal compris. Notre laboratoire a montré par la technique du peignage moléculaire que l'instabilité du site *FRA3B* chez les lymphocytes est due à la présence d'une grande région pauvre en origines de réplication actives (cœur) dont la réplication ne peut pas se terminer en cas de stress réplicatif. Dans le but de généraliser ces conclusions, j'ai étudié par la technique de « Répli-Seq », le timing de réplication à l'échelle du génome entier chez des cellules lymphoblastoïdes humaines traitées ou non à l'aphidicoline. Les résultats confirment clairement que les SFC connus dans ce type cellulaire correspondent à des régions incluant un grand gène et qui sont déficitaires en réplication sous un stress réplicatif.

D'autre part, j'ai montré que retarder l'entrée des cellules en mitose par l'utilisation du RO-3306, un inhibiteur de CDK1, abolit complètement la fragilité des SFC induite par l'aphidicoline chez des lymphocytes et des fibroblastes humains. De plus, j'ai montré que cette suppression des cassures au niveau de *FRA3B* et *FRA16D* chez les lymphocytes et de *FRA1L* chez les fibroblastes ne s'accompagne pas d'une diminution de la transcription de *FHIT*, *WWOX* et *NEGR1*, les trois grands gènes contenus respectivement dans ces trois SFC. Par ailleurs, j'ai corrélé l'expression des gènes *FHIT*, *WWOX*, *NEGR1* et *LSAMP* à la fragilité des sites *FRA3B*, *FRA16D*, *FRA1L* et *FRA3L* dans 7 lignées de fibroblastes et 7 lignées de lymphocytes. J'ai aussi mis en évidence une expression atypique de ces gènes dans certaines de ces lignées et montré que cette expression atypique s'accompagne d'une fragilité atypique du SFC correspondant. Cependant, j'ai montré que l'inhibition de la transcription par deux inhibiteurs différents, le triptolide et l' $\alpha$ -amanitine, ne diminue pas la fragilité induite par l'aphidicoline. Je propose un modèle pour expliquer ces résultats apparemment contradictoires.

Enfin, j'ai corrélé la disparition des cassures au niveau de *FRA3B* dans les cellules traitées à l'aphidicoline et au RO-3306 à une augmentation des événements d'initiation dans la région cœur. J'ai montré que l'inhibition de CDK1 par le RO-3306 permet l'accumulation à la chromatine des facteurs CDC6, CDT1, ORC1 et MCM7 dans les cellules bloquées en phase G2 et que l'accumulation de CDT1 et de CDC6 est indispensable à la diminution de la fragilité sous RO-3306. Je propose un modèle permettant de rendre compte de ces observations qui postule l'existence d'origines de réplication latentes partiellement chargées en complexes de pré-réplication.

# Liste des abréviations

---

**2AP** : 2-aminopurine  
**53BP1** : tumor protein p53 binding protein 1  
**ACS** : ARS consensus sequence  
**ADN** : acide désoxyribonucléique  
**AMPD2** : adénosine monophosphate déaminase 2  
**Aph** : aphidicoline  
**ARN** : acide ribonucléique  
**ARNm** : acide ribonucléique messenger  
**ARO** : aphidicoline + RO-3306  
**ARS** : autonomously replicating sequence  
**ATM** : ataxia telangiectasia mutated  
**ATR** : ataxia telangiectasia and Rad3 related  
**BAC** : bacterial artificial chromosome  
**BLM** : Bloom syndrome, RecQ helicase-like  
**BRCA1** : breast cancer 1  
**BrdU** : bromo-déoxyuridine  
**BSA** : bovin serum albumin  
**C-FOS** : c-FBJ murine osteosarcoma viral oncogene homolog  
**C-JUN** : c- jun proto-oncogene  
**CDC** : cell division cycle  
**CDK** : cyclin-dependent kinases  
**CDT1** : cdc10 dependent transcript  
**ChIP** : chromatin immunoprecipitation  
**CHK1** : checkpoint kinase 1  
**CldU** : chlorodéoxyuridine  
**CMG** : cdc45-MCM2-7-GINS  
**CMYC** : v-myc avian myelocytomatosis viral oncogene homolog  
**CNTNAP2** : contactin associated protein-like 2  
**Cot-1 DNA** : placental DNA enriched with repetitive sequences of length 50bp-100bp  
**CpG** : cytosine-phosphate-guanine  
**Ctf4** : chromosome transmission fidelity factor 4  
**DAPI** : 4',6'-diamidino-2-phénylindole  
**Dbf4** : dumb bell former 4  
**DDK** : Dbf4-dependent kinase  
**DHFR** : dihydrofolate réductase  
**DIO** : distance inter-origines  
**DMD** : dystrophin  
**DNA-PK** : DNA-dependent protein kinase  
**DRB** : 5,6-dichloro-1-β-d-ribofuranosylbenzimidazole  
**DUE-B** : DNA unwinding element binding  
**EDTA** : éthylène diamine tétra-acétique  
**EME1** : essential meiotic structure-specific endonuclease 1

**ENCODE** : the encyclopedia of DNA elements  
**ERCC1** : excision repair cross-complementation group 1  
**ERFs** : early replication fragile sites  
**EU** : 5-éthynyl uridine  
**EXCO4** : exocyst complex component 4  
**FACS** : fluorescence activated cell sorting  
**FANCD2** : Fanconi anemia, complementation group D2  
**FANCI** : Fanconi anemia, complementation group I  
**FHIT** : fragile histidine triad  
**FISH** : fluorescent *in situ* hybridization  
**FPKM** : fragments per kilobase per million reads  
**G1** : Gap 1  
**G2** : Gap 2  
**G4** : G-quadruplexe  
**GEMC1** : geminin coiled-coil domain containing  
**GINS** : acronyme pour Go-Ichi-Ni-San - 5, 1, 2,3 en japonais - tirés du nom des 4 protéines du complexe : Sld5, Psf1, Psf2, Psf3  
**HBO1** : human acetylase binding to ORC1  
**HPV** : human papillomavirus  
**IdU** : iodo-déoxyuridine  
**IgG** : immunoglobuline G  
**IMMP2L** : IMP2 inner mitochondrial membrane peptidase-like  
**kb** : kilobase  
**Ki** : inhibitor constant  
**LINE** : long-interspersed repetitive elements  
**LSAMP** : limbic system-associated membrane protein  
**M** : mitose  
**MACROD2** : MACRO domain containing 2  
**MAR** : matrix attachment region  
**Mb** : mégabase  
**MCM** : mini chromosome maintenance  
**MLL1** : myeloid/lymphoid leukemia  
**NBEA** : neurobeachin  
**NEGR1** : neuronal growth regulator 1  
**OGRE** : origin G-rich repeated element  
**ORC** : origin recognition complex  
**ORCA** : origin recognition complex associated  
**PARK2** : parkin RBR E3 ubiquitin protein ligase  
**PBS** : phosphate buffer saline  
**PCNA** : proliferating cell nuclear antigen  
**PCR** : polymerase chain reaction  
**pH** : potentiel hydrogène  
**PICH** : PLK1 interacting checkpoint helicase  
**PIP** : PCNA-interacting peptide  
**PMSF** : phényl-méthyl-sulfonyl-fluoride  
**Pré-IC** : pre-initiation complex



**Pré-RC** : pre-replication complex  
**qRT-PCR** : quantitative reverse transcription PCR  
**Rb** : retinoblastoma  
**RecQ4** : RecQ protein-like 4  
**RIF1** : replication timing regulatory factor 1  
**RMI**: RecQ mediated genome instability  
**RNA Pol II** : RNA polymérase II  
**Rpd3** : reduced potassium deficiency 3  
**S** : synthesis  
**SAR** : scaffold attachment regions  
**SFC** : site fragile commun  
**siRNA** : silencing RNA  
**Sld** : synthetic lethal with dpb11-1  
**SSC** : standard saline citrate  
**SSS** : small single strand  
**Swi** : mating type switching  
**TDP** : timing decision point  
**Ticrr** : TOPBP1-interacting checkpoint and replication regulator  
**TOP1** : topoisomerase (DNA) I  
**TopBP1** : topoisomerase (DNA) II binding protein 1  
**TSA** : trichostatine A  
**WWOX** : WW domain containing oxidoreductase  
**λ-SNS** : λ resistant short nascent strands

# Liste des figures

---

	<u>Titre</u>	<u>Page</u>
<b>Figure 1</b>	Modèle de stress répliatif induit par l'activation d'un oncogène	20
<b>Figure 2</b>	Modèle expliquant le mécanisme d'instabilité des sites fragiles communs	22
<b>Figure 3</b>	Représentation schématique de la dynamique de réplication de la région cœur de FRA3B chez les lymphocytes et les fibroblastes.	24
<b>Figure 4</b>	Modèle expliquant le mécanisme d'instabilité des sites fragiles communs	27
<b>Figure 5</b>	Niveaux d'expression des grands gènes (> 300 kb) dans les cellules HCT116.	28
<b>Figure 6</b>	Modèle représentant les mécanismes moléculaires de prise en charge d'un SFC incomplètement répliqué	31
<b>Figure 7</b>	Chargement et déclenchement des origines de réplication.	38
<b>Figure 8</b>	Cartographie des événements d'initiation et de terminaison au niveau du gène FHIT par la technique du peignage moléculaire.	43
<b>Figure 9</b>	Représentation schématique de la régulation des origines latentes en réponse à un stress répliatif.	52
<b>Figure 10</b>	Principe de la méthode du Repli-Seq.	67
<b>Figure 11</b>	Profils de réplication dans la région du gène FHIT obtenus par Repli-Seq dans des cellules lymphoblastoïdes humaines.	69
<b>Figure 12</b>	Profils de réplication obtenus par Repli-Seq dans les régions des gènes WWOX, NBEA et EXOC4 chez les cellules JEFF	70
<b>Figure 13</b>	Action du RO-3306 sur le cycle cellulaire.	71
<b>Figure 14</b>	Effet du RO-3306 et de l'aphidicoline sur le cycle des cellules JEFF	72
<b>Figure 15</b>	Cinétique d'accumulation en G2 des cellules JEFF après traitement au RO-3306	73
<b>Figure 16</b>	Cassures induites par l'aphidicoline au niveau des SFC.	74
<b>Figure 17</b>	Effet du RO-3306 sur les cassures induites par l'aphidicoline	75
<b>Figure 18</b>	Cinétique d'action du RO-3306 sur les cassures induites par l'aphidicoline	77
<b>Figure 19</b>	Niveaux d'expression du gène FHIT dans plusieurs lignées de lymphocytes et de fibroblastes humains.	80
<b>Figure 20</b>	Niveaux d'expression du gène WWOX dans plusieurs lignées de lymphocytes et de fibroblastes humains.	81
<b>Figure 21</b>	Niveaux d'expression de NEGR1 dans plusieurs lignées de lymphocytes et de fibroblastes	82
<b>Figure 22</b>	Niveaux d'expression de LSAMP dans plusieurs lignées de lymphocytes et de fibroblastes	83
<b>Figure 23</b>	Fragilité des SFC dans les cellules présentant une expression atypique du grand gène associé	85

<b>Figure 24</b>	Analyse de l'unité de transcription incluse dans le gène FHIT chez les cellules MRC5SV2.	86
<b>Figure 25</b>	Expression du gène FHIT dans les cellules lymphoblastoïdes JEFF traitées au RO-3306	88
<b>Figure 26</b>	Expression du gène WWOX dans les cellules lymphoblastoïdes JEFF traitées au RO-3306	89
<b>Figure 27</b>	Expression du gène NEGR1 dans les fibroblastes MRC5 traitées au RO-3306	90
<b>Figure 28</b>	Expression du gène LSAMP dans les fibroblastes MRC5 traitées au RO-3306	91
<b>Figure 29</b>	Effet de l' $\alpha$ -amanitine sur la transcription et sur les cassures chromosomiques induites par l'aphidicoline	92
<b>Figure 30</b>	Effet du triptolide sur la transcription et sur les cassures chromosomiques induites par l'aphidicoline	93
<b>Figure 31</b>	Effet du DRB sur la transcription et sur les cassures chromosomiques induites par l'aphidicoline	95
<b>Figure 32</b>	Effet du RO-3306 sur la vitesse de progression des fourches de réplication et sur les distances inter-origines.	97
<b>Figure 33</b>	Cartographie des initiations et des terminaisons dans la région du gène FHIT chez les cellules lymphoblastoïdes JEFF.	99
<b>Figure 34</b>	Effet du RO-3306 sur les cassures induites au cycle suivant	100
<b>Figure 35</b>	Accumulation de protéines du complexe de pré-réplication dans les cellules traitées à l'aphidicoline et/ou au RO-3306	104
<b>Figure 36</b>	Effet de la carence en facteurs impliqués dans la formation des complexes de pré-RC sur la prévention de cassures chromosomiques par le RO-3306.	106
<b>Figure 37</b>	Effet de la carence en facteur RIF1 sur l'instabilité des SFC	116
<b>Figure 38</b>	Effet de la trichostatine A sur les cassures chromosomiques	119
<b>Figure 39</b>	Carte du gène LSAMP humain.	121
<b>Figure 40</b>	Expression des deux formes du gène LSAMP dans des lymphocytes et des fibroblastes humains.	122

# Sommaire

---

<b>Introduction</b>	<b>p13</b>
<b>I. <u>Les sites fragiles communs</u></b>	<b>p13</b>
1. Définition et généralités	p13
2. Les SFC sont conservés au cours de l'évolution	p13
3. Les SFC ont une réplication tardive	p14
4. Les SFC sont spécifiques de tissu	p15
5. Les SFC referment des grands gènes	p16
6. Les SFC sont impliqués dans le cancer et d'autres maladies	p17
7. Trois modèles pour expliquer le mécanisme de l'instabilité des SFC	p21
7.1. <u>Tout est dans la séquence</u>	p22
7.2. <u>Complétion tardive de la réplication</u>	p23
7.3. <u>Réplication versus transcription</u>	p25
8. <b>Prise en charge des SFC et réparation</b>	<b>p28</b>
<b>II. <u>La réplication de l'ADN</u></b>	<b>p32</b>
1. <b>Le modèle du réplicon</b>	<b>p32</b>
2. <b>Les origines de réplication chez les eucaryotes</b>	<b>p32</b>
3. <b>Mécanismes de déclenchement des origines de réplication chez les eucaryotes</b>	<b>p34</b>
3.1. <u>Chargement des origines par les complexes de pré-réplication</u>	<b>p34</b>
3.2. <u>Démarrage de la réplication</u>	<b>p35</b>
3.2.1. Déclenchement des origines chez <i>S. cerevisiae</i>	<b>p36</b>
3.2.2. Déclenchement des origines chez les eucaryotes supérieurs	<b>p36</b>
3.3. <u>Inhibition de la re-réplication</u>	<b>p37</b>
3.3.1. Chez <i>S. cerevisiae</i>	<b>p38</b>
3.3.2. Chez les eucaryotes supérieurs	<b>p39</b>
4. <b>Déterminants génétiques et/ou épigénétiques des origines de réplication</b>	<b>p40</b>
4.1. <u>Cartographies des origines de réplication dans les cellules humaines</u>	<b>p40</b>
4.2. <u>Déterminants génétiques</u>	<b>p43</b>
4.3. <u>Marques épigénétiques</u>	<b>p45</b>
4.4. <u>Transcription et démarrage de la réplication</u>	<b>p46</b>
4.5. <u>Association des origines actives avec la matrice nucléaire</u>	<b>p47</b>
5. <b>Programme temporel de réplication</b>	<b>p48</b>
6. <b>Les origines latentes</b>	<b>p50</b>

<b><u>Matériels et Méthodes</u></b>	<b>p53</b>
I. Culture cellulaire	p53
II. Cytométrie en flux	p53
III. Etalement des métaphases	p54
IV. Hybridation <i>in situ</i> en fluorescence sur métaphases	p54
V. Marquage des fourches de réplication et préparation des blocs d'ADN pour le peignage	p56
VI. Immunodétection des signaux de réplication et mesure de la vitesse de progression des fourches de réplication	p57
VII. Hybridation <i>in situ</i> en fluorescence sur ADN peigné et immunodétection des signaux de réplication	p57
VIII. Transfection des cellules JEFF	p59
IX. Immunoprécipitation de la chromatine	p59
X. La technique de Repli-Seq	p60
XI. Transfert Western	p62
XII. Capture des ARN naissants	p63
<b><u>Résultats</u></b>	<b>p65</b>
I. <u>Programme temporel de la réplication sous aphidicoline</u>	p65
II. <u>Effet du RO-3306 sur l'instabilité des SFC</u>	p70
1. Effet du RO-3306 sur le cycle cellulaire des lymphocytes et des fibroblastes	p70
2. Effet du RO-3306 sur l'instabilité des SFC chez les lymphocytes et les fibroblastes	p73
3. Cinétique d'action du RO-3306 sur le taux de cassures au niveau des SFC	p76
III. <u>Relation entre transcription des grands gènes et fragilité des SFC</u>	p78
1. Corrélation entre fragilité des SFC et transcription des grands gènes	p78
2. Effet du RO-3306 sur la transcription	p86
3. Inhibition de la transcription et effet sur la fragilité des SFC	p91
IV. <u>Effet du RO-3306 sur la dynamique de réplication</u>	p95
1. La technique du peignage moléculaire.	p96
2. Effet du RO-3306 sur la dynamique de réplication au niveau pan-génomique	p96
3. Effet du RO-3306 sur la dynamique de réplication au niveau de <i>FRA3B</i>	p97
V. <u>Effet du RO-3306 sur le programme temporel de réplication des SFC</u>	p101
VI. <u>Mécanisme d'action du RO-3306 dans le déclenchement des origines latentes du cœur</u>	p101

1. Effet du RO-3306 sur les protéines du complexe de pré-réplication	p103
2. Effet du RO-3306 sur la dynamique de réplication au niveau de <i>FRA3B</i> dans des cellules appauvries en protéines du complexe de pré-réplication	p105
<u>Discussion et perspectives</u>	p107
I. <u>Cartographie des SFC par la technique de Repli-Seq</u>	p107
II. <u>Lien entre fragilité et transcription</u>	p110
III. <u>Compléter le chargement d'origines de réplication pour finir la réplication</u>	p112
<u>Annexes</u>	p115
I. <u>Effet d'une invalidation de RIF1 sur les cassures chromosomiques induites par l'aphidicoline</u>	p115
II. <u>Effet de la trichostatine A sur les cassures chromosomiques</u>	p118
III. <u>Les deux formes du gène <i>LSAMP</i> humain</u>	p120
IV. <u>Article « Replication Dynamics : Biases and Robustness of DNA Fiber Analysis »</u>	p123
<u>Références</u>	p135

# Introduction

---

## I. Les sites fragiles communs

### **1. Définition et généralités**

Les sites fragiles communs (SFC) sont des loci présentant des cassures récurrentes visibles au niveau des chromosomes métaphasiques dans des conditions où la réplication de l'ADN est perturbée (stress réplicatif). Ils sont présents chez tous les individus, d'où leur nom. L'aphidicoline, un inhibiteur des ADN polymérase répliquatives ( $\alpha$ ,  $\delta$  et  $\epsilon$ ), est l'agent le plus largement utilisé pour induire des cassures au niveau des SFC. Les SFC se caractérisent par une réplication tardive au cours du cycle cellulaire. Ils sont conservés chez plusieurs espèces ce qui suggère un rôle biologique. Il a été montré que ces sites constituent des points préférentiels de recombinaison homologue, d'échanges entre chromatides sœurs, d'intégration d'ADN exogène ainsi que de réarrangements chromosomiques dans les cancers tels que les délétions, les amplifications et les translocations. Ces sites ont été, pour des raisons historiques, d'abord caractérisés par des cytogénéticiens dans les lymphocytes humains. Actuellement, chez l'homme, les SFC sont cartographiés chez les lymphocytes, les fibroblastes, les cellules épithéliales du colon et du sein et dans les cellules érythroïdes. Les différentes cartes des SFC montrent qu'ils sont spécifiques de tissu et ne présentent pas tous la même fréquence de cassures. On distingue ainsi des SFC majeurs présentant une fréquence de cassures élevée et d'autres, mineurs, qui cassent moins fréquemment. Les SFC sont associés de façon significative à des grands gènes (>300 kb). Parmi ceux-ci on trouve des gènes oncosuppresseurs. C'est le cas de *FRA3B* et *FRA16D*, les deux SFC majeurs dans les lymphocytes qui contiennent respectivement les deux grands gènes *FHIT* (1,5 Mb) et *WWOX* (1,1 Mb). Trois principaux modèles ont été proposés pour expliquer le mécanisme responsable de la fragilité des SFC.

### **2. Les SFC sont conservés au cours de l'évolution**

Les SFC sont conservés au cours de l'évolution chez les mammifères. Des orthologues des SFC humains ont été trouvés chez d'autres primates (Durkin and Glover, 2007), ainsi que

chez le chat, le chien, le porc, le cheval, le bœuf, le rat et la souris (Elder and Robinson, 1989). Cette dernière espèce a été la plus étudiée en dehors de l'homme. En effet, huit SFC ont été identifiés chez les lymphocytes de souris et sont des orthologues de sites humains : *Fra14A2 (FRA3B)* (Glover et al., 1998), *Fra8E1 (FRA16D)* (Krummel et al., 2002), *Fra6C1 (FRA4F)* (Rozier et al., 2004), *Fra12C1 (FRA7K)*, *Fra2D (FRA2G)*, *Fra6A3 (FRA7G)*, *Fra6B1 (FRA7H)*, *Fra4C2 (FRA9E)* (Helmrich et al., 2006). Cette conservation concerne aussi les grands gènes hébergés par les SFC. Il semble probable que la conservation des SFC soit la conséquence de la conservation des grands gènes qu'ils renferment. C'est le cas des gènes *Fhit* et *Wwox* murins. Par ailleurs, il a été montré que, chez *S. cerevisiae*, des cassures bicaténaires se produisent de façon récurrente au niveau de régions spécifiques qui ont été appelé *replication slow zones*, et certains auteurs ont suggéré que ces zones soient des analogues des SFC. La conservation des SFC suggère un rôle biologique et une fonction commune à ces sites. Il a d'abord été suggéré que ces régions étaient conservées pour favoriser la recombinaison homologue (Glover and Stein, 1987). Plus récemment, il a été proposé par notre laboratoire que la réplication des SFC constitue un signal qui marque la fin de la phase S et le passage en mitose (Debatisse et al., 2006). En effet, les SFC sont les dernières régions à être répliquées dans le génome. Ce point sera développé dans le prochain paragraphe.

### **3. Les SFC ont une réplication tardive**

En 1998, Le Beau et coll. ont proposé pour la première fois un lien entre réplication tardive et fragilité. En étudiant le programme temporel (timing ; voir II-5) de réplication de *FRA3B* dans des lymphocytes humains par la technique de FISH, ces auteurs ont montré que ce site est répliqué tardivement pendant la phase S et que sa réplication est encore retardée sous aphidicoline. Ils ont trouvé que 16,5% des sites *FRA3B* n'ont pas achevé leur réplication dans les cellules en phase G2 en présence d'aphidicoline (Le Beau et al., 1998). Un an plus tard, Wang et coll. ont montré que les deux allèles de *FHIT* dans les lymphocytes humains n'ont pas le même timing de réplication et que l'allèle le plus tardif est l'allèle le plus fragile (Wang et al., 1999). De même, il a été montré que le timing de réplication de *FRA7H* est spécifique pour chaque allèle et qu'un traitement à l'aphidicoline retarde la complétion de la réplication de cette région par rapport à des régions de référence (Hellman et al., 2000). Plus tard, l'étude du timing de réplication de *WWOX* par PCR semi-quantitative sur des cellules



triées selon différentes phases du cycle cellulaire (Palakodeti et al., 2004) a montré que *FRA16D* présente aussi une réplication tardive dans des lymphocytes humains. Une étude réalisée dans notre laboratoire a généralisé le caractère tardif de la réplication des SFC (El Achkar et al., 2005). Dans cette étude, il a été montré que la condensation précoce de l'ADN dans des cellules en phase G2, induite par un traitement à la calyculine A, révèle des cassures au niveau des SFC. La calyculine A est un inhibiteur spécifique des phosphatases à sérine-thréonine qui induit la condensation prématurée des chromosomes à tous les stades du cycle. Plus récemment, des données obtenues par la technique de Repli-Seq, qui permet l'étude pan-génomique du timing de réplication, ont mis en évidence le timing de réplication tardif des SFC d'une façon globale (Hansen et al., 2010). Les cassures au niveau des SFC ont donc été attribuées à une entrée des cellules en mitose alors que la réplication de ces sites n'est pas achevée.

Dans une étude récente, Barlow et coll. (Barlow et al., 2013), ont décrit une nouvelle catégorie de sites fragiles, les *Early Replicating Fragile Sites* (ERFS) qui contrairement aux SFC se répliquent précocement. Ces sites ont été identifiés par immunoprécipitation de la chromatine comme des régions enrichies en protéine RPA chez des lymphocytes B murins bloqués en début de S par des doses élevées d'hydroxyurée. La protéine RPA recouvre l'ADN monocaténaire au niveau des fourches de réplication. Ces ERFs présentent des cassures au niveau des chromosomes métaphasiques après traitement des cellules par l'hydroxyurée suivi d'une reprise du cycle cellulaire consécutif au retrait de la drogue. Ces cassures ne sont pas localisées au niveau des SFC. Les ERFS sont préférentiellement localisés dans des régions riches en gènes fortement transcrits. Enfin, les auteurs rapportent que plus de 50% des amplifications et délétions identifiées chez des patients ayant un lymphome diffu à grandes cellules B par Lenz et coll. (Lenz et al., 2008) se produisent au niveau des ERFS. Dans l'étude de Barlow et coll. l'instabilité des ERFS a été attribuée à des conflits entre la machinerie de transcription et les fourches de réplication au niveau de ces sites.

#### **4. Les SFC sont spécifiques de tissu**

Pendant longtemps il a été implicitement admis que les SFC caractérisés chez les lymphocytes humains étaient universels, c'est-à-dire présents chez tous les types cellulaires. Récemment notre laboratoire a montré que *FRA3B* et *FRA16D*, les deux sites majeurs des lymphocytes sont peu ou pas fragiles dans les fibroblastes. Ceci suggérait que les SFC étaient

spécifiques de tissu et que leur position dépendait de caractères épigénétiques (Letessier et al., 2011). Des analyses plus exhaustives portant sur une lignée de fibroblastes humains (MRC5) (Le Tallec et al., 2011), trois lignées de cellules épithéliales de cancer du colon (HCT116, LoVo, LS174T), deux lignées de cellules épithéliales normales (MCF10A) ou tumorales de sein (CAL51) et une lignée de cellules érythroïdes (K562) (Le Tallec et al., 2013) ont été réalisées. Elles ont montré que deux types cellulaires différents partagent moins de 20% de leurs SFC et que, même si un SFC est partagé entre deux types cellulaires, sa fréquence de cassures n'est pas la même entre les deux types. Néanmoins, ces études ont aussi montré que l'ensemble des SFC caractérisés correspond seulement à environ 50 loci différents susceptibles de devenir des SFC cassant avec une fréquence supérieure à 1% dans un de ces différents types cellulaires. Ces résultats indiquent qu'il existe un ensemble de régions chromosomiques qui peuvent devenir des SFC dans un type cellulaire donné selon le contexte épigénétique. Ceci soulève la question de l'existence d'une (ou plusieurs) caractéristique(s) commune(s) à ces loci qui les prédisposent à devenir fragiles. Ces résultats soulèvent aussi la question de l'existence et de la localisation de SFC dans les cellules souches embryonnaires ainsi que de leur devenir au cours de la différenciation cellulaire. Cette intéressante question n'a pas été documentée à ce jour.

## **5. Les SFC renferment des grands gènes**

Il avait été remarqué que les SFC étaient fréquemment associés à des grands gènes couvrant plusieurs centaines de kb (Helmrich et al., 2006). Cette observation a été confirmée et généralisée dans notre laboratoire (Le Tallec et al., 2013). En effet, il a été montré que 80 à 100% des SFC humains et 100% des SFC murins hébergent un grand gène (>300 kb). Sachant que la taille moyenne des gènes humains est de 20 kb environ, les gènes de plus de 300 kb ont donc une taille plus de 15 fois supérieure à la moyenne. Ces grands gènes ne représentent que 3% du génome humain. Etant donné que la carte des gènes annotés dans le génome humain varie constamment du fait de la découverte de nouvelles unités de transcription, il est tout à fait possible qu'une partie des SFC (voire tous) non encore associés à des grands gènes ne le devienne ultérieurement. Une question essentielle reste pour l'instant sans réponse : quelle est la base de cette association?

## 6. Les SFC sont impliqués dans le cancer et d'autres maladies

L'instabilité génétique constitue une caractéristique bien établie des cellules tumorales. Néanmoins, nous pouvons distinguer plusieurs catégories d'instabilité génétique chez ces cellules. La forme la plus répandue dans les cancers humains est l'instabilité chromosomique. Elle se manifeste par des délétions ou des amplifications de très grande taille (de quelques centaines de kb à plusieurs Mb) initiées par des cassures bicaténares et est très majoritaire dans les cancers sporadiques. D'autres formes d'instabilité se traduisent par l'accumulation de mutations ponctuelles ou de microdélétions et de microduplications. La forme la mieux documentée de cette dernière est l'instabilité des microsatellites. Ces types d'instabilité sont fréquents parmi les cancers héréditaires. Ils sont dus à la transmission génétique d'une copie défectueuse d'un des gènes impliqués dans la réparation de l'ADN par excision de bases ou de nucléotides ou dans la réparation des mésappariements de bases.

De nombreuses études montrent que les SFC sont des régions préférentielles de réarrangements chromosomiques dans toute une panoplie de cancers: cancers de l'œsophage (Fang et al., 2000), cancers gastriques (Lee et al., 2001), cancers des voies aérodigestives supérieures (Virgilio et al., 1996), cancers des poumons (Sozzi et al., 1996), lymphomes (Kameoka et al., 2004), cancers du sein (Negrini et al., 1996), cancers pancréatiques (Shridhar et al., 1996), cancers rénaux (Ohta et al., 1996). Ces réarrangements correspondent majoritairement à des délétions et entraînent souvent l'inactivation du gène qui se trouve au sein des SFC. De plus, notre laboratoire a montré que les SFC sont des sites préférentiels de cassures au cours de l'amplification chromosomique par des cycles de cassure-fusion-pont (Coquelle et al., 1997). Par exemple, l'oncogène *CMYC* humain est localisé entre deux sites fragiles (*FRA8C* et *FRA8D*) et est amplifié dans de nombreux de cancers (Ferber et al., 2004; Ferber et al., 2003). Les SFC sont aussi des sites préférentiels d'intégration de l'ADN de virus qui sont impliqués dans le développement de certains cancers. C'est le cas du papillomavirus HPV16 dont on trouve l'ADN intégré au niveau de SFC dans plus de 50% des cas de cancers du col de l'utérus (Thorland et al., 2003).

Une relation supplémentaire entre SFC et cancer vient du fait qu'il a été rapporté que plusieurs grands gènes inclus dans des SFC avaient des propriétés de gènes oncosuppresseurs (Drusco et al., 2011). C'est notamment le cas des gènes *FHIT* et *WWOX* hébergés par les sites *FRA3B* et *FRA16D*. Ces deux grands gènes ont été les plus étudiés et

leur caractère oncosuppresseur a été discuté par certains auteurs (Bignell et al., 2010). Ces gènes sont effectivement fréquemment altérés dans plusieurs types de cancers. Cependant il n'est pas actuellement clairement établi si c'est une simple conséquence de l'instabilité des SFC ou si leur inactivation a activement participé à la progression tumorale.

Il faut noter que toutes les études portant sur l'implication des SFC dans les réarrangements chromosomiques associés aux cancers ne prenaient en compte que les SFC cartographiés dans les lymphocytes. Or nous savons aujourd'hui que d'autres SFC existent dans d'autres types cellulaires et les tumeurs les plus prévalentes chez l'homme sont des carcinomes (sein, colon, prostate) qui dérivent de cellules épithéliales. Il faudrait donc reprendre ces études en tenant compte des SFC qui sont fragiles dans le tissu d'origine de chaque type de tumeur. Ainsi, en reprenant les données de deux études récentes (Beroukhim et al., 2010; Bignell et al., 2010) portant sur les délétions récurrentes cartographiées dans les tumeurs humaines et dans des lignées cancéreuses et en incluant les SFC nouvellement caractérisés dans notre laboratoire, nous avons constaté que certaines régions préférentiellement délétées et pour lesquelles les auteurs n'avaient pas d'explication correspondaient en fait à ces nouveaux SFC (Le Tallec et al., 2013).

Dans l'étude de Bignell et coll., des régions préférentielles de délétions ont été identifiées dans 746 lignées cancéreuses (Bignell et al., 2010). Ces régions ont été classées selon qu'elles correspondent à des SFC connus au moment de l'étude, à des gènes oncosuppresseurs ou à des régions connues pour être réarrangées dans les lymphocytes afin d'assurer la diversité du système immunitaire. Cependant, 82% de ces régions restaient « inexplicables » parce qu'elles ne correspondaient à aucune de ces trois catégories. Dans cette étude, des signatures ont été attribuées à ces régions selon la proportion de délétions de grande taille, de taille réduite hémizygote ou homozygote. Les auteurs notaient que la majorité des régions de délétions « inexplicables » présentaient une signature comparable à celle des régions correspondant à des SFC.

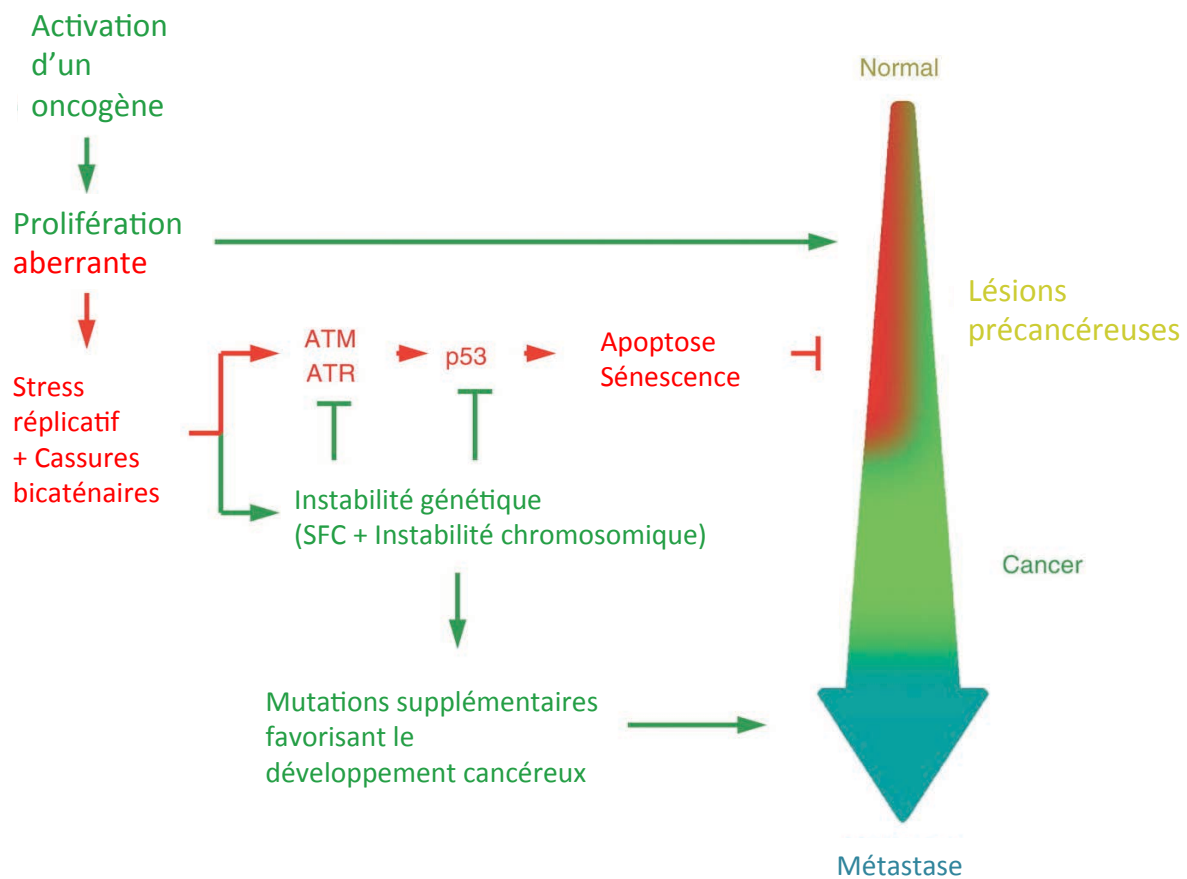
Beroukhim et coll. ont étudié en majorité des échantillons tumoraux provenant de tissus cancéreux de patients (Beroukhim et al., 2010). Ces auteurs ont identifié 82 régions de délétions localisées. Ils ont constaté que, parmi les 73 régions ne correspondant pas à un gène oncosuppresseur connu, un tiers était localisé au niveau d'un grand gène (>750 kb). En analysant ces mêmes données, d'autres auteurs ont souligné que 8 des 10 régions de

délétions les plus fréquemment trouvées correspondaient à des grands gènes et que tous ces gènes, sauf un, étaient maintenant associés à un SFC (Dereli-Oz et al., 2011).

Dans ces deux études, 70-80% des régions de délétions restaient inexplicées. La caractérisation moléculaire de 15 SFC supplémentaires dans notre laboratoire a permis d'expliquer 10% de ces délétions (Le Tallec et al., 2013). C'est le cas des deux SFC majeurs des fibroblastes *FRA1L* (*NEGR1*) et *FRA3L* (*LSAMP*) et de *FRA20B* (*MACROD2*), un SFC fragile dans les cellules épithéliales. La caractérisation moléculaire de nouveaux SFC dans un grand nombre d'autres types cellulaires permettrait certainement d'expliquer un nombre supplémentaire de délétions, mais ceci nécessiterait une énorme quantité de travail et serait très difficile à mettre en œuvre pour des cellules qu'on ne peut pas cultiver *in vitro*. Pour contourner le problème, Benoit Le Tallec dans notre laboratoire a émis l'hypothèse suivante : puisque la très grande majorité des SFC connus correspondent à des grands gènes, les régions de délétions inexplicées devraient aussi majoritairement correspondre à des grands gènes si elles correspondent à des SFC non encore cartographiés. Il a pu montrer que plus de 50% des délétions mises en évidence dans les deux études sont effectivement associées à des grands gènes.

Les délétions homozygotes qu'on trouve fréquemment au niveau des gènes oncosuppresseurs sont très majoritairement constituées d'au moins une grande délétion (>1 Mb) affectant un des deux allèles ou de deux grandes délétions inactivant les deux allèles. Dans le premier cas, le second allèle est inactivé par une délétion de petite taille (<1 Mb) ou par mutation. Ces événements ont fourni un avantage prolifératif à la cellule porteuse et ont été sélectionnés au cours de la progression tumorale. Logiquement, on trouve très peu de délétions hémizygotiques au niveau de ces gènes. Au niveau des sites fragiles connus ou supposés (présence d'un grand gène), la proportion des différents types de délétion est très différente. La présence d'au moins une grande délétion n'est trouvée que dans moins d'un tiers des cas. Vingt-cinq à 30% des cas présentent deux petites délétions homozygotes et le reste (souvent plus de 50% des cas) présente une petite délétion hémizygotique. Dans cette dernière situation, on comprend difficilement comment ceci a pu conférer un avantage sélectif à la cellule. De plus, il faut noter que le nombre de grands gènes associés à des SFC pour lesquels des propriétés oncosuppressives ont été rapportées est très faible. On est donc amené à se demander pourquoi les SFC sont fréquemment affectés par des délétions dans les cellules tumorales. Les mutations au niveau des gènes oncosuppresseurs ou des

oncogènes sont des mutations initiatrices qui offrent un avantage sélectif à la cellule. Par contre, les délétions au niveau des SFC et des grands gènes qu'ils renferment seraient plutôt des altérations fréquentes mais passives, traduisant une instabilité de ces régions dans les cellules tumorales et qui sont conservées du fait de la nature clonale des tumeurs (Dereli-Oz et al., 2011). Le modèle de stress répliatif induit par les oncogènes propose une explication à l'accumulation de ces délétions (Negrini et al., 2010). Dans ce modèle, l'activation d'un oncogène induit une prolifération incontrôlée et un stress répliatif. Les SFC étant particulièrement sensibles à ce type de stress, ils seraient plus fréquemment touchés que d'autres régions du génome. Le stress répliatif induirait donc l'instabilité génétique, une propriété commune à toutes les cellules cancéreuses (Figure 1).



**Figure 1 : Modèle de stress répliatif induit par l'activation d'un oncogène.** L'activation d'un oncogène conduit à une prolifération aberrante et à un stress répliatif endogène. Deux scénarios sont possibles: 1) le point de contrôle des dommages à l'ADN (ATR, ATM) est activé ce qui active la voie p53 et conduit à l'apoptose; 2) le stress répliatif conduit à l'instabilité génétique au niveau des SFC notamment; la cellule échappe ainsi au point de contrôle, continue à cycler et accumule des mutations qui favorisent son développement tumoral. Adapté de (Halazonetis et al., 2008).

Comment l'activation d'un oncogène peut engendrer un stress réplicatif ? Deux types de mécanismes ont été proposés. L'activation d'un oncogène conduit à une phosphorylation de la protéine Rb en absence de stimulus mitogène externe et à une activation du facteur de transcription E2F, provoquant un passage forcé des cellules en phase S. Ceci se produirait sans que la cellule n'ait eu le temps d'accumuler suffisamment de précurseurs nucléotidiques et créerait un stress réplicatif endogène. Le stress réplicatif pourrait aussi résulter d'un chargement et/ou d'un déclenchement perturbé des origines de réplication. Le chargement des origines en complexe de pré-réplication (pré-RC) a normalement lieu uniquement en phase G1 quand les Kinases Dépendantes des Cyclines (CDK) ne sont pas actives. L'activation des origines durant la phase S nécessite des CDK actives. Les oncogènes et les gènes oncosuppresseurs contrôlent l'activité des CDK et pourraient ainsi perturber le chargement et le déclenchement des origines (Macheret and Halazonetis, 2015).

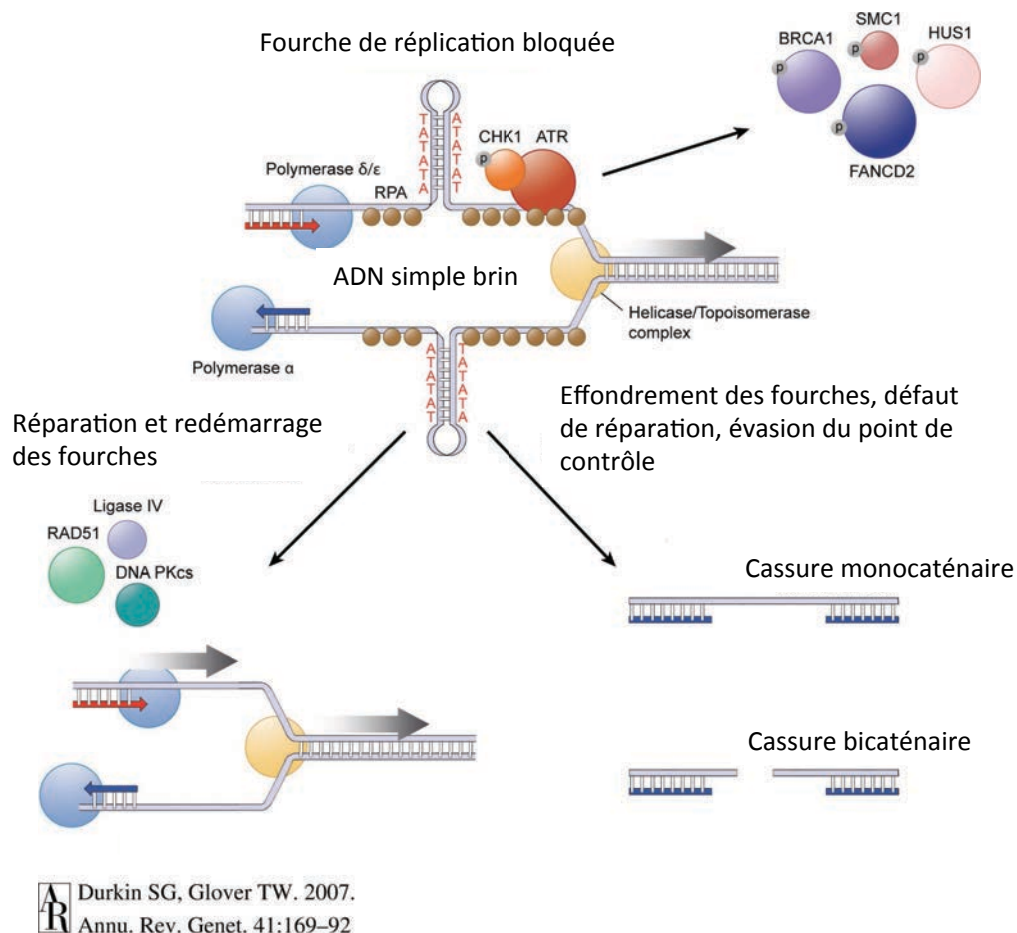
L'implication des SFC dans d'autres maladies que le cancer a aussi été avancée. C'est le cas du site *FRA6E* qui héberge le grand gène *PARK2*. Des délétions hétérozygotes et homozygotes au niveau de *PARK2* ont été mises en évidence dans la forme autosomale récessive juvénile de la maladie de Parkinson (Denison et al., 2003). Le gène *NBEA* inclus dans le site *FRA13A* a été impliqué dans certaines formes d'autisme (Savelyeva et al., 2006).

### **7. Trois modèles pour expliquer le mécanisme de l'instabilité des SFC**

Le mécanisme responsable de fragilité des SFC fait l'objet d'un intense débat dans la communauté scientifique. A ce jour, trois modèles ont été proposés. Le plus ancien attribue la fragilité des SFC à des séquences particulières riches en AT et qui auraient tendance à former des structures secondaires et à bloquer les fourches de réplication (Durkin and Glover, 2007). Dans le deuxième modèle, les SFC sont des régions dont la réplication se termine très tardivement (fin de S-G2) et qui présentent une zone très pauvre en initiations de la réplication (Letessier et al., 2011). Dans le troisième modèle, la fragilité des SFC résulte de collisions entre les fourches de réplication et les complexes de transcription qui parcourent les grands gènes hébergés (Helmrich et al., 2011).

### 7.1. Tout est dans la séquence

D'après ce modèle, la fragilité des SFC serait due aux caractéristiques intrinsèques de leurs séquences qui favoriseraient le blocage des fourches de réplication dans des conditions de stress répliatif (Durkin and Glover, 2007) (Figure2).



**Figure 2 : Modèle expliquant le mécanisme d'instabilité des sites fragiles communs.** D'après ce modèle, la fragilité des SFC résulte de structures secondaires formées après accumulation d'ADN monocaténaire au niveau d'une fourche de réplication bloquée. La protéine RPA se fixe au niveau de l'ADN monocaténaire, ce qui recrute les facteurs de la réponse aux dommages (CHK1, ATR). L'activation de ce point de contrôle active les voies de réparation (RAD51, Ligase IV, DNA PKcs). Si la réparation ne se fait pas, ceci engendrera des cassures mono- ou bicaténaires au niveau des SFC. Adapté de (Durkin and Glover, 2007).

Les auteurs qui ont proposé ce modèle se sont basés sur plusieurs observations : 1) les séquences de certains SFC sont riches en AT. C'est le cas de *FRA3B* (Boldog et al., 1997), *FRA7H* (Mishmar et al., 1998), *FRA16D* (Ried et al., 2000), *FRA3B* (Arlt et al., 2002). Ces séquences auraient la capacité de former des structures secondaires (structure en tige-boucle ou en cruciforme) qui bloqueraient la progression des fourches de réplication. 2) les mesures d'angles de torsion des séquences d'ADN au niveau de certains SFC montrent que



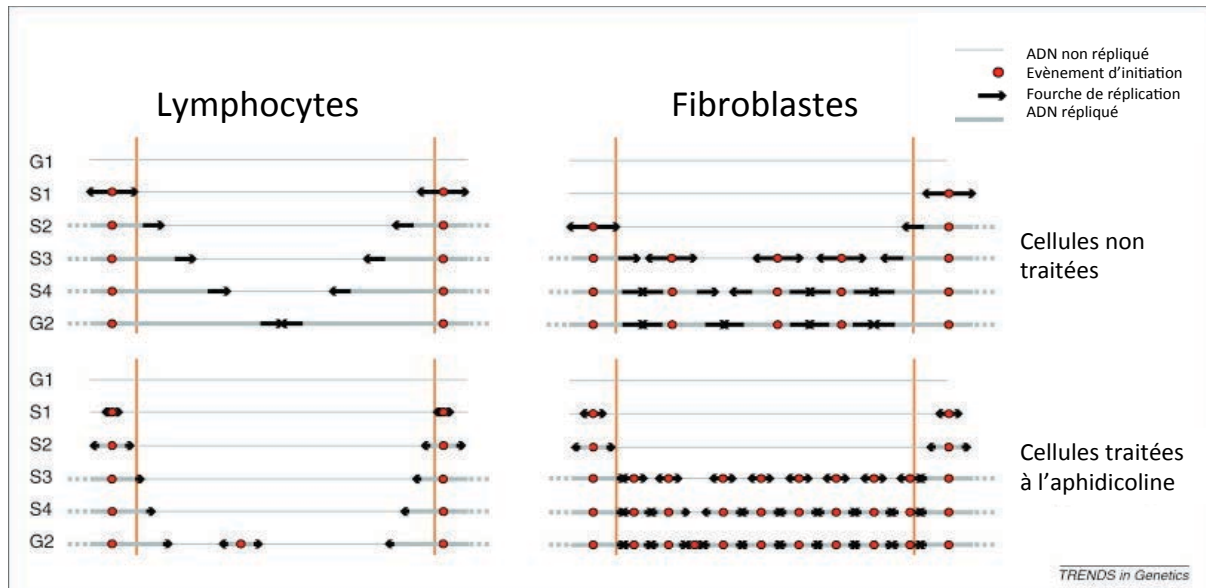
leur flexibilité est élevée par rapport à des séquences contrôles non fragiles. C'est le cas pour *FRA2G* (Limongi et al., 2003), *FRA3B* (Mimori et al., 1999), *FRAXB* (Arlt et al., 2002), *FRA7H* (Mishmar et al., 1998), *FRA8C* (Ferber et al., 2004), et *FRA16D* (Ried et al., 2000). 3) la délétion dans deux lignées tumorales d'une région de 500 kb incluant les séquences riches en AT supprime la fragilité du site *FRAXB* (Arlt et al., 2002). A l'appui à ce modèle, il a été rapporté que l'intégration ectopique d'une partie de la séquence de *FRA3B* s'accompagne d'une augmentation de la fragilité de la région d'insertion (Ragland et al., 2008). D'autres auteurs ont affirmé que le nombre de fourches bloquées au niveau des séquences riches en AT de *FRA16C* (Ozeri-Galai et al., 2011) ou de *FRA16D* (Zhang and Freudenreich, 2007) augmente sous stress réplcatif.

Cependant des arguments ont été avancés contre ce modèle. En effet, d'autres études n'ont pas pu mettre en évidence une aptitude particulière des séquences riches en AT des SFC à former des structures secondaires par rapport à celle de régions non fragiles ayant la même composition en bases (Helmrich et al., 2006). De plus, il a été montré que la délétion des séquences riches en AT à flexibilité élevée de *FRA3B* (Corbin et al., 2002) ou de *FRA16D* (Finnis et al., 2005) ne diminue pas leur fragilité. L'étude de la dynamique de réplication au niveau de *FRA3B* (Letessier et al., 2011) ou de *FRA6E* (Palumbo et al., 2010) n'a pas pu mettre en évidence un blocage des fourches significativement plus élevé au niveau de ces sites. Enfin, la spécificité de tissu des sites fragiles plaide d'une façon évidente contre le fait que la fragilité soit dictée par la séquence de l'ADN.

## 7.2. Complétion tardive de la réplication

Ce deuxième modèle a été proposé par notre laboratoire après étude de la dynamique de réplication au niveau de *FRA3B* par la technique du peignage moléculaire couplé à la technique du FISH dans des cellules lymphoblastoïdes et des fibroblastes humains (Letessier et al., 2011). *FRA3B* est le SFC majeur chez les lymphocytes mais il n'est pas fragile chez les fibroblastes. Ce site inclut le grand gène *FHIT* dans lequel se trouve la région la plus fragile du site. Les différents paramètres de la dynamique de réplication (vitesse de réplication, cartographie des événements d'initiation et de terminaison, blocage des fourches) ont été étudiés dans ces deux types cellulaires dans des conditions normales de culture et dans des conditions de stress réplcatif. Les résultats obtenus montrent que la vitesse de réplication au niveau de *FHIT* est la même qu'au niveau du reste du génome et

qu'il n'y a pas plus de fourches bloquées au niveau de *FHIT* que par rapport au reste du génome sous stress réplcatif. Ces résultats ne sont pas en accord avec le premier modèle présenté ci-dessus qui prédit des ralentissements et/ou des blocages de fourches au niveau de la région fragile.



**Figure 3 : Représentation schématique de la dynamique de réplication de la région cœur de *FRA3B* chez les lymphocytes et les fibroblastes.** La région cœur est délimitée par les barres orange verticales. A gauche sont indiquées les phases du cycle cellulaire avec la phase S divisée en quatre fractions. A droite sont mentionnées les conditions de culture des cellules, avec ou sans stress réplcatif. Chez les lymphocytes, la région cœur est répliquée par des fourches de réplication provenant des régions flanquantes qui ne se rejoignent qu'en G2. En présence d'aphidicoline, les fourches convergentes ne se rejoignent pas et la région reste non répliquée. Dans le cas des fibroblastes, le cœur présente une densité normale en événements d'initiation de la réplication qui se déclenchent pendant la deuxième moitié de la phase S. En présence d'aphidicoline, la densité des événements d'initiation dans le cœur augmente en réponse au ralentissement des fourches. Adapté de (Debatisse et al., 2012).

La carte des événements d'initiation et de terminaison montre que, chez les lymphocytes, le gène *FHIT* possède une région cœur de 700 kb très pauvre en événements d'initiation et qui est répliquée grâce à des fourches progressant à partir des deux régions flanquantes. Ce n'est pas le cas chez les fibroblastes où la région cœur n'existe pas et où une densité normale d'événements initiation a pu être mise en évidence tout le long du gène. En conditions de stress réplcatif, les événements de terminaison sont exclus de la région cœur chez les lymphocytes alors qu'ils sont présents tout le long du gène chez les fibroblastes. La fragilité a donc été corrélée à l'absence de terminaison de la réplication sous stress

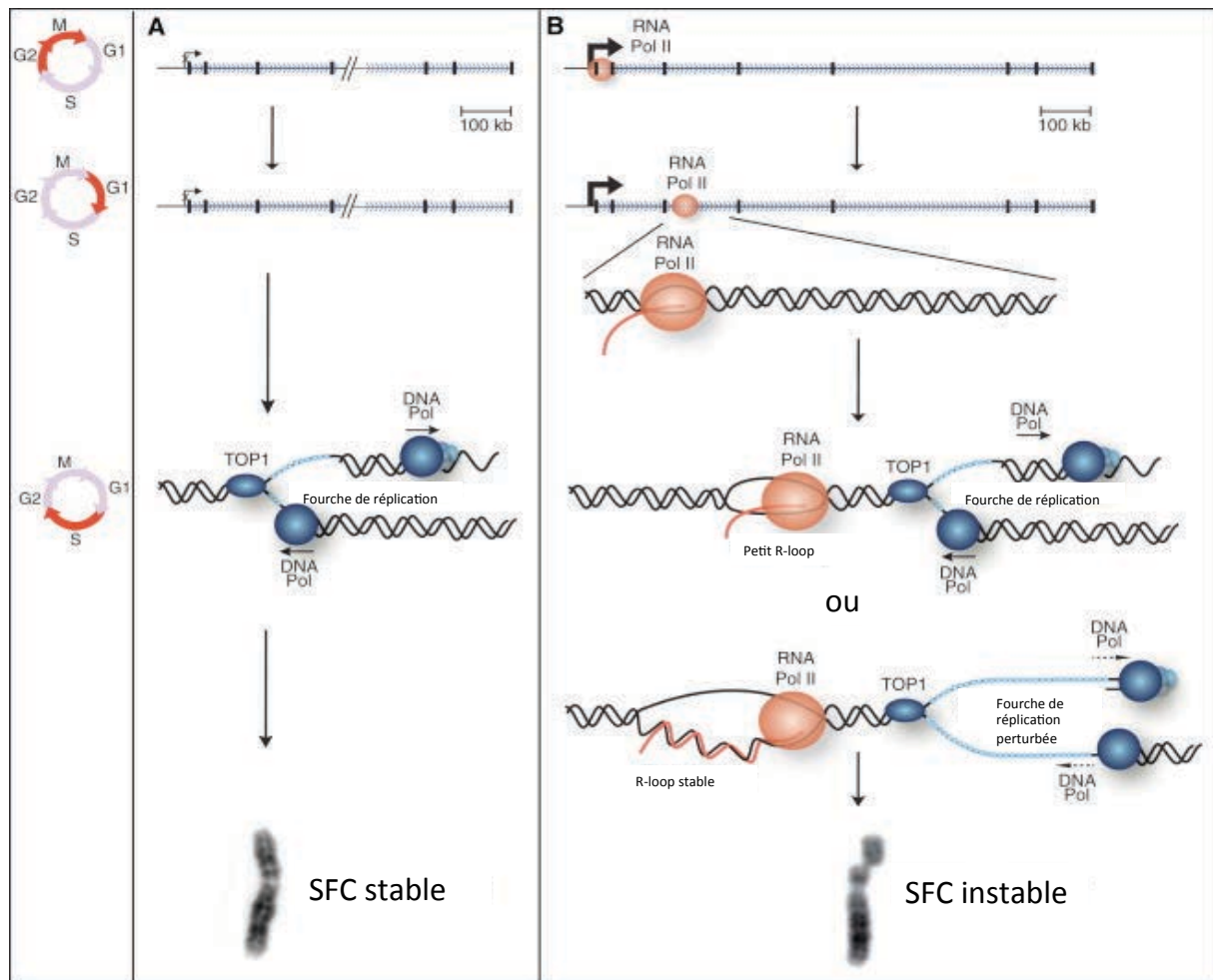
réplicatif. Comme nous l'avons déjà évoqué, la complétion de la réplication de *FRA3B* se fait tardivement pendant la phase S, voire en G2. Le modèle proposé est le suivant : la fragilité des SFC est due à la présence d'une région pauvre en événements d'initiations qui est répliquée par des fourches progressant à partir des régions adjacentes. Quand ces fourches sont ralenties par un stress répliatif, elles n'ont pas le temps de se rejoindre et de finir la réplication de la région, ce qui engendre sa fragilité (Figure 3). A l'appui à ce modèle, l'étude ultérieure des deux sites fragiles majeurs des fibroblastes *FRA1L* et *FRA3L* par la technique de Repli-Seq a montré que ces deux sites ont le même profil de réplication que *FRA3B* dans les lymphocytes, à savoir un cœur répliqué tardivement pendant la phase S par des fourches progressant à partir des régions encadrant le site. Chez les lymphocytes où ces sites ne sont pas fragiles, la région cœur n'existe pas. La fragilité est donc attribuée de la même façon à un déficit d'initiations et une complétion tardive de la réplication. Le timing de réplication et le positionnement des origines de réplication actives sont mis en place cours de la différenciation cellulaire par des mécanismes épigénétiques. Ce modèle explique donc parfaitement pourquoi la fragilité de différents loci varie selon le type cellulaire.

### 7.3. Réplication versus transcription

L'ADN sert de matrice pour la réplication et la transcription. Afin d'éviter des collisions entre la machinerie de transcription et les fourches de réplication, ces deux processus ont généralement lieu séparément dans le temps et dans l'espace (Aguilera and Garcia-Muse, 2013). Contrairement à la situation chez les procaryotes, les ADN polymérases et les ARN polymérases avancent à des vitesses comparables chez les eucaryotes: 17-33 nt/s pour la fourche de réplication et 17-72 nt/s pour la machinerie de transcription, ce qui limite la probabilité de rencontre entre les deux. Un des mécanismes réduisant encore cette probabilité est l'existence de points de pause de la fourche de réplication dans le génome. Ces points de pause résultent de l'existence de barrières empêchant la progression de fourches de réplication à travers des séquences hautement transcrites telles que les gènes qui codent pour les ARN ribosomiaux (Lopez-estrano et al., 1998). Une régulation temporelle existe aussi : les gènes transcrits tendent à être répliqués précocement pendant la phase S (Ryba et al., 2010). Une autre forme de régulation concerne les gènes à multiples copies. Leurs différentes copies n'ont pas les mêmes séquences régulatrices et la transcription peut donc avoir lieu au niveau de certaines copies pendant que d'autres sont répliquées. C'est le

cas des gènes codant pour l'histone H4 (Holmes et al., 2005). Néanmoins, malgré ces systèmes de régulation, de multiples études montrent que des conflits entre la transcription et la réplication existent *in vivo* et que l'absence de leur résolution conduit à des dommages de l'ADN et à des cassures bicaténares (Kim and Jinks-Robertson, 2012; Tuduri et al., 2009).

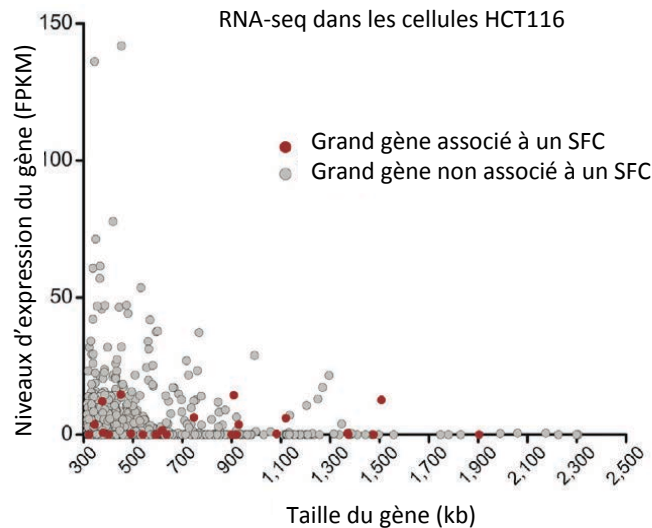
Le troisième modèle proposé pour expliquer le mécanisme de l'instabilité des SFC l'a été par Helmrich et coll. (Helmrich et al., 2011). Il attribue la fragilité des SFC à des collisions non résolues entre les fourches de réplication ralenties par un stress réplicatif et la machinerie de transcription. Comme nous l'avons déjà mentionné, une large majorité des SFC héberge des grands gènes. Beaucoup de ces gènes sont exprimés au cours de la différenciation terminale des cellules. Si la transcription de ces gènes joue un rôle dans l'instabilité des SFC, ceci expliquerait la spécificité de tissu de ces derniers. Dans leur article, ces auteurs décrivent d'abord une corrélation entre le niveau d'ARNm de 5 grands gènes (*CNTNAP2*, *DMD*, *FHIT*, *WWOX*, *IMMP2L*) dans deux types cellulaires (lymphocytes et myoblastes) et le niveau de fragilité du SFC correspondant. Ils rapportent ensuite que la transcription de ces grands gènes, à cause de leurs tailles, dure plus d'un cycle cellulaire, ce qui doit conduire à des collisions entre les fourches de réplication et la machinerie de transcription au niveau de ces sites. Ces collisions engendreraient la formation de structures appelées R-loops, qui sont des hybrides entre l'ADN qui a servi de matrice et l'ARN néosynthétisé sortant de l'ARN polymérase II. Il a été rapporté que de telles structures peuvent être responsables d'une instabilité génétique (Lin and Pasero, 2012). Les auteurs affirment enfin que l'élimination de ces structures par la RNase H réduit la fragilité des SFC induite par l'aphidicoline. Ils ont donc proposé le modèle suivant : quand la fourche de réplication est ralentie par un stress réplicatif, il y aurait formation et stabilisation de R-loops ce qui perturberait la progression des fourches de réplication et engendrerait la fragilité des SFC (Figure 4).



**Figure 4 : Modèle expliquant le mécanisme d'instabilité des sites fragiles communs.** A gauche sont indiquées les phases du cycle cellulaire. En A) est représentée la réplication d'un SFC dans des conditions normales (DNA pol : ADN polymérase, TOP1 : topoisomérase I). En B) est présentée la réplication d'un SFC : une fourche de réplication perturbée rencontre la machinerie de transcription (RNA pol II : ARN polymérase II). Il y a formation d'une structure hybride entre l'ARN naissant et l'ADN monocaténaire (R-loop) puis extension et stabilisation de la R-loop, ce qui bloque l'avancement de la fourche et conduit à l'instabilité du site. Adapté de (Helmrich et al., 2011).

Cependant, ce modèle semble peu compatible avec des résultats obtenus dans notre laboratoire. Les résultats d'une analyse du transcriptome de cellules HCT116 obtenus par la technique de RNA-seq ont été comparés par Benoit Le Tallec avec la fragilité des SFC dans cette lignée. Cette analyse montre que, dans ces cellules, il n'y a pas de corrélation entre le niveau d'ARNm des grands gènes et la fragilité des SFC (Le Tallec et al., 2013) (Figure 5). Notamment, il existe des grands gènes transcrits à un niveau élevé qui sont stables et ne correspondent pas à des SFC. Le modèle des collisions ne semble pas pouvoir expliquer à lui seul le mécanisme de la fragilité. De plus, si des R-loops sont formés fréquemment au niveau des SFC sous stress réplcatif, cela doit se traduire par une augmentation très significative du

nombre de fourches bloquées. Or, comme nous l'avons déjà mentionné, l'étude de la dynamique de réplication au niveau de *FRA3B* dans des lymphocytes JEFF, où le gène *FHIT* est exprimé, a montré qu'il n'y a pas plus de fourches bloquées au niveau de ce site que dans le reste du génome (Letessier et al., 2011).



**Figure 5: Niveaux d'expression des grands gènes (> 300 kb) dans les cellules HCT116.** Le niveau d'expression des gènes est donnée en Fragments Per Kilobase per Million reads (FPKM) d'après les données tirées du projet ENCODE 2011. Adapté de (Le Tallec et al., 2013).

## 8. Prise en charge des SFC et réparation

Il a été montré que sous stress réplcatif, les SFC colocalisent avec des foyers  $\gamma$ H2AX, RAD51 et DNA-PK (Schwartz et al., 2006) ce qui suggère que des cassures bicaténares sont formées au niveau des SFC. L'activation de la réponse aux dommages à l'ADN a de multiples conséquences: la stabilisation des fourches de réplication, le redémarrage des fourches bloquées, le déclenchement d'origines de réplication supplémentaires, l'activation des différents voies de réparation ainsi que la restriction de passage des cellules en mitose. Un nombre important de facteurs de cette réponse a été impliqué dans la protection des SFC. Il a été montré que l'invalidation de ces facteurs augmente l'instabilité des SFC. C'est le cas d'ATR (Casper et al., 2002), de CHK1 (Durkin et al., 2006), de BRCA1 (Arlt et al., 2004), de RAD51, de la Claspine (Focarelli et al., 2009), de FANCD2 (Howlett et al., 2005), de BLM (Naim et al., 2013) et d'autres. Cependant pour interpréter correctement ces données, il faut prendre en considération un ensemble de résultats qui montrent que la carence ou l'inhibition de certains de ces facteurs induisent par elles mêmes un stress réplcatif.

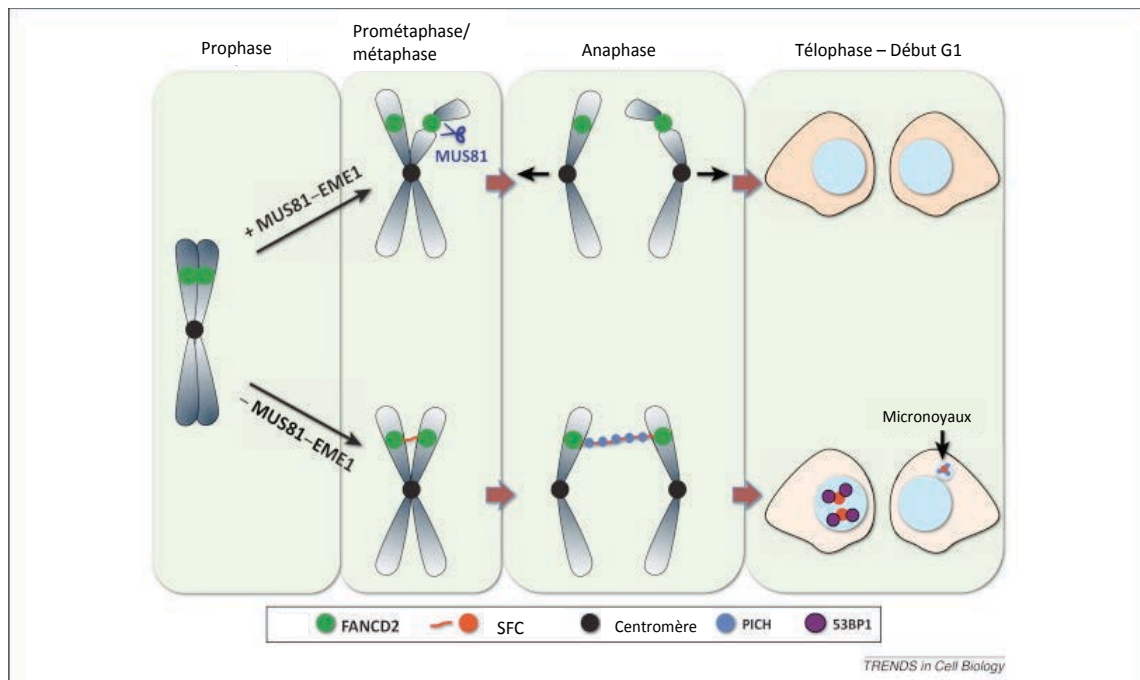
Notamment, il a été montré que la protéine BLM contribue à maintenir l'équilibre des pools de précurseurs pyrimidiques dans les cellules et que son invalidation réduit la vitesse de réplication par cette voie (Chabosseau et al., 2011). Il a aussi été montré que l'invalidation de CHK1 (Koundrioukoff et al., 2013; Petermann et al., 2006), de RAD51 (Daboussi et al., 2008) ou de la Claspine (Petermann et al., 2008) diminue la vitesse de réplication et donc que leurs effets sur l'instabilité des SFC peut s'expliquer, au moins en partie, par le stress réplicatif induit par leur inhibition. De plus, il a été montré au laboratoire que l'invalidation de CHK1 n'a pas d'effet propre sur la stabilité des SFC, en dehors de celui résultant de la diminution de la vitesse de réplication. Par contre, l'invalidation d'ATR affecte spécifiquement la stabilité des SFC en plus de son effet *via* la réduction de vitesse. Cet effet passe par une stabilisation des fourches et non pas par l'activation de la réponse aux dommages (Koundrioukoff et al., 2013).

Vu le danger que des régions non répliquées et des cassures bicaténaires représentent pour la cellule, une question se pose : comment se fait-il que, sous stress réplicatif modéré tel que celui induit par les doses modérées d'aphidicoline qui sont utilisées pour induire la fragilité des SFC, la réponse aux dommages ne soit pas activée et que la cellule puisse passer en mitose avec des sites sous-répliqués? D'une part, il est possible que, comme chez la levure, la réponse aux dommages ne se déclenche pas lorsqu'une très faible fraction du génome est non répliquée (Torres-Rosell et al., 2007). De plus, plusieurs résultats suggèrent fortement que les lésions résultant d'un non-achèvement local de la réplication peuvent être prises en charge par la cellule jusqu'au cycle suivant.

A l'heure actuelle, nous ne connaissons pas le mécanisme exact responsable de la création des cassures au niveau des SFC. Deux hypothèses non exclusives sont proposées : les cassures se créent mécaniquement au moment de la condensation des chromosomes et/ou les structures intermédiaires résultant des régions sous-répliquées sont résolues par des nucléases spécifiques. Des études récentes commencent à apporter une réponse à cette question. En effet, il a été montré que les endonucléases MUS81/EME1 et ERCC1 qui participent à la résolution des structures inappropriées dans l'ADN telles que les fourches bloquées ou les jonctions de Holliday, sont recrutées au niveau des SFC en fin de phase G2 et en début de mitose. L'invalidation de ces nucléases conduit à une diminution des cassures au niveau des SFC, ce qui suggère qu'elles participent activement à la création des cassures (Naim et al., 2013; Ying et al., 2013). De même, il a été montré que FANCD2 et FANCI sont

recrutées au niveau des SFC et que, si les structures aberrantes résultant d'une réplication inachevée ne sont pas résolues par les nucléases, il y a formation, au cours de l'anaphase, de ponts interchromosomiques (ponts anaphasiques). Ces ponts sont protégés par l'hélicase BLM en association avec la topoisomérase IIIa, les facteurs RMI1 et RMI2 (RecQ mediated genome instability 1 and 2) et PICH (PLK1 interacting checkpoint helicase) et leurs extrémités sont associées avec les facteurs FANCD2 et FANCI (Chan et al., 2009; Naim and Rosselli, 2009). Ces structures permettent à la ségrégation chromosomique d'avoir lieu. Néanmoins, des anomalies structurales persistent et se manifestent par une augmentation pendant la phase G1 suivante, du nombre de micronoyaux PITCH qui sont des marqueurs des erreurs de ségrégation, et des foyers 53BP1. On trouve alors les SFC préférentiellement associés à ces structures. La protéine 53BP1 est un facteur clé dans la réparation des cassures bicaténares. Ceci suggère que les SFC pourront être réparés pendant la phase G1 suivante (Figure 6).





**Figure 6 : Modèle représentant les mécanismes moléculaires de prise en charge d'un SFC incomplètement répliqué.** Les chromatides sœurs sont marquées au niveau du site fragile non répliqué par le recrutement pendant la phase G2 du facteur FANCD2 qui reste associé pendant la mitose. La coupure de l'ADN par l'endonucléase MUS81 en complexe avec la protéine EME1 est nécessaire pour la ségrégation des deux chromatides sœurs pendant l'anaphase. Si le complexe MUS81-EME1 est invalidé, la coupure n'a pas lieu et il y a formation d'un pont anaphasique pendant la ségrégation des chromatides sœurs qui est marqué au niveau de ses extrémités par FANCD2 et recouvert de la protéine PICH. Ceci conduit pendant la phase G1 du cycle suivant à la formation de micro-noyaux contenant le facteur PICH ainsi qu'à une augmentation du nombre de foyers contenant le facteur de réparation 53BP1. Adapté de (Minocherhomji and Hickson, 2014).

Dans la première partie de l'introduction nous avons exposé les principales caractéristiques des sites fragiles communs. Nous avons vu qu'il existe une relation très étroite entre leur fragilité et leur profil de réplication. En effet, les SFC se caractérisent par une réplication tardive et une pauvreté en origines de réplication actives. Pendant ma thèse, j'ai entrepris d'explorer le rôle de ces paramètres dans le mécanisme de l'instabilité des SFC. La deuxième partie de l'introduction va donc porter sur la réplication de l'ADN chez les eucaryotes.

## II. La réplication de l'ADN

Avant de se diviser, une cellule doit dupliquer son génome afin de pouvoir transmettre une copie conforme de son ADN à chacune de ses deux cellules filles. La réplication de l'ADN a lieu pendant la phase S du cycle cellulaire. Elle comporte trois étapes : l'initiation, l'élongation et la terminaison. Ce processus constitue un enjeu majeur pour la cellule et il est extrêmement régulé afin d'assurer la fidélité de la réplication aux niveaux quantitatif et qualitatif.

### **1. Le modèle du réplicon**

En 1963, Cairns a inventé l'autoradiographie des fibres d'ADN néosynthétisé en leur incorporant de la thymidine radioactive tritiée (Cairns, 1963). Il a ainsi pu montrer que la réplication du chromosome circulaire de la bactérie *Escherichia Coli* commence à partir d'un seul point et qu'elle est bidirectionnelle. En 1963, François Jacob, Sydney Brenner et François Cuzin ont proposé le modèle du réplicon pour expliquer la régulation de la réplication chez les bactéries. Selon ce modèle, la réplication nécessite l'interaction d'une protéine (initiateur), codée par un gène de structure, avec une séquence spécifique de l'ADN (réplicateur, appelé aujourd'hui origine de réplication) (Jacob and Brenner, 1963). Des années plus tard ces deux éléments, l'initiateur (*oriC*) et le réplicateur (*DnaA*) ont été isolés, ce qui a permis la reconstitution du réplicon *in vitro* (Bramhill and Kornberg, 1988).

### **2. Les origines de réplication chez les eucaryotes**

L'étude de la réplication chez les eucaryotes par la technique d'autoradiographie a montré qu'il n'y a pas un seul réplicon mais plusieurs, espacés de 20-400 kb (Huberman, 1968). Les points de déclenchement de la réplication (origines de réplication) étant multiples, le réplicon a été redéfini pour désigner la séquence d'ADN répliquée à partir d'une seule origine de réplication.

Les origines de réplication ont été d'abord identifiées chez la levure *Saccharomyces cerevisiae* dans les années 1970 (Stinchcomb et al., 1979) en isolant des séquences de 100-200 pb qui permettent la réplication autonome de plasmides et ont été donc appelées *Autonomously Replicating Sequence* (ARS). Les ARS ont alors été considérées comme l'équivalent du réplicateur chez les bactéries. Elles se composent d'une séquence consensus

de onze nucléotides riche en AT, l'ACS (ARS consensus sequence) et de séquences moins conservées en 3' de l'ACS.

Le facteur initiateur a été d'abord identifié chez *S. cerevisiae* comme étant un complexe hétéro-hexamérique, appelé *Origin Recognition Complex* (ORC) (Bell et al., 1993). Ensuite, il a été montré que le complexe ORC est fixé en permanence sur la chromatine pendant le cycle cellulaire et que d'autres facteurs se fixent au niveau d'ORC pendant la phase G1 pour former ce qui a été appelé le complexe de pré-réplication (pré-RC) (Diffley et al., 1994). Tous les ACS ne portent pas de complexe ORC. Seulement 400 environ parmi les 20000 ACS sont chargés en complexe ORC et peuvent ainsi fonctionner comme des origines de réplication. De plus, tous les ARS chargées en complexe ORC ne se déclenchent pas à chaque cycle cellulaire. Chaque ARS se déclenche avec une probabilité variant de 0 à 100% par cycle. Cette valeur exprime son efficacité: plus la probabilité est élevée, plus l'ARS est dite efficace. Les ARS ne se déclenchent pas toutes au même moment au cours de la phase S. Le temps de déclenchement des origines est déterminé par le programme temporel de la réplication ou timing de réplication. On distingue ainsi des origines précoces et d'autres tardives. Il faut noter que le caractère exclusif attribué pendant des longues années aux ARS comme seul élément faisant fonction d'initiateur chez *S. cerevisiae* a été récemment remis en question. En effet, il a été montré *in vitro* que les ARS ne sont pas absolument requises pour le démarrage de la réplication mais que ces séquences deviennent indispensables à la réplication en présence d'ADN compétiteur ou lorsque certains facteurs de réplication sont limitant (Gros et al., 2014).

Le caractère bien défini des origines de réplication chez *S. cerevisiae* est apparu au fil du temps comme une exception parmi les eucaryotes. En effet, toutes les études visant à chercher une séquence consensus reconnue par le complexe ORC chez les autres eucaryotes, à commencer par la levure *S. pombe*, ont échoué. Ainsi, il a été montré chez le xénope et chez l'homme que toute séquence peut faire fonction de réplicateur (Takeda et al., 2005). Malgré de multiples études à grande échelle utilisant diverses techniques, jusqu'à aujourd'hui nous n'avons pas de réponse claire sur ce qui définit le réplicateur chez les eucaryotes supérieurs.

### **3. Mécanismes de déclenchement des origines de réplication chez les eucaryotes**

Le fonctionnement des origines de réplication est régulé à trois niveaux : elles sont chargées avec le complexe de pré-RC en phase G1 ; elles sont activées en phase S pour démarrer la réplication et enfin, dans les régions où la réplication a déjà eu lieu, elles sont réprimées pour éviter la re-réplication. Le chargement des complexes de pré-RC sur les origines a lieu quand les activités des CDK sont basses. Par contre, l'initiation et l'inhibition de la re-réplication ont lieu quand les activités des CDK sont élevées (Figure 7).

#### 3.1. Chargement des origines par les complexes de pré-RC

Cette étape a lieu strictement pendant la phase G1 du cycle cellulaire. Son but final est de charger les hélicases réplicatives au niveau des origines de réplication. Le chargement commence avec le recrutement du facteur ORC1 sur le complexe ORC2-6 en fin de mitose pour former un complexe ORC complet. Durant la phase G1, CDC6, une protéine codée par un gène paralogue de celui de ORC1, se fixe au niveau du complexe ORC. CDC6, ainsi que les facteurs ORC 1 à 6 font partie de la famille ATPase AAA+. Leur fixation à la chromatine est dépendante de l'ATP (Speck et al., 2005). Le complexe ORC/CDC6 fonctionne alors comme une plateforme de recrutement des protéines MCM 2 à 7. Le complexe MCM2-7 constitue l'hélicase responsable de l'ouverture de l'ADN pendant la réplication. Son recrutement nécessite la protéine CDT1. Celle-ci possède au moins deux domaines d'interaction avec les MCM : un au niveau de son domaine N-terminal et un autre au niveau de son domaine C-terminal. La formation du complexe CDT1/MCM2-7 dans le cytoplasme permet son importation dans le noyau (Tanaka and Diffley, 2002). Il a été montré récemment que l'interaction entre CDT1 et MCM6, lève une inhibition au niveau du domaine C-terminal de cette dernière (Fernandez-Cid et al., 2013) et permet ainsi la formation du complexe ORC/CDC6/CDT1/MCM2-7 (OCCM). La formation de ce complexe déclenche les activités ATPase d'ORC1 et de CDC6, ce qui induit un changement conformationnel indispensable à la stabilisation du complexe ORC/CDC6/MCM2-7 (OCM) et libère CDT1. La formation d'un complexe OCM stable enclenche le recrutement d'un second complexe CDT1/MCM2-7 afin de former un dimère du complexe MCM2-7 nécessaire à une réplication bidirectionnelle à partir de l'origine. Théoriquement, un double hexamère des complexes MCM2-7 suffit au démarrage de la réplication. Néanmoins les complexes MCM2-7 sont en très large excès dans la cellule par rapport aux autres protéines impliquées dans la construction des

complexes de pré-RC et au nombre d'origines de réplication et un grand nombre d'hexamères sont chargés au niveau de chaque origine, probablement par répétition du cycle de charge décrit ci-dessus et glissement des hexamères déjà chargés le long de l'ADN. La raison du chargement de plusieurs complexes MCM2-7 par origine est mal comprise.

Chez les eucaryotes supérieurs, il a été montré que d'autres facteurs peuvent accompagner cette étape de chargement des complexes MCM sans lui être indispensables. C'est le cas de MCM9, un facteur paralogue des autres MCM et de l'acétyltransférase HBO1 (Lutzmann and Mechali, 2008) qui assiste CDT1 dans le recrutement des hélicases. C'est aussi le cas de ORCA qui a été proposé comme facteur assistant le positionnement de ORC1 sur le complexe ORC (Shen et al., 2012).

### 3.2. Démarrage de la réplication

Pendant la phase G1, on ne peut pas détecter d'ADN monocaténaire, ni chez la levure (Geraghty et al., 2000) ni dans un système *in vitro* basé sur l'utilisation d'extraits de xénope (Walter, 2000). Ceci montre que les hélicases sont chargées en G1 sous une forme inactive, contrairement à ce qui a lieu chez les bactéries. Leur activation nécessite l'addition d'une trentaine de protéines au niveau du complexe de pré-RC qui se transforme en complexe de pré-initiation (pré-IC) (Gambus et al., 2006). Cette étape a été largement étudiée chez *S. cerevisiae*, néanmoins ses grandes lignes sont conservées chez tous les eucaryotes. Deux types d'activité kinase contrôlent le déclenchement des origines : les kinases dépendantes des cyclines de phase S (S-CDK) et la kinase dépendante de DBF4 (DDK) (Siddiqui et al., 2013). Les kinases agissent de façon locale au niveau des origines. Elles déclenchent donc les origines précoces au début de la phase S et les origines plus tardives en avançant dans la phase S. Au cours de la transition G1/S, l'activité des complexes S-CDK (CDK2/cycline E et CDK2/cycline A) et des DDK augmentent. Le complexe DDK est composée de la sous-unité régulatrice DBF4 (DumbBell Former 4) et de la kinase Cdc7 (Cell Division Cycle 7). Une fois activée, DDK phosphoryle les protéines MCM2, MCM4 et MCM6. Cette phosphorylation modifie la conformation spatiale du complexe MCM et permet le recrutement du facteur Cdc45. Les complexes S-CDK sont également indispensables à l'activation des origines. Chez les mammifères, les mécanismes d'action des S-CDK dans l'activation des origines restent peu documentés. Des travaux récents chez *S. cerevisiae* ont pu mettre en évidence l'importance des S-CDK dans la phosphorylation de deux protéines essentielles à l'initiation

de la réplication, Sld2 et Sld3 (Muramatsu et al., 2010; Zegerman and Diffley, 2007)

### 3.2.1. Déclenchement des origines chez *S. cerevisiae*

En fin de phase G1, l'activité de DDK permet l'association des facteurs Sld3, Sld7 et Cdc45 aux complexes de pré-RC. Ensuite, les activités CDK déclenchent, *via* une interaction entre Sld3 et Dbp11, le recrutement des facteurs Sld2, Dpb11, du complexe GINS et de l'ADN polymérase  $\epsilon$ . Ceci aboutit à la formation du complexe CMG (Cdc45-MCM2-7-GINS). L'activation de l'activité hélicase du complexe CMG requiert le facteur MCM10. Enfin, par l'intermédiaire du facteur Ctf4, il y a recrutement de l'ADN polymérase  $\alpha$  qui permet le démarrage de la synthèse de l'ADN par son activité primase. Cette polymérase synthétise une amorce d'ARN de 10 pb sur chacun des deux brins de l'ADN. Cependant, ayant une très faible processivité, cette polymérase ne peut synthétiser qu'une vingtaine de nucléotides et sera ensuite remplacée par la polymérase  $\epsilon$  sur le brin direct et par la polymérase  $\delta$  sur le brin retardé (Labib, 2010; Siddiqui et al., 2013).

### 3.2.2. Déclenchement des origines chez les eucaryotes supérieurs

La recherche de facteurs homologues à ceux du complexe de pré-IC connus chez la levure a permis d'identifier les facteurs Treslin/Ticrr, RecQ4 et TopBP1 comme étant les homologues respectifs de Sld3, Sld2 et Dbp11. Ces facteurs présentent des divergences notables dans la séquence ainsi que dans la structure par rapport à leurs homologues de levure. Comme c'est le cas chez *S. cerevisiae*, Treslin/Ticrr (Sld3) recrute Cdc45 sur les origines. Le domaine responsable de l'interaction entre Sld3 et Cdc45 est conservé chez Treslin/Ticrr. De plus, chez l'homme et le xénope, Treslin/Ticrr est phosphorylée par les S-CDK sur les mêmes sites conservés et est capable de se lier au domaine N-terminal de TopBP1 (Dbp11). Un argument supplémentaire en faveur de l'homologie fonctionnelle de ces facteurs est fourni par la réponse aux dommages de l'ADN. Chez la levure, la réponse aux dommages implique l'activation du facteur Rad53 et réduit l'interaction entre Sld3 et Dbp11. Il est apparu que la même réponse existe chez les eucaryotes supérieurs : une fois activée le facteur CHK1 impliqué dans la réponse aux dommages inhibe l'interaction entre Treslin et TopBP1.

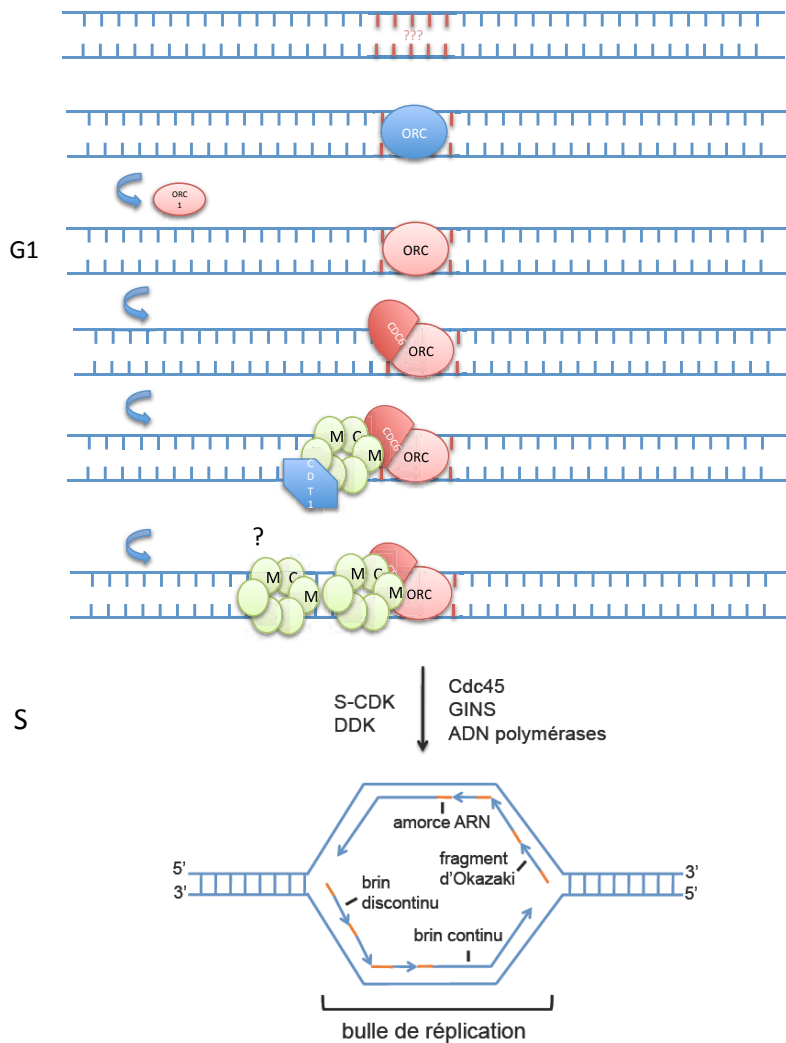
L'homologie fonctionnelle entre RecQ4 et Sld2, est moins claire. En effet, les deux protéines présentent très peu de similitude de séquence. L'association de RecQ4 aux

origines est dépendante de TopBP1. RecQ4 comporte un domaine C-terminal à activité hélicase qui n'est pas requis pour le recrutement de TopBP1, Cdc45 et GINS sur les origines. Ceci suggère que RecQ4 joue un rôle supplémentaire. De plus, TopBP1 et RecQ4 peuvent interagir sans être phosphorylées par les S-CDK, ce qui n'est pas le cas pour Sld2-Dpb11. Contrairement à Sld2, la liaison de RecQ4 au complexe CMG est dépendante de MCM10. Ceci suggère que, contrairement à la levure où MCM10 n'intervient que pour activer le complexe CMG, chez les eucaryotes supérieurs MCM10 semble participer à la formation de ce complexe. Enfin, concernant l'homologie entre TopBP1 et Dbp11, le recrutement de TopBP1 est dépendant des S-CDK, comme c'est le cas pour Dbp11.

Parmi tous les facteurs qui interviennent dans le déclenchement des origines chez la levure, il n'y a que Sld7 qui n'a pas d'homologue connu chez les eucaryotes supérieurs. Il a été proposé récemment que le facteur MTBP pourrait être son homologue. En effet, MTBP interagit avec la Treslin et son absence empêche la formation du complexe CMG et inhibe la réplication. Enfin, deux facteurs participant au recrutement de Cdc45 chez les eucaryotes supérieurs n'ont pas d'homologues connus chez la levure. Il s'agit de GEMC1 et DUE-B (Labib, 2010; Siddiqui et al., 2013).

### 3.3. Inhibition de la re-réplication

Afin que l'ADN ne soit dupliqué qu'une fois et une seule au cours de la phase S, plusieurs mécanismes inhibent le chargement des origines en dehors de la phase G1 et empêchent la re-réplication.



**Figure 7 : Chargement et déclenchement des origines de réplication.** Durant la phase G1, ORC1 se fixe sur le complexe ORC2-6 au niveau des origines de réplication et recrute CDC6. Ensuite, il y a fixation du premier complexe MCM par l'intermédiaire de CDT1. Le mécanisme de recrutement du second complexe MCM est assez mal connu. La fixation de deux complexes MCM signe la formation du complexe de pré-réplication. Ce processus est inhibé pendant les autres phases du cycle. Pendant la phase S, il y a chargement du facteur Cdc45 et du complexe GINS et formation du complexe de pré-initiation. Les activités kinases S-CDK et DDK permettent le déclenchement de l'origine.

### 3.3.1. Chez *S. cerevisiae*

Tous les mécanismes connus pour empêcher la re-réplication chez la levure impliquent les S-CDK qui ciblent différents composants du complexe de pré-RC. CDC6 est inhibée de trois façons différentes : 1) elle est phosphorylée, ce qui permet son ubiquitination par la ligase E3 SCF<sup>cdc4</sup> et sa dégradation par le protéasome (Drury et al., 1997) ; 2) les S-CDK inhibent l'importation dans le noyau du facteur de transcription Swi5, ce qui inhibe la transcription du gène codant pour CDC6 (Jans et al., 1995) ; 3) la phosphorylation de CDC6 au niveau de son domaine N-terminal stabilise son interaction avec les CDK pendant la mitose



et inhibe son recrutement au niveau des origines (Mimura et al., 2004). Le complexe MCM2-7 est phosphorylé par les S-CDK et exporté vers le cytoplasme (Liku et al., 2005). Etant donné que CDT1 est associée au complexe MCM2-7, l'exportation de ce dernier vers le cytoplasme aboutit également à celle de CDT1 pendant la phase S. Le complexe ORC est aussi phosphorylé par les S-CDK, ce qui l'empêche de se fixer à la chromatine. Il a été montré qu'un seul des différents mécanismes que nous venons d'exposer ne suffit pas à inhiber la re-réplication, ce qui suggère qu'ils ne sont pas complètement redondants.

### 3.3.2. Chez les eucaryotes supérieurs

L'inhibition de la re-réplication passe par plusieurs voies faisant intervenir les S-CDK: la destruction de CDT1, l'exportation de CDC6 vers le cytoplasme, la dégradation d'ORC1, la phosphorylation des MCM. La différence majeure par rapport à la levure est l'inhibition de CDT1 d'une façon indépendante des S-CDK *via* son interaction avec la géminine. CDT1 est ubiquitinée et dégradée pendant la phase S. Deux voies de protéolyse de CDT1 sont connues. La première dépend de PCNA. CDT1 possède un domaine d'interaction avec PCNA (PIP box) dans son domaine N-terminal. L'interaction de PCNA liée à la chromatine avec CDT1 induit le recrutement de l'ubiquitine ligase E3 Cul4-Ddb1<sup>Cdt2</sup> qui ubiquitine CDT1 et induit sa protéolyse. Cette voie est favorisée pendant la phase S où PCNA est abondante. Une autre ubiquitine ligase E3 est aussi impliquée dans la destruction de CDT1 dans les cellules humaines. Il s'agit de SCF<sup>Skp2</sup> qui interagit avec CDT1 après la phosphorylation de cette dernière par le complexe CDK2/cycline A. Cette voie est favorisée pendant la phase G2.

CDT1 est aussi inhibée par la géminine. La géminine est une protéine de petite taille initialement identifiée dans des extraits d'œufs de xénope comme étant un facteur inhibant le chargement des MCM (McGarry and Kirschner, 1998). Aujourd'hui, nous savons que cette inhibition est assurée par la liaison de la géminine à CDT1, ce qui empêche cette dernière d'interagir avec le complexe MCM (Lutzmann et al., 2006). La géminine est présente en phase S et G2 et pendant la mitose. Elle joue un double rôle : elle inhibe le recrutement des MCM par CDT1 mais elle protège cette dernière de la dégradation par SCF<sup>SKP2</sup> en G2 afin de maintenir un niveau suffisant de CDT1 et permettre le bon déroulement du chargement des origines quand les cellules entrent de nouveau en phase G1. Cette voie d'inhibition de la re-réplication est spécifique aux eucaryotes supérieurs, étant donné qu'aucun homologue de la

gémينية n'a été identifié chez la levure. C'est aussi la seule voie connue à nos jours qui ne fait pas intervenir les S-CDK.

Concernant CDC6, il a d'abord été montré que ce facteur est phosphorylé par les S-CDK et exportée du noyau pendant la phase S chez le xénope et chez l'homme (Petersen et al., 1999). Cependant, il a été aussi montré que l'expression d'une forme non phosphorylable de CDC6 n'induit pas de re-réplication et que l'exportation du noyau ne concerne que la fraction soluble de CDC6 et non pas la fraction chromatinienne (Oehlmann et al., 2004). Par ailleurs, il a été montré que CDC6 est dégradée en fin de mitose et en début de phase G1. Elle recommence à s'accumuler en phase G1 et pendant les phases S et G2. Néanmoins, l'expression d'une forme non dégradable de CDC6 n'induit pas de re-réplication (Petersen et al., 2000). Il apparaît donc que le rôle de CDC6 dans les mécanismes qui empêchent la re-réplication n'est pas clairement établi.

En ce qui concerne le complexe ORC, il semble exister des différences entre espèces de mammifères. Dans les cellules humaines, ORC1 est ubiquitinylé par SCF<sup>SKP2</sup> et dégradé par le protéasome pendant la phase S. Par contre, il a été montré que, chez les cellules de Hamster, ORC1 n'est pas dégradé mais hyperphosphorylé au cours de la transition S/G2M ce qui diminue son affinité pour la chromatine (DePamphilis, 2005).

#### **4. Déterminants génétiques et/ou épigénétiques des origines de réplication**

Comme nous l'avons déjà mentionné, l'élément « réplicateur » chez les eucaryotes supérieurs n'a pas pu être clairement identifié par sa séquence jusqu'à présent. Néanmoins le passage de la cartographie de quelques origines (moins d'une trentaine) à la cartographie des origines à grande échelle grâce aux puces d'ADN et au séquençage à haut débit a permis de mettre en évidence des caractéristiques génétiques et épigénétiques communes à ces origines. Actuellement, les origines de réplication ont été cartographiées chez la souris, la drosophile, l'arabette des dames (*Arabidopsis thaliana*) et l'homme.

##### 4.1. Cartographies des origines de réplication dans les cellules humaines

Nous allons d'abord passer en revue les résultats obtenus par différentes techniques dans des études indépendantes. Dans une des premières études cartographiant les origines de réplication dans des cellules lymphoblastoïdes humaines (Lucas et al., 2007), les auteurs ont utilisé la technique d'isolement des petits fragments d'ADN monocaténaire (Short Single

Strands : SSS) en considérant que ces fragments correspondent à l'ADN néosynthétisé. Néanmoins, cette technique a été remise en question après une étude montrant que 99% des SSS sont éliminés par une digestion avec l'enzyme exonucléase 5' du phage  $\lambda$  qui ne peut digérer que les fragments d'ADN n'ayant pas d'amorce ARN en 5'. Les fragments d'ADN néosynthétisés possèdent une telle amorce et sont résistants à l'action de cette enzyme (Cadoret et al., 2008). Dans cette étude, les fragments obtenus à partir de cellules Hela après digestion des SSS par l'exonucléase 5' du phage  $\lambda$  ( $\lambda$ -SNS  $\lambda$ -resistant Short Nascent Strands) ont été analysés par hybridation sur une puce ne portant qu'une sélection de séquences couvrant 1% du génome humain (projet ENCODE) et 283 origines ont été identifiées. D'autres auteurs ont confirmé par la suite que la digestion à l'exonucléase 5' du phage  $\lambda$  était nécessaire pour obtenir des résultats concluants (Cayrou et al., 2011). Cependant, ceci a été remis en question par un autre groupe qui a rapporté ne pas avoir trouvé de différence entre les résultats obtenus avec les fragments SSS et  $\lambda$ -SNS (Valenzuela et al., 2011). Pour expliquer cette absence de différence, ces auteurs se sont appuyés sur les travaux d'un autre groupe (Gomez, 2008). Ces derniers ont montré qu'au moment de l'initiation, il y a une surproduction de fragments de 200 bp à cause d'un phénomène de re-réplication locale au niveau des origines de réplication. Cette surproduction de petits fragments due à ces cycles de réplication abortive expliquerait l'enrichissement de la population de fragments isolés en fragments correspondant à des origines de réplication. Deux autres groupes ont cartographié les origines de réplication dans les cellules Hela en utilisant des puces ENCODE :

- 1) Karnani et coll. (Karnani et al., 2010) ont utilisé deux techniques différentes pour purifier les fragments d'ADN naissant: celle des  $\lambda$ -SNS et celle des BrdU-SNS. Dans ce dernier cas, les fragments ont été isolés grâce au BrdU qu'ils renferment après un marquage bref des cellules par cet analogue de la thymidine ;
- 2) Mesner et coll. (Mesner et al., 2011) ont utilisé la technique de piège à bulles d'ADN. La comparaison des résultats obtenus dans chacune de ces trois études indépendantes montre un chevauchement de 11-35% entre les ensembles d'origines cartographiées (Hyrien, 2015). Les origines de réplication ont été aussi cartographiées dans les cellules HeLa au niveau pan-génomique par la technique des  $\lambda$ -SNS (Besnard et al., 2012). La comparaison de ces résultats avec les trois cartographies antérieures, a montré une concordance de 51% avec les résultats de Mesner et coll., de 6-14% avec les résultats de Karnani et coll. et de 80% avec ceux de Cadoret et coll. Par ailleurs, une cartographie pan-génomique des origines de réplication dans des érythroblastes

humains primaires utilisant les techniques  $\lambda$ -SNS et BrdU-SNS a montré une concordance de 70% entre les données obtenues avec les deux techniques (Mukhopadhyay et al., 2014).

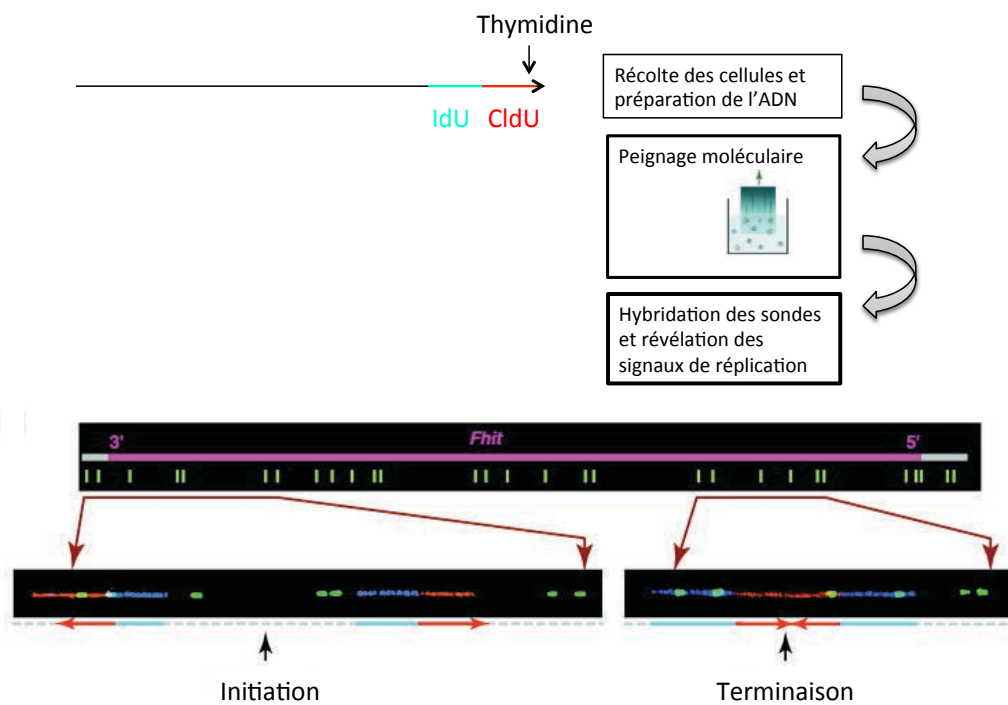
Intéressons nous maintenant à la comparaison des résultats obtenus dans différentes lignées cellulaires. Mesner et coll. ont trouvé 28-43% de chevauchement entre l'ensemble des origines cartographiées dans les cellules Hela et celles cartographiées dans les cellules lymphoblastoïdes humaines GM066990. Dans l'étude de Besnard et coll., quatre lignées cellulaires différentes ont été étudiées et 65-84% de concordance ont été trouvés entre les différents résultats. D'après les auteurs de cette étude, 50% des origines identifiées sont ubiquitaires. Parmi les origines ubiquitaires, ils ont retrouvé 91% des origines cartographiées par Cadoret et coll. et Martin et coll. (Martin et al., 2011) et 81% de celles cartographiées par Lucas (2007). Enfin, la ré-analyse des données de Besnard et coll. par Picard et coll (2014) (Picard et al., 2014) a montré une concordance de 65% entre ces données et celles de Mesner et coll. (Mesner et al., 2013).

A ce jour une seule étude a cartographié les origines de réplication dans les cellules HeLa en utilisant la technique d'immunoprécipitation de la chromatine avec un anticorps dirigé contre ORC1 (Dellino et al., 2013). Ces auteurs ont trouvé respectivement 11%, 30%, et 47% de concordance avec les résultats de Karnani et coll., Cadoret et coll. et Mesner et coll.

D'autres techniques de cartographie des origines existent mais ne peuvent être utilisées qu'à petite échelle. C'est le cas de la technique du peignage moléculaire (Figure 8). J'ai utilisé cette technique pendant ma thèse pour l'étude de la dynamique de réplication au niveau du site fragile *FRA3B*. Elle sera développée dans la partie résultats.

Une technique très largement utilisée pour l'étude du timing de réplication est celle de Repli-Seq. Cette technique ne donne que des indications sur les zones où se produisent des initiations. Elle n'est pas assez résolutive pour cartographier précisément les origines. Néanmoins, les résultats obtenus dans ce cas apparaissent beaucoup plus robustes et reproductibles.

Malgré toutes les limitations exposées ci-dessus, la combinaison de différentes approches a permis d'identifier des caractéristiques génétiques et épigénétiques communes aux origines de réplication que nous allons exposer maintenant.



**Figure 8 : Cartographie des événements d'initiation et de terminaison au niveau du gène *FHIT* par la technique du peignage moléculaire.** L'ADN en cours de synthèse est marqué consécutivement par deux analogues de la thymidine, l'IdU, qui sera révélé en bleu, et le CldU qui sera révélé en rouge. L'incorporation des deux analogues est arrêtée par un excès de thymidine. Les cellules sont récupérées et l'ADN est préparé et peigné sur des lames silanisées. Les lames sont ensuite hybridées avec un mélange de 33 sondes réparties sur 1,5 Mb représentées par des traits verts et couvrant la totalité du gène *FHIT*. La révélation des signaux de réplication permet de reconnaître les fourches divergentes et les fourches convergentes et de placer ainsi les initiations et les terminaisons.

#### 4.2. Déterminants génétiques

Les séquences d'ADN riches en G avec des motifs de type  $G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$  sont capables de former des structures en quadruplexe G (G4) (Huppert and Balasubramanian, 2005). Il a été montré que ces structures sont fréquemment présentes au niveau des promoteurs, des régions pauvres en nucléosomes ou des îlots CpG et qu'elles sont impliquées dans la régulation de la transcription (Halder et al., 2010). L'étude de Besnard et coll. sur quatre lignées cellulaires humaines différentes a montré que 90% des origines de réplication cartographiées par la technique des brins naissants comportent des motifs G4 et que l'efficacité de ces origines et leur déclenchement précoce sont corrélés au nombre de motifs G4 qu'elles contiennent. Des résultats similaires ont été obtenus en étudiant par la même technique les origines de réplication chez la souris, la drosophile et dans des cellules

humaines (Cayrou et al., 2012). D'après ces auteurs, les origines de réplication sont localisées à des endroits précis sans partager une séquence consensus stricte mais plutôt un élément riche en G4 qu'ils l'ont appelé *OGRE (Origin G-rich Repeated Element)*. Les auteurs rapportent que 90% des origines cartographiées chez la souris, 81% de celles cartographiées chez la drosophile et 90% de celles cartographiées dans les cellules humaines comportent un élément *OGRE*. A l'appui à cette observation, il a été montré que le complexe ORC humain, qui se fixe d'une façon aléatoire à l'ADN bicaténaire, se fixe préférentiellement au niveau des motifs G4 sur de l'ADN monocaténaire (Hoshina et al., 2013). De plus, dans une étude réalisée chez le poulet, il a été montré que les motifs G4 sont requis pour l'accumulation des brins naissants au niveau de deux origines de réplication (Valton et al., 2014). Dans cette même étude, il a aussi été rapporté que les motifs G4 doivent coopérer avec une séquence cis-régulatrice qui, d'après ces auteurs, pourrait potentiellement fixer des facteurs de transcription, introduisant un niveau de complexité supplémentaire dans la reconnaissance des origines par le complexe de pré-RC. Enfin, un autre point commun entre les motifs G4 et les origines de réplication est leur localisation préférentielle dans des régions pauvres en nucléosomes.

Malgré ces résultats associant la présence de motifs G4 et les origines de réplication, le sujet est toujours débattu. En effet, il a été rapporté que les séquences riches en GC sont résistantes à la digestion par l'exonucléase 5' du phage  $\lambda$  et pourrait donc être artificiellement surreprésentées dans les préparations de brins naissants (Conroy et al., 2010). De plus, la cartographie des origines par Mesner et coll. (Mesner et al., 2013) montre que 36% des origines ne contiennent pas de motifs G4 et que le taux d'enrichissement en motifs G4 des séquences identifiées dans cette étude n'est que de 1,05 par rapport à des régions contrôles. Il faut souligner qu'à nos jours les motifs G4 (ou l'élément *OGRE*) sont les seuls éléments de séquence qui ont été associés aux origines de réplication chez les eucaryotes supérieurs. Tous les autres éléments décrits sont de nature épigénétique.

#### 4.3. Marques épigénétiques

Les origines de réplication sont fréquemment localisées au niveau de promoteurs de transcription. Il est donc difficile de discriminer les marques épigénétiques associées à la transcription de celles associées à la réplication. Il est tentant de penser qu'une chromatine ouverte est plus accessible aussi bien aux protéines de réplication qu'aux facteurs de

transcription. Il est connu qu'une chromatine transcriptionnellement active est caractérisée par la présence de formes acétylées des histones H3 et H4 (H3ac, H4ac) ainsi que par la mono-, di- ou tri-méthylation de la lysine 4 de l'histone 3 (H3K4me1,2,3). Un tiers des origines cartographiées par Cadoret et coll. sont localisées dans des régions enrichies en au moins une des trois formes d'histone H3 modifiée : H3ac, H3K4me2 et H3K4me3 (Cadoret et al., 2008). Sequeira-Mendes et coll. ont trouvé une corrélation entre la localisation des origines de réplication chez la souris et celles des histones modifiées H3K9ac, H3K14ac et H3K4me3 (Sequeira-Mendes et al., 2009). De même, Martin et coll. ont trouvé une corrélation entre la localisation des origines chez l'homme et celles des histones H3K4me3 et H3K9ac (Martin et al., 2011). Karnani et coll. ont rapporté un enrichissement en histone H3K4me2, H3K4me3 et H3ac au niveau des origines (Karnani et al., 2010).

Le rôle de l'acétylation des histones dans le positionnement et l'efficacité des origines de réplication est un sujet débattu. Il a été montré que l'invalidation de Rpd3, une histone déacétylase, induit un déclenchement plus précoce des origines tardives chez *S. cerevisiae* (Vogelauer et al., 2002; Yoshida et al., 2014). Il a été aussi proposé que l'acétyltransférase HBO1 participerait au chargement des origines dans les cellules humaines. En effet, pendant la phase G1, HBO1 se lie à ORC1 et CDT1 et acétyle les lysines 5 et 12 de l'histone H4, ce qui engendre une décondensation de la chromatine et favorise le recrutement des complexes MCM (Miotto and Struhl, 2008). Un rôle similaire a été attribué à Chameau, l'orthologue de HBO1 chez la drosophile et à d'autres acétyltransférases (McConnell et al., 2012). L'explication proposée pour expliquer l'effet de l'acétylation est qu'elle provoque l'ouverture de la chromatine. Néanmoins, d'autres auteurs ont montré que HBO1 est capable d'acétyler ORC2, MCM2, CDC6 et la géminine *in vitro*, ce qui suggère un rôle de HBO1 n'impliquant pas une action sur l'état de la chromatine (Iizuka et al., 2006). De plus, les résultats de plusieurs études sont en contradiction avec l'explication avancée. Ces études montrent que beaucoup d'origines se trouvent dans des régions hypoacétylées (Cadoret et al., 2008). Enfin, une étude réalisée dans notre laboratoire a montré que la position et l'efficacité des origines au niveau du locus *AMPD2* dans des fibroblastes de hamster est sans rapport avec l'état d'acétylation local des histones (Gay et al., 2010).

Concernant le rôle de la méthylation des histones dans le positionnement et l'efficacité des origines, un certain nombre d'observations, parfois contradictoires, ont été rapportées. Les origines cartographiées par Martin et coll sont préférentiellement localisées

dans des régions enrichies en histone H3K79me2 (Martin et al., 2011). Néanmoins, une étude ultérieure a montré que la baisse de la méthylation de H3K79 suite à l'inactivation de DOT1L, la méthylase qui dépose cette marque, ne modifie pas la distance inter-origines mesurée par peignage moléculaire (Fu et al., 2013). Un enrichissement en histone H4K20me1 a été corrélé avec une efficacité élevée des origines cartographiées dans quatre types cellulaires différents (Picard et al., 2014). Il a été aussi montré que l'inactivation de l'histone méthyltransférase PR-Set7, responsable de la mono-méthylation de H4K20, diminue le recrutement à la chromatine de certaines protéines impliquées dans la construction des complexes de pré-RC (CDC6, MCM5, ORC2). Le recrutement forcé de PR-Set7 fusionné à un domaine Gal4 augmente localement le recrutement de certains facteurs des complexes de pré-RC (MCM2, MCM5, ORC1). De plus, il a été montré que l'accumulation d'une forme non dégradable de PR-Set7 induit de la re-réplication (Tardat et al., 2010). Une fois mono-méthylée par PR-set7, la position H4K20 est ensuite di- puis tri-méthylée par la méthyltransférase Suv4-20h. Il a été montré que le domaine BAH de ORC1 reconnaît spécifiquement la forme H4K20me2 (Kuo et al., 2012). Néanmoins, cette modification semble être trop répandue (80% des histones H4 la possède) pour que l'on puisse lui attribuer un rôle spécifique dans le positionnement des origines de réplication (Schotta et al., 2008). Il a été aussi montré que ORCA, un facteur qui interagit avec ORC1 et participe à son recrutement sur la chromatine, présente une affinité élevée pour la forme H4K20me3 (Beck et al., 2012).

#### 4.4. Transcription et démarrage de la réplication

En 1977, McKnight et coll. ont montré par microscopie électronique que la réplication démarre dans les régions non transcrites séparant les gènes des ARNr chez des embryons de drosophile (McKnight and Miller, 1977). Plus tard, cette même observation a été étendue aux cellules de xénope et aux cellules humaines (Bozzoni et al., 1981). Le locus *DHFR* chez le hamster a été utilisé comme modèle dans plusieurs études pour explorer le lien entre transcription et réplication. Le démarrage de la réplication de ce locus a lieu dans une région de 55 kb localisée entre les gènes *DHFR* et *2BE2121*. Il a été montré que, suite à la délétion de la région promotrice de la transcription du gène *DHFR*, la réplication démarre dans une zone plus large incluant la région normalement transcrite du gène *DHFR* (Saha et al., 2004). De plus, la délétion du terminateur de ce même gène permet d'étendre la région transcrite à



la région intergénique et supprime l'activation des origines dans cette dernière (Mesner and Hamlin, 2005).

Vingt pourcent des origines cartographiées par Karnani et coll. (Karnani et al., 2010) sont localisées à proximité d'un site de liaison pour un facteur de transcription connu. De même, les origines cartographiées par Cadoret et coll. (Cadoret et al., 2008) et Martin et coll. (Martin et al., 2011) sont préférentiellement localisées à proximité de sites de fixation pour les facteurs de transcription C-JUN et C-FOS. La ré-analyse des origines de réplication cartographiées par Cadoret et coll., par Swarnalatha et coll. (Swarnalatha et al., 2012) a montré que 43% des origines sont situées à proximité d'un site de fixation pour la protéine CMYC. Ces auteurs montrent que la fixation de CMYC au niveau de son site (E-box) dans le promoteur du gène de la lamine B2, permet le recrutement de la méthyltransférase MLL1 et augmente localement le niveau de l'histone H3K4me3. Ceci permet le recrutement de l'acétyltransférase HBO1 et favorise l'acétylation et l'ouverture de la chromatine et le recrutement des complexes MCM (Swarnalatha et al., 2012). Il a été aussi montré que la restriction de la réplication aux régions intergéniques est mise en place après la transition mid-blastuléenne chez l'embryon de xénope ou de drosophile (Hyrien et al., 1995).

#### 4.5. Association des origines actives avec la matrice nucléaire

La matrice nucléaire est un réseau protéique qui participerait au maintien de l'architecture nucléaire et à la régulation de la transcription et la réplication de l'ADN. Son existence et sa nature sont encore l'objet de discussions. Opérationnellement, elle correspond à la structure protéique qui persiste après élimination des histones de préparations de chromatine. Dans ces conditions, l'ADN nu reste attaché sous forme de boucles à la matrice par des séquences nommées MAR/SAR (Matrix/Scaffold attachment region). Une relation fonctionnelle entre la matrice ainsi définie opérationnellement et la réplication est maintenant bien documentée. Il a été montré que certaines origines très efficaces sont attachées à la matrice. C'est le cas de l'origine OriGNAI3 du locus *AMPD2* chez le hamster (Courbet et al., 2008), de l'origine de réplication de *CMYC* dans les cellules de xénope (Girard-Reydet et al., 2004), de celle de la lamine B2 chez l'homme (Lagarkova et al., 1998), de celle du locus *DHFR* chez le hamster (Dijkwel and Hamlin, 1988) et de trois des quatre origines du locus du gène de la dystrophine chez l'homme (Iarovaia et al., 2004). Dans notre laboratoire, il a été montré que OriGNAI3, l'origine majeure du locus *AMPD2*, est

associée à la matrice durant la phase G1. Cette origine est délocalisée dans la boucle en phase G2, très vraisemblablement suite à la progression de la réplication au cours de la phase S (Courbet et al., 2008). Il a aussi été montré que les cohésines, qui participent à la formation des boucles, sont localisées au niveau des complexes ORC chez la drosophile (MacAlpine et al., 2010) et que, dans les cellules humaines, les cohésines interagissent avec les protéines du complexe de pré-RC et sont associées à certaines origines (Guillou et al., 2010).

## 5. Programme temporel de réplication

Le programme temporel (timing) de réplication est un autre niveau de régulation du fonctionnement des origines de réplication. Il concerne le moment de déclenchement pendant la phase S des origines chargées en complexe de pré-RC. Ce timing est établi en fin de mitose et en début de phase G1 pendant un processus appelé le *timing decision point* (TDP). Le timing de réplication est sous la dépendance de plusieurs facteurs agissant en *trans* ou en *cis*. La régulation en *trans* implique les facteurs d'initiations de la réplication et celle en *cis* met en jeu la structure chromatinienne. Les facteurs agissant en *trans* sont les kinases de phase S (CDK et DDK) responsables du déclenchement des origines, ainsi que certains facteurs du complexe de pré-IC. En effet, il a été montré, aussi bien chez la levure que chez l'homme, que certains facteurs d'initiation sont présents en quantité limitante. Après leur intervention au niveau des origines précoces, ils doivent être recyclés pour une réutilisation au niveau des régions tardives. C'est le cas de Sld2, Sld3, Sld7 et Cdc45 (Mantiero et al., 2011; Tanaka et al., 2011). Il a aussi été proposé que plus il y a des complexes MCM chargés sur une origine, plus son déclenchement sera précoce (Yang et al., 2010).

Pour ce qui est du rôle de la structuration de la chromatine, les études du timing de réplication à grande échelle chez la souris et chez l'homme ont permis de mettre en évidence des domaines de l'ordre de 0,4 à 1 Mb au sein desquels le timing de réplication est coordonné. Ces domaines contiennent plusieurs réplicons qui se déclenchent en même temps. Ils sont séparés par des régions de transition qui sont de l'ordre de 0,1 à 0,6 Mb. Il a été proposé que ces zones de transition soient des zones sans origine de réplication, traversées par des fourches unidirectionnelles. Néanmoins, d'autres auteurs ont montré qu'il y a des origines dans ces zones qui s'activent en cascade au cours de la phase S (Guilbaud et al., 2011). D'une façon générale, les domaines de réplication précoce ont été

corrélés à l'existence d'une chromatine ouverte, accessible aux nucléases, à une richesse en gènes et en GC, ainsi qu'à une pauvreté en séquences répétées de type LINE. Par opposition, les domaines à réplication tardive correspondent plutôt à l'hétérochromatine pauvre en GC et en gènes et riche en séquences répétées. Néanmoins, ces caractéristiques de séquences ne dictent pas le timing de réplication puisque les deux allèles d'un même gène peuvent avoir des timing de réplication différents (Hansen et al., 2010).

Par ailleurs, la réplication précoce et la réplication tardive n'ont pas lieu dans les mêmes territoires nucléaires. Ceci a été montré par marquage de foyers de réplication chez des cellules isolées à différents stades de la phase S. Des données récentes d'analyse de l'organisation tridimensionnelle de la chromatine à l'intérieur du noyau par la technique de Hi-C ont confirmé que les domaines précoces interagissent entre eux et les domaines tardifs aussi, mais que les deux types de domaines n'interagissent pas entre eux (Lieberman-Aiden et al., 2009).

Il a été montré que l'inactivation d'un chromosome X dans les cellules femelles s'accompagne d'un décalage du timing de réplication vers un profil tardif, avec une augmentation de la compaction de la chromatine et un déplacement vers la périphérie nucléaire (Hiratani and Gilbert, 2010). Cette localisation périphérique ne concerne pas que le chromosome X. En effet, d'une façon générale, la réplication tardive a lieu à la périphérie nucléaire ou nucléolaire alors que la réplication précoce se fait à l'intérieur du noyau. Le rôle de ce positionnement dans le déterminisme du timing n'est pas encore clairement compris.

Environ la moitié du génome des mammifères est sujet à des changements de timing de réplication au cours de la différenciation cellulaire. Ce changement est accompagné d'un changement au niveau de la transcription (Ryba et al., 2010). Il faut noter qu'étant donné la taille des domaines de timing, plusieurs gènes peuvent avoir le même timing sans pour autant avoir la même régulation transcriptionnelle, ce qui rend l'analyse plus difficile (Hiratani et al., 2010). Une question reste sans réponse : est-ce que le timing de réplication est établi d'une façon passive par une certaine architecture de la chromatine ou est-ce qu'un mécanisme activement contrôlé parce que son existence a une fonction biologique ? Pourquoi la phase S dure-t-elle aussi longtemps alors qu'elle peut être beaucoup plus courte ? En effet, la phase S est de 20 minutes chez les embryons de xénope et de drosophile alors qu'elle est de 6 à 8 heures dans les cellules différenciées de l'adulte avec la mise en place du timing. Plusieurs hypothèses ont été proposées. Réduire le nombre de fourches

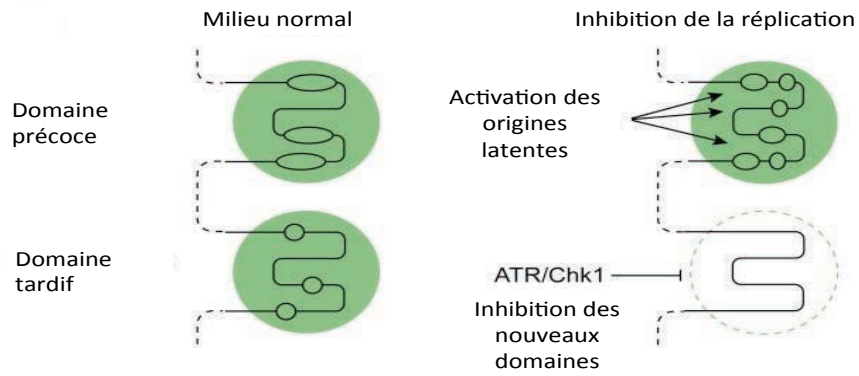
fonctionnant simultanément permettrait de limiter la quantité de protéines de réplication à synthétiser par unité de temps. Les cellules embryonnaires qui vivent sur un important stock de protéines maternelles avant le démarrage de la transcription à la transition mid-blastuléenne ne seraient pas confrontées à ce problème. La réduction du nombre de fourches permettrait aussi de réduire la dépense énergétique instantanée. Toutefois, la différence avec les cellules embryonnaire n'apparaît pas immédiatement évidente à ce niveau et il n'est pas sûr que l'économie énergétique réalisée soit significative par rapport aux dépenses énergétiques totales de la cellule. Réduire le nombre de fourches pourrait aussi permettre de réduire la quantité de dommages à l'ADN à réparer et donc réduire la charge des systèmes de réparation.

## **5. Les origines latentes**

Il n'y a que 10-20% des origines chargées en complexe de pré-RC qui se déclenchent pendant la phase S. Les origines chargées en pré-RC sont donc en très large excès par rapport aux besoins de la cellule. En accord avec cette idée, il est possible de réduire les niveaux des MCM de 3 à 10 fois sans gêner le déroulement de la phase S (Alver et al., 2014). En situation normale, la majorité des origines chargées sont répliquées passivement par le passage de fourches provenant d'origines proximales qui se sont déclenchées plus tôt. En condition de stress réplicatif, les origines latentes, qui ne se déclenchent pas pendant une phase S normale, deviennent indispensables à l'accomplissement de la réplication. L'impossibilité de charger des origines en complexe de pré-RC pendant la phase S justifie l'existence d'un excédant d'origines chargées en G1 afin de pouvoir en recruter en cas de stress réplicatif. En accord avec cette idée, il a été montré qu'il existe un point de contrôle du chargement des origines avant l'entrée en phase S. En effet, l'appauvrissement des cellules en protéines impliquées dans la construction des complexes de pré-RC (CDC6, MCM2-7 (Feng et al., 2003), CDT1, ORC (Machida et al., 2005) retarde l'entrée des cellules en phase S. La perte de ce point de contrôle peut engendrer une instabilité génétique. C'est le cas chez certaines cellules cancéreuses. Il a été trouvé que, dans les cellules qui ont activé le point de contrôle de chargement des origines, l'activité kinase des complexes CDK4-6/cycline D et CDK2/cycline E est basse et que le facteur Rb est hypophosphorylé (Liu et al., 2009). Ceci suggère que l'activation du point de contrôle de chargement des origines active

le point de restriction. Le mécanisme par lequel le défaut de chargement est détecté et signalé n'est pas encore connu. Néanmoins, il a été montré que l'absence de p53 abolit ce point de contrôle (Nevis et al., 2009).

En réponse à un stress réplicatif, la densité d'origines au sein d'un domaine de réplication actif augmente par recrutement d'origines latentes (Taylor, 1977) alors que, inversement, le déclenchement des origines dans les domaines plus tardifs qui n'ont pas encore démarré la réplication, est inhibé (Ge and Blow, 2010) (Figure 9). L'inhibition des domaines plus tardifs semble être une nécessité. Comme nous l'avons mentionné, certains facteurs de réplication sont limitant, comme la protéine RPA, et une activation excessive d'origines engendre des dommages dans l'ADN (Toledo et al., 2013). En ce qui concerne le mécanisme qui contrôle le déclenchement des origines latentes, il reste très mal connu. L'hypothèse la plus simple pour expliquer le déclenchement des origines latentes au sein d'un domaine en réplication est que ces origines vont avoir le temps de se déclencher suite au ralentissement des fourches provenant des origines flanquantes. Néanmoins, d'autres mécanismes plus actifs ont été proposés faisant intervenir ATR. En effet, ATR phosphoryle les complexes MCM (Cortez et al., 2004) mais aucun rôle d'ATR dans le déclenchement des origines n'a été établi à ce jour.



**Figure 9 : Représentation schématique de la régulation des origines latentes en réponse à un stress répliatif.** Les cercles verts représentent des domaines de réplication précoce ou tardive. A l'intérieur des domaines, les origines de réplication actives sont représentées par des bulles. Sous stress répliatif, la densité des origines actives au sein du domaine précoce augmente après activation des origines latentes. Par contre, les origines de réplication au sein du domaine tardif sont inhibées par ATR et CHK1. Adapté de (Ge and Blow, 2010).

# Matériels et Méthodes

---

## I. Culture cellulaire

Les cellules JEFF (lymphocytes B humains immortalisés par le virus d'Epstein Barr) ainsi que toutes les autres lignées lymphoblastoïdes (LLD138, BJ-Lympho, LLD43, LLD86, HeDe, NaDe) sont cultivées en suspension à 37°C en atmosphère humide contenant 5% de CO<sub>2</sub> en milieu RPMI1640 additionné de glutamax (GIBCO Invitrogen), de 10% de sérum de veau foetal (SVF), et de 1% pénicilline et streptomycine. Les cellules MRC5 (fibroblastes humains primaires de poumon) ainsi que toutes les autres lignées de fibroblastes (MRC5SV2, BJ-fibro, BJ-hTERT, ICIG7, IMR90, WI38) sont cultivées à 37°C en atmosphère humide contenant 5% de CO<sub>2</sub> et 3% O<sub>2</sub> en milieu MEM contenant 10% de sérum de veau foetal, des acides aminés non essentiels, 1 mM de pyruvate de sodium, 2 mM de glutamine et 1% de pénicilline et streptomycine.

## II. Cytométrie en flux

Avant récupération, les cellules sont incubées à 37°C pendant 20 min en présence de 30 µM de Bromo-déoxyUridine (BrdU). Elles sont ensuite centrifugées 5 min à 2800 rpm et resuspendues dans 1 ml de tampon Phosphate Buffered Saline (PBS : 20 mM tampon phosphate pH 7,4, 150 mM NaCl). La suspension cellulaire est fixée par ajout de 3 ml d'éthanol froid. Après avoir été conservées 24 h à 4°C, les cellules sont centrifugées puis resuspendues dans une solution de pepsine (0,5 mg/ml pepsine, 0,03 N HCl). Après 20 min d'incubation à 37°C, les cellules sont centrifugées puis resuspendues dans 500 µl de HCl 2 N et incubées 20 min à température ambiante. Elles sont alors lavées en PBS, centrifugées 10 min à 2800 rpm puis resuspendues dans 1 ml de tampon Bu (0,5% Tween 20, 0,5% sérum de chèvre, 20 mM Hepes pH 7,3). Les cellules sont récupérées par centrifugation et incubées pendant 45 min dans une solution d'anticorps anti-BrdU (AbD Serotec OBT 300 dilué au 4/100<sup>ème</sup> dans du tampon Bu) puis pendant 30 min dans une solution d'anticorps secondaire (poulet anti-IgG de rat-Alexa 488, Invitrogen, au 2/100<sup>ème</sup> dans du tampon Bu). A la fin, les cellules sont resuspendues dans une solution d'iodure de propidium (25 µg/ml) avant d'être analysées par cytométrie en flux (FACS Sort). Les résultats sont analysés à l'aide du logiciel

FlowJo (Tree Star, Inc).

### III. Etalement des métaphases

Après 2 ou 3 h de traitement au nocodazole (0,2  $\mu$ M), les cellules sont collectées par centrifugation (3 min, 1500 rpm à 20°C). Elles sont ensuite incubées 10 min à 37°C dans une solution hypotonique (14% de sérum de veau fœtal, 11 mM KCl), puis fixées par resuspension dans une solution contenant 25% d'acide acétique et 75% d'éthanol (fixateur) et conservées la nuit à -20°C. Les cellules fixées sont suspendues dans 500  $\mu$ l de fixateur et les métaphases sont étalées en faisant tomber des gouttes de la suspension sur des lames de microscope. Après séchage à température ambiante, les lames qui servent au comptage des cassures totales sont incubées pendant 10 min dans une solution de coloration (Giemsa R solution, Réactif RAL, dilué à 6% dans du tampon phosphate pH 7,0), puis lavées à l'eau distillée et séchées à l'air.

### IV. Hybridation *in situ* en fluorescence sur métaphases

*Préparation des BAC* : les ADN des BAC sont préparés à partir de cultures bactériennes à l'aide du kit NucleoBond XtraMidi Plus (Macherey-Nagel).

*Préparation des sondes* : Les ADN des BAC sont marqués par la biotine à l'aide du kit Bioprime DNA labeling system (Invitrogen). Un mélange des ADN des BAC (200 ng de chaque) correspondant à chaque site est préparé (3 BAC pour *FRA3B* et 2 BAC pour *FRA16D*, voir tableau). A ce mélange sont ajoutés 200  $\mu$ g d'ADN compétiteur (human Cot-1 DNA, Invitrogen). Le mélange est précipité à l'éthanol pendant la nuit à -20°C. Les sondes précipitées sont centrifugées (30 min à 13000 rpm et 4°C), rincées à l'éthanol 70% et puis centrifugées de nouveau 15 min dans les mêmes conditions. Le culot est séché afin d'éliminer toute trace d'éthanol et resuspendu dans 20  $\mu$ l de tampon TE (Tris-Cl 10 mM pH 8,0 et EDTA 1 mM). Afin de révéler les centromères des chromosomes d'intérêt, 8  $\mu$ l (13 ng) de sondes correspondant aux séquences alpha satellite du chromosome 1 (Aquarius probes LPE 01 R), du chromosome 3 (Aquarius probes LPE 03 R) ou du chromosome 16 (Aquarius probes LPE 16 R) (Cytocell) sont ajoutées. Le mélange est ajouté à 68  $\mu$ l de solution d'hybridation contenant 10% dextran sulfate, 50% formamide, 2xSSC et 1% de Tween 20 (SSC : Standard Saline Citrate ; 15 mM citrate de sodium, 150 mM NaCl, pH 5,6). Après dénaturation des ADN par chauffage à 75°C pendant 5 min, le mélange est incubé pendant



20 min à 37°C pour permettre la renaturation des séquences répétées des BAC avec celles de l'ADN compétiteur.

*Traitement des lames* : Les lames portant les métaphases sont incubées pendant une heure à 37°C dans une solution de 2xSSC contenant 0,1 mg/ml de RNase A. Après 2 lavages en 2xSSC, elles sont déshydratées par 3 bains successifs d'éthanol à 70%, 85% et 100%. Une fois sèches, les lames sont incubées dans une solution contenant 70% formamide et 2xSSC pendant 2 min à 72°C pour dénaturer l'ADN des chromosomes. La dénaturation est suivie de 3 bains successifs d'éthanol froid à 70%, 85% et 100%. Après séchage des lames, 30 µl des sondes sont déposés sur chaque lame. Les lames sont incubées en chambre humide à 37°C pendant la nuit. Après 3 lavages en 50% de formamide et 2xSSC à 45°C suivis de 3 autres en 2xSSC à 45°C et d'un lavage en 4xSSC et 0,05% Tween 20, les lames sont incubées 30 min à 37°C dans un tampon contenant 4xSSC, 0,05% Tween 20 et 5% de lait en poudre écrémé afin de prévenir la fixation des anticorps sur des sites aspécifiques. L'immunodétection des sondes se fait grâce à la liaison streptavidine-biotine. Cinq couches sont réalisées en utilisant alternativement de la streptavidine couplée au fluorochrome Alexa 488 (Invitrogen, 1/100<sup>ème</sup>) et de l'anti-streptavidine de lapin biotinylé (Rockland, 1/50<sup>ème</sup>) dilués dans le tampon de blocage. Pour chaque couche, les lames sont incubées en chambre humide 20 min à 37°C puis lavées trois fois pendant 5 min dans une solution contenant 4xSSC et 0,05% Tween 20. Enfin, les lames sont montées avec du Vectashield (Vector labs) contenant du DAPI pour marquer l'ADN des chromosomes.

<b>Site fragile</b>	<b>BAC</b>
<b>FRA3B</b>	RP11-32J15 (AC104161)
	RP11-641C17 (AC104164)
	RP11-147N17 (AC104300)
<b>FRA16D</b>	RP11-105F24 (AC106741)
	RP11-57106 (AC092724)
<b>FRA1L</b>	RP11-88425
	RP11-729G19
<b>FRA3L</b>	RP11-59m6
	RP11-10915
	RP11-739D3

## V. Marquage des fourches de réplication et préparation des blocs d'ADN pour le peignage

Les cellules JEFF sont traitées ou non à l'aphidicoline (0,6  $\mu\text{M}$ ) en présence ou en absence de RO-3306 (5  $\mu\text{M}$ ) pendant 16 h. Pendant le temps de traitement, deux marquages courts (30 min ou 2 h) de l'ADN en cours de synthèse sont réalisés grâce à deux analogues de la thymidine, l'IododéoxyUridine (IdU : 20  $\mu\text{M}$ ) et la ChlorodéoxyUridine (CldU : 100  $\mu\text{M}$ ). En absence d'aphidicoline, le temps de marquage pour chaque analogue est de 30 min mais en sa présence, il est de 2 h. Cette adaptation du temps de marquage est nécessaire dans les conditions de ralentissement important de la vitesse des fourches afin d'avoir des signaux de réplication analysables. A la fin des 16 h, l'incorporation d'IdU et de CldU est arrêtée par un excès de Thymidine (T : 300  $\mu\text{M}$ ). Les cellules sont ensuite récupérées, comptées, suspendues dans du PBS et additionnées d'un volume d'une solution d'agarose low melting à 2% (Sigma-Aldrich) dans du PBS conservée à 45°C. Le mélange cellules/agarose est distribué dans des puits spéciaux à raison de 300 000 cellules/puit (80  $\mu\text{l}$ ). Les puits sont ensuite placés 15 min à 4°C afin de permettre la solidification de l'agarose. Les blocs d'ADN sont transférés dans une solution de protéinase K (Boldog et al.) (1 mg/ml dans 1% SDS et 0,25 M EDTA) puis incubés pendant 48 h à 42°C. Les blocs sont ensuite récupérés et lavés dans une solution de PMSF (Phenyl MethaneSulfonyl Fluoride - Sigma) (1 mM dans du TE), un inhibiteur de la protéinase K, pendant 20 min à 42°C, puis deux fois dans du tampon TE pendant 20 min à température ambiante. L'ADN est ensuite coloré par incubation pendant 2 h dans une solution de YOYO-1 à 3  $\mu\text{M}$  dans du TE (Invitrogen). Après fusion de l'agarose par incubation 15 min à 68°C, il est digéré par addition d'une solution de  $\beta$ -agarase (2,5 unités/ml) (Biolabs) pendant 48 h à 42°C. Enfin, l'ADN solubilisé est dilué 5 fois avec une solution de MES (2-(N-Morpholino)-ethane-sulfonic acid, Research organics) 250 mM à pH 5,5 et stocké à 4°C pendant au moins une semaine avant de procéder au peignage. L'ADN est peigné à l'aide d'une machine Molecular Combing System (Genomic Vision). Des lamelles silanisées sont plongées verticalement dans la solution d'ADN, puis ressorties à vitesse constante (300  $\mu\text{m/s}$ ). Ce système conduit à un étirement régulier des molécules d'ADN (2 kb/ $\mu\text{m}$ ) sur les lamelles au passage du ménisque formé à l'interface air/eau.

## **VI. Immunodétection des signaux de réplication et mesure de la vitesse de progression des fourches de réplication**

L'ADN peigné est dénaturé par immersion des lamelles dans un bain de NaOH 1 N pendant 5 min. Elles sont ensuite traitées par la solution de blocage (15 mg/ml de blocking reagent (Boehringer Mannheim) dans du PBS contenant 0,05% Tween 20). Les analogues incorporés sont ensuite révélés dans l'ADN peigné par immunodétection spécifique. Les lamelles sont incubées pendant 1 h en présence d'anticorps de souris anti-BrdU (AbD serotec, 1/5<sup>ème</sup>) détectant l'IdU et de rat anti-BrdU (Oxford Technology, 1/50<sup>ème</sup>) détectant la CldU. Les anticorps sont dilués dans la solution de blocage. Après lavage en 0,5 M NaCl, 20 mM Tris-Cl pH 7,8 et 0,05% Tween 20, les lamelles sont incubées avec les anticorps secondaires anti-IgG de souris couplés à l'Alexa 488 (Molecular Probes, 1/50<sup>ème</sup>) et anti-IgG de rat couplés à la Cyanine 3 (Molecular Probes, 1/50<sup>ème</sup>). Les molécules d'ADN monocaténaire sont ensuite révélées par la succession de trois couches d'anticorps : un anticorps de souris anti-ADN monocaténaire (Millipore, 1/25<sup>ème</sup>), suivi d'un anti-IgG de souris fabriqué chez la chèvre et couplé à la Cyanine 5 (Abcam, 1/100<sup>ème</sup>) puis d'un anti-IgG de chèvre préparé chez le singe, couplé à la Cyanine 5 (Abcam, 1/100<sup>ème</sup>). Toutes les couches d'anticorps (20 min à 37°C en chambre humide) sont séparées par 3 lavages de 5 min en PBS. La vitesse de progression des fourches est déterminée en mesurant la longueur des signaux de réplication. Pour chaque signal, la vitesse de réplication correspondante est calculée en kb/min selon la formule suivante :  $Vitesse = (longueur\ du\ signal\ (\mu m) \times 2) (kb) / temps\ de\ marquage\ (min)$ .

## **VII. Hybridation *in situ* en fluorescence sur ADN peigné et immunodétection des signaux de réplication**

*Préparation des sondes:* Les ADN des BAC sont préparés comme décrit ci-dessus puis les 32 sondes du code morse sont préparées par PCR à l'aide du kit Expand 20 kb plus PCR system dNTP (Boldog et al.) comme décrit dans Letessier et coll. (Letessier et al., 2011). Les sondes du motif i ont été préparées à partir du BAC RP11-641C17, en utilisant 3 couples d'amorces :  
CATCACTCCAAGACATAGTTTCAGGT  
CATCAACATGGAGATTACAGGACCT  
GCCTGAGGTGACTCACTTAAGCA  
TCTGGTCTCTTACAAGTGAAGGGTT

GGCCTCAGCAAGTGCTAAGAGA

GGTGAAGTCTCCTAGCGTTAGCTCT

Les produits de PCR sont marqués par la biotine à l'aide du Kit Bioprime DNA labeling system (Invitrogen). Un mélange des 32 produits de PCR représentant tous les motifs (a - e, i) (50 ng de chaque sonde) est préparé. A ce mélange sont ajoutés 10 µg d'ADN compétiteur (human Cot-1 DNA, Invitrogen). Le mélange est précipité à l'éthanol pendant la nuit à -20°C. Les sondes précipitées sont centrifugées (30 min à 13000 rpm et 4°C), rincées à l'éthanol 70% puis centrifugées de nouveau 15 min dans les mêmes conditions. Le culot est séché afin d'éliminer toute trace d'éthanol et resuspendu dans 8,8 µl de tampon TE. Le mélange est ajouté à 21,2 µl de solution d'hybridation (10% dextran sulfate, 50% formamide, 2xSSC et 1% de Tween 20). L'ADN est dénaturé par chauffage à 75°C pendant 10 min.

*Traitement des lames:* Les molécules d'ADN peignées sont dénaturées dans un bain de NaOH 1 N pendant 5 min. La dénaturation est suivie d'un bain de PBS froid, puis de 3 bains successifs d'éthanol froid à 70%, 85% et 100%. Après séchage des lames, 30 µl du mélange des sondes sont déposés sur chaque lame. Les lames sont incubées en chambre humide à 37°C pendant la nuit. Après 3 lavages en 50% formamide et 2xSSC pH 5,6 suivis de 3 autres en 2xSSC pH 7,0 et d'un lavage en PBS, les lames sont incubées 1 h à 37°C dans la solution de blocage (15 mg/ml de blocking reagent (Boehringer Mannheim) en PBS contenant 0,05% de Tween 20). L'immunodétection des sondes se fait grâce à la liaison streptavidine-biotine. Cinq couches sont réalisées en utilisant alternativement de la streptavidine couplée au fluorochrome Alexa 488 (Invitrogen, 1/100<sup>ème</sup>) et de l'anti-streptavidine de lapin biotinylé (Rockland, 1/50<sup>ème</sup>) dilué dans le tampon de blocage. A partir de la troisième couche, les analogues de la thymidine (IdU, CldU) incorporés sont révélés par immunodétection spécifique. Les lamelles sont incubées en chambre humide pendant 1 h en présence d'anticorps de souris anti-BrdU (BD biosciences, 1/5<sup>ème</sup>) et de rat anti-BrdU (Abd serotec, 1/25<sup>ème</sup>). Après lavage en 0,5 M NaCl, 20 mM Tris-Cl pH 7,8 et 0,05% Tween 20, les lamelles sont incubées avec les anticorps secondaires anti-IgG de souris fabriqué chez la chèvre couplés à l'Alexa 350 (Invitrogen, 1/25<sup>ème</sup>) et anti-IgG de rat couplés au Texas Red (Jackson Immunoresearch, 1/50<sup>ème</sup>), puis en présence d'un anticorps anti-IgG de chèvre couplé à l'Alexa 350 (Invitrogen, 1/25<sup>ème</sup>). L'ADN monocaténaire est ensuite révélé par la succession de trois couches d'anticorps : un anticorps de souris anti-ADN monocaténaire (Millipore, 1/50<sup>ème</sup>), suivi d'un anti-IgG de souris fabriqué chez la chèvre couplé à la Cyanine

5 (Abcam, 1/100<sup>ème</sup>) et d'un anticorps anti-IgG de chèvre préparé chez le singe couplé à la Cyanine 5 (Abcam, 1/100<sup>ème</sup>). Enfin, les lames sont montées avec du Vectashield (Vector labs) sans DAPI et examinées au microscope à fluorescence. Les images des fibres d'ADN portant des motifs du code morse et des signaux de réplication sont ensuite traitées avec Photoshop (Adobe) et modélisées avec Illustrator (Adobe).

### VIII. Transfection des cellules JEFF

Des cellules JEFF en croissance exponentielle sont centrifugées puis suspendues à 20x10<sup>6</sup> cellules par ml dans du tampon C du kit Cell Line Nucleofector (Lonza). Deux millions de cellules sont ajoutées au mélange des siARN (0,6 nmol de chaque, voir tableau) et électroporées à l'aide de l'Amaxa nucleofector II en utilisant le programme Z-001 de la machine. Après électroporation, les cellules sont remises immédiatement dans 10 ml de milieu de culture à 37°C.

siARN	Référence
<b>Contrôle</b>	All Stars Negative Control siRNA (SI03650318) <b>Qiagen</b>
<b>CDT1</b>	Stealth siRNAs (Set of 3) HSS129994, HSS188618, HSS188619, <b>Invitrogen</b>
<b>ORC1</b>	Stealth siRNAs (Set of 3) HSS107467, HSS107468, HSS107469, <b>Invitrogen</b>
<b>CDC6</b>	Stealth siRNAs (Set of 3) HSS101647, HSS101648, HSS101649, <b>Invitrogen</b>
<b>RIF1</b>	FlexiTube GeneSolution GS55183 (set of 4) SI04349037, SI04186616, SI04172910, SI04150657, <b>Qiagen</b>

### IX. Immunoprécipitation de la chromatine

*Préparation de la chromatine* : Des cellules JEFF non traitées, traitées à l'aphidicoline (0,6 µM), au RO-3306 (5 µM) ou à l'aphidicoline et au RO-3306 pendant 16 h sont récupérées, comptées et resuspendues dans du PBS (5x10<sup>6</sup> cellules dans 500 µl). Elles sont ensuite fixées en présence de 1 % de formaldéhyde pendant 10 min à température ambiante avec agitation. Le formaldéhyde est ensuite neutralisé en ajoutant 1,25 M de glycine. Les cellules sont ensuite lavées deux fois avec du PBS froid, puis resuspendues à 100x10<sup>6</sup> cellules/ml dans le tampon de lyse (50 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0,1 % déoxycholate de sodium et 0,5 % SDS) en présence d'inhibiteurs de protéases. Après 30 min à 4°C, le lysat est dilué avec 500 µl de tampon de lyse et les tubes sont soumis à une sonication à froid pendant 15 min. Une partie de la chromatine est traitée à la protéinase K et à la RNase A afin d'analyser par électrophorèse dans un gel d'agarose 1,5 %

la taille des fragments d'ADN obtenus (200-500 pb). Une autre partie de la chromatine préparée ne subit pas l'immunoprécipitation. Elle servira de référence pour déterminer l'abondance des séquences d'intérêt dans la chromatine de départ (input).

*Immunoprécipitation de la chromatine* : Dans un premier temps, les billes sont couplées aux anticorps qui vont être utilisés. Pour cela 50 µl de billes portant de la protéine A (Millipore 16-157) sont resuspendues dans du PBS contenant 2% de BSA puis 5 µg d'anticorps de lapin non-immun (ab37415-5) sont ajoutés dans les tubes des contrôles négatifs et 3 µg d'anticorps dirigés contre la forme phosphorylée de la sérine 5 du domaine C-terminal de l'ARN polymérase II (ab 5131) sont ajoutés dans les tubes de test. Les billes sont incubées en présence des anticorps pendant 2 h à 4°C. La chromatine correspondant à  $2 \times 10^6$  cellules est ensuite ajoutée aux billes couplées aux anticorps et les tubes sont incubés sous agitation à 4°C pendant la nuit. Les complexes (billes + anticorps + chromatine) ainsi formés subissent une série de lavages à 4°C : 1) dans le tampon de lyse ; 2) dans le tampon de lyse contenant 500 mM NaCl au lieu de 150 mM ; 3) dans le tampon LiCl (0,25 M LiCl, 0,5 % NP40, 1 mM EDTA, 10 mM Tris-Cl pH8,0 0,5 % sodium déoxycholate) ; 4) dans le tampon TE.

*Réversion du couplage au formaldéhyde* : les tubes issus de l'immunoprécipitation et les tubes contenant la chromatine de départ (input) sont additionnés de 125 µl de tampon de réversion (25 mM Tris-Cl pH8,0 5 mM EDTA, 0,5% SDS) puis incubés sous agitation pendant 20 min à 65°C. Après centrifugation pour éliminer les billes, les surnageants sont incubés en présence de 100 µg/ml de protéinase K pendant 2 h à 65°C puis en présence de 15 µg/ml de RNase A pendant 30 min à 37°C. Les ADN sont ensuite purifiés grâce au *Quiaquick PCR purification Kit* de Qiagen et élués dans 100 µl d'eau. Les séquences d'intérêt sont quantifiées par PCR quantitative (qPCR) en utilisant des amorces introniques au niveau des gènes d'intérêt. Les quantités sont déterminées grâce à une droite de référence obtenue en se servant de quantités croissantes connues d'ADN soniqué extrait de cellules JEFF. Elles sont normalisées par rapport aux quantités déterminées dans les *input* correspondants. Le résultat final est exprimé en taux d'enrichissement, calculé comme étant le rapport de la quantité normalisée obtenue pour le test à celle obtenue pour le contrôle négatif.

## **X. La technique de Repli-Seq**

*Marquage au BrdU et tri cellulaire* : Les cellules JEFF non traitées sont incubées

pendant 1 h en présence de 50  $\mu\text{M}$  de BrdU et les cellules traitées avec 0,6  $\mu\text{M}$  d'aphidicoline pendant 16 h sont incubées en présence de 50  $\mu\text{M}$  de BrdU pendant les trois dernières heures du traitement. Les cellules sont ensuite récupérées, comptées et resuspendues à  $10 \times 10^6$  cellules/ml dans du milieu RPMI sans sérum. Elles seront incubées pendant 20 min à 37°C en présence de 10  $\mu\text{g/ml}$  du Hoechst 33342 (life technologies H3570). Les cellules sont ensuite reconcentrées à  $75 \times 10^6$  cellules/ml par centrifugation avant de passer au trieur. Le tri se fait à vitesse élevée à l'aide d'une machine BD FACS Aria II. L'ensemble des cellules à trier est distribué en 3 tubes, chaque tube servant pour le tri de 2 fractions (G1b et S1, S2 et S3, S4 et G2M). Les fenêtres de tri sont positionnées par rapport à la distance qui sépare le pic G1 de celui du pic G2. La largeur des fenêtres de tri des fractions G1b, S1, S2, S3, S4 correspond à la division de cette distance par 6. La largeur de la fenêtre de la fraction G2M correspond au double de cette valeur. Les cellules sont récupérées dans du PBS.

*Extraction et immunoprécipitation de l'ADN* : Après centrifugation, les culots cellulaires sont resuspendus dans un tampon contenant 50 mM Tris-Cl pH 8,0, 10mM EDTA, 0,1% SDS et 50  $\mu\text{g/ml}$  RNase A puis incubés pendant une heure à 37 °C. Une digestion à la protéinase K (100  $\mu\text{g/ml}$ ) est ensuite effectuée pendant la nuit à 37°C en présence de 1 % SDS. L'ADN des différentes fractions est ensuite extrait par un mélange phénol-chloroforme (50/50) et précipité par addition de 100 mM NaCl et 2 volumes d'éthanol et maintenu pendant au moins 2 h à -20°C. L'ADN précipité est collecté par centrifugation, lavé une fois à l'éthanol 70%, séché et dissout dans du tampon TE pour le doser au NanoDrop. L'ADN est ensuite fragmenté par sonication pendant 10 min. La taille des fragments d'ADN après sonication est vérifiée par migration sur un gel d'agarose 1,5% (fragments entre 200-500 pb). Deux  $\mu\text{g}$  d'ADN de chaque fraction subit une réaction de ligation des adaptateurs. Cette étape est réalisée par l'équipe de Claude Thermes. Les ADN issus de la ligation sont ensuite dilués dans du PBS contenant 0,1% de Tween 20 et dénaturés par chauffage pendant 3 min à 98°C afin de rendre accessible à l'immunoprécipitation l'ADN néosynthétisé. Celle-ci est réalisée en incubant, pendant 30 min à température ambiante, l'ADN dénaturé en présence de 2  $\mu\text{g}$  d'anticorps dirigé contre le BrdU (BD 347580). Ensuite, 30  $\mu\text{g}$  d'anticorps secondaires dirigés contre les IgG de souris et produit chez le lapin (Sigma M7023) sont ajoutés et une nouvelle incubation de 30 min à température ambiante est effectuée. Les complexes ainsi formés entre les séquences d'ADN contenant du BrdU et les deux anticorps sont récupérés par

centrifugation, lavés dans du PBS contenant 0,1% de Tween 20 puis redissouts dans le tampon de digestion à la protéinase K (10 mM Tris-Cl pH 8,0, 10 mM EDTA, 0,5% SDS, 100 µg/ml protéinase K) et incubés la nuit à 37°C. L'ADN des différentes fractions est extrait par un mélange phénol-chloroforme (50/50) après addition de 2 µg de glycogène (Boehringer Mannheim, 901 393). Il est ensuite précipité par addition de 100 mM NaCl et 2 volumes d'éthanol et maintenu pendant au moins 2 h à -20°C. Après centrifugation, les culots d'ADN sont lavés à l'éthanol 70%, séchés et redissouts dans du tampon TE puis envoyés à la plateforme (Imagif) à Gif-Sur-yvette pour le séquençage. L'alignement des séquences et l'analyse des données sont réalisés par l'équipe de Claude Thermes.

## **XI. Transfert Western**

Des cellules JEFF non traitées, ou traitées à l'aphidicoline (0,6 µM), au RO-3306 (10 µM) ou à l'aphidicoline et au RO-3306 sont récupérées après 16 h de traitement. Elles sont soit préparées pour être triées en quatre fractions (G1, S1, S2, G2M), soit utilisées pour la préparation des extraits chromatinien de cellules non triées. En ce qui concerne le tri cellulaire, les cellules sont préparées de la même façon que dans le cas du Repli-Seq. Des extraits totaux ou chromatinien sont préparés à partir des quatre fractions récupérées après tri.

Pour préparer la chromatine (cellules triées et cellules non triées), le culot cellulaire est resuspendu dans un tampon contenant 0,25 % NP-40, 10 mM Tris-Cl pH 8,0, 10 mM NaCl, 10 mM MgCl<sub>2</sub> et des inhibiteurs de protéases puis incubé pendant 5 min dans la glace. Ce tampon provoque la lyse de la membrane plasmique. Après centrifugation, les noyaux sont resuspendus dans un tampon contenant 1 % Triton-X100, 20 mM Hepes, 0,35 M NaCl, 10 mM KCl, 1 mM EDTA, 20 % glycérol et des inhibiteurs de protéases puis incubés pendant 30 min dans la glace, ce qui provoque leur lyse. Une centrifugation permet de séparer la chromatine de la fraction nucléaire soluble. Les culots chromatinien sont resuspendus dans le tampon de charge (Biolabs B7709S) (1% SDS, 10 mM DTT, 10 mM MgCl<sub>2</sub>) puis incubés pendant 30 min à température ambiante en présence de 25 U/µl de benzonase (Millipore D0017) pour digérer l'ADN. Avant d'être déposé sur gel de polyacrylamide-SDS, les échantillons sont chauffés pendant 5 min à 95°C.

Pour préparer des extraits protéiques totaux à partir des quatre fractions des cellules triées, les culots cellulaires récupérés après le tri sont dissouts directement dans le tampon



de charge et traités et analysés comme décrit ci-dessus.

Pour l'électrophorèse des protéines, nous utilisons des gels pré-coulés de *Life Technologies* (gels 4-12% NP0322) et le tampon de migration correspondant (MOPS NP0001). Le transfert sur membrane de nitrocellulose est réalisé à l'aide de la machine *blot* de *Life Technologies* en utilisant un kit de transfert (*Life Technologies* IB301001) et le programme 3 de la machine. Les membranes sont ensuite saturées par incubation pendant 1 h à température ambiante en PBS contenant 0,05% de Tween 20 et 5% de BSA. L'incubation avec les différents anticorps primaires (voir tableau) est effectuée à 4°C pendant la nuit. Celle avec les anticorps secondaires couplés à la peroxydase de Raifort est faite pendant une heure à température ambiante. Tous les anticorps (primaires et secondaires) sont dilués dans du PBS contenant 0,05% de Tween 20 et 5% de BSA. L'incubation avec les anticorps (primaires ou secondaires) est suivie de trois lavages de 10 min en PBS contenant 0,05% de Tween 20. La révélation du signal de chimiluminescence est faite à l'aide du kit WesternBright ECL de Advansta (K-12045).

	<b>Anticorps</b>	<b>Référence</b>	<b>Dilution</b>
<b>Anticorps primaires</b>	ORC 1	Abcam ab85830	1/2000
	CDC6	Cell signaling 3387	1/1000
	CDT1	Cell Signaling 8064	1/1000
	MCM7	Santa Cruz sc-56324	1/1000
	Histone H4	Millipore 07-108	1/1000
	β-actine-HRP	Santa Cruz sc-47778	1/20000
	RIF1	Bethyl A300-569A	1/1000
<b>Anticorps secondaires</b>	Anti-Souris-HRP	Dako P0447	1/5000
	Anti-Lapin-HRP	Dako P0448	1/5000

## **XII. Capture des ARN naissants**

L'ARN naissant est isolé grâce à la technique de Click-It en utilisant le kit *Click-iT Nascent RNA Capture* d'Invitrogen (10365). Les cellules non traitées ou traitées 16 h à l'aphidicoline (0,6 μM), au RO-3306 (10 μM) ou à l'aphidicoline et au RO-3306 sont incubées

pendant les dernières 30 min en présence de 1 mM d'un analogue de l'uridine : le 5-éthynyl-uridine (EU). Après extraction de l'ARN total, ce dernier subit une réaction catalysée par le cuivre qui lie la biotine par l'intermédiaire d'un groupement azide à l'EU. Les ARN naissants sont alors purifiés par chromatographie d'affinité grâce à des billes magnétiques couplées à la streptavidine (Dynabeads MyOne Streptavidin T1 (Invitrogen 65601)). La réaction de réverse transcription pour la synthèse de l'ADNc est faite à l'aide du kit *Superscript VILO cDNA Synthesis* (Invitrogen 11754) directement sur l'ARN capturé sur les billes. La PCR quantitative est réalisée en utilisant des amorces introniques. Pour l'ARN total, la réaction de réverse transcription est réalisée en utilisant le Kit SuperScript III First-Strand (Invitrogen 18080).

# Résultats

---

## **I. Programme temporel de la réplication sous aphidicoline**

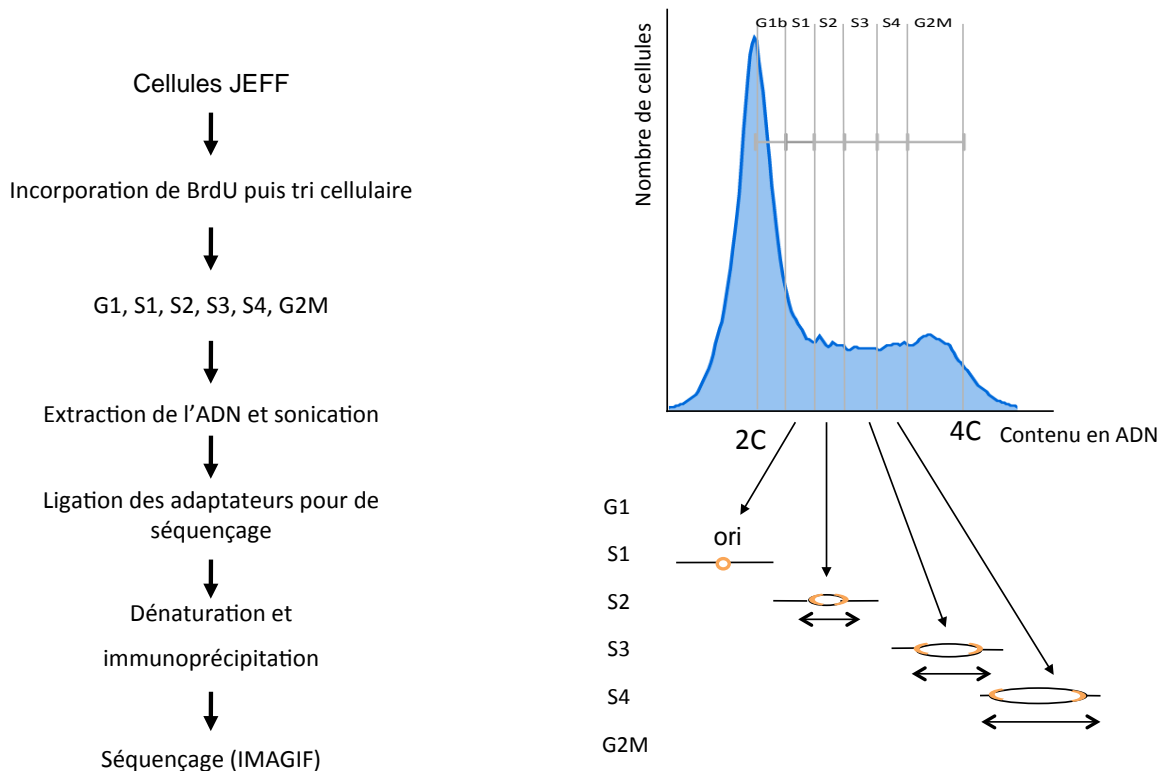
Les résultats obtenus par peignage moléculaire dans le laboratoire montrent qu'en absence de stress réplicatif, les évènements d'initiation de la réplication au niveau du gène *FHIT* sont exclus d'une région longue d'environ 700 kb (région cœur) chevauchant la zone la plus instable de *FRA3B* dans les cellules lymphoblastoïdes JEFF, alors qu'ils surviennent tout au long du locus dans les cellules fibroblastiques MRC5. Par contre, des évènements de terminaison sont visibles tout au long du locus dans les deux types cellulaires. En présence d'aphidicoline, les évènements de terminaison ne sont plus visibles dans la région cœur des cellules JEFF, mais ils sont toujours présents chez les fibroblastes (Letessier et al., 2011). La technique de peignage moléculaire permet de cartographier les évènements d'initiation et de terminaison mais elle ne permet pas de déterminer à quel moment de la phase S ces évènements ont eu lieu (programme temporel ou timing). Cette information peut être obtenue grâce à la technique de Repli-Seq mise au point par Hansen et coll. (Hansen et al., 2010). Les profils de timing de réplication de *FRA3B* et de *FRA16D* obtenus par ces auteurs dans deux autres lignées de lymphocytes humains, ont montré que ces deux régions sont répliquées tardivement pendant la phase S par des fourches de réplication venant des régions situées de part et d'autre de la région fragile et que la réplication ne se termine qu'en phase G2. Ceci engendre un profil de réplication en « V » de la région fragile (voir figure 11). Ce profil en « V » a aussi été mis en évidence au niveau des deux SFC majeurs des fibroblastes : *FRA1L* et *FRA3L*. Il est important de signaler que les régions possédant des profils en « V » ne sont pas toutes des SFC. En se basant sur nos résultats de peignage moléculaire au niveau de *FRA3B* et sur les résultats de Repli-Seq obtenus par Hansen et coll., la fragilité des SFC a donc été expliquée par la présence d'une grande région centrale pauvre en évènements d'initiation dont la réplication est assurée par des fourches provenant des régions adjacentes. Ces fourches devant parcourir plusieurs centaines de kilobases, la terminaison de la réplication de cette région a lieu très tardivement (fin de S, G2). En cas de perturbation de la progression de ces fourches, la réplication de la région n'est pas terminée

lorsque les cellules entrent en mitose, ce qui entraîne l'apparition de cassures chromosomiques.

Ainsi que nous l'avons déjà mentionné, l'étude de la dynamique de réplication par la technique du peignage moléculaire n'a été effectuée que pour un seul SFC (*FRA3B*) dans deux lignées cellulaires (lymphocytes JEFF et fibroblastes MRC5). Cette technique étant extrêmement chronophage, nous avons pensé que la technique de Repli-Seq serait une bonne alternative pour étendre les observations faites au niveau de *FRA3B* à d'autres SFC. Bien que les résultats de Repli-Seq soient disponibles pour différents types cellulaires dans les bases de données, aucune analyse n'avait été faite sur des cellules traitées par l'aphidicoline. Nous ne pouvions donc pas nous référer à ces résultats pour déterminer l'effet d'un stress répliatif au niveau d'autres SFC. Pour cette raison, nous avons entrepris d'utiliser au laboratoire la technique de Repli-Seq, en collaboration avec l'équipe de Claude Thermes à Gif-sur-Yvette, afin de visualiser la dynamique de réplication au niveau des SFC en présence d'aphidicoline.

Notre hypothèse de travail était la suivante : si la plupart des SFC (tous ?) présentent au niveau du cœur un profil en « V » qui reflète une terminaison de la réplication en fin de S ou en G2, une réplication incomplète en présence d'aphidicoline devrait se traduire par un profil en forme de V incomplet « \ / » (voir figure 11). Dans ce cas, seuls les loci ayant un profil en «V» qui se transforme en un V incomplet sous aphidicoline pourraient être considérés comme des SFC. Nous voulions donc voir dans un premier temps si ces prédictions étaient vérifiées au niveau de SFC déjà caractérisés. Si tel était le cas, nous pourrions étendre notre modèle à une majorité des SFC. De plus, ceci nous permettrait de caractériser de nouveaux SFC en repérant des régions ayant ces mêmes caractéristiques et non classées actuellement comme tel.

La technique de Repli-Seq (Figure 10) consiste à marquer pendant un temps court l'ADN néosynthétisé par un analogue de la thymidine, le BrdU, dans les cellules d'une population asynchrone puis à trier les cellules selon différentes phases du cycle cellulaire (G1, S1, S2, S3, S4 et G2M). L'ADN est ensuite extrait de ces différentes fractions, fragmenté par sonication et dénaturé avant d'être immunoprécipité grâce à un anticorps dirigé contre le BrdU. L'ADN ainsi récupéré à partir de chaque fraction est séquencé par la plateforme de séquençage à haut débit « Imagif ». Les analyses bioinformatiques et les alignements sur le génome sont réalisés par l'équipe de Claude Thermes.

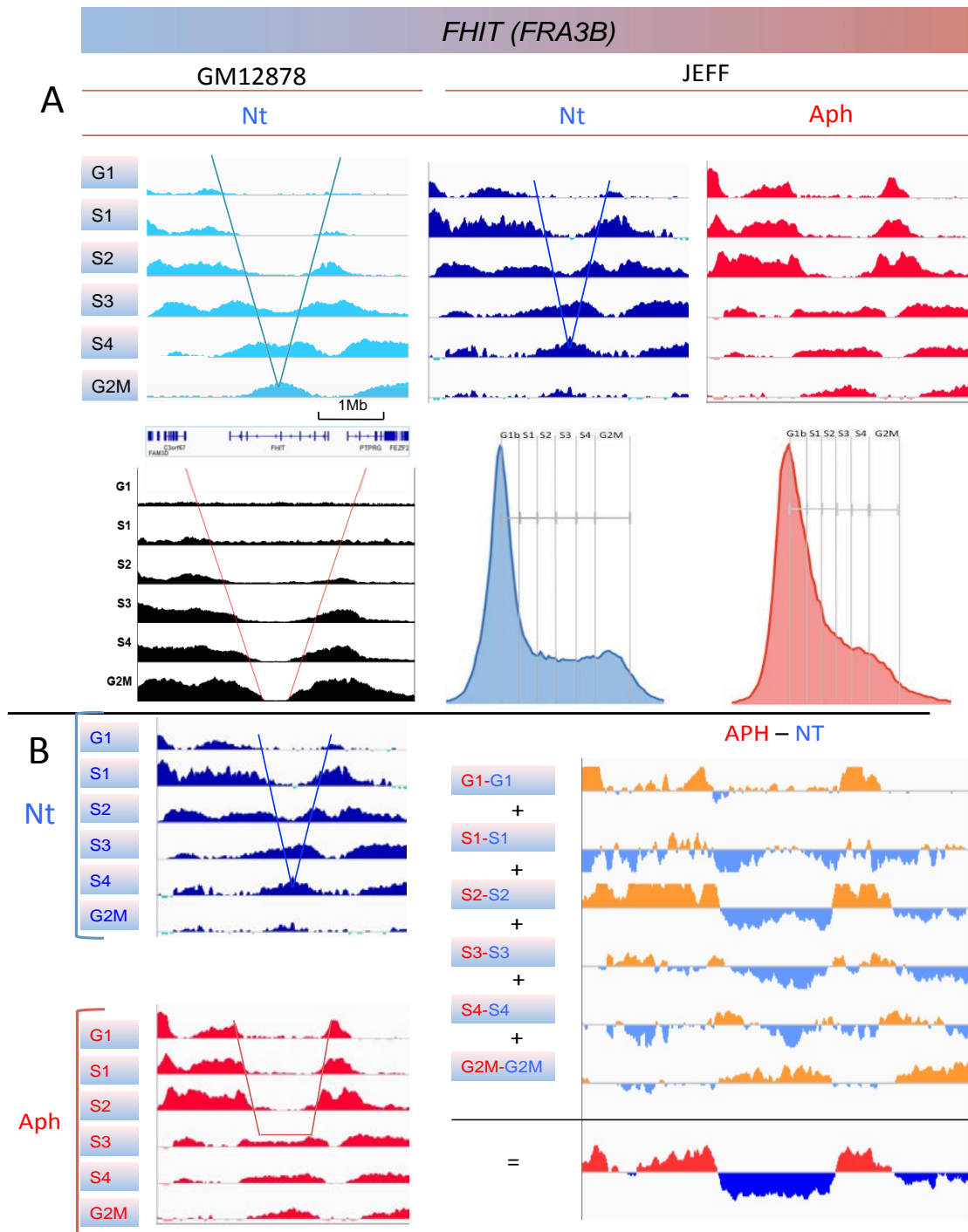


**Figure 10 : Principe de la méthode du Repli-Seq.** Les différentes étapes sont résumées à gauche. A droite est présenté un histogramme typique de distribution des cellules lors du tri au FACS selon le contenu en ADN. La position des différentes fractions récoltées est indiquée. Seule la seconde moitié du pic G1 (G1b) est utilisée car selon Hansen et coll. (Hansen et al., 2010), la première (G1a) est inutilisable car elle génère un bruit de fond très élevé. Sous l’histogramme est schématisé le cas d’une origine de réplication se déclenchant précocement (S1) et la progression des fourches de réplication au cours de la phase S (S2, S3, S4). L’incorporation de BrdU est symbolisée en orange. On voit que les séquences distantes de l’origine sont répliquées (et donc marquées par le BrdU) plus tard que celles qui sont proches.

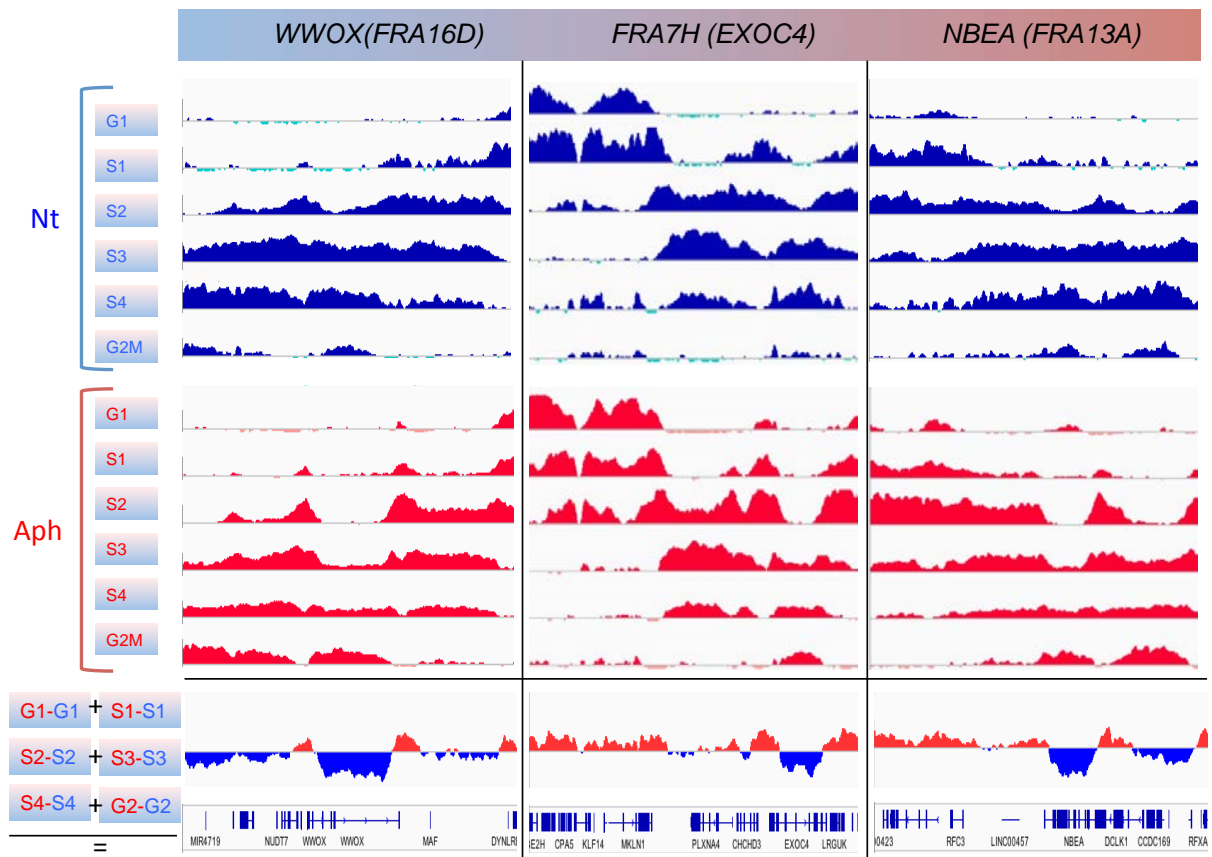
Nous avons réalisé cette expérience avec des cellules JEFF non traitées ou traitées avec 0,6  $\mu$ M d’aphidicoline pendant 16 h (Figure 11). Ce traitement représente la condition standard utilisée pour induire la fragilité des SFC. Les résultats obtenus avec les cellules JEFF non traitées montrent un profil en « V » au niveau des sites *FRA3B* et *FRA16D* et d’autres SFC connus pour être instables dans les lymphocytes (Mrasek et al., 2010). Ces résultats confirment donc ceux obtenus par Hansen et ses collaborateurs sur les cellules lymphoblastoïdes humaines GM12878. Néanmoins le « V » obtenu dans les cellules JEFF apparaît décalé d’un stade par rapport aux cellules GM12878. Ainsi, la complétion de la réplication du cœur de *FRA3B* a lieu en S4 dans les cellules JEFF au lieu de G2 dans les cellules GM12878.

Les résultats obtenus avec les cellules traitées à l'aphidicoline montrent effectivement un déficit de réplication aux niveaux de SFC connus (Figure 12). Par contre le profil ne ressemble pas à un « V » inachevé mais plutôt à un profil en « U ». En effet, des signaux de réplication sont détectés dans la région cœur de *FRA3B* à partir du stade S3. Néanmoins, conformément aux résultats du peignage au niveau de *FRA3B*, les fourches de réplication provenant des deux régions flanquantes ne se rejoignent pas sous aphidicoline ce qui affecte la terminaison de la réplication. La présence d'ADN néosynthétisé dans le cœur des sites implique que des événements d'initiation y surviennent en présence d'aphidicoline. Ceci peut être expliqué par le recrutement des origines latentes (phénomène de compensation) que nous avons déjà évoqué dans l'introduction. Ce phénomène n'avait pas été observé précédemment (Letessier et al., 2011), possiblement parce que les cellules avaient été traitées à l'aphidicoline pendant 5 h pour l'analyse par peignage alors que pour le Repli-Seq, elles ont été traitées pendant 16 h. Nous reviendrons sur ce point ultérieurement. En tout état de cause, les résultats montrent que le déclenchement de ces origines de compensation ne suffit pas à terminer la réplication de la région.

Une analyse plus exhaustive au niveau du génome entier est en train d'être réalisée dans le but de valider notre démarche comme étant une nouvelle méthode pour cartographier par séquençage tous les SFC d'une lignée cellulaire donnée. Pour ce faire, nous essayons d'identifier toutes les régions génomiques qui présentent un déficit de réplication significatif sous aphidicoline et de voir si ces régions correspondent à des SFC déjà caractérisés ou pas.



**Figure 11: Profils de réplication dans la région du gène *FHIT* obtenus par Repli-Seq dans des cellules lymphoblastoïdes humaines.** A) Profils de réplication dans les cellules GM12878 et dans les cellules JEFF non traitées (Nt) ou traitées à l’aphidicoline (Aph; 0,6µM). Pour chaque phase du cycle (S1 à G2M) est porté en ordonnées le nombre normalisé de séquences obtenues par fenêtre de 50 kb et en abscisse la position sur le génome dans la région du gène *FHIT*. Une carte de cette région est présentée sous les résultats de GM12878 avec la position des gènes (exons et sens de la transcription) et l’échelle. L’aspect en “V” des profils chez les cellules non traitées (Nt) est symbolisé par des traits bleus. Les profils en noir (en bas à gauche) sont théoriques. Ils correspondent à ce qu’on attendrait de lymphocytes traités à l’aphidicoline : un aspect en V incomplet (traits rouges). Dans le cas des cellules JEFF les histogrammes de tri sont présentés comme dans la Figure 10. B) Mise en évidence de séquences présentant un retard de réplication dans la région du gène *FHIT* chez des cellules JEFF traitées à l’aphidicoline. A gauche sont présentés les mêmes histogrammes que dans la partie A. L’aspect en “U” des profils chez les cellules traitées par l’aphidicoline (Aph) est symbolisé par des traits rouges. A droite, sont présentés les profils obtenus par soustraction des valeurs des cellules non traitées (Nt) de celles des cellules traitées (Aph) pour chaque fenêtre de 50 kb (G1-G1 à G2M-G2M). La somme sur l’ensemble du cycle des valeurs obtenues est présentée dans le dernier panneau (=). Les séquences présentant un retard de réplication sous aphidicoline sont indiquées en bleu.



**Figure 12 : Profils de réplication obtenus par Repli-Seq dans les régions des gènes *WWOX*, *NBEA* et *EXOC4* chez les cellules JEFF. Le nom des sites fragiles correspondant est indiqué entre parenthèses. Les résultats sont présentés comme dans la Figure 11.**

## **II. Effet du RO-3306 sur l'instabilité des SFC**

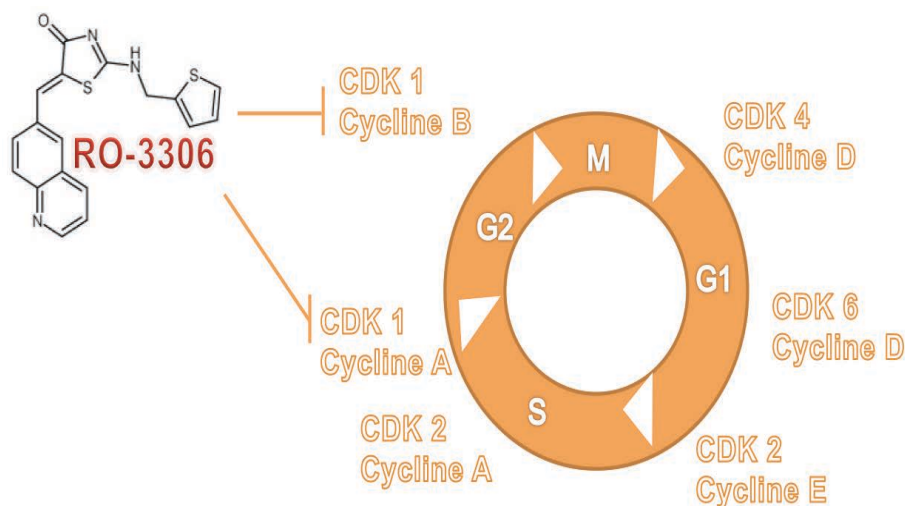
### **1. Effet du RO-3306 sur le cycle cellulaire des lymphocytes et des fibroblastes**

Dans cette partie nous avons voulu allonger la durée de la phase G2 afin de tester l'hypothèse suivante : retarder l'entrée des cellules en mitose laissera le temps à la réplication des SFC pour s'achever et abolira donc leur fragilité. Afin d'allonger la phase G2, nous avons eu recours à un inhibiteur de CDK. Les CDKs sont membres de la famille des sérine/thréonine kinases et leur activité dépend de leur liaison avec des cyclines. Il existe 21 gènes codant pour des CDKs dans le génome humain. Quatre parmi elles contrôlent le cycle cellulaire. Il s'agit de CDK1, CDK2, CDK4 et CDK6. CDK1 est la kinase responsable de la transition entre la phase G2 et la mitose. De nos jours plus d'une vingtaine d'inhibiteurs de CDKs ont été caractérisés. Ce sont essentiellement des analogues de l'ATP. Néanmoins, vu la

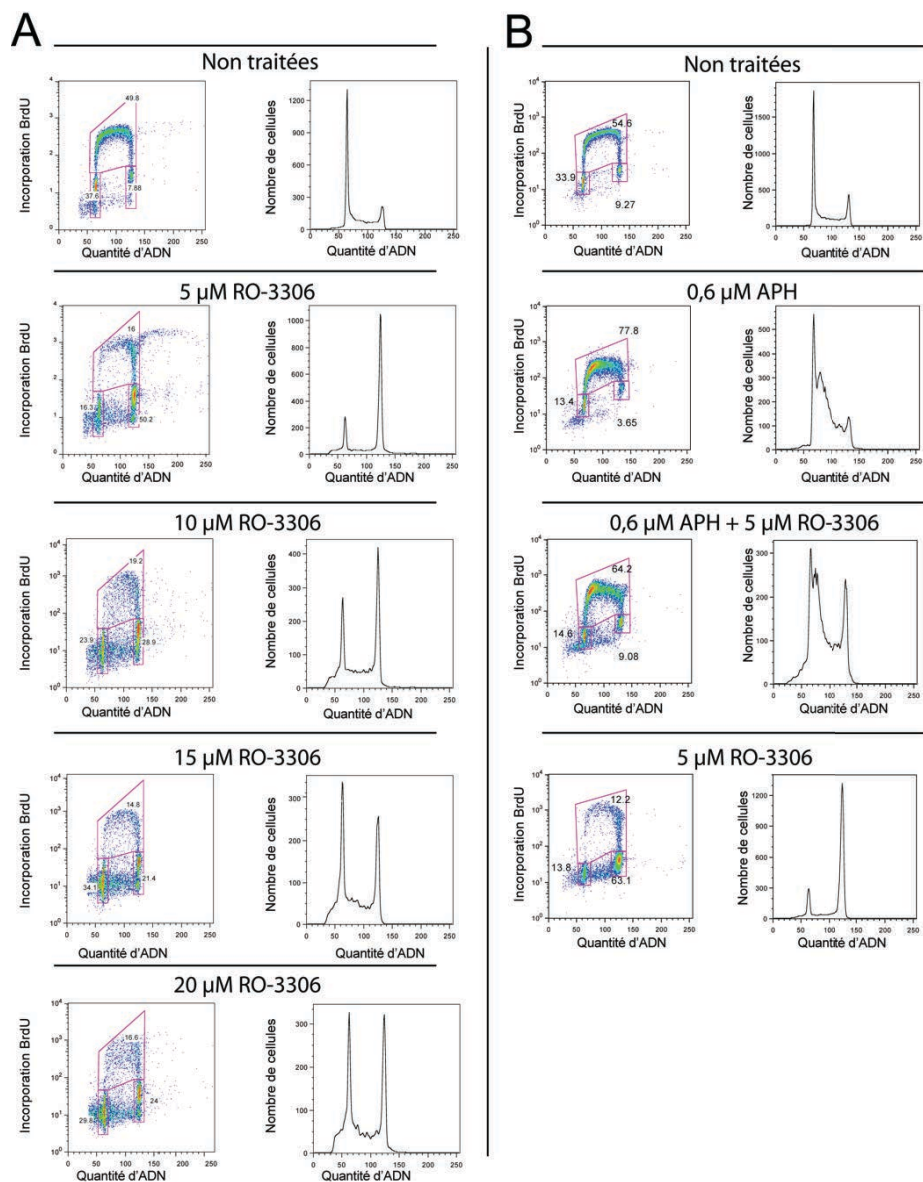


ressemblance entre les différents CDKs, il est difficile de trouver un inhibiteur spécifique de l'une par rapport aux autres (Cicenas et al., 2014).

Le RO-3306 inhibe *in vitro* l'activité kinase du complexe CDK1/cycline B1, qui contrôle l'entrée en mitose, avec un  $K_i$  de 35 nM (Vassilev et al., 2006) (Figure 13). Il est trois fois moins sélectif vis-à-vis du complexe CDK1/cycline A ( $K_i = 110$  nM) et il est dix fois plus sélectif vis-à-vis de CDK1 que de la kinase la plus apparentée : CDK2 ( $K_i = 340$  nM). Dans un premier temps, nous avons voulu déterminer la concentration de RO-3306 minimale pour obtenir une accumulation des cellules lymphoblastoïdes JEFF et des fibroblastes humains normaux (MRC5) à la frontière G2/M sans affecter d'autres kinases, donc sans perturber l'avancement des cellules dans le reste du cycle cellulaire. Pour ce faire, la répartition de la population cellulaire dans les différentes phases du cycle a été analysée par FACS après 16 h de traitement par différentes concentrations de RO-3306. Les résultats ont montré qu'une concentration de 5  $\mu$ M permettait une accumulation maximale des cellules en G2. Au delà de cette concentration, les cellules s'accumulent dans d'autres phases du cycle (Figure 14). C'est donc cette concentration que nous avons initialement choisie pour la suite des expériences.



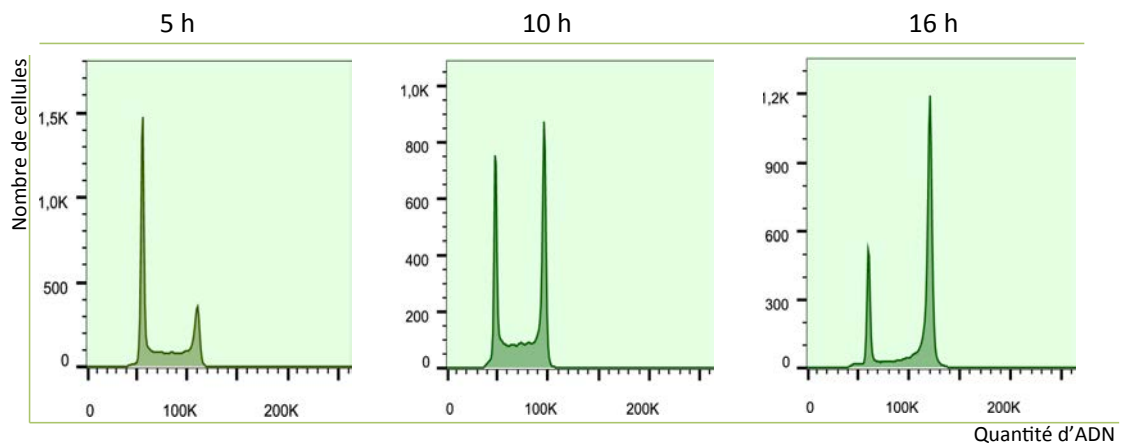
**Figure 13 : Action du RO-3306 sur le cycle cellulaire.** Le RO-3306 est un analogue de l'ATP qui inhibe spécifiquement l'activité kinase de CDK1. Les principaux acteurs du contrôle du cycle cellulaire sont indiqués: CDK4/cycline D et CDK6/cycline D qui contrôlent le déroulement de la phase G1; CDK2/cycline E qui contrôlent le passage en S; CDK2/cycline A qui contrôlent la complétion de la phase S; CDK1/cycline A qui contrôlent le passage en G2 ; CDK1/cycline B qui assurent le passage G2/M.



**Figure 14 : Effet du RO-3306 et de l'aphidicoline sur le cycle des cellules JEFF.** La répartition des cellules dans les différentes phases du cycle a été analysée par FACS après 16 h de culture dans les conditions indiquées suivie d'un marquage au BrdU pendant 20 min. Ordonnées : incorporation de BrdU (panneaux de gauche) ou nombre de cellules (panneaux de droite). Abscisses : quantité d'ADN. **A)** Différentes concentrations de RO-3306 ont été testées. **B)** Effet du RO-3306 à 5  $\mu\text{M}$  en présence d'aphidicoline (0,6  $\mu\text{M}$ ). Il est apparu ultérieurement qu'en fonction de l'état physiologique des cellules et/ou du lot de RO-3306, il pouvait être nécessaire d'utiliser 10  $\mu\text{M}$  de RO-3306 pour obtenir les mêmes effets.

Ces premières expériences avaient été réalisées après 16 h de traitement, ce qui représente la durée d'un cycle cellulaire complet pour les cellules JEFF. Nous avons ensuite voulu déterminer quelle était la cinétique d'accumulation des cellules en G2. Pour ce faire, la répartition de la population cellulaire dans les différentes phases du cycle a été analysée par

FACS après 5 h, 10 h ou 16 h de traitement au RO-3306. Les résultats montrent que le temps 16 h est effectivement le plus adéquat pour une accumulation optimale des cellules en G2 (Figure 15).



**Figure 15 : Cinétique d'accumulation en G2 des cellules JEFF après traitement au RO-3306.** La répartition des cellules dans les différentes phases du cycle a été analysée par FACS après 5 h, 10 h ou 16 h de culture en présence de 10 µM de RO-3306.

## 2. Effet du RO-3306 sur l'instabilité des SFC dans les lymphocytes et les fibroblastes

Nous avons ensuite déterminé l'effet du RO-3306 sur l'apparition de cassures totales (sur l'ensemble des chromosomes) induites par l'aphidicoline dans les cellules lymphoblastoïdes JEFF. Ces cassures sont induites de façon standard en traitant ces cellules avec 0,6 µM d'aphidicoline. Dans ces conditions, la vitesse de progression des fourches de réplication est réduite d'environ 8 fois, aussi bien globalement qu'au niveau du site fragile *FRA3B* (Letessier et al., 2011). Des analyses par FACS ont montré qu'en aphidicoline seule, les cellules s'accumulent en phase S, très probablement à cause du ralentissement des fourches. En présence d'aphidicoline et de RO-3306, la fraction des cellules qui parviennent à terminer leur réplication s'accumule en G2 (Figure 14).

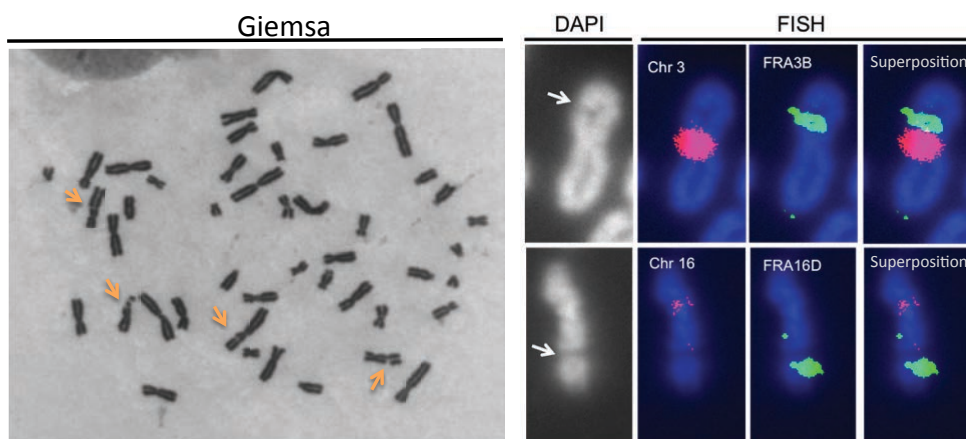
Des cellules JEFF ont été traitées pendant 16 h à l'aphidicoline en présence ou en absence de RO-3306. Elles ont ensuite été placées pendant trois heures dans du milieu avec ou sans aphidicoline et sans RO-3306 (pour leur permettre d'entrer en mitose) et avec du nocodazole, un poison du fuseau mitotique qui les bloque en métaphase (Figure 16). Les résultats obtenus montrent qu'après 16 h de traitement par l'aphidicoline et le RO-3306, le taux de cassures totales dans les cellules JEFF devient comparable à celui observé dans les cellules non traitées ou traitées avec le RO-3306 seul. Nous avons ensuite analysé les taux de

cassures au niveau des deux SFC majeurs des lymphocytes : *FRA3B* et *FRA16D*. Comme dans le cas des cassures totales, les cassures induites par l'aphidicoline au niveau de ces deux sites disparaissent en présence de RO-3306 dans le milieu de culture (Figure 17).

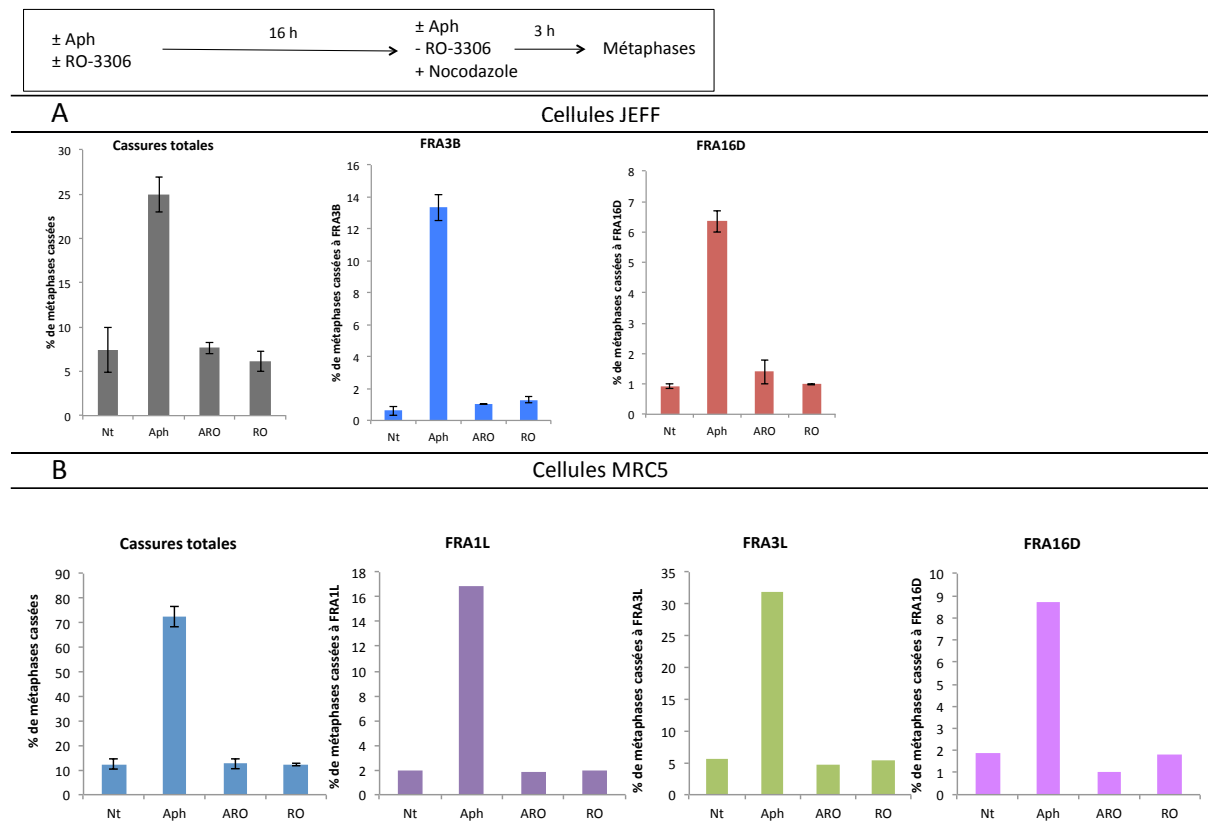
Nous avons ensuite souhaité étendre cette observation en réalisant cette même expérience avec des fibroblastes humains normaux (MRC5). Comme dans le cas des lymphocytes, l'analyse des chromosomes métaphasiques montre que les cassures induites par l'aphidicoline sont supprimées en présence du RO-3306. Ceci est vrai aussi bien pour les cassures totales que pour les cassures induites à *FRA1L* et *FRA3L*, les deux SFC majeurs dans ce type cellulaire (Figure 17).

En utilisant le RO-3306, nous avons bloqué les cellules en phase G2 afin de permettre l'achèvement de la réplication au niveau des SFC avant l'entrée des cellules en mitose. Nous avons pensé qu'un autre moyen permettant l'achèvement de la réplication avant l'entrée des cellules en mitose serait d'avancer le timing de réplication des SFC en invalidant le facteur RIF1. Les résultats sont présentés en Annexe I.

Des auteurs (Jiang et al., 2009) ont rapporté que le traitement de cellules à la trichostatine A (TSA) diminuait les cassures au niveau des SFC. Etant donné que nous avons obtenu ce même phénotype en traitant les cellules au RO-3306, nous avons voulu vérifier que le traitement de nos cellules à la TSA produisait effectivement l'effet décrit. Les résultats de ces expériences sont présentés en Annexe II.



**Figure 16 : Cassures induites par l'aphidicoline au niveau des SFC.** Des cellules JEFF ont été incubées 16 h en présence d'aphidicoline (0,6  $\mu$ M) puis 3 h en présence d'aphidicoline et de nocodazole. Des exemples d'une métaphase complète (panneau de gauche) avec plusieurs cassures (flèches), ainsi que de cassures à *FRA3B* et *FRA16D* sont montrés (flèches panneaux de droite). Pour l'analyse des cassures totales, les chromosomes ont été colorés au Giemsa. Pour le comptage des cassures à *FRA3B* et *FRA16D*, la technique de FISH a été utilisée : les lames ont été incubées en présence de sondes correspondant aux sites qui sont révélées par un anticorps fluorescent (vert). Les chromosomes 3 et 16 ont été marqués par des sondes centromériques (rouge). Les chromosomes sont contre-colorés au DAPI (bleu).



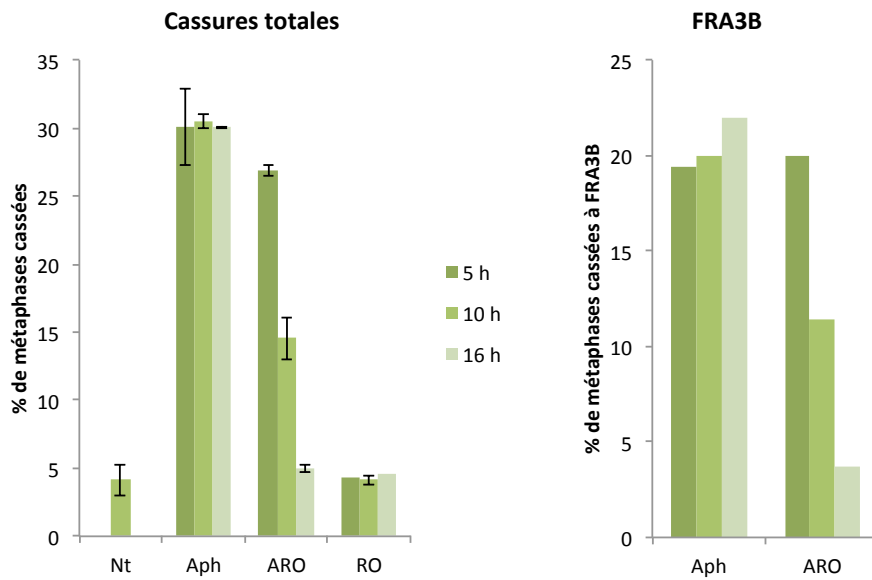
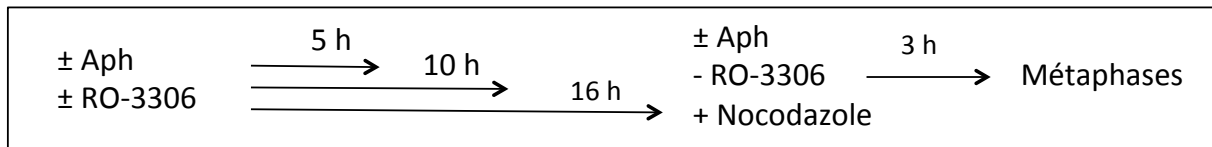
**Figure 17 : Effet du RO-3306 sur les cassures induites par l'aphidicoline.** **A)** Des cellules lymphoblastoïdes JEFF ont été incubées pendant 16 h dans les conditions indiquées dans le panneau du haut (aphidicoline 0,6  $\mu$ M, RO-3306 5  $\mu$ M) puis pendant 3 h en absence de RO-3306, en présence ou en absence d'aphidicoline et en présence de nocodazole. Les résultats sont présentés en pourcentage des métaphases présentant des cassures sur n'importe quel chromosome (cassures totales) ou spécifiquement à *FRA3B* ou *FRA16D* (sites majeurs des lymphocytes) (Nt : non traitées ; Aph : aphidicoline ; ARO: aphidicoline + RO-3306 ; RO : RO-3306) Les barres d'erreur représentent l'erreur standard de la moyenne de deux expériences. **B)** Les mêmes expériences ont été réalisées avec des fibroblastes MRC5 et les résultats sont présentés comme en A. Les cassures totales ainsi que celles à *FRA1L*, *FRA3L* (sites majeurs des fibroblastes), ou *FRA16D* (site mineur) ont été analysées.

Comme nous l'avons déjà mentionné dans l'introduction, tous les SFC n'ont pas la même fréquence de cassures. Le fait que le RO-3306 abolisse apparemment complètement les cassures totales laisse penser qu'il agit de façon similaire sur les SFC majeurs et sur ceux qui cassent à des fréquences plus faibles. Cependant, étant donné que les cassures à *FRA3B* et *FRA16D* représentent plus que 80% des cassures totales dans les cellules JEFF (Letessier et al., 2011), l'effet du RO-3306 sur un site à fragilité modérée pourrait être masqué. Nous

avons donc choisi de tester son effet au niveau de *FRA16D* dans les cellules fibroblastiques MRC5. Ce site représente 3,7% des cassures totales dans ce type cellulaire après traitement avec 0,6  $\mu$ M d'aphidicoline. Les résultats obtenus montrent que le RO-3306 abolit aussi la faible fragilité de ce site (Figure 17).

### **3. Cinétique d'action du RO-3306 sur le taux de cassures au niveau des SFC**

Ayant montré que le traitement des cellules JEFF par le RO-3306 pendant 16 h supprime les cassures induites par l'aphidicoline, nous nous sommes demandé si cette drogue a besoin de 16 h pour réduire le taux de cassures ou si son effet peut être visible après des temps plus courts. Nous avons donc déterminé comme précédemment le taux de cassures totales ainsi que les cassures à *FRA3B* après 5 h, 10 h ou 16 h de traitement des cellules JEFF à l'aphidicoline et au RO-3306 (Figure 18). Nos résultats montrent que 5 h de traitement par le RO-3306 ne prévient pas l'apparition des cassures induites par l'aphidicoline. Après 10 h, la fréquence de cassures baisse d'environ 50%. Comme montré précédemment, les cassures sont complètement supprimées après 16 h. L'effet du RO-3306 ne se manifeste donc qu'après un traitement prolongé des cellules. Il faut remarquer que l'allongement du temps de traitement par le RO-3306 fait varier deux paramètres : le temps que les cellules passent à la frontière G2/M et la proportion des cellules collectées en mitose qui ont traversé une partie de la phase S en présence de la drogue par rapport à celles qui ont subi son effet exclusivement en G2. Nous ne pouvons pas, à ce stade, savoir le taux de contribution de chacun des deux paramètres dans la diminution des cassures après 16 h de traitement au RO-3306.



**Figure 18: Cinétique d'action du RO-3306 sur les cassures induites par l'aphidicoline.** Les expériences ont été réalisées et les résultats sont présentés comme dans la Figure 17.

### **III. Relation entre transcription des grands gènes et fragilité des SFC**

En Décembre 2011, Tora et ses collaborateurs ont proposé un nouveau modèle pour expliquer le mécanisme d'instabilité des SFC (Helmrich et al., 2011). Les auteurs affirment que, étant donné leur taille, la transcription des grands gènes contenus dans les SFC dure plus d'un cycle cellulaire et donc se poursuit en phase S contrairement à ce qui est décrit pour la majorité des gènes. Ceci conduirait à des collisions entre la machinerie de transcription et celle de réplication. En situation normale, la cellule gère ces collisions. Par contre, quand la réplication est ralentie, la collision provoquerait un arrêt des deux machineries, ce qui stabiliserait des structures appelées R-loops (hybrides ADN-ARN) qui résultent de la transcription. Une fois stabilisées, ces structures bloqueraient l'avancement des fourches de réplication qui ne pourraient plus terminer la réplication de la région concernée, la rendant fragile au cours de la mitose. A l'appui de leur modèle, les auteurs indiquent que pour 5 grands gènes dans deux types cellulaires, les niveaux d'ARNm et la fragilité sont corrélés.

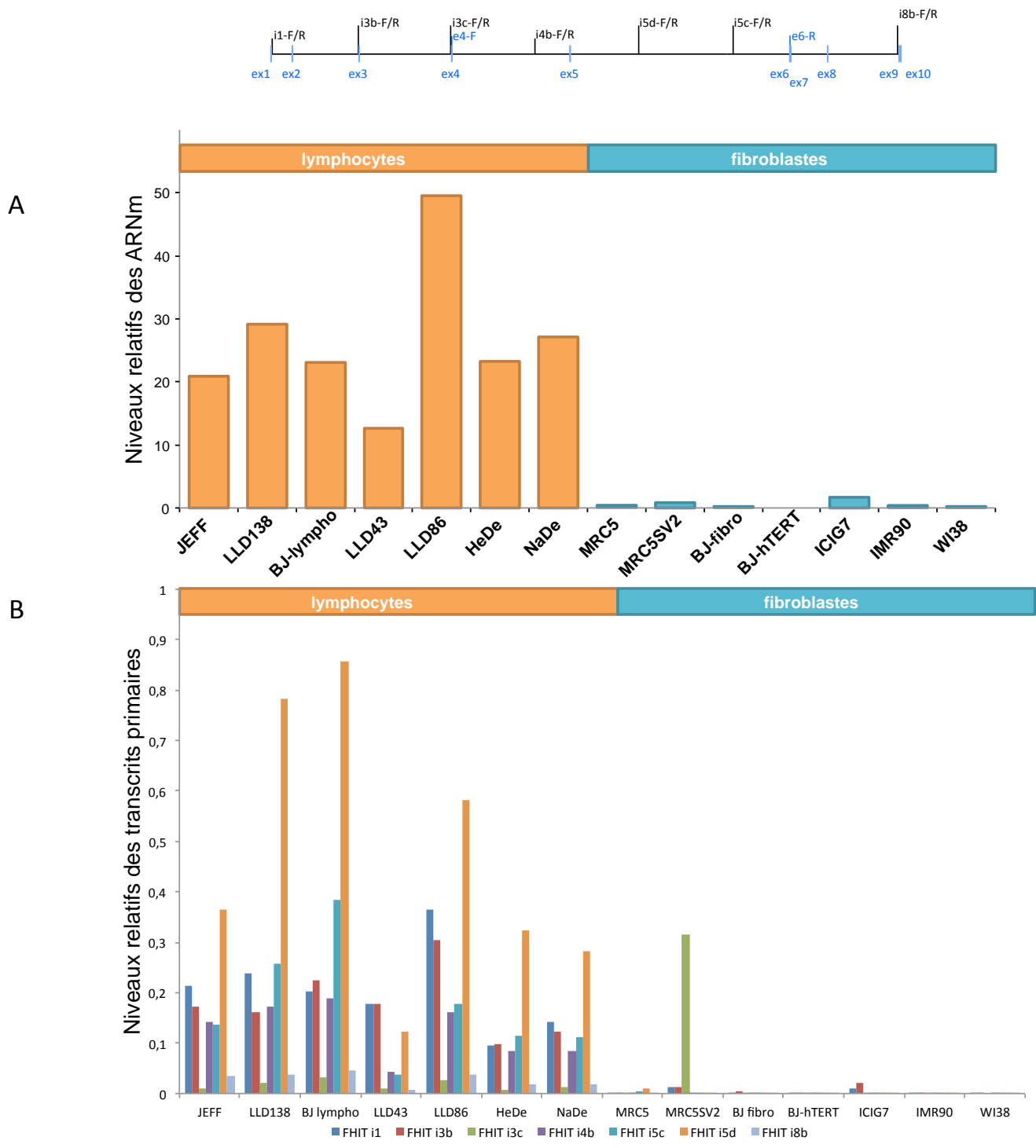
Nous avons donc voulu déterminer s'il existait une corrélation entre expression des grands gènes et fragilité pour les SFC majeurs des deux systèmes cellulaires sur lesquels nous travaillons : les cellules lymphoblastoïdes JEFF et les fibroblastes MRC5. D'autre part, si le modèle des collisions est vrai, une variation du niveau de transcription devrait influencer le niveau de fragilité. Nous nous sommes donc demandé si la diminution du taux de cassures au niveau des SFC que nous observons en traitant les cellules au RO-3306 est due, au moins en partie, à une diminution de l'expression des grands gènes nichés dans ces SFC.

#### **1. Corrélation entre fragilité des SFC et transcription des grands gènes**

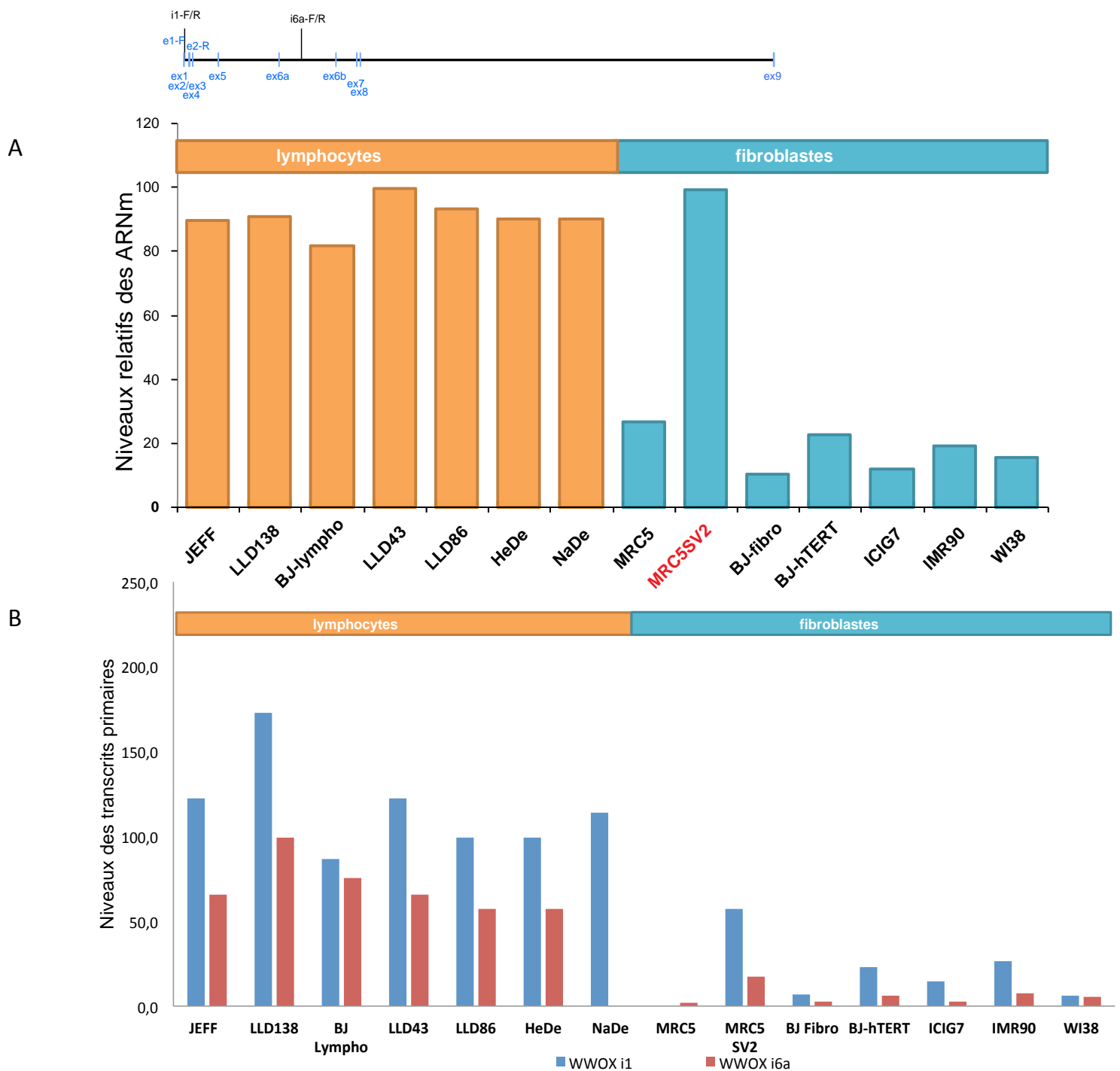
Afin de déterminer s'il existe une corrélation entre expression des gènes et fragilité des SFC dans nos systèmes cellulaires, nous avons mesuré dans 7 lignées de lymphocytes et 7 lignées de fibroblastes humains la quantité d'ARNm des 4 grands gènes localisés dans les SFC majeurs de chacun de ces deux types cellulaires : *FHIT* (*FRA3B*), *WVOX* (*FRA16D*), *LSAMP* (*FRA3L*) et *NEGR1* (*FRA1L*). Les résultats obtenus montrent effectivement une corrélation : les niveaux des messagers *FHIT* et *WVOX* sont en moyenne 25 et 4 fois plus élevés dans les lignées de lymphocytes, où *FRA3B* et *FRA16D* sont fragiles, par rapport aux lignées de fibroblastes où les deux sites ne sont pas ou peu fragiles (Figure 19, Figure 20).

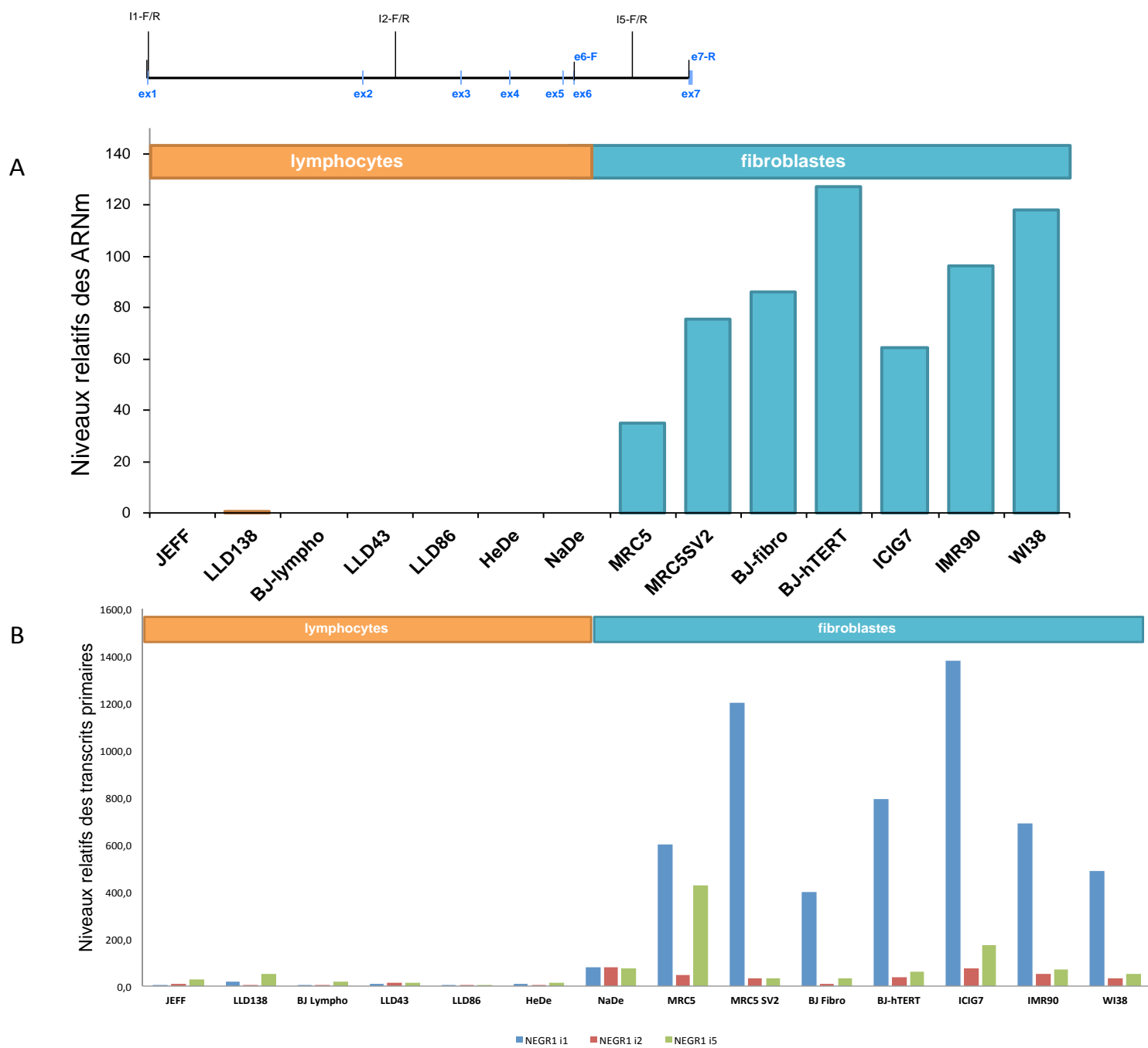


C'est aussi le cas pour les niveaux des messagers *LSAMP* et *NEGR1* qui sont en moyenne 20 et 100 fois plus élevés dans les lignées de fibroblastes, où *FRA3L* et *FRA1L* sont fragiles, par rapport aux lignées de lymphocytes où les deux sites ne sont pas fragiles (Figure 21, Figure 22). Sachant que les niveaux d'ARNm ne reflètent pas nécessairement le taux de transcription, nous avons aussi quantifié les transcrits primaires de chaque gène en utilisant des couples d'amorces introniques. Les résultats obtenus sont complètement en accord avec les résultats obtenus pour les ARNm.

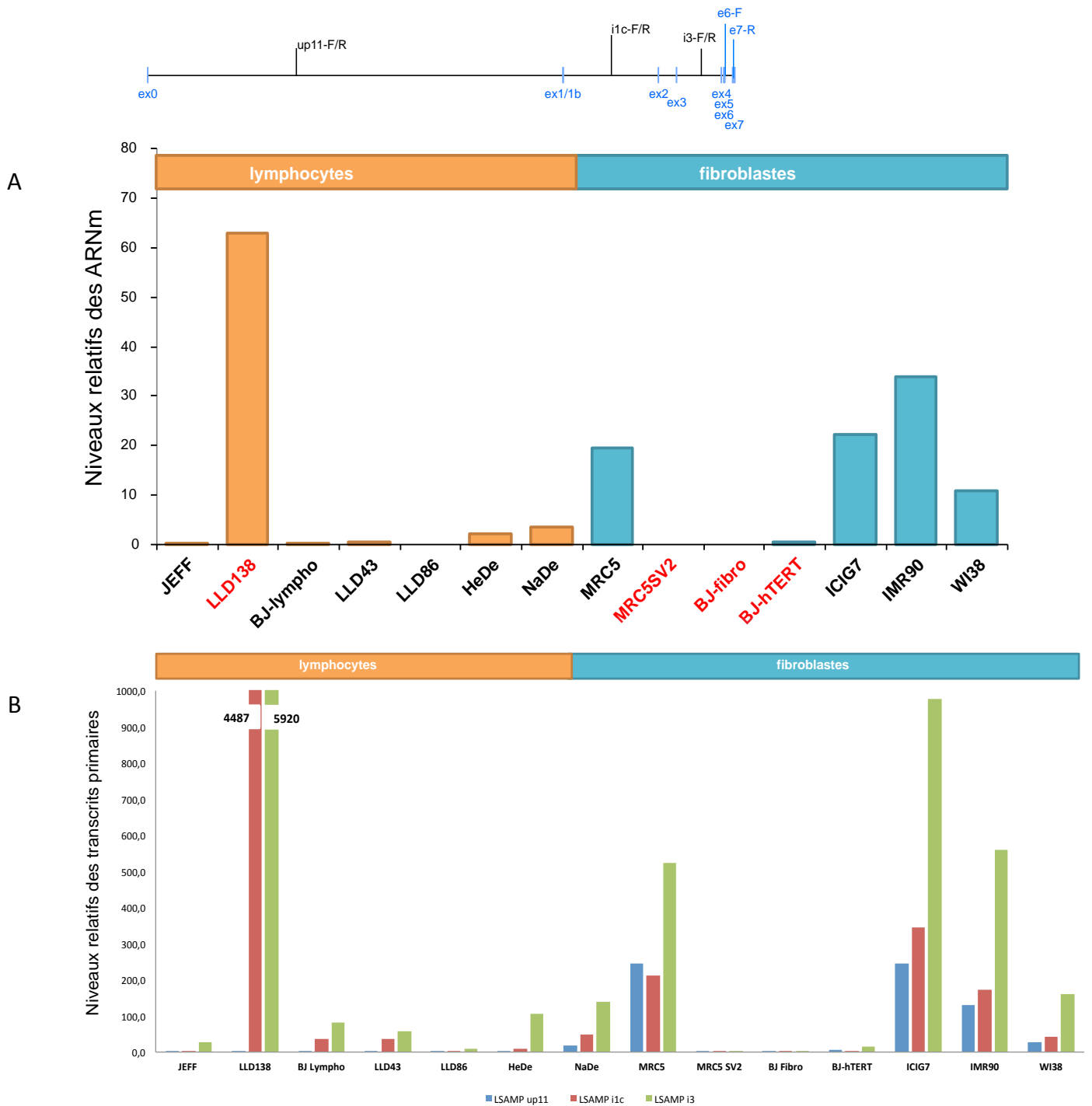


**Figure 19 : Niveaux d'expression du gène *FHIT* dans plusieurs lignées de lymphocytes et de fibroblastes humains.** Les niveaux d'ARNm (A) et de transcrits primaires (B) ont été mesurés par qRT-PCR avec un couple d'amorces exoniques pour les ARNm (e4-F/e6-R) et différents couples d'amorces introniques pour les transcrits primaires (i1-F/R à i8-F/R). La carte du gène *FHIT* avec la position des exons et des couples d'amorce est montrée au dessus. Dans les deux cas, les niveaux de transcrits sont normalisés par rapport à celui de l'ARNm de la cyclophiline B.





**Figure 21: Niveaux d'expression de NEGR1 dans plusieurs lignées de lymphocytes et de fibroblastes.** Les niveaux d'ARNm (A) et de transcrits primaires (B) sont mesurés par qRT-PCR avec un couple d'amorces exoniques pour les ARNm et différents couples d'amorces introniques pour les transcrits primaires. La carte du gène NEGR1 et la position des couples d'amorce sont montrés au dessus. Dans les deux cas, les niveaux de transcrits sont normalisés par rapport à celui de l'ARNm de la cyclophiline B.

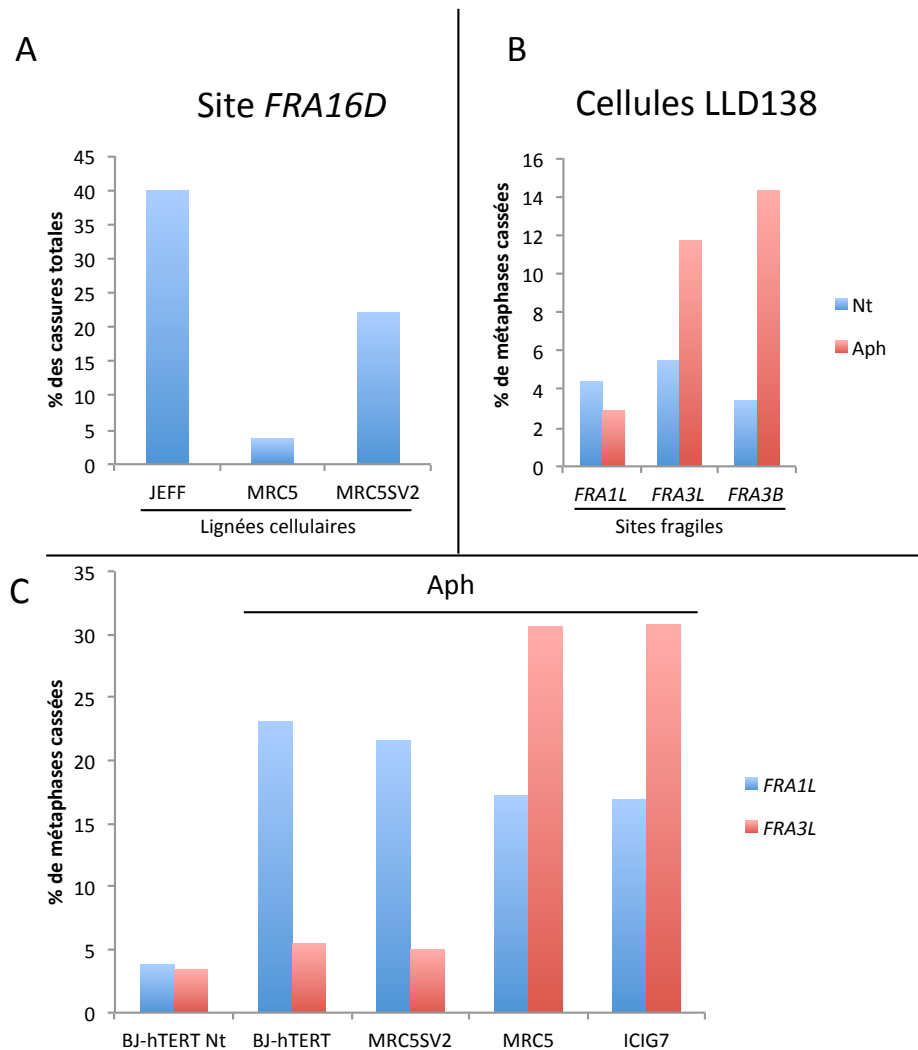


**Figure 22 : Niveaux d'expression de LSAMP dans plusieurs lignées de lymphocytes et de fibroblastes.** Les niveaux d'ARNm (A) et de transcrits primaires (B) sont mesurés par qRT-PCR avec un couple d'amorces exoniques pour les ARNm et différents couples d'amorces introniques pour les transcrits primaires. Dans les deux cas, les niveaux des transcrits sont normalisés par rapport à celui de la cyclophyliline B. Les lignées présentant un niveau d'expression atypique sont indiquées en rouge.

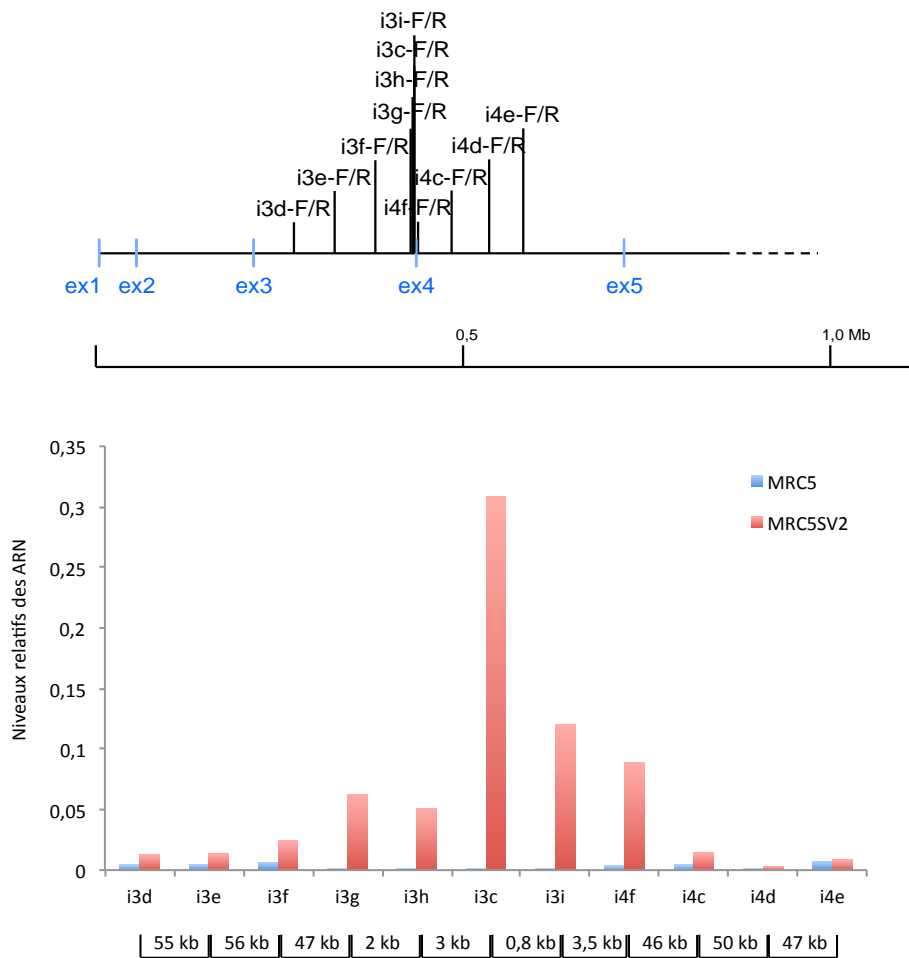
Cependant cette règle n'est pas sans exception : pour les gènes *WWOX* et *LSAMP* nous avons pu mettre en évidence une variation de leur expression dans certaines lignées. C'est le cas d'une lignée de fibroblastes (MRC5SV2, dérivée de MRC5 par immortalisation) dans laquelle le niveau de messagers *WWOX* est comparable à celui des lignées de lymphocytes. Par contre, le messager *LSAMP* est indétectable dans ces cellules ainsi que dans deux autres lignées de fibroblastes apparentées (BJ fibro et BJ-hTERT). Une lignée de lymphocytes (LLD138) présente un taux de messager *LSAMP* comparable à celui des fibroblastes. Ces exceptions nous ont fourni un outil pour confirmer la corrélation entre transcription et fragilité. Pour cela, nous avons déterminé si ces lignées qui ont un niveau d'expression atypique de certains gènes, présentent simultanément un niveau atypique de fragilité des SFC correspondants (Figure 23). Nos résultats montrent que dans ces systèmes particuliers, la corrélation est conservée : quand le gène présente un niveau anormalement bas ou élevé d'ARNm par rapport aux autres lignées du même type cellulaire, le site qui lui correspond présente une fragilité anormalement basse ou élevée, respectivement. Nous avons vérifié que ce changement de fragilité est spécifique du site concerné en analysant la fragilité d'autres sites pour lesquels le niveau d'expression du grand gène associé est resté typique. Nous avons vérifié que *FRA1L (NEGR1)* est toujours fragile chez les cellules MRC5SV2 et BJ-hTERT. De même *FRA3B (FHIT)* est toujours fragile dans les cellules LLD138 alors que *FRA1L (NEGR1)* n'est pas fragile dans cette lignée.

Il reste le cas particulier de l'intron 3 du gène *FHIT* dans la lignée MRC5-SV2. Le niveau d'ARNm *FHIT* est très bas dans ces cellules comme dans les autres lignées fibroblastiques (Figure 19) et le site n'est pas fragile. Cependant des transcrits produits à un haut niveau sont détectables avec un couple d'amorces situées à l'extrémité 3' de l'intron 3 (i3c- F/R) (Figure 19B). Ils ne sont pas détectés avec des couples d'amorces situés environ 200 kb en amont (i3b-F/R) ou en aval (i4b-F/R). Ces transcrits ne correspondent très vraisemblablement pas à une expression du gène *FHIT* mais plutôt à une unité de transcription incluse dans le gène et qui est active dans ces cellules. Le haut niveau d'expression de cette unité n'est apparemment pas corrélé à une fragilité de *FRA3B*. Nous nous sommes donc demandé quelle était la taille de cette unité de transcription. Nous l'avons donc étudiée avec des couples d'amorces distribués autour de l'exon 4 de *FHIT*

(Figure 24). Un haut niveau de transcrits est détecté avec les couples d'amorces i3g-F/R à i4f-F/R. Ces résultats indiquent que cette unité de transcription couvre moins de 10 kb.



**Figure 23 : Fragilité des SFC dans les cellules présentant une expression atypique du grand gène associé.**  
**A)** Fragilité de *FRA16D* dans les cellules fibroblastiques MRC5SV2. Les cassures à *FRA16D* ont été déterminées par FISH et leur nombre est exprimé par rapport aux cassures totales déterminées après coloration au Giemsa des chromosomes. Cette normalisation est nécessaire quand on veut comparer la fragilité d'un site dans des fibroblastes et des lymphocytes qui ont des sensibilités très différentes à une même concentration d'aphidicoline.  
**B)** Fragilité de *FRA3L* dans les cellules lymphoblastoïdes LLD138. Les cassures à *FRA3L*, *FRA3B* (contrôle positif) et *FRA1L* (contrôle négatif) ont été déterminées par FISH dans des cellules non traitées (Nt) ou traitées par l'aphidicoline (Aph 0,6  $\mu$ M).  
**C)** Faible fragilité de *FRA3L* dans les cellules fibroblastiques MRC5SV2 et BJ-hTERT. Les cassures à *FRA3L* et *FRA1L* (contrôle positif) ont été déterminées par FISH dans les cellules BJ-hTERT non traitées (Nt) ou dans les cellules BJ-hTERT, MRC5SV2, MRC5 et ICIG7 traitées à l'aphidicoline (Aph).



**Figure 24: Analyse de l'unité de transcription incluse dans le gène *FHIT* chez les cellules MRC5SV2.** Les niveaux des transcrits primaires ont été déterminés par qRT-PCR chez les cellules MRC5 (parentales) et MRC5SV2 (dérivées par immortalisation) avec les différents couples d'amorces introniques indiqués sur la carte partielle du gène *FHIT*. Les niveaux sont normalisés par rapport à celui de l'ARNm de la cyclophiline B. Sous l'histogramme sont indiquées les distances séparant les différents couples d'amorces.

## 2. Effet du RO-3306 sur la transcription

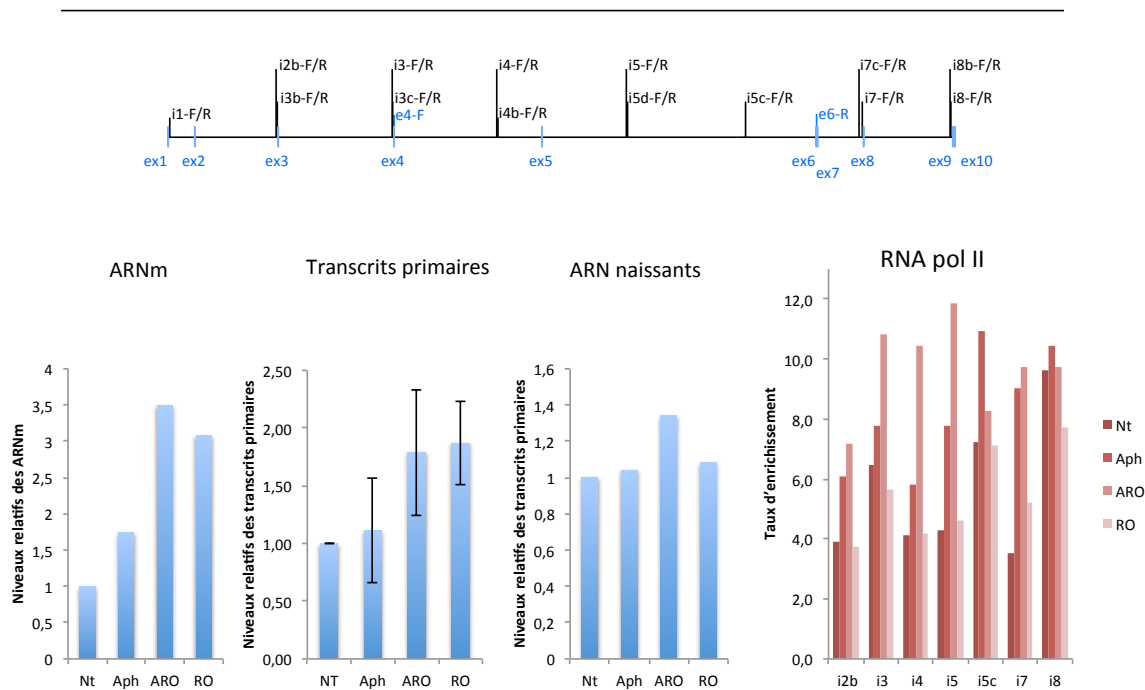
Afin de savoir si la diminution du taux de cassures au niveau des SFC que nous observons en traitant les cellules au RO-3306 est due, au moins en partie, à une diminution de l'expression des grands gènes qui se trouvent dans ces SFC, nous avons étudié par plusieurs approches la transcription des gènes *FHIT* et *WWOX* dans des cellules JEFF traitées au RO-3306. Nous avons, dans un premier temps, quantifié les ARNm par qRT-PCR en utilisant des amorces exoniques. Nous avons ensuite étudié le niveau de transcription en



quantifiant les transcrits primaires nucléaires par utilisation de différents couples d'amorces introniques, répartis tout le long des 2 gènes. Nous avons aussi déterminé la densité de l'ARN polymérase II le long des gènes *FHIT* et *WWOX* par la technique d'immunoprécipitation de la chromatine avec un anticorps ciblant les sérines 5 phosphorylées du domaine carboxy-terminal de la grande sous unité de la polymérase. Cette phosphorylation est spécifique de l'ARN polymérase II en élongation. Enfin nous avons quantifié le taux d'ARN néosynthétisé grâce à la technique de Click-It qui consiste à marquer les cellules pendant un temps court avec un analogue de l'uridine, l'éthynyluridine (EU), à extraire l'ARN puis à capturer les ARN naissants sur des billes afin de le quantifier par qRT-PCR. Les différents résultats obtenus montrent clairement que la transcription des gènes *FHIT* et *WWOX* dans les cellules JEFF n'est pas diminuée par le traitement au RO-3306 (Figure 25, Figure 26). De même, la transcription du gène *NEGR1* n'est pas réduite dans les MRC5 traitées au RO-3306 (Figure 27). La transcription du gène *LSAMP* (forme longue du gène, voir Annexe III) dans les cellules MRC5 est diminuée (- 40%) par le traitement au RO-3306 (Figure 28). Ces résultats montrent qu'il est possible de faire varier fortement la fragilité d'un SFC sans affecter la transcription du grand gène résident dans les mêmes proportions.

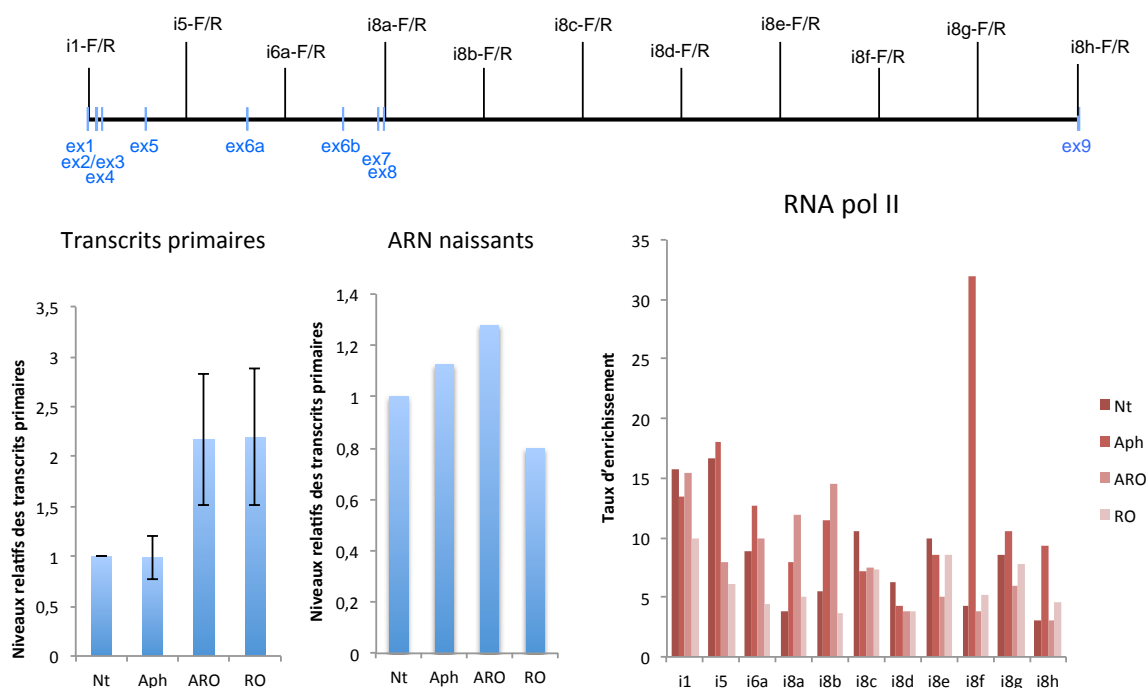
L'ensemble de ces résultats nous permet de conclure d'une part, qu'il existe une excellente corrélation entre fragilité et transcription et d'autre part, que nous pouvons supprimer la fragilité des SFC sans diminuer le niveau de transcription des grands gènes qui leur sont associés. L'ensemble de ces résultats montre donc clairement que le lien entre transcription et fragilité est plus complexe que celui proposé par Tora et collaborateurs (Helmrich et al., 2011) et que les mécanismes mis en jeu restent à élucider.

## Gène *FHIT*



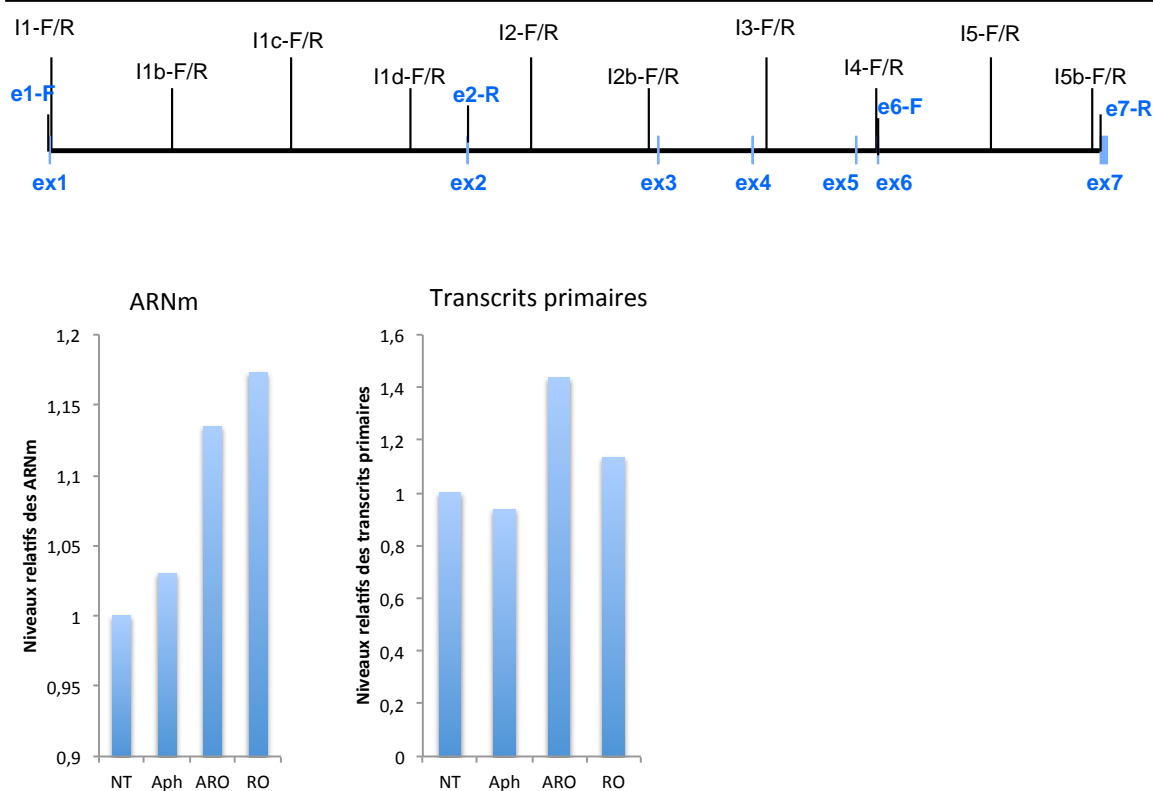
**Figure 25: Expression du gène *FHIT* dans les cellules lymphoblastoïdes JEFF traitées au RO-3306.** Les cellules n'ont pas été traitées (Nt) ou ont été traitées à l'aphidicoline (Aph) ou au RO-3306 (RO) ou à l'aphidicoline en présence de RO-3306 (ARO) comme décrit dans la Figure 14. Pour l'analyse des ARNm (une expérience) et des transcrits primaires (3 expériences) les ARN totaux ont été préparés puis étudiés par qRT-PCR avec, respectivement, des couples d'amorces exoniques (e4-F/e6-R) ou introniques (i1-F/R, i2b-F/R, i3-F/R, i4-F/R, i5-F/R, i5c-F/R, i7-F/R, i8-F/R). Dans ce dernier cas, les valeurs présentées correspondent pour chaque condition à la moyenne des valeurs obtenues avec les différents couples d'amorces introniques et les barres d'erreur représentent la déviation standard de ces valeurs. Pour l'étude des ARN naissants, les cellules ont été incubées 30 min en présence d'éthynyluridine (1 mM) puis les ARN totaux ont été extraits. Les ARN néosynthétisés ont été purifiés par chromatographie d'affinité (technique Click-It) puis analysés par qRT-PCR avec des couples d'amorces introniques (i1-F/R, i3b-F/R, i3c-F/R, i4b-F/R, i5d-F/R, i5c-F/R, i7c-F/R, i8b-F/R). Dans les trois cas, les niveaux de transcrits sont normalisés par rapport à celui de l'ARNm de la cyclophiline B. Pour la détermination de la densité des ARN polymérase II le long du gène, les cellules ont été fixées au formaldéhyde et la chromatine fragmentée a été préparée. Elle a été analysée par la technique d'immunoprécipitation de la chromatine avec un anticorps dirigé contre la forme phosphorylée de la sérine 5 du domaine C-terminal de l'ARN polymérase II ou des immunoglobulines d'animal non-immun. Les résultats présentés correspondent aux taux d'enrichissement de la séquence testée dans l'immunoprécipité par rapport à la valeur obtenu avec les immunoglobulines de contrôle. L'expérience a été réalisée deux fois avec des résultats comparables.

## Gène *WWOX*



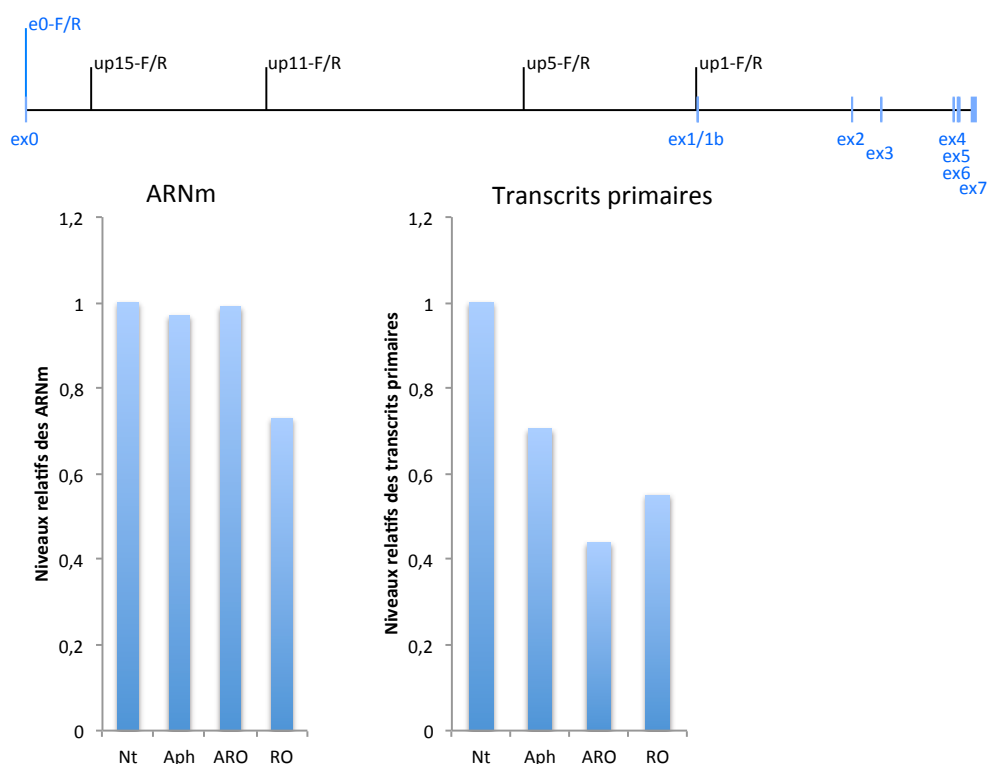
**Figure 26 : Expression du gène *WWOX* dans les cellules lymphoblastoïdes JEFF traitées au RO-3306.** Les cellules ont été traitées comme décrit dans la Figure 14. Pour l'analyse des transcrits primaires (2 expériences), les ARN totaux ont été préparés puis étudiés par qRT-PCR avec des couples d'amorces introniques (i1-F/R, i5-F/R, i6a-F/R, i8a-F/R, i8b-F/R, i8c-F/R, i8d-F/R, i8e-F/R, i8f-F/R, i8g-F/R, i8h-F/R). Les valeurs présentées correspondent pour chaque condition à la moyenne des valeurs obtenues avec les différents couples d'amorces introniques et les barres d'erreur représentent la déviation standard de ces valeurs. Pour l'étude des ARN naissants, les cellules ont été incubées 30 min en présence d'éthynyluridine (1 mM) puis les ARN totaux ont été extraits. Les ARN néosynthétisés ont été purifiés par chromatographie d'affinité (technique Click-It) puis analysés par qRT-PCR avec des couples d'amorces introniques (i1-F/R, i5-F/R, i6a-F/R, i8a-F/R, i8b-F/R, i8c-F/R, i8d-F/R, i8e-F/R). Dans les deux cas, les niveaux de transcrits sont normalisés par rapport à celui de l'ARNm de la cyclophylène B. Pour la détermination de la densité des ARN polymérase II le long du gène, la chromatine a été préparée et analysée par la technique d'immunoprécipitation de la chromatine comme décrit dans la Figure 25.

## Gène *NEGR1*



**Figure 27 : Expression du gène *NEGR1* dans les fibroblastes MRC5 traitées au RO-3306.** Les cellules ont été traitées comme décrit dans la Figure 14. Pour l'analyse des ARNm et des transcrits primaires, les ARN totaux ont été préparés puis étudiés par qRT-PCR, respectivement, avec des couples d'amorces exoniques (e1-F/e2-R, e6-F/e7-R) ou introniques (i1-F/R, i1b-F/R, i1c-F/R, i1d-F/R, i2-F/R, i2b-F/R, i3-F/R, i4-F/R, i5-F/R, i5b-F/R). Les valeurs présentées correspondent, pour chaque condition, à la moyenne des valeurs obtenues avec les différents couples d'amorces exoniques ou introniques. Dans les deux cas les niveaux de transcrits sont normalisés par rapport à celui de l'ARNm de la cyclophyliline B.

## Gène *LSAMP*

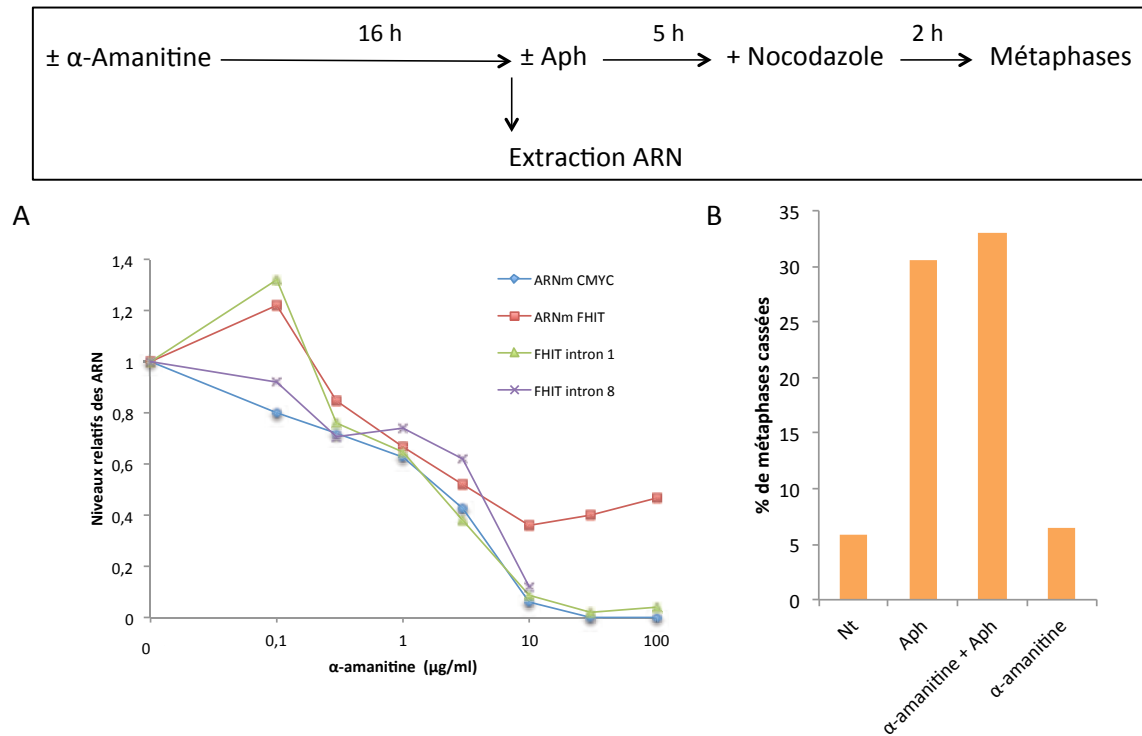


**Figure 28 : Expression du gène *LSAMP* dans les fibroblastes MRC5 traitées au RO-3306.** Les cellules ont été traitées comme décrit dans la Figure 14. Pour l'analyse des ARNm et des transcrits primaires, les ARN totaux ont été préparés puis étudiés par qRT-PCR, respectivement, avec des couples d'amorces exoniques (e0-F/R) ou introniques (up15-F/R, up11-F/R, up5-F/R, up1-F/R). Les valeurs présentées pour les transcrits primaires correspondent pour chaque condition à la moyenne des valeurs obtenues avec les différents couples d'amorces introniques. Dans les deux cas, les niveaux de transcrits sont normalisés par rapport à celui de l'ARNm de la cyclophylène B.

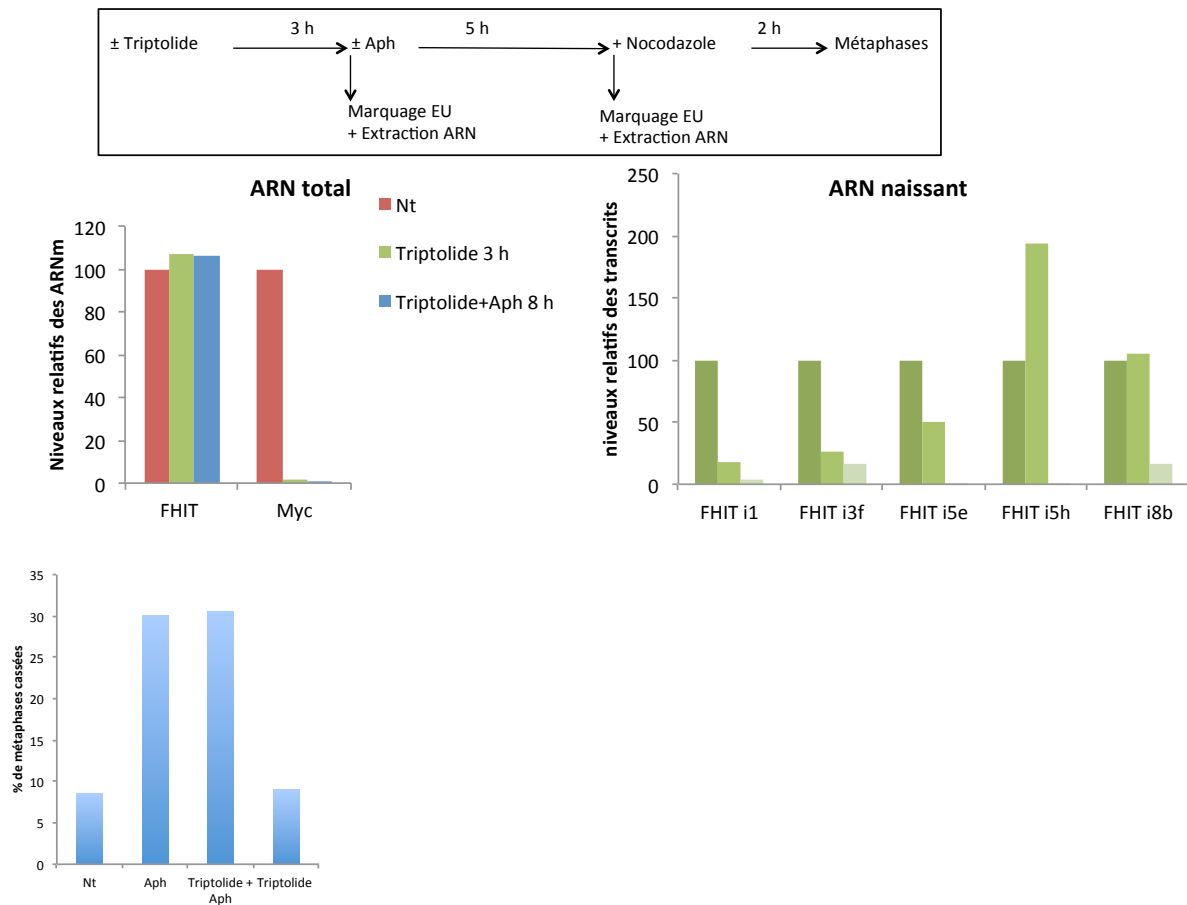
### 3. Inhibition de la transcription et effet sur la fragilité des SFC

Afin d'aller plus loin dans l'investigation du lien entre transcription et fragilité, nous avons entrepris de déterminer si l'inhibition de la transcription des grands gènes réduit la fragilité des SFC qui les hébergent. Nous avons eu recours à plusieurs inhibiteurs de la transcription : 1) l' $\alpha$ -amanitine qui, à faibles concentrations, bloque la progression des ARN polymérase II sur l'ADN (Bensaude, 2011). La transcription est inhibée et les ARN polymérase bloquées sont finalement dégradées au bout de quelques heures (Anindya et al., 2007) ; 2) le triptolide qui, en moins de trois heures, inhibe la transcription en bloquant les ARN polymérase au niveau du promoteur puis induit leur polyubiquitination et leur dégradation (Manzo et al., 2012). Ces inhibiteurs vidangent donc rapidement les ARN

polymérase des gènes et éliminent ainsi d'éventuelles collisions avec la machinerie de réplication. Nous avons d'abord confirmé l'action de ces inhibiteurs sur la transcription du gène *FHIT* et sur celle du gène *CMYC* dont l'ARNm à une très courte demi-vie (30 min) (McCormack et al., 1984). Nous avons ensuite montré que l'inhibition de la transcription par l'un ou l'autre de ces deux inhibiteurs est sans effet sur le taux de cassures totales induit par l'aphidicoline (Figure 29, Figure 30).



**Figure 29 : Effet de l'α-amanitine sur la transcription et sur les cassures chromosomiques induites par l'aphidicoline.** **A)** Des cellules JEFF ont été incubées pendant 16 h en présence des doses indiquées d'α-amanitine puis les ARN totaux ont été préparés. Le niveau des ARNm (*CMYC*, *FHIT*) ou des transcrits introniques (*FHIT* introns 1 et 8) a été déterminé par qRT-PCR avec des couples d'amorces exoniques (*CMYC* F/R, *FHIT* e4-F/e6-R) ou introniques (*FHIT* i1-F/R; i8-F/R) (voir Figure 25). Les niveaux de transcrits sont donnés relativement à celui des cellules non traitées. Remarquer que la transcription du gène *FHIT* est très sensible à l'α-amanitine mais que le niveau de l'ARNm ne décroît que modérément, vraisemblablement à cause de sa longue demi-vie. La demi-vie de l'ARNm de *CMYC* est d'environ 30 min. **B)** Effet de l' α-amanitine sur les cassures chromosomiques totales. Des cellules JEFF ont été pré-incubées 16 h en absence ou en présence d' α-amanitine (10 µg/ml) puis incubées 5 h en présence ou en absence d' α-amanitine et d'aphidicoline (Aph 0,6 µM). Des métaphases ont été préparées par traitement au nocodazole et les cassures chromosomiques totales ont été dénombrées après coloration au Giemsa.

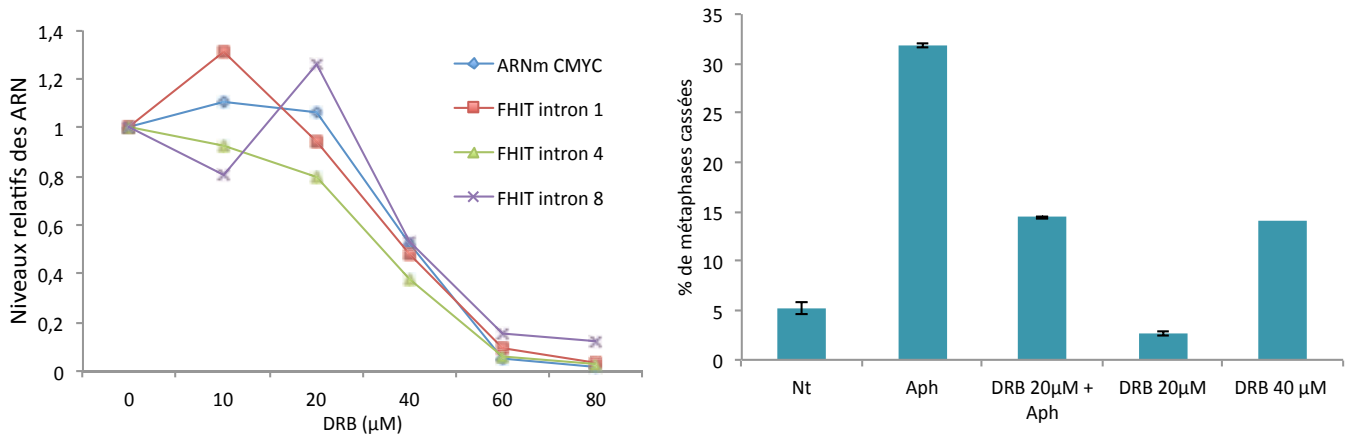
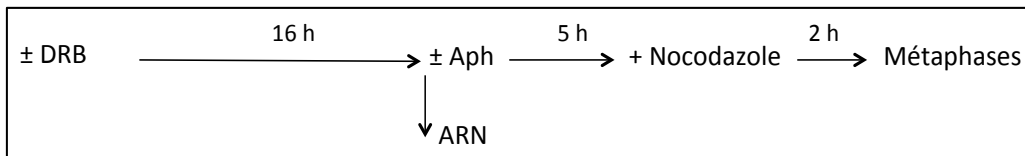


**Figure 30 : Effet du triptolide sur la transcription et sur les cassures chromosomiques induites par l'aphidicoline. A)** Des cellules JEFF ont été pré-incubées 3 h en absence ou en présence de triptolide (1  $\mu$ M) puis incubées 5 h en présence ou en absence de triptolide et d'aphidicoline (Aph 0,6  $\mu$ M). A la fin de chaque période d'incubation, des cellules ont été marquées pendant 30 min en présence d'éthynyluridine (1 mM) puis les ARN totaux ont été extraits. Les ARN néosynthétisés ont été purifiés par chromatographie d'affinité (technique Click-It). Les niveaux des ARNm (*CMYC* et *FHIT*) ou des transcrits primaires (*FHIT* introns 1, 3, 5 et 8) ont été déterminés par qRT-PCR avec des couples d'amorces exoniques (*CMYC* F/R, *FHIT* e4-F/e6-R) ou introniques (i1-F/R, i3f-F/R, i5e-F/R, i5h-F/R, i8b-F/R). Pour chaque couple d'amorces, les niveaux de transcrits sont donnés relativement au niveau obtenu dans les cellules non traitées. Remarquer que le niveau d'ARNm de *FHIT* ne varie pas dans les cellules traitées, vraisemblablement à cause de sa longue demi-vie, alors que celui de l'ARNm de *CMYC* s'effondre (demi-vie de 30 min). Noter aussi que des transcrits primaires correspondant aux introns situés en 3' du gène *FHIT* (intron 5 et 8) sont encore détectables après 3 h de traitement mais disparaissent après 8 h. Ce résultat est complètement en accord avec le mode d'action décrit du triptolide : blocage des ARN polymérase sur le promoteur et vidange du gène des polymérase qui terminent la transcription. **B)** Effet du triptolide sur les cassures chromosomiques. Des cellules JEFF ont été traitées ou non au triptolide et à l'aphidicoline comme décrit en A. Des métaphases ont été préparées par traitement au nocodazole et les cassures chromosomiques totales ont été dénombrées après coloration au Giemsa.

Le troisième inhibiteur que nous avons voulu utiliser est le DRB (5,6-dichloro-1-beta-ribofuranosyl benzimidazole). C'est un inhibiteur de la phosphorylation activatrice des serines 5 du domaine carboxy-terminal de la grande sous unité de l'ARN polymérase II (Bensaude, 2011). En présence de cette drogue, les polymérases fixées au niveau du promoteur ne peuvent pas entrer en élongation productive. Par contre, les polymérases déjà engagées dans le gène peuvent terminer la transcription. Après un certain temps, le gène est donc vidé des ARN polymérases. Nous avons découvert que le prétraitement de cellules JEFF avec de faibles doses de DRB (20  $\mu$ M) réduit des 2/3 le taux de cassures induit par l'aphidicoline. Cependant, nous avons montré qu'à ces faibles doses, le DRB n'a d'effet ni sur la transcription du gène *FHIT* ni sur celle de *CMYC*. Nous n'avons pas d'explication à cette réduction des cassures sous DRB qui constitue une observation inédite et intéressante. Nous pouvons cependant conclure qu'elle est indépendante de l'état transcriptionnel du gène. Le traitement des cellules avec des doses plus élevées de DRB (40  $\mu$ M) et plus, qui permettent l'inhibition de la transcription, induit par lui même des cassures chromosomiques ce qui rend l'expérience ininterprétable (Figure 31).

La limite de ces inhibiteurs est qu'ils ont un effet général sur la transcription et n'agissent pas uniquement sur les gènes localisés dans les SFC. Pour éviter ce problème, le projet a été poursuivi par d'autres membres du laboratoire en utilisant une méthode qui permet de moduler spécifiquement la transcription d'un gène. Elle implique l'utilisation d'activateurs transcriptionnels dérivés des TALE (transcription activator-like effector) qui permettent de cibler spécifiquement un gène. De tels activateurs ciblant le promoteur du gène *FHIT* ont été construits grâce à une collaboration avec la société Celectis. Je n'ai pas participé à la suite de ce projet.





**Figure 31 : Effet du DRB sur la transcription et sur les cassures chromosomiques induites par l'aphidicoline.** **A)** Des cellules JEFF ont été incubées pendant 16 h en présence des doses indiquées de DRB puis les ARN totaux ont été préparés. Le niveau des ARNm (*CMYC*) ou des transcrits introniques (*FHIT* introns 1, 4 et 8) a été déterminé par qRT-PCR avec des couples d'amorces exoniques (*CMYC* F/R) ou introniques (*FHIT* i1-F/R; i4-F/R; i8-F/R). Les niveaux de transcrits sont donnés relativement à celui des cellules non traitées. **B)** Effet du DRB sur les cassures chromosomiques totales. Des cellules JEFF ont été pré-incubées 16 h en absence ou en présence de DRB (20 μM ou 40 μM) puis incubées 5 h en présence ou en absence de DRB et d'aphidicoline. Des métaphases ont été préparées par traitement au nocodazole et les cassures chromosomiques totales ont été dénombrées après coloration Giemsa. Les barres d'erreur représentent l'écart type de la moyenne de deux expériences, sauf pour le point DRB 40 μM (une seule expérience).

#### IV - Effet du RO-3306 sur la dynamique de réplication

Nous avons montré que le traitement par le RO-3306 supprime les cassures induites par l'aphidicoline, ce qui suggère que le blocage en prolongé G2 permet à la réplication de se terminer au niveau des SFC. Afin de tester cette hypothèse, nous avons utilisé la technique du peignage moléculaire de l'ADN pour cartographier les événements de terminaison et d'initiation de la réplication le long du gène *FHIT* dans des cellules JEFF traitées à l'aphidicoline et au RO-3306. La technique du peignage moléculaire consiste à marquer séquentiellement l'ADN néosynthétisé par 2 analogues différents de la thymidine, l'iododéoxyuridine (IdU) et la chlorodéoxyuridine (CldU). L'ADN est ensuite extrait par une

méthode permettant de conserver des molécules de grande taille (>200 Kb) qui sont ensuite fixées par une extrémité et étirées (peignage) sur des lamelles silanisées. Les deux analogues incorporés *in vivo* sont ensuite révélés par deux anticorps différents. Ce double marquage permet de déterminer la direction de propagation des fourches et d'identifier les origines (fourches divergentes) et les terminaisons (fourches convergentes). Cette technique permet aussi de calculer la vitesse de progression des fourches de réplication, grâce à la régularité de l'étirement des molécules d'ADN, ainsi que la densité des fourches de réplication. Elle peut être combinée à la technique de FISH afin d'étudier la dynamique de réplication au niveau d'un locus donné et de cartographier les événements d'initiation et de terminaison qui ont lieu à son niveau (Lebofsky and Bensimon, 2005).

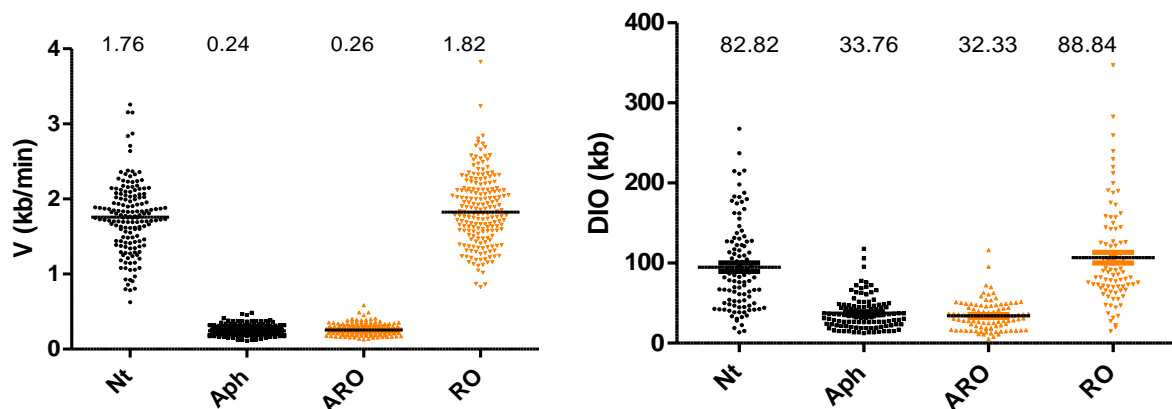
### **1. La technique du peignage moléculaire**

Dans un article publié en 2013 auquel j'ai participé (Techer et al., 2013), nous avons discuté les aspects techniques du peignage moléculaire. En effet, nous avons mis en évidence des points critiques qui doivent être respectés afin de pouvoir obtenir des résultats reproductibles et tirer des conclusions correctes avec cette technique. Il y a notamment une ambiguïté dans l'analyse des signaux mono-couleur correspondant au premier marquage. Certains auteurs considèrent qu'ils correspondent à des fourches bloquées pendant ce marquage et d'autres considèrent qu'ils correspondent à des terminaisons qui se produisent pendant ce premier marquage. Nous avons montré que ces signaux doivent être considérés comme des terminaisons ayant lieu pendant le premier marquage. Par ailleurs, nous avons montré, que pour une bonne application du peignage moléculaire, les fibres d'ADN doivent avoir une taille minimum de 200 kb et que 100 à 200 mesures doivent être effectuées pour chaque condition analysée. Il est également important de contre-colorer les molécules d'ADN par un anticorps dirigé contre l'ADN monocaténaire. Une copie de l'article est fournie en Annexe IV pour les lecteurs désirant de plus amples explications.

### **2. Effet du RO-3306 sur la dynamique de réplication au niveau pan-génomique**

Dans un premier temps nous avons voulu vérifier que le traitement de cellules JEFF par le RO-3306 ne change pas la dynamique de réplication globale et, notamment, que l'action de cet inhibiteur ne compense pas (par un mécanisme inconnu) la réduction de vitesse de réplication provoquée par le traitement à l'aphidicoline, éliminant ainsi le stress

réplicatif induit par cette dernière. Nos résultats montrent que la vitesse de réplication et la distance inter-origines sont réduites de 8 et 3 fois, respectivement, dans des cellules JEFF traitées avec 0,6  $\mu$ M d'aphidicoline seul par rapport à des cellules non traitées, ce qui est en accord avec les résultats précédemment obtenus dans le laboratoire pour ces cellules (Letessier et al., 2011). Ces mêmes taux de réduction sont trouvés pour les cellules traitées simultanément à l'aphidicoline et au RO-3306, montrant que cette drogue n'a pas d'action générale sur la dynamique de réplication (Figure 32).



**Figure 32 : Effet du RO-3306 sur la vitesse de progression des fourches de réplication et sur les distances inter-origines.** Des cellules JEFF ont été traitées comme indiqué dans la Figure 14 et les vitesses des fourches de réplication (V) ainsi que les distances inter-origines ont été déterminées par peignage moléculaire. Chaque point représente une valeur individuelle. Pour chaque condition, la barre horizontale représente la valeur médiane qui est donnée en haut des graphiques. Les différences entre cellules non traitées (Nt) et cellules traitées au RO-3306 et entre cellules traitées à l'aphidicoline (Aph) et cellules traitées à l'aphidicoline et au RO-3306 (ARO) ne sont pas significatives selon le test de Mann-Whitney.

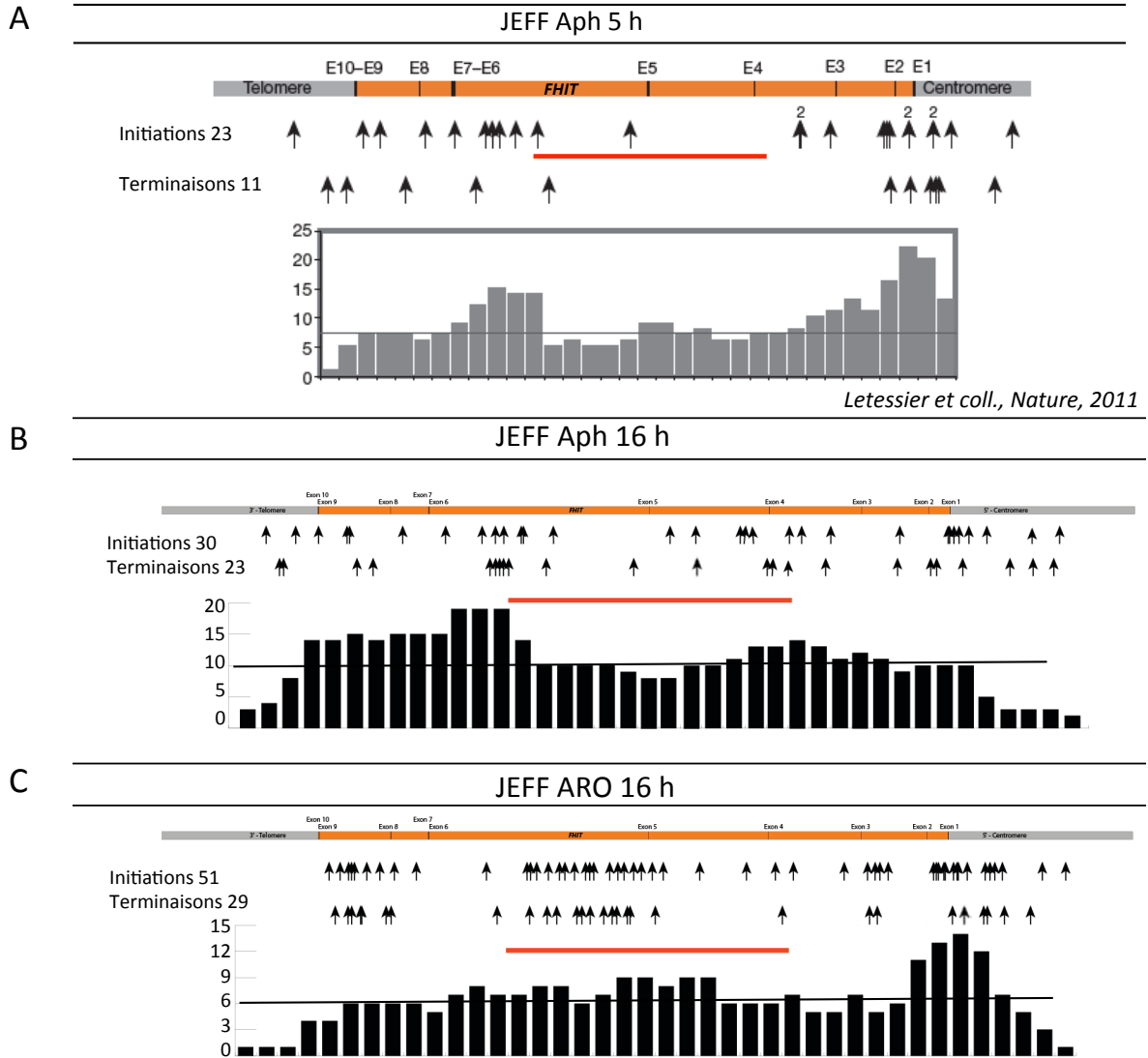
### 3. Effet du RO-3306 sur la dynamique de réplication au niveau du site *FRA3B*

Nous avons ensuite analysé les événements d'initiation et de terminaison de la réplication dans une région de 1,6 Mb contenant le gène *FHIT*, chez des cellules JEFF traitées pendant 16 h à l'aphidicoline en présence ou en absence de RO-3306. Les molécules d'ADN peignées portant le locus sont repérées par un ensemble de 33 sondes formant des motifs d'hybridation uniques (a-e, i) répartis sur tout le locus (code morse). Les molécules portant tout ou partie de cette région et présentant des signaux de réplication sont alignées sur la carte du locus grâce au code morse. Les signaux de réplication sont ensuite analysés de façon à déterminer la position des événements d'initiation et de terminaison. Il faut noter

que pour pouvoir comparer deux conditions différentes par cette technique, il faut corriger les résultats en fonction de la couverture le long du locus. Ce paramètre correspond au nombre de fibre d'ADN analysées qui couvrent entièrement chaque fenêtre de 50 kb du locus (Figure 33).

Les résultats obtenus montrent d'abord que la région cœur, pauvre en événements d'initiation et de terminaison est moins aisément discernable dans les cellules traitées pendant 16 h à l'aphidicoline seule qu'elle ne l'était après 5 h de traitement. En effet, quelques événements d'initiation et de terminaison y sont détectés. Ceci est vraisemblablement dû à un mécanisme de compensation qui entraîne le déclenchement d'origines latentes lorsque les fourches sont ralenties. Néanmoins, ce phénomène n'entraîne pas une réduction de la fragilité du site. Quand nous avons commencé ce projet, notre hypothèse était qu'en bloquant les cellules en G2, le RO-3306 laisse le temps à la réplication des SFC de se terminer avant l'entrée en mitose malgré le ralentissement des fourches en présence d'aphidicoline. Les résultats montrent effectivement une augmentation très nette de la densité des événements de terminaison dans la région cœur après traitement au RO-3306. Ceci indique que, dans les cellules traitées à l'aphidicoline en présence de RO-3306, la réplication de la région cœur arrive à se terminer malgré la faible vitesse de progression des fourches, ce qui explique la perte de la fragilité du site dans ces conditions. Cependant, par peignage moléculaire, nous avons aussi mis en évidence une augmentation importante des événements initiation dans la région cœur après traitement au RO-3306. Le profil obtenu rappelle fortement celui observé à ce même locus dans les fibroblastes MRC5 où le site n'est pas fragile (Letessier et al., 2011). Dans ce type cellulaire, la région cœur n'existe pas, les événements d'initiation et de terminaison étant régulièrement distribués tout au long du gène *FHIT*.

Ce résultat indique qu'une grande partie des événements de terminaison observés dans le cœur résultent du déclenchement de ces origines et non de fourches hautement processives venant des régions flanquantes. Ceci suggère que le déclenchement d'origines dans le cœur du gène *FHIT* chez les cellules JEFF traitées au RO-3306 est pour une large part responsable de la suppression des cassures à *FRA3B*.



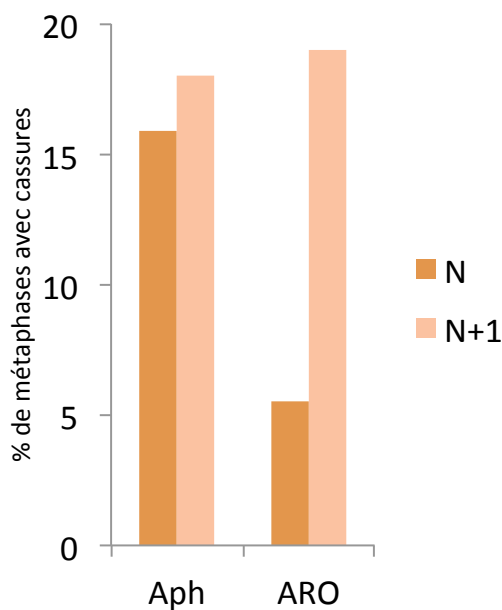
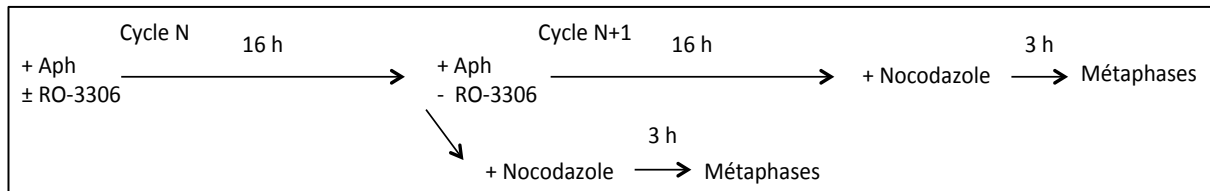
**Figure 33 : Cartographie des initiations et des terminaisons dans la région du gène *FHIT* chez les cellules lymphoblastoïdes JEFF.** Les cellules ont été traitées 5 h (A) ou 16 h (B) par l'aphidicoline (Aph; 0,6  $\mu$ M) ou 16 h par l'aphidicoline (0,6  $\mu$ M) et le RO-3306 (10  $\mu$ M) (ARO) (C). Le panneau (A) est repris de Letessier et coll. (Letessier et al., 2011). Les positions des initiations et des terminaisons ont été déterminées par peignage moléculaire et sont indiquées sur la carte du gène *FHIT* avec ses exons. La position de la région cœur pauvre en initiations déterminée par Letessier et coll. est indiquée par une barre rouge. Les histogrammes représentent la couverture avec la valeur médiane (trait horizontal), c'est-à-dire le nombre de fibres d'ADN qui couvrent entièrement chaque fenêtré de 50 kb.

Nous avons donc décidé d'étudier plus en détail ces origines et le mécanisme de leur déclenchement en posant trois questions :

- 1) les origines latentes du cœur dont le RO-3306 a provoqué le déclenchement gardent-elles une « marque » de cet événement leur permettant de se déclencher au cours des cycles suivants en absence de RO-3306 ?

2) Le RO-3306 est-il capable, comme dans le cas de *FRA3B*, de déclencher des origines latentes au niveau d'autres SFC?

3) Par quel mécanisme le traitement au RO-3306 déclenche-t-il les origines latentes du cœur de *FRA3B* ?



**Figure 34: Effet du RO-3306 sur les cassures induites au cycle suivant.** Des cellules JEFF ont été incubées pendant un cycle complet (16 h) en présence d'aphidicoline (0,6  $\mu$ M) avec ou sans RO-3306 (10  $\mu$ M) puis pendant le cycle suivant en présence d'aphidicoline. Des métaphases ont été préparées par traitement au nocodazole après le premier ou le second cycle et les cassures totales ont été déterminées après coloration au Giemsa

Afin de répondre à la première question, nous avons réalisé l'expérience suivante. Des cellules JEFF ont été traitées à l'aphidicoline en présence ou en absence de RO-3306 pendant un cycle (cycle N), puis le RO-3306 a été retiré du milieu de culture et les cellules ont été de nouveau mises sous aphidicoline pendant un deuxième cycle (cycle N+1). Nous avons alors déterminé le taux de cassures totales dans les cellules en métaphases. Le résultat obtenu montre que le taux de cassures dans les cellules qui ont été traitées par le RO-3306 pendant le cycle N n'est pas réduit par rapport à celui des cellules qui ne l'ont pas été (Figure 34). Ceci suggère que les origines latentes dont le RO-3306 a permis le déclenchement au cycle N antérieur n'ont pas pu se déclencher au cycle N+1 en absence de

la drogue. Ceci montre que ces origines n'ont pas conservé de marque transmissible à la génération suivante.

Nous allons essayer de répondre aux questions 2 et 3 dans les deux prochains paragraphes.

#### **V. Effet du RO-3306 sur le programme temporel de réplication des SFC**

Un moyen de savoir si le RO-3306 est capable, comme dans le cas de *FRA3B*, de déclencher des origines latentes au niveau d'autres SFC, serait de cartographier les événements d'initiation et de terminaison au niveau d'autres SFC dans des cellules traitées à l'aphidicoline et au RO-3306. Répondre à cette question par la technique du peignage moléculaire nécessiterait de mettre au point un code morse pour chaque locus et serait extrêmement lourd et chronophage. Nous avons donc envisagé d'utiliser la technique de Repli-Seq pour déterminer le timing de réplication dans ces conditions à différents loci. Nous avons conscience que cette technique ne permet pas de cartographier les événements d'initiation et de terminaison avec la même résolution que le peignage moléculaire, mais elle permet de vérifier si la réplication a pu se terminer au niveau des autres SFC chez les cellules traitées au RO-3306 ou si elle est restée inachevée comme chez les cellules traitées uniquement à l'aphidicoline. Par ailleurs, cette technique donne une information supplémentaire par rapport au peignage moléculaire. Elle devrait permettre de déterminer le moment de déclenchement des origines latentes au cours de la phase S dans le cœur de *FRA3B* et peut être au niveau d'autres SFC. Elle va permettre surtout de confirmer que la complétion de la réplication au niveau de *FRA3B* ainsi qu'au niveau des autres SFC a bien lieu. Les expériences sont actuellement en cours, en collaboration avec l'équipe de Claude Thermes.

#### **VI. Mécanisme d'action du RO-3306 dans le déclenchement des origines latentes du cœur de *FRA3B***

En abordant cette question nous avons envisagé deux possibilités : 1) comme chez les fibroblastes, des origines chargées en complexe de pré-réplication existeraient dans le cœur de *FHIT* chez les lymphocytes, mais leur déclenchement serait très difficile (très tardif) dans ces cellules, empêchant leur activation dans les conditions de normales culture. Dans ce cas,

le RO-3306 aurait un rôle passif dans leur activation en leur laissant suffisamment de temps pour se déclencher ; 2) les origines du cœur du gène *FHIT* ne se déclenchent pas en conditions normales parce qu'elles n'ont pas été chargées en complexes de pré-RC, ou l'ont été seulement partiellement. L'inhibition de CDK1 permettrait la mise en place en G2 d'un complexe de pré-RC fonctionnel et ainsi le déclenchement de ces origines par un mécanisme restant à déterminer. En situation normale, les origines de réplication sont chargées en fin de mitose et en G1. Comme nous l'avons déjà expliqué dans l'introduction, plusieurs mécanismes inhibent ce processus pendant les phases S et G2, évitant ainsi la re-réplication. Ce processus nous paraît cependant compatible avec notre deuxième hypothèse si l'on considère les nombreuses données bibliographiques impliquant CDK1 dans l'inhibition du chargement des origines au cours des phases S et G2.

En 1997 l'étude d'un mutant conditionnel de CDK1 dans des cellules humaines a suggéré que CDK1 aurait un rôle dans l'inhibition de la formation des complexes de pré-RC au cours des phases G2/M (Itzhaki et al., 1997). En bon accord avec ce modèle, il avait préalablement été montré que CDK1 phosphoryle le facteur MCM4 et diminue son affinité pour la chromatine dans des extraits embryonnaires de xénope (Hendrickson et al., 1996). D'autres auteurs ont montré ensuite que les protéines MCM2 et MCM4 sont des substrats du complexe CDK1/cycline B dans les cellules humaines Hela et que l'inactivation de ce complexe chez un mutant murin thermosensible de CDK1 empêche la phosphorylation de ces deux MCM et leur relargage de la chromatine (Fujita et al., 1998). Grâce à ce même mutant, il a été ensuite établi que le facteur CDT1 est aussi phosphorylé par le complexe CDK1/cycline B et que son inhibition conduit à la refixation de CDT1 à la chromatine dans des cellules murines synchronisées à la frontière G2/M (Sugimoto et al., 2004). En se basant sur un autre modèle murin présentant une délétion dans CDT1 ( $\Delta^6$ PEST), d'autres auteurs ont plus récemment montré que la protéine CDT1 murine est phosphorylée en G2/M au niveau de son domaine PEST, ce qui empêche sa fixation à la chromatine. Chez d'autres vertébrés, CDT1 possède un domaine apparenté à PEST. Bien que ce domaine puisse présenter des différences de taille ou de séquence, les sites de phosphorylation potentiels sont bien conservés et leur phosphorylation pourrait être assurée par CDK1 (Coulombe et al., 2013). De plus, l'inhibition de CDK1 par la roscovitine dans des cellules humaines U2OS conduit au recrutement de la forme non phosphorylée de CDT1 à la chromatine. L'inhibition de CDK1 seule pourrait donc suffire à la formation prématurée de complexes de pré-RC (Ballabeni et

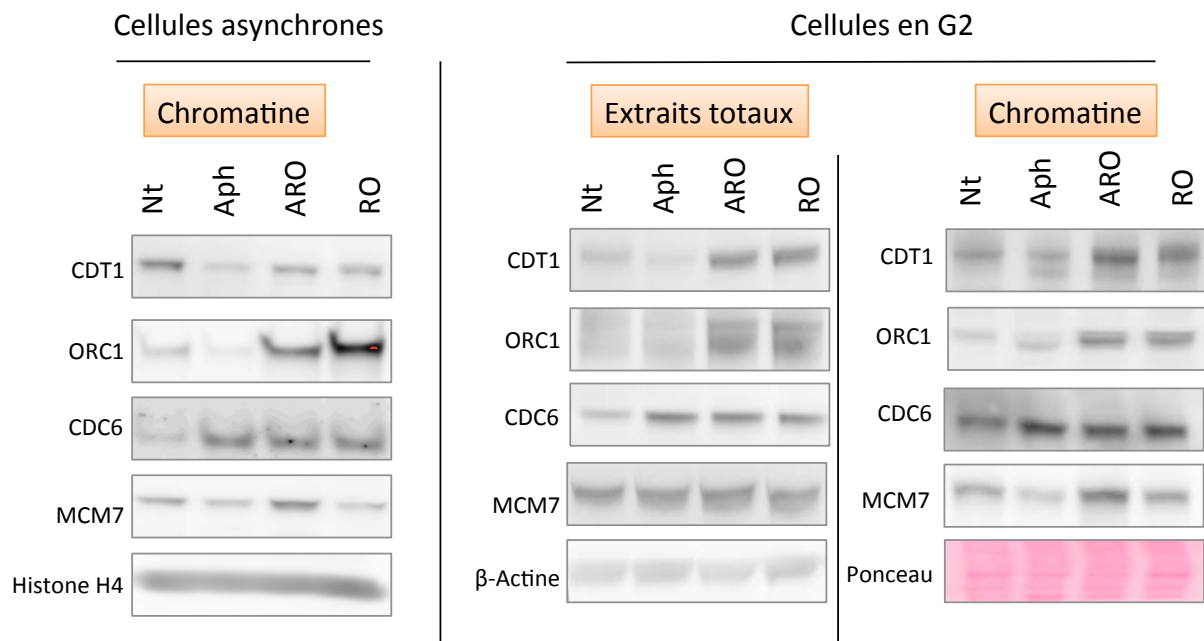


al., 2004). Par ailleurs, il a été montré chez des cellules de hamster que l'inhibition du complexe CDK1/cycline A par des inhibiteurs de kinase (2AP ou olomoucine) empêche la phosphorylation d'ORC1 et permet une fixation rapide et stable de ce dernier à la chromatine (Li et al., 2004). En 2014, une étude portant sur des cellules U2OS traitées au RO-3306 a montré que CDK1 participe à l'exportation de CDC6 du noyau dans les cellules qui progressent de S en G2 (Clijsters and Wolthuis, 2014). Finalement, une étude récente du protéome de cellules leucémiques promyélocyaires humaines NB4 a mis en évidence une surexpression des protéines de pré-RC telles que ORC1, ORC3, ORC5, MCM6 et CDT1 après traitement au RO-3306 (Ly et al., 2015).

### **1. Effet du RO-3306 sur l'accumulation des protéines du complexe de pré-réplication**

Dans le but de déterminer si CDK1 régule les niveaux de certains facteurs du complexe de pré-RC dans les cellules JEFF, nous avons évalué la quantité de certaines de ces protéines (ORC1, CDC6, CDT1, MCM7) par Western dans des cellules traitées ou non à l'aphidicoline en présence ou en absence de RO-3306. Nous avons étudié des extraits protéiques totaux ainsi que des extraits chromatinien. Afin d'étudier de façon plus spécifique les cellules en phase G2, où l'activité de CDK1 est maximale, nous avons trié des cellules JEFF en fonction de leur position dans le cycle cellulaire (G1, S1, S2, G2M) et nous avons déterminé le niveau de ces protéines dans des extraits protéiques totaux ainsi que dans la chromatine.

Nos résultats montrent que les protéines ORC1, CDT1 et MCM7 s'accumulent à la chromatine dans les cellules traitées à l'aphidicoline en présence de RO-3306 par rapport aux cellules traitées uniquement à l'aphidicoline. Le facteur CDC6 s'accumule à la chromatine sous aphidicoline seule, sous RO-3306 seul ainsi qu'en présence des deux inhibiteurs. Nous avons obtenus les mêmes résultats pour la chromatine des cellules en phase G2, à savoir une augmentation des facteurs ORC1, CDT1, CDC6 et MCM7 dans les cellules traitées au RO-33 06 (Figure 35).



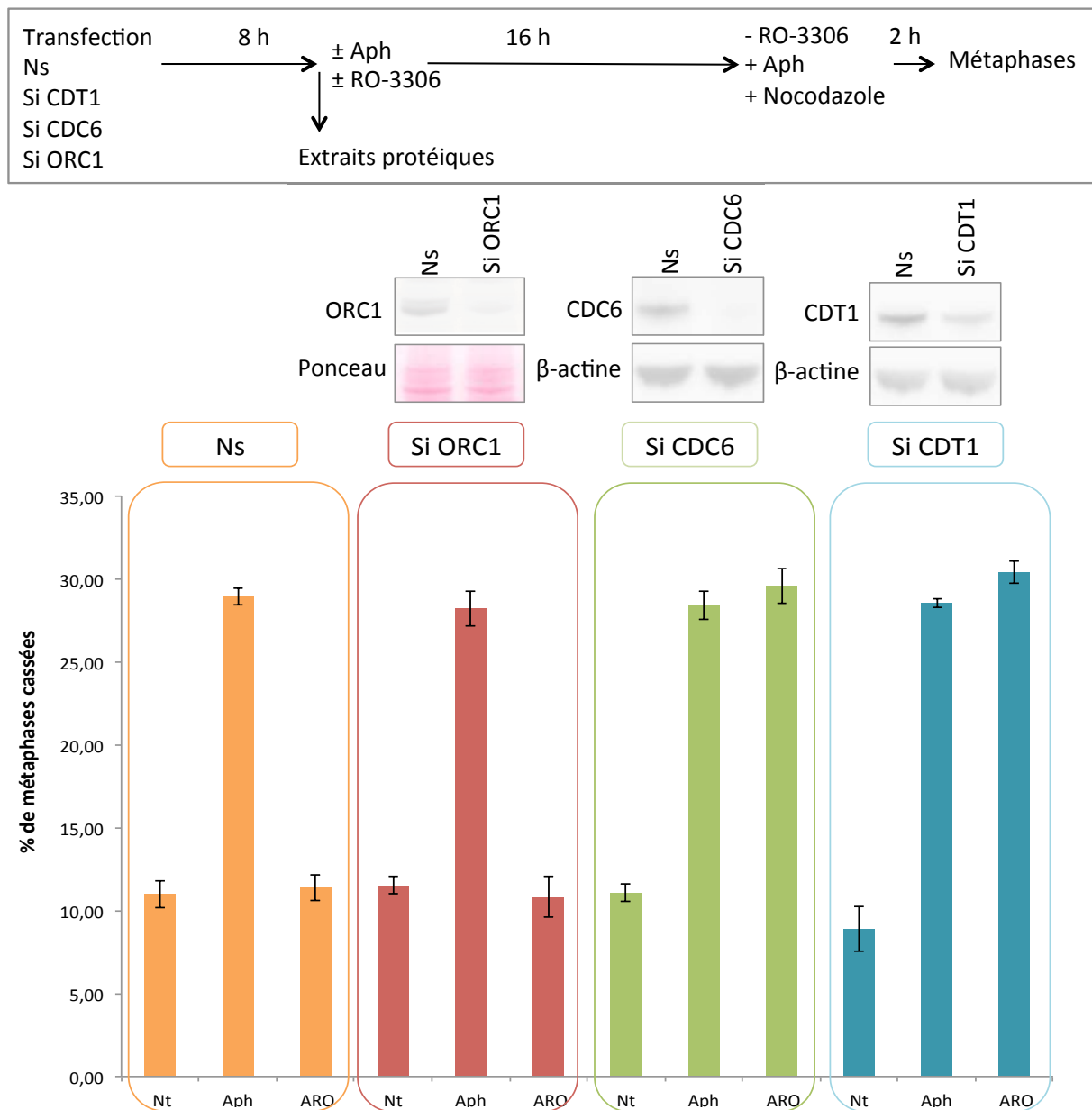
**Figure 35 : Accumulation de protéines du complexe de pré-réplication dans les cellules traitées à l'aphidicoline et/ou au RO-3306.** A) Des cellules JEFF ont été traitées comme indiqué dans la Figure 14 puis des extraits protéiques totaux ont été préparés et analysés par Western à l'aide d'anticorps dirigés contre les protéines indiquées. L'histone H4 est utilisée comme contrôle de charge. B) Des cellules JEFF ont été traitées comme indiqué dans la Figure 14 puis les cellules en phase G2 ont été triées au FACS. Des extraits protéiques totaux ou chromatiniens ont ensuite été préparés à partir des cellules triées et analysés par Western comme décrit en A. La  $\beta$ -actine ou une coloration de la membrane au rouge ponceau ont été utilisés comme contrôles de charge.

Afin de déterminer si les effets du RO-3306 sur la suppression des cassures induites par l'aphidicoline dépendent ou non de l'accumulation à la chromatine de CDT1, CDC6 ou ORC1, nous avons entrepris de bloquer leur accumulation. En pratique, nous avons transfectées des cellules JEFF avec des siRNA dirigés contre les ARNm d'ORC1, de CDC6 ou de CDT1. Huit heures après transfection, les cellules ont été traitées ou non à l'aphidicoline avec ou sans RO-3306 pendant 16 h et les métaphases ont été collectées par traitement au nocodazole. Nous avons ajouté les inhibiteurs après un temps relativement court (8 h) après la transfection pour éviter la récupération de cellules ayant traversé une phase G1 en absence de protéines impliquées dans la construction des complexes de pré-RC. En effet, pour que l'expérience soit concluante, nous ne devons pas analyser des cellules qui ont subi un défaut de chargement des origines en G1 mais nous focaliser sur les cellules ont été carencées plus tardivement au cours du cycle, en phase S et G2, où ces facteurs ne sont normalement plus indispensables. Nos résultats montrent que le RO-3306 est incapable de réduire les cassures induites par l'aphidicoline dans les cellules carencées en protéines CDC6

ou CDT1. Ce n'est pas le cas des cellules carencées en ORC1 pour lesquelles le taux de cassures sous RO-3306 est comparable à celui des cellules non traitées. Ces résultats indiquent que l'accumulation de CDT1 et de CDC6 induite par le RO-3306 est indispensable à l'abolition des cassures induites par l'aphidicoline. En revanche l'accumulation d'ORC1 ne semble pas être indispensable à l'action du RO-3306 sur les cassures (Figure 36).

## **2. Effet du RO-3306 sur la dynamique de réplication au niveau de *FRA3B* dans des cellules appauvries en protéines du complexe de pré-réplication**

Nous sommes en train d'étudier, par peignage moléculaire, la dynamique de réplication au niveau de *FRA3B* dans des cellules carencées en CDT1 ou en ORC1 comme contrôle, et traitées à l'aphidicoline et au RO-3306. Ces expériences sont délicates à réaliser car il est difficile d'obtenir des fibres d'ADN en bonne qualité et en nombre suffisant à partir de cellules ayant subi une transfection. Si *FRA3B* retrouve son profil de réplication initial, à savoir un cœur pauvre en initiations ainsi qu'en terminaisons en absence de CDT1, cela confirmerait que le déclenchement d'origines dans le cœur est nécessaire à l'action du RO-3306 sur les cassures et cela supporterait un rôle actif de l'inhibition de CDK1 dans le chargement (ou la complétion de chargement) de complexes de pré-RC sur des origines latentes du cœur. Dans le cas d'ORC1, les résultats obtenus sur les cassures totales suggèrent que son absence n'aura pas d'incidence sur la distribution des initiations et des terminaisons dans le cœur de *FRA3B*.



**Figure 36: Effet de la carence en facteurs impliqués dans la formation des complexes de pré-RC sur la prévention de cassures chromosomiques par le RO-3306.** Des cellules JEFF ont été transfectées avec des siARN sans ARNm cible connu (Ns) ou dirigés contre les ARNm de ORC1 (SiORC1), CDC6 (SiCDC6), CDT1 (SiCDT1). Huit heures après transfection, des extraits protéiques totaux ont été préparés. En parallèle, les cellules ont été incubées 16 h sans traitement (Nt), en présence d'aphidicoline (Aph 0,6  $\mu$ M) ou en présence d'aphidicoline et de RO-3306 (10  $\mu$ M) (ARO). Des métaphases ont été alors préparées par traitement au nocodazole et les cassures totales ont été dénombrées après coloration au Giemsa. Les barres d'erreur représentent les écarts types (Ns et SiCDC6 : 5 expériences, SiORC1 et SiCDT1 : 3 expériences) Les extraits protéiques ont été analysés par Western pour contrôler l'efficacité des siARN utilisés. Les résultats sont présentés comme dans la Figure 35.

# Discussion et Perspectives

---

Pendant ma thèse je me suis intéressée à tester et, éventuellement, à approfondir et généraliser le modèle proposé par notre laboratoire pour expliquer la fragilité des sites fragiles communs. En effet, suite à l'étude de la dynamique de réplication au niveau de quatre SFC dans deux types cellulaires, il a été proposé par notre laboratoire que ces sites se caractérisent par un timing de réplication très tardif associé à une pauvreté en événements d'initiation de réplication.

Je me suis donc consacrée d'une part à l'extension de ce modèle à un grand nombre de SFC (possiblement tous) par une nouvelle démarche impliquant la technique de Repli-Seq. J'ai d'autre part exploré la relation entre fragilité des SFC et transcription des grands gènes qu'ils renferment. Enfin, j'ai étudié l'effet d'une entrée retardée en mitose sur la fragilité des SFC et sur leur profil de réplication afin de confirmer la contribution de ces paramètres à la fragilité des SFC.

Dans cette dernière partie du manuscrit je vais souligner les principaux résultats obtenus pendant la thèse, discuter leur interprétation et leur contribution à notre compréhension actuelle du mécanisme de la fragilité des SFC et aborder les questions et les perspectives qui peuvent en découler.

## **I. Cartographie des SFC par la technique de Repli-Seq**

Les SFC sont actuellement cartographiés par des méthodes de cytogénétique classique et moléculaire. Cette cartographie consiste à identifier sur chromosomes métaphasiques les bandes cytogénétiques qui présentent des cassures récurrentes sous stress réplicatif, puis à affiner la localisation des SFC ainsi repérés grâce à la technique de FISH en utilisant comme sondes des BAC distribués le long de la région d'intérêt. Cette analyse prend beaucoup de temps et nécessite une quantité considérable de travail puisqu'au moins une centaine de métaphases doit être analysée à chaque fois. De plus, nous avons remarqué au laboratoire que des divergences peuvent surgir entre la cartographie des sites au niveau des bandes cytogénétiques suivant que le banding révèle les bandes R ou G (El Achkar et al., 2005). La cartographie des sites par FISH diffère également de celle obtenue

par banding. Selon les cellules étudiées, les cassures à certains SFC sont en effet plus faciles à repérer sur les chromosomes traités pour la cytogénétique classique que celles survenant au niveau d'autres SFC, alors que par FISH les cassures à un locus donné sont beaucoup mieux recensées.

Une partie de ma thèse a donc porté sur la validation d'une nouvelle démarche pour cartographier les sites fragiles communs au niveau pan-génomique avec l'hypothèse qu'ils correspondent à des régions qui n'achèvent pas leur réplication sous stress réplicatif. Nous avons donc comparé les profils de timing de réplication obtenus par Repli-Seq dans des lymphocytes non traités, où les SFC présentent un profil de timing en « V », à ceux de lymphocytes traités à l'aphidicoline où on s'attendait à mettre en évidence un profil en « V » incomplet qui reflèterait l'absence d'achèvement de la réplication. Les résultats obtenus dans les cellules lymphoblastoïdes JEFF montrent que le profil en « V » observé au niveau de SFC déjà connus chez les lymphocytes, se transforme en fait en un profil en « U » sous aphidicoline et non pas en un profil en « V » incomplet. Néanmoins, par des opérations de traitement de données nous avons pu mettre en évidence un déficit de réplication au niveau de ces sites dans les cellules traitées à l'aphidicoline par rapport aux cellules non traitées. Le point fort de ces résultats réside dans le fait que nous avons pu confirmer d'une façon plus globale ce qui avait été observé au niveau de *FRA3B* par peignage moléculaire (Letessier et al., 2011): les SFC sont des régions sous-répliquées en conditions de stress réplicatif.

Nous avons montré au laboratoire que la carte des sites fragiles varie d'un type cellulaire à l'autre mais aussi qu'elle varie, dans de moindres proportions, d'une lignée cellulaire à une autre au sein du même type cellulaire (reflétant probablement des différences interindividuelles) (Le Tallec et al., 2013). L'interprétation fine des résultats obtenus par Repli-Seq dans les cellules JEFF nécessitera donc de connaître la carte des SFC dans ces cellules. Cette carte n'est actuellement pas établie et nous nous référons aux SFC déjà décrits dans les lymphocytes humains. Nous sommes donc en train de réaliser une cartographie des SFC dans les cellules JEFF avec l'aide du cytogénéticien Bernard Dutrillaux (Muséum d'Histoire Naturelle de Paris). La technique de Repli-Seq sera ensuite appliquée aux fibroblastes MRC5 dont nous connaissons la carte des SFC, afin de vérifier que les SFC peuvent être repérés et se comportent de la même façon dans ce type cellulaire que dans les lymphocytes. Au cours de ces analyses des cellules JEFF et MRC5, nous serons attentifs à l'existence éventuelle de régions présentant un déficit de réplication mais ne correspondant

pas à un SFC déjà caractérisé. Nous vérifierons alors par la technique de FISH que cette région candidate se comporte effectivement comme un SFC. Lorsque l'approche que nous avons développée sera complètement validée par analyse des cellules JEFF et MRC5, elle pourra être appliquée pour établir la carte des SFC dans un grand nombre de types cellulaires actuellement inexplorés, en particulier les cellules souches embryonnaires.

Ainsi que nous l'avons déjà mentionné, le profil en « U » observé sous aphidicoline dans les cellules JEFF au niveau des SFC connus reflète le déclenchement d'origines dans la région cœur après des temps longs de traitement. Ce profil est totalement en accord avec l'augmentation des événements d'initiations observé par peignage moléculaire dans la région cœur de *FHIT* chez des cellules traitées à l'aphidicoline pendant 16 h. Un tel phénomène n'était pratiquement pas observable dans les cellules qui avaient été traitées pendant 5 h (Letessier et al., 2011). Nous sommes en train de réaliser une analyse par Repli-Seq de cellules JEFF traitées à l'aphidicoline pendant 5 h afin de vérifier que, dans ce cas, nous obtenons un profil en « V » incomplet. Les initiations que nous observons au niveau de *FHIT* par les deux approches (peignage et Repli-Seq) sont probablement dues au déclenchement des origines latentes en réponse à la réduction de vitesse induite par l'aphidicoline. La présence de ces initiations ne permet pas pour autant de supprimer la fragilité de *FRA3B* et des SFC en général. Ceci peut être expliqué par la faible efficacité et le nombre insuffisant et/ou le déclenchement très tardif de ces origines.

Outre la cartographie des SFC à grande échelle, la technique de Repli-Seq appliquée à des cellules sous stress répliatif pourra servir à étudier d'autres questions. Par exemple, comme nous l'avons déjà mentionné dans l'introduction, il a été montré par étude des foyers de réplication que, dans des conditions de stress répliatif, la réplication est régulée différemment au niveau local (il y a plus de fourches de réplication par foyer donc déclenchement d'origines supplémentaires) et au niveau global (inhibition des domaines tardifs donc répression des origines correspondantes) (Ge and Blow, 2010). L'étude du timing de réplication en conditions de stress répliatif devrait permettre de visualiser la régulation du déclenchement des origines latentes au niveau de la séquence nucléotidique plutôt qu'au niveau des foci de réplication. De plus, il serait intéressant de comparer les timing de réplication sous stress répliatif en présence ou en absence de certains facteurs de la voie de réponse aux dommages de l'ADN tel que CHK1 ou ATR. Ceci permettrait de

préciser le rôle de cette voie dans le contrôle temporel du déclenchement des origines de réplication.

## II. Lien entre fragilité et transcription

Nous avons montré qu'un blocage prolongé des cellules à la frontière G2/M par le RO-3306, un inhibiteur de l'activité kinase de CDK1, permet de supprimer la fragilité des SFC dans les fibroblastes et les lymphocytes. Pour les sites analysés, *FRA3B*, *FRA16D*, *FRA1L* et *FRA3L*, cette abolition de la fragilité par le RO-3306 n'est pas accompagnée d'une diminution (*FHIT*, *WWOX* et *NEGR1*) ou est accompagnée d'une faible diminution (*LSAMP*) de la transcription des grands gènes qui leur sont associés. Nous avons aussi montré qu'inhiber la transcription des grands gènes par l' $\alpha$ -amanitine ou le triptolide, dans des conditions où ces gènes sont vidés des complexes de transcription, n'aboutit pas à une diminution de la fragilité induite par l'aphidicoline au niveau des SFC correspondant. Cependant, en apparente contradiction avec cet ensemble de résultats, nous avons mis en évidence une corrélation entre expression de quatre grands gènes (*FHIT*, *WWOX*, *LSAMP*, *NEGR1*) et fragilité des quatre SFC correspondants (*FRA3B*, *FRA16D*, *FRA3L*, *FRA1L*) dans sept lignées de fibroblastes et sept lignées de lymphocytes. De plus, nous avons mis en évidence des expressions atypiques des gènes *WWOX* et *LSMAP* dans certaines de ces lignées et nous avons montré que ces expressions s'accompagnent presque systématiquement d'une fragilité atypique du site concerné. Le seul contre-exemple est le haut niveau d'expression d'une petite unité (< 10 kb) de transcription incluse dans le gène *FHIT* chez les cellules MRC5SV2 qui n'est pas corrélé à une fragilité du site *FRA3B* dans ces cellules. Une explication plausible est que la petite taille de cette unité ne lui permet pas d'influencer la fragilité. Rappelons que seuls des gènes de plus de 300 kb sont significativement associés à la présence d'un SFC (Le Tallec et al., 2013).

Comme je l'ai déjà mentionné, l'étude de la relation entre transcription et réplication a été poursuivie au laboratoire en modulant spécifiquement la transcription du gène *FHIT* et en évaluant l'impact de ces manipulations sur la fragilité du site correspondant. Des activateurs transcriptionnels dérivés des TALE et ciblant spécifiquement le promoteur de ce gène ont été développés puis introduits dans les fibroblastes MRC5SV2 chez lesquels le gène *FHIT* n'est pas exprimé et *FRA3B* n'est pas fragile. Les résultats obtenus montrent que l'activation de la transcription de *FHIT* dans les transfectants stables ainsi obtenus



s'accompagne d'une augmentation très importante de la fragilité de *FRA3B* (O. Brison et A-M. Lachagès, résultats non publiés). D'autres outils (TALEN, CRISPR-Cas) ciblant le promoteur du gène *FHIT*) sont en train d'être développés au laboratoire pour permettre d'inhiber la transcription de *FHIT* dans les cellules JEFF et de déterminer l'effet de cette inhibition sur la fragilité de *FRA3B*.

Nos résultats, pris dans leur ensemble, montrent que la relation entre fragilité et transcription est beaucoup plus complexe qu'initialement suggéré par Tora et collaborateurs (Helmrich et al., 2011) Des études supplémentaires et plus approfondies seront nécessaires pour élucider la nature de cette relation. Pour le moment nous proposons l'hypothèse de travail suivante pour rendre compte de l'ensemble des résultats obtenus: la dynamique de réplication et la fragilité des SFC dépendraient de la structuration de grands domaines chromatinien qui serait elle même en relation avec l'état transcriptionnel du/des grand(s) gène(s) correspondant(s). Cette organisation chromatinienne pourrait par exemple être sous le contrôle d'éléments régulateurs de la transcription des grands gènes, éventuellement situés à distance de l'unité de transcription qu'ils contrôlent. En particulier, la capacité de facteurs se fixant sur les insulateurs, les séquences activatrices (enhancers) ou les terminateurs à interagir entre eux a depuis longtemps suggéré que ces interactions conduisent à la formation de boucles de chromatine définissant des domaines où la régulation de la transcription est coordonnée. Nous pensons que cette organisation en boucles pourrait aussi dicter localement la capacité de déclenchement des origines de réplication. A l'appui de cette hypothèse, il faut rappeler que des changements de programme temporel de réplication pendant le développement et la différenciation cellulaire sont fréquemment corrélés à des changements transcriptionnels. Dans cette hypothèse, la fragilité serait liée à une organisation chromatinienne correspondant à un état de « compétence transcriptionnelle » et non à la transcription effective du grand gène. La corrélation apparente entre transcription et profil de réplication serait due au fait que les deux phénomènes sont sous le contrôle des mêmes facteurs épigénétiques, et n'impliquerait donc pas un lien direct entre eux. Remarquons que les situations où nous n'avons pas trouvé de corrélation entre fragilité et transcription (RO-3306,  $\alpha$ -amanitine, triptolide) correspondent à des expériences où nous avons induit un changement rapide et transitoire d'un des deux paramètres. Au contraire, dans le cas des lignées de lymphocytes et de fibroblastes présentant une expression atypique des grands gènes ainsi que dans le cas des

transfectants obtenus avec les constructions dérivées des TALE, il s'agit de situations établies depuis de nombreuses générations cellulaires dans lesquelles un éventuel changement de la structure chromatinienne a eu le temps de se mettre en place et d'être pérennisé.

Pour tenter de valider cette hypothèse de travail, il a été entrepris au laboratoire d'étudier l'organisation tridimensionnelle du domaine contenant le gène *FHIT* dans les cellules JEFF et MRC5 par la technique 4C et par celle de FISH en 3D sur noyaux interphasiques. Enfin notons que l'analyse de la dynamique de réplication et du timing par Repli-Seq que nous avons développée avec l'équipe de Claude Thermes permettra d'analyser ces paramètres au niveau pan-génomique dans différentes conditions de traitement des cellules ( $\alpha$ -amanitine, le triptolide ou le DRB) et dans différentes lignées cellulaires ainsi que dans les transfectants TALE.

### **III. Compléter le chargement d'origines de réplication pour finir la réplication**

Dans les cellules traitées au RO-3306, nous avons corrélié la disparition des cassures induites par l'aphidicoline au niveau de *FRA3B* avec l'augmentation des événements d'initiation et de terminaison dans la région cœur de *FHIT*. Ces résultats suggèrent que l'augmentation des événements d'initiation participe à l'achèvement de la réplication au niveau des SFC avant l'entrée des cellules en mitose, en bon accord avec l'hypothèse que la fragilité résulte d'une pauvreté en événements d'initiation dans la région cœur du site.

Afin d'étendre cette dernière observation aux autres SFC, dont la fragilité est aussi supprimée par le RO-3306, nous sommes en train d'étudier par Repli-Seq le timing de réplication dans des cellules lymphoblastoïdes JEFF traitées à l'aphidicoline et au RO-3306. Ceci sera aussi fait chez des fibroblastes MRC5. Malgré la faible résolution de cette technique pour la cartographie des événements d'initiation, nous espérons pouvoir ainsi mettre en évidence l'achèvement de la réplication au niveau d'autres SFC dans des cellules traitées à l'aphidicoline en présence du RO-3306 et ainsi généraliser le mécanisme proposé pour *FRA3B*.

Nous avons de plus montré que l'inhibition de CDK1 par le RO-3306 s'accompagne dans les cellules en phase G2 d'une accumulation à la chromatine de protéines participant à la construction des complexes de pré-RC (ORC1, CDT1, CDC6 et MCM7). Nous avons enfin montré que la suppression par le RO-3306 des cassures induites par l'aphidicoline dépend de l'accumulation de CDC6 et CDT1, mais pas de celle de ORC1. Ces résultats constituent des

arguments forts en faveur d'un rôle actif des initiations détectées au niveau de *FHIT* dans la suppression de la fragilité de *FRA3B* sous RO-3306. Pour définitivement établir ce point, nous sommes en train d'étudier le profil de réplication de *FHIT* dans des cellules traitées à l'aphidicoline et au RO-3306 et carencées en CDT1. Il faut être prudent dans l'interprétation des résultats concernant l'apparente dispensabilité de l'accumulation de ORC1 à l'action du RO-3306. Cette absence d'effet peut être expliquée soit par le fait qu'ORC1 est déjà chargé sur les origines activables par le traitement au RO-3306 soit parce que la carence induite par le SiORC1 n'est que partielle, la fraction d'ORC1 restante étant suffisante pour assurer le déclenchement des origines.

L'ensemble de ces résultats suggère que l'inhibition de CDK1 par le RO-3306 permet de charger ou de compléter le chargement d'origines de réplication dans le cœur des SFC. D'après le modèle actuellement admis de la régulation de la réplication, le recrutement de complexes de pré-RC sur les origines ne peut s'effectuer que pendant la phase G1. Cependant, comme je l'ai déjà mentionné, il existe un certain nombre d'arguments laissant penser que ce processus pourrait survenir en dehors de la phase G1 si CDK1 est inhibée.

Néanmoins, si ce chargement concernait l'ensemble des origines du génome plutôt qu'un sous groupe d'entre elles localisées au cœur des SFC, il conduirait vraisemblablement à une re-réplication extensive. Nous avons des observations faites au laboratoire qui indiquent que ce n'est pas le cas. L'analyse par FACS de cellules maintenues 16 h ou plus sous RO-3306 ne montre pas l'apparition d'une population cellulaire présentant un contenu en ADN supérieur à 4C. De plus, des cellules accumulées en phase G2 par traitement au RO-3306 puis remises en milieu normal passent la mitose et rentrent normalement en phase G1 avec un contenu en ADN de 2C. Enfin, aucun foyer  $\gamma$ H2AX traduisant la présence de cassures de l'ADN n'est détecté dans les cellules maintenues 16 h sous RO-3306 ou dans ces mêmes cellules après remise en cycle en milieu normal (O. Brison, résultats non publiés).

Si, malgré l'inhibition de CDK1, toutes les origines non chargées en complexe de pré-RC au cours de la phase G1 ne peuvent l'être en phase S ou G2, on peut envisager une hypothèse alternative. Le chargement des origines en complexe de pré-RC se faisant en plusieurs étapes (voir Introduction) il est envisageable qu'il existe en fin de phase G1 des origines de réplication possédant des complexes de pré-RC incomplets, peut être parce que ces origines sont difficilement accessibles aux facteurs de réplication. Dans les cellules traitées au RO-3306, l'inhibition de CDK1 et l'accumulation de facteurs impliqués dans la

construction des complexes de pré-RC (CDT1, CDC6, MCM) permettraient de terminer le chargement de ces origines en phase S ou G2 et leur déclenchement dans les régions non encore répliquées comme le cœur des SFC. Dans les régions déjà répliquées, le passage des fourches de réplication pourrait avoir inactivé ces origines comme c'est le cas pour les origines latentes complètement chargées. Cette hypothèse permet d'expliquer pourquoi l'action du RO-3306 nécessite l'accumulation de CDT1 et/ou CDC6 (et apparemment pas celle de ORC1) sans induire de re-réplication notable. Remarquons que si des origines partiellement chargées existent, il n'y a pas de raison pour qu'elles soient spécifiques des SFC et pourraient donc être présentes dans d'autres régions du génome, difficilement accessibles aux facteurs du pré-RC.

Outre l'accumulation de facteurs tels que CDC6, CDT1 et MCM, le chargement complet de ces origines en phase G2 pourrait être facilité par d'autres facteurs. Si des marques épigénétiques jouent un rôle dans le positionnement des origines de réplication (voir Introduction), il est possible qu'un changement de ces marques au cours du cycle cellulaire facilite le recrutement d'un pré-RC complet sur ces origines. Il est aussi possible qu'un changement de l'organisation des boucles de chromatine facilite la charge complète des origines partiellement chargées. Comme je l'ai déjà mentionné, il a été montré au laboratoire que la taille des boucles de chromatine change au cours de la phase S et qu'il y a un repositionnement de certaines origines depuis la matrice nucléaire vers les boucles (Courbet et al., 2008).

Lorsque la réponse aux dommages de l'ADN est activée dans des cellules en prolifération, elle aboutit, entre autres, à un blocage du passage en mitose (point de contrôle G2/M). Ce blocage est consécutif à une inhibition du complexe CDK1/cycline B par phosphorylation de CDK1. Nous nous sommes donc demandés si l'inhibition de CDK1 par le RO-3306 et le blocage en G2 ne mimait pas le déclenchement du point de contrôle G2/M. Dans cette hypothèse, la charge complète en fin de phase S ou en G2 d'origines partiellement chargées, pourrait être vue, non pas comme une conséquence bizarre du traitement par le RO-3306, mais comme faisant partie intégrante de la réponse de la cellule à un stress répliatif. Ce phénomène permettrait de recruter cette nouvelle classe d'origines silencieuses qui, additionnées aux origines latentes classiques (entièrement chargées), fournirait un réservoir d'origines supplémentaires pour terminer la réplication au niveau de régions non encore répliquées avant l'entrée en mitose.

# Annexes

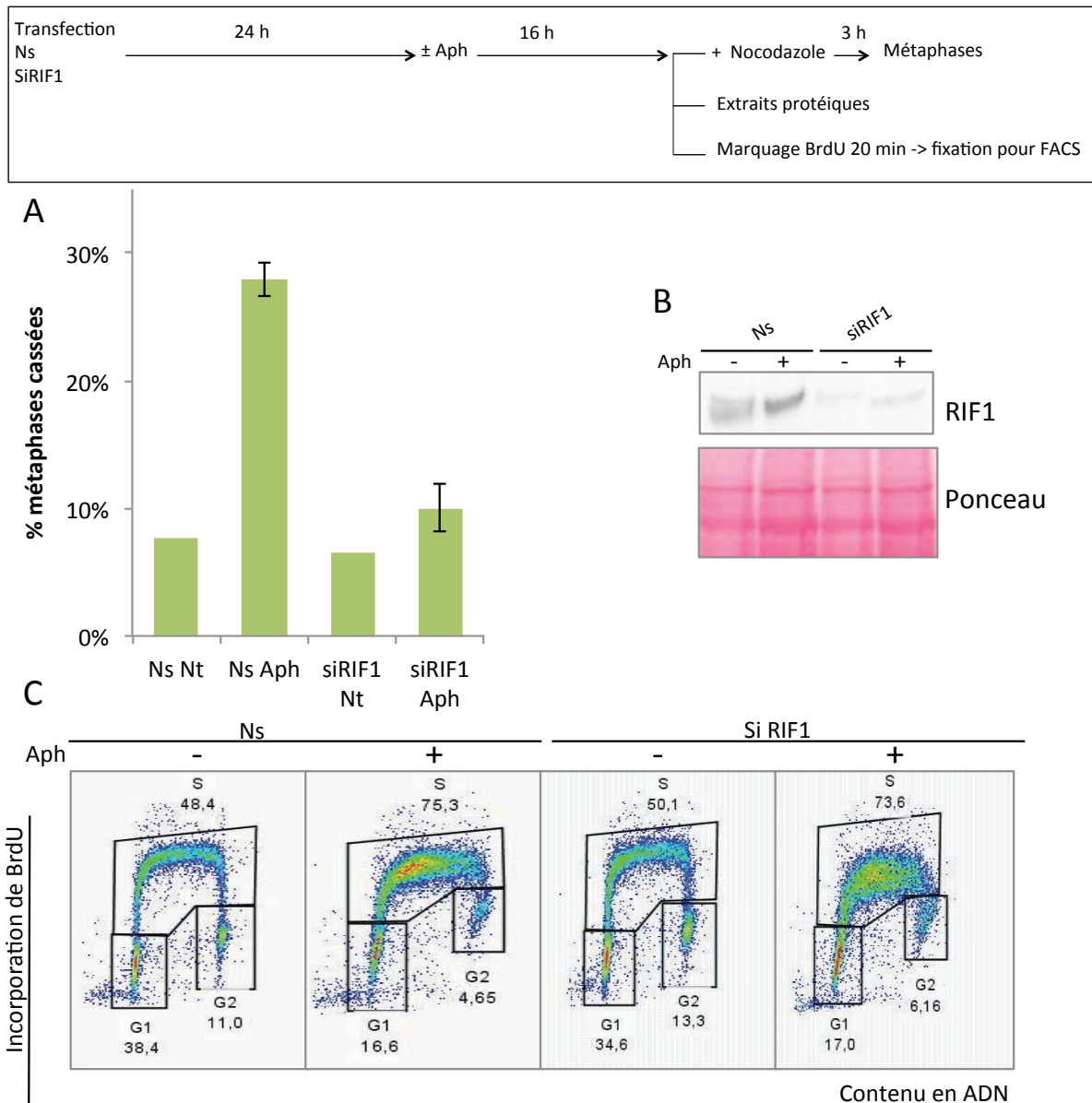
---

## I. Effet de l'inactivation de RIF1 sur les cassures chromosomiques induites par l'aphidicoline

En utilisant le RO-3306 pour bloquer les cellules en phase G2, nous avons permis la complétion de la réplication sous aphidicoline, ce qui se traduit par une disparition de la fragilité des SFC. Une autre façon de permettre la complétion de la réplication avant l'entrée des cellules en mitose, sans allonger le temps de la phase G2, serait d'avancer le timing de la réplication des SFC qui est très tardive. Afin de tester cette hypothèse, nous nous sommes basés sur des études récentes impliquant la protéine RIF1 dans la régulation du timing de réplication chez les cellules de mammifères.

RIF1 a été dans un premier temps impliquée dans la régulation de la réplication des régions subtélomériques chez la levure. Ensuite, il a été montré que, chez les levures carencées en RIF1, les origines tardives se déclenchent plus précocement et qu'inversement les origines précoces deviennent plus tardives (Hayano et al., 2012). Ces dérèglements engendrent un profil de timing de réplication plus homogène correspondant à un profil de milieu de phase S. D'après ces travaux, l'inactivation de RIF1 n'affecte pas la formation des complexes de pré-RC mais l'accessibilité des facteurs d'initiation tel que Cdc45 aux origines de réplication. Chez la souris ainsi que dans des cellules tumorales humaines, l'inactivation de RIF1 conduit au même phénotype que chez la levure, à savoir une altération du timing de réplication qui devient comparable à celui de cellules en milieu de phase S (Cornacchia et al., 2012; Yamazaki et al., 2012).

Le modèle actuel du mode d'action de RIF1 est le suivant : ce facteur participerait au maintien de l'organisation des boucles de chromatine et à la formation des nucléosomes. Il maintiendrait les régions de réplication tardive dans des zones inaccessibles aux facteurs d'initiation en attendant que la réplication des régions précoces soit accomplie. En absence de RIF1, les régions séquestrées seraient libérées et certains facteurs de réplication ayant accès à un plus grand nombre de régions chromosomiques deviendraient limitant. Ceci expliquerait que les origines précoces se déclenchent plus tardivement et que les origines tardives s'activent plus précocement.



**Figure 37 : Effet de la carence en facteur RIF1 sur l'instabilité des SFC.** Des cellules JEFF ont été transfectées avec des siARN sans cible connue (Ns) ou dirigés contre l'ARNm de RIF1. Vingt quatre heures après transfection, les cellules ont été incubées 16 h sans traitement (Nt) ou en présence d'aphidicoline (Aph; 0,6  $\mu$ M). Des métaphases ont été préparées après traitement au nocodazole et les cassures totales ont été dénombrées après coloration au Giemsa. Les barres d'erreur représentent l'écart-type lorsque l'expérience a été faite deux fois (A). En parallèle, des extraits protéiques ont été préparés et analysés par Western pour contrôler l'efficacité des siARN utilisés. Une coloration au rouge ponceau de la membrane a été utilisée comme contrôle de charge (B). En parallèle, des cellules ont été incubées 20 min en présence de BrdU puis fixées et analysées par FACS. Les pourcentages de cellules dans chaque phase du cycle sont indiqués (C).

Afin de tester l'effet de la carence de RIF1 sur la fréquence des cassures induites par l'aphidicoline, nous avons transfecté des cellules lymphoblastoïdes JEFF avec un siRNA dirigé contre l'ARNm de RIF1 ou un siRNA sans cible connue (contrôle). Vingt-quatre heures après la transfection, les cellules ont été traitées à l'aphidicoline pendant 16 h puis mises en présence de nocodazole pour collecter les cellules en métaphase et les cassures

chromosomiques ont été dénombrées. Les résultats obtenus montrent que, dans les cellules carencées en RIF1, le taux de cassures totales induites par l'aphidicoline a fortement baissé et est à un niveau comparable à celui de cellules non traitées (Figure 37). Ce résultat suggère que l'inactivation de RIF1 a permis la complétion de la réplication des SFC avant l'entrée en mitose, peut être en favorisant un déclenchement plus précoce des origines de réplication dont dépend la réplication des SFC et/ou en permettant le déclenchement des origines normalement latentes au sein des SFC. Cependant, comme nous n'avons pas étudié le timing de réplication dans les cellules carencées en RIF1, nous ne pouvons pas conclure quant au mécanisme à l'origine de l'absence de cassures dans ces cellules.

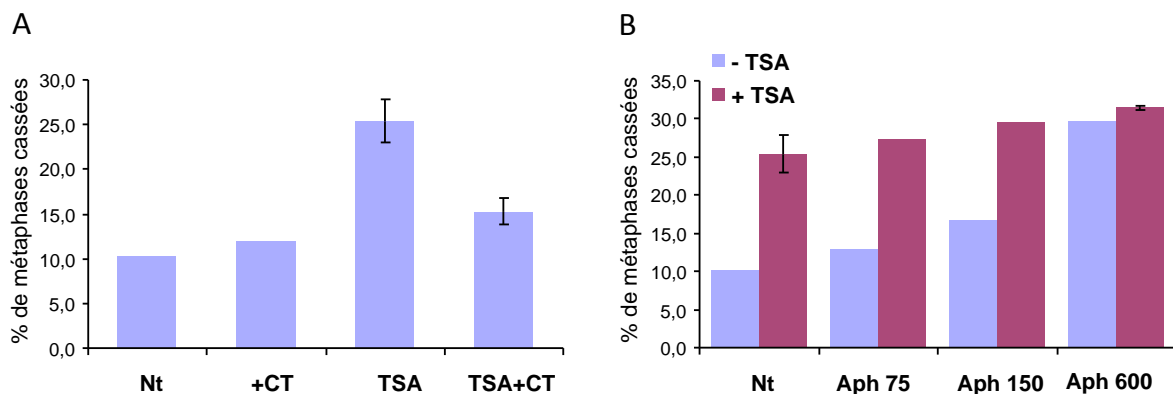
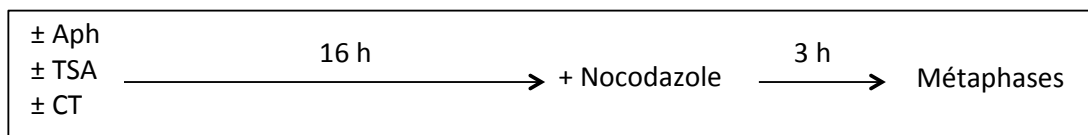
## **II. Effet de la trichostatine A sur les cassures chromosomiques induites par l'aphidicoline**

La trichostatine A est un inhibiteur des histones déacétylases. Les cellules traitées par cette drogue présentent une hyperacétylation généralisée des histones car l'activité des histones acétyltransférases n'est plus contre balancée par celle des déacétylases. Dans une étude publiée en 2009, Le Beau et collaborateurs (Jiang et al., 2009) affirment que le traitement de lymphocytes humains par la TSA conduit à une réduction importante de la fragilité des SFC. Ces auteurs suggèrent que l'hyperacétylation de la chromatine résultant du traitement par la TSA faciliterait la réplication des SFC. Ces résultats, qui décrivaient un phénomène analogue à celui que nous obtenions avec le RO-3306, nous ont interpellé pour plusieurs raisons. En effet, des travaux effectués précédemment dans notre laboratoire ont montré que, dans des cellules de hamster chinois, l'efficacité des origines de réplication au niveau du locus *AMPD2* est sans rapport avec le niveau local d'acétylation des histones (Gay et al., 2010). Le traitement de ces cellules par la TSA induit un changement d'efficacité relative des origines du locus mais nous avons montré que c'est une conséquence de la réduction de la vitesse de propagation des fourches de réplication. En effet, le traitement par la TSA réduit le niveau d'expression de deux gènes codants pour des enzymes impliquées dans la biosynthèse des pyrimidines : la CTP synthétase et la thymidylate synthase. La réduction de vitesse induite par la TSA est due à une carence des cellules en précurseurs pyrimidiques. Elle peut être supprimée et l'effet sur les origines aboli en ajoutant de la cytidine et de la thymidine dans le milieu de culture des cellules. La réduction de la fragilité des SFC par la TSA décrite dans l'article de Le Beau et collaborateurs et l'explication proposée étaient donc surprenantes.

Afin de déterminer l'effet de la TSA sur la fragilité des SFC dans notre système cellulaire, nous avons mesuré le taux de cassures totales dans des cellules lymphoblastoïdes JEFF traitées à la TSA, à l'aphidicoline ou avec les deux. Nos résultats montrent que, conformément à son effet sur la vitesse des fourches documenté précédemment au laboratoire, la TSA induit par elle-même des cassures chromosomiques. De plus, il n'y a aucune réduction du taux de cassures dans les cellules traitées à l'aphidicoline en présence de TSA. A faibles doses d'aphidicoline (75, 150 et 300 nM), les effets de la TSA et de l'aphidicoline sont additifs. A 600 nM d'aphidicoline, où la vitesse de réplication est réduite de 8 fois, les deux inhibiteurs n'ont plus cet effet additif sur le taux de cassures,



probablement parce que la réduction de vitesse due à l'aphidicoline est très supérieure à celle due à la TSA à la concentration utilisée (trois fois) (Gay et al., 2010). Enfin, afin de vérifier si les cassures induites par la TSA sont bien dues à son effet sur la vitesse de réplication, nous avons compté les cassures chromosomiques dans des cellules traitées à la TSA en ajoutant ou pas dans le milieu de culture de la cytidine et de la thymidine. Nos résultats montrent que la fréquence des cassures induites par la TSA est réduite de 2/3 par l'ajout des précurseurs pyrimidiques, confirmant que les cassures qui apparaissent sont en grande partie dues à la réduction de la vitesse de réplication (Figure 38).



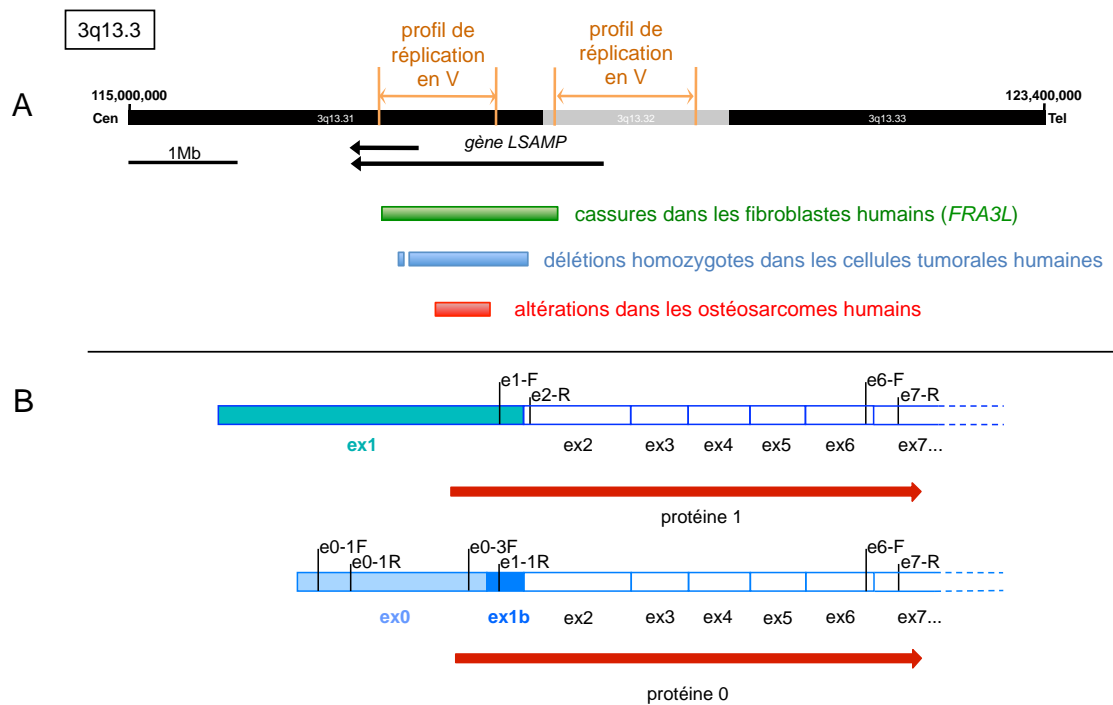
**Figure 38: Effet de la trichostatine A sur les cassures chromosomiques.** **A)** Des cellules JEFF ont été incubées 16 h en absence (Nt) ou en présence de cytidine et thymidine (CT; 10  $\mu$ M chaque) ou de trichostatine A (TSA; 20 ng/ml) ou de trichostatine A, de cytidine et de thymidine (TSA + CT) puis des métaphases ont été préparées après traitement au nocodazole et les cassures totales ont été dénombrées après coloration au Giemsa. Les barres d'erreur représentent l'écart-type lorsque l'expérience a été faite deux fois. **B)** Des cellules JEFF ont été traitées ou non comme décrit en **A**, excepté que la concentration d'aphidicoline dans le milieu était de 75 nM, 150 nM ou 600 nM comme indiqué. Les cassures ont été analysées et les résultats sont présentés comme en **A**.

### III. Les deux formes du gène *LSAMP* humain

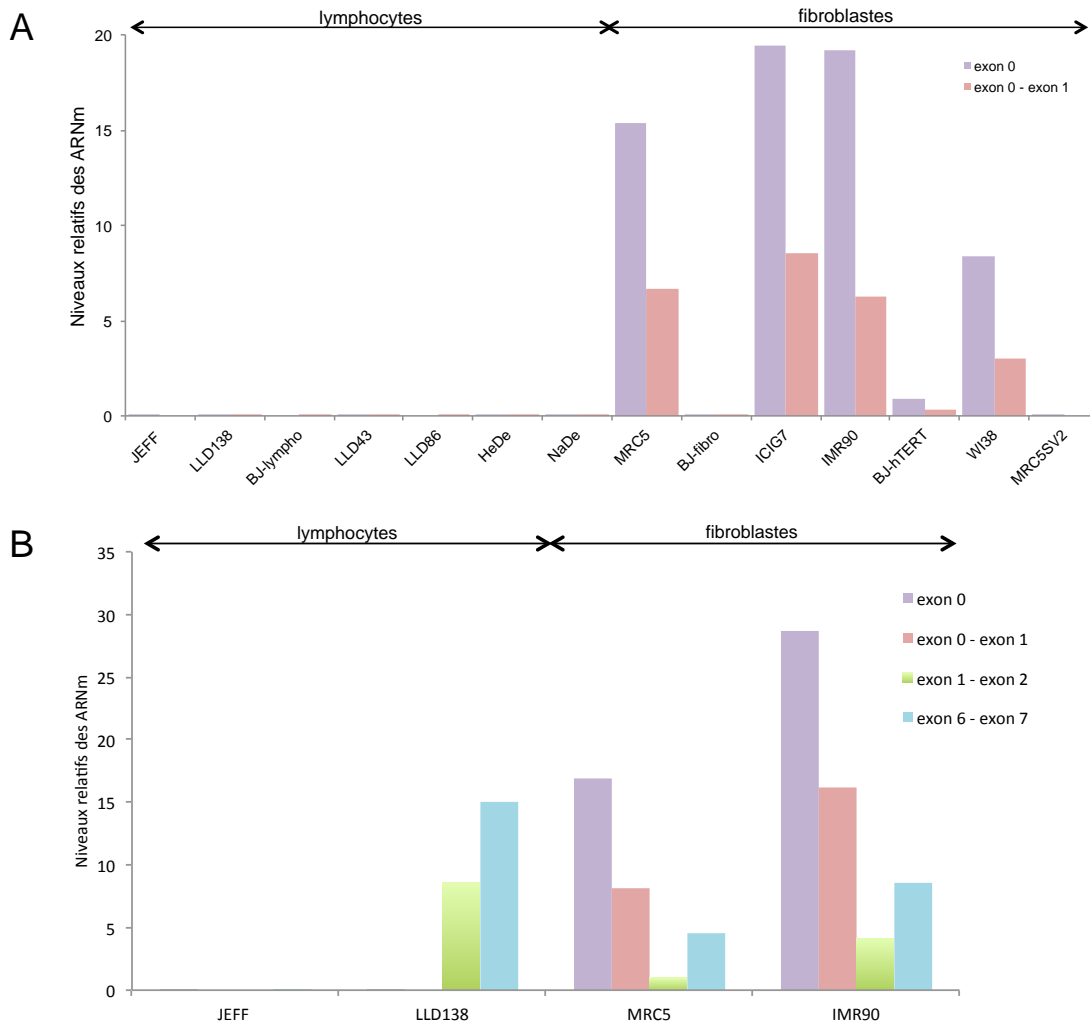
Le grand gène *LSAMP* est inclus dans le site fragile *FRA3L* qui est le site majeur des fibroblastes MRC5. C'est un gène paralogue du gène *NEGR1* qui réside dans le second site fragile de ces cellules : *FRA1L*. Cependant, il a été montré au laboratoire (Le Tallec et al., 2011) que plus de 80% des cassures chromosomiques au site *FRA3L* se produisent en amont du gène *LSAMP* tel qu'il est annoté dans les banques de données (7 exons s'étendant sur 643 176 pb). Il en est de même de la grande majorité des délétions observées dans les cellules tumorales humaines (Bignell et al., 2010) et des altérations chromosomiques décrites dans les ostéosarcomes (Pasic et al., 2010) (Figure 39). De plus, deux régions présentant un profil de réplication en V sont localisées en amont du gène et l'une d'elles correspond à la région fragile dans les fibroblastes et altérée dans les cellules tumorales. Benoît Le Tallec dans notre laboratoire a remarqué qu'une forme 3,5 fois plus longue du gène *LSAMP* (8 exons, 2232 654 pb) avait été mentionnée dans la version 19 de la banque de données GENCODE puis retirée de la version 20. Cette forme beaucoup plus étendue en 5' engloberait alors la région fragile qui est altérée dans les cancers et qui présente un profil en V. Elle comprendrait un exon supplémentaire (exon 0) épissé sur une forme variante de l'exon 1 (1b). Nous nous sommes donc demandés si cette forme du gène *LSAMP* existait réellement et si on pouvait détecter son expression dans des fibroblastes humains avec, en particulier, l'épissage prédit de l'exon 0 sur l'exon 1b. Nous avons donc conçu des couples d'amorces pour répondre à ces questions par qRT-PCR (Figure 39).

Les résultats montrent que des transcrits contenant l'exon 0 (amorces e0-1F/e0-1R) apparemment épissé sur l'exon 1 (amorces e0-3F/e1-1R) sont facilement détectables dans les lignées de fibroblastes MRC5, ICIG7, IMR90 et WI38 mais très peu ou pas dans les lignées BJ-fibro, BJ-hTERT et MRC5SV2 (Figure 40). Ces résultats sont totalement en accord avec ceux qui avaient été obtenus précédemment avec un couple d'amorces situées dans les exons 6 et 7 (e6-F/e7-R). Il apparaît donc que les lignées de fibroblastes qui expriment le gène *LSAMP* le font très vraisemblablement à partir d'un promoteur situé en amont de l'exon 0. Dans ces cellules, c'est donc bien la forme de 2,2 Mb du gène qui s'exprime. Dans les lignées lymphoblastoïdes, aucun transcrit n'est détecté avec les amorces e0-1F/R et e0-3F/e1-1R y compris chez la lignée LLD138 qui présentait un niveau d'expression supérieur à celui des lignées fibroblastiques avec un couple d'amorces situées dans les exons 6 et 7.

Nous avons réanalysé cette lignée avec l'ensemble des amorces et confirmé les résultats. Il apparaît donc que l'expression atypique de *LSAMP* dans cette lignée se fait à partir d'un promoteur situé en amont de l'exon 1 (forme de 643 kb du gène) et non à partir du promoteur situé en amont de l'exon 0 comme chez les fibroblastes.



**Figure 39 : Carte du gène *LSAMP* humain.** **A)** Carte de la région 3q13.3. La position des bandes chromosomiques (en noir et gris), les coordonnées chromosomiques, les directions du centromère et du télomère (Tel), les régions présentant un profil de réplication en V, la localisation des cassures dans les fibroblastes et les altérations dans les cellules tumorales et la position du gène *LSAMP* sont indiqués. Les flèches courtes et longues représentent les deux formes du gène. **B)** Cartes de l'extrémité 5' des ARNm du gène *LSAMP*. Les deux isoformes d'ARNm contenant l'exon 1 ou les exons 0 et 1b sont représentées avec la position des exons, des amorces utilisées pour les expériences de qRT-PCR et les cadres de lecture des protéines. Noter que la protéine 0 codée par l'ARNm issu de la forme longue du gène diffère de la protéine *LSAMP* classique (protéine 1) dans sa partie N-terminale.



**Figure 40 : Expression des deux formes du gène *LSAMP* dans des lymphocytes et des fibroblastes humains. A)** Les ARN totaux des lignées indiquées ont été préparés et le niveau des transcrits issus de la forme longue du gène a été déterminé par qRT-PCR en utilisant les amorces e0-1F/e0-1R (exon 0) ou e0-3F/e1-1R (exon0-exon1). Ces niveaux sont exprimés relativement à celui de l'ARNm de la cyclophiline B utilisé comme gène de référence. **B)** Analyse détaillée des isoformes d'ARNm *LSAMP* produites par les cellules lymphoblastoïdes LLD138. Les expériences ont été réalisées et les résultats sont présentés comme en A. Les lignées MRC5, IMR90, et JEFF ont été incluses dans l'analyse comme références. Les amorces e1-F/e2-R ont été utilisées pour l'analyse des transcrits exon1-exon2.

**IV. Article « Replication Dynamics : Biases and Robustness of DNA Fiber Analysis »**

# Replication Dynamics: Biases and Robustness of DNA Fiber Analysis

Hervé Têcher<sup>1,2,3,†</sup>, Stéphane Koundrioukoff<sup>1,2,3,†</sup>, Dana Azar<sup>1,2,3</sup>, Therese Wilhelm<sup>1,2,3</sup>, Sandra Carignon<sup>1,2,3</sup>, Olivier Brison<sup>1,2,3</sup>, Michelle Debatisse<sup>1,2,3</sup> and Benoît Le Tallec<sup>1,2,3</sup>

**1 - Institut Curie, Centre de Recherche, 26 rue d'Ulm, 75248 Paris Cedex 05, France**

**2 - Université Pierre et Marie Curie Paris 06, 4 Place Jussieu, 75005 Paris, France**

**3 - Centre National de la Recherche Scientifique, Unité Mixte de Recherche 3244, 75248 Paris Cedex 05, France**

**Correspondence to Michelle Debatisse and Benoît Le Tallec:** M. Debatisse and B. Le Tallec are to be contacted at Institut Curie, Centre de Recherche, 26 rue d'Ulm, 75248 Paris Cedex 05, France. [michelle.debatisse@curie.fr](mailto:michelle.debatisse@curie.fr); [benoit.letallec@curie.fr](mailto:benoit.letallec@curie.fr)

<http://dx.doi.org/10.1016/j.jmb.2013.03.040>

**Edited by S. Kowalczykowski**

## Abstract

The factors that govern replication programs are still poorly identified in metazoans, especially in mammalian cells. Thanks to molecular combing, the dynamics of DNA replication can be assessed at the genome-scale level from the cumulative analysis of single DNA fibers. This technique notably enables measurement of replication fork speed and fork asymmetry and that of distances separating either initiation or termination events. The results presented here aim to evaluate requirements critical to accurate measurement of replication parameters by molecular combing. We show that sample size, fiber length and DNA counterstaining are crucial to gain robust information concerning replication dynamics. Our results thus provide a methodological frame to investigate the DNA replication program through molecular combing analyses.

© 2013 Elsevier Ltd. All rights reserved.

## Introduction

DNA replication ensures faithful duplication of the genome at each cell cycle. Replication of eukaryotic chromosomes relies primarily on two parameters, namely, the coordinated firing of multiple replication origins and the speed of forks emanating from fired origins.<sup>1</sup> Over the past few years, genetic and biochemical methods were teamed up to study the cellular mechanisms controlling DNA replication. However, the factors governing the location, the choice and the efficiency of mammalian replication origins remain poorly understood.<sup>2,3</sup> Analysis of DNA fibers stretched by DNA spreading<sup>4</sup> or molecular combing<sup>5,6</sup> proved to be among the most powerful tools to address these issues because, while other techniques provide parameters obtained from a population of cells, these techniques reveal the replication dynamics along individual DNA molecules upon direct visualization of neo-synthesized DNA *via* immunodetection of thymidine analogs incorporated in cells grown *in vitro*. In particular,

consecutive pulse labeling with two thymidine analogs, iododeoxyuridine (IdU) and chlorodeoxyuridine (CldU), permits measurement of the speed of individual forks and determination of their frequency of stalling, as well as identification of initiation and termination events.<sup>7–9</sup> DNA fiber analysis is therefore the method of choice to accurately monitor replication and has been successfully used in a variety of organisms.<sup>10,11</sup>

Genome-wide estimation of fork progression and origin activation is commonly used to give a snapshot of replication dynamics, which has contributed to identify factors governing the replication program. Notably, the analysis of DNA fibers has shown that initiation events cluster in domains along which multiple initiation events occur in close proximity and in an almost synchronous fashion.<sup>7–9,12–14</sup> This observation set the basis for the estimation of origin density by measuring the distance between adjacent initiation signals along a DNA fiber, named inter-origin distance (IOD).<sup>12,15,16</sup> As every beginning has an end, one can also

measure inter-termination distance (ITD, also known as inter-termini distance).<sup>17</sup> It is worth noting that the IOD estimated by DNA fiber analysis likely overestimates the actual distance between all origins that fire during S phase because (i) adjacent origins that fire during a single pulse cannot be resolved and (ii)

origins that have fired long before or that would have fired after the labeling periods are missed.<sup>18,19</sup> Similar limitations apply to ITD. IOD is also known as origin-to-origin (ori to ori)<sup>20,21</sup> or *i2i*<sup>22</sup> distances. DNA fiber analysis has also shown that progression of replication forks ranges from 0.5 to 2 kb/min

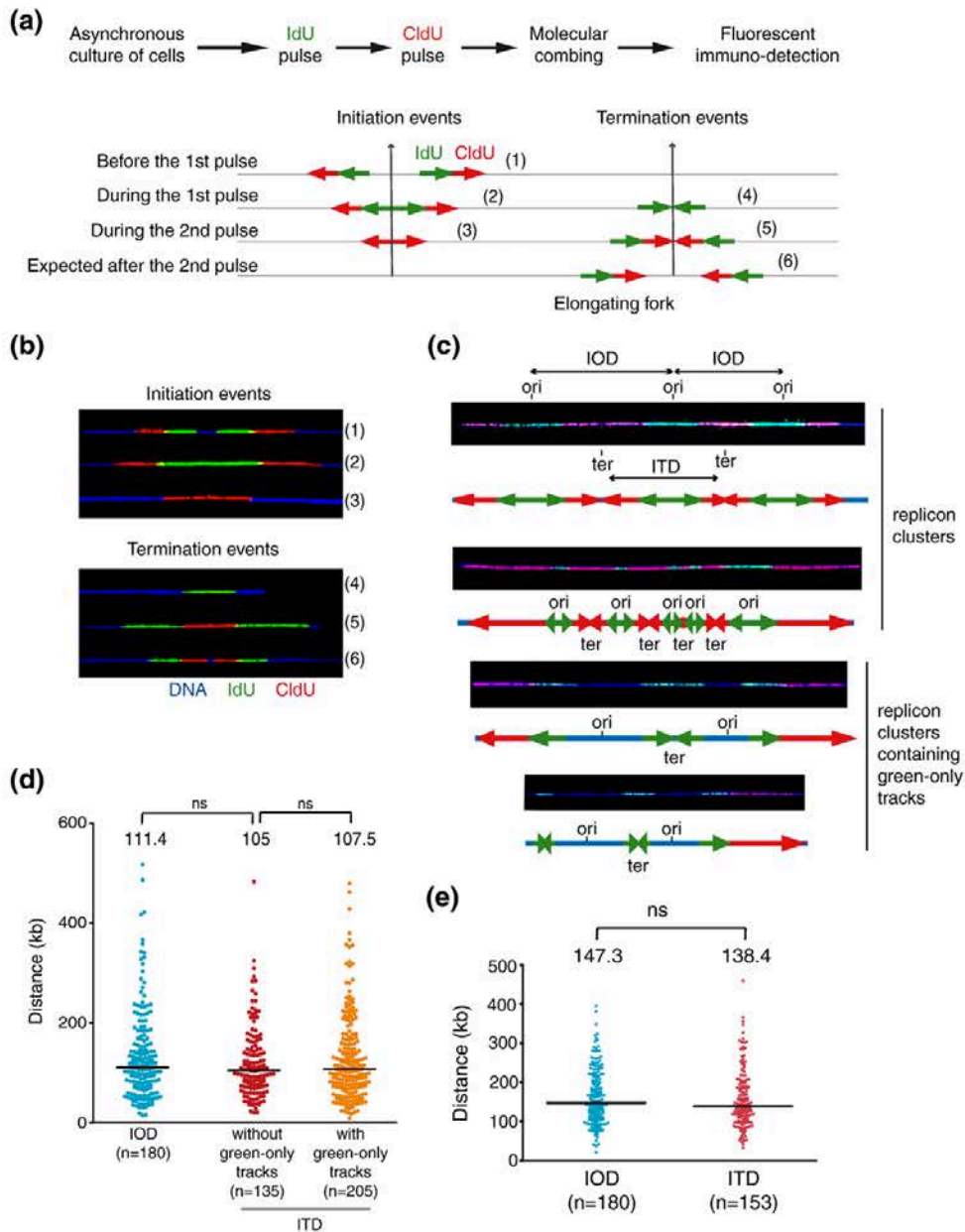


Fig. 1 (legend on next page)



in mammals, depending on the cell lines analyzed, which indicates that genetic and/or epigenetic factors influence fork movement.<sup>17</sup> Surprisingly, fork speed can show pronounced variability even in a given cell line. For example, in HeLa cells, fork speed has been estimated either  $\sim 2$  kb/min<sup>23,24</sup> or  $\sim 0.7$  kb/min.<sup>17,25,26</sup> Heterogeneous IOD values have also been reported for this cell line, ranging from 40 kb<sup>17</sup> to 144 kb.<sup>7</sup> Similarly, forks have been reported to travel at 2.16 kb/min,<sup>27</sup> 1.05 kb/min<sup>28</sup> and 0.37 kb/min<sup>20</sup> in U2OS cells. Although physiological variations may explain these discrepancies, it remains possible that differences in the preparation and the analysis of DNA fibers between laboratories may account, at least in part, for these differences.

In this study, we aim to provide a robust methodology to assess replication dynamics through molecular combing. We therefore investigated which biases and limitations affect DNA fiber analysis. We first intended to standardize the interpretation of replication patterns obtained following a double pulse labeling of neo-synthesized DNA. We then assessed the impact of the number and the length of DNA fibers analyzed in each experiment and that of DNA counterstaining on the determination of replication parameters.

## Results and Discussion

### Recognition of replication patterns obtained by double pulse labeling of forks

In order to determine the direction in which a given fork progresses along a DNA molecule and thus to recognize initiation and termination events (see below), we sequentially label asynchronously growing cells by two consecutive pulses of halogenated thymidine analogs (Fig. 1a). We will consider below that cells are labeled first with IdU, then with CldU (Fig. 1a). Each pulse routinely lasts 20–30 min but it has been shown that pulse length has no impact on the estimation of replication parameters, notably, fork speed.<sup>28</sup> A detailed protocol for molecular combing has recently been described.<sup>29</sup> Briefly, following pulse labeling, cells are harvested and embedded in agarose plugs to avoid shearing during

DNA purification. Plugs are melted and DNA molecules are combed onto silanized coverslips, leading to regular DNA stretching (2 kb/ $\mu$ m).<sup>5,6</sup> Alternatively, stretching of DNA molecules can also be achieved through the DNA spreading technique.<sup>4</sup> Neo-synthesized DNA is then revealed by indirect immunofluorescence.<sup>7,9</sup> Replication patterns can be classified into three categories: (i) individual elongating forks that are commonly used to assess fork speed (namely, ratio of fork length/pulse time; see [Materials and Methods](#)), (ii) initiations and (iii) terminations (Fig. 1a). Initiation events give rise to diverging forks, and termination events give rise to merging forks. In both cases, different replication patterns (numbered 1–6 in Fig. 1a and b) can be observed depending on when an origin fires or when replication forks merge relative to the labeling period.

### Single-color tracks from the first pulse correspond to termination events

Surprisingly, while there is a large consensus to record single-color signals corresponding to labeling during the second pulse only (Fig. 1a and b, pattern 3: red only) as initiation events,<sup>9,14,15,22,30–32</sup> the interpretation of single-color signals corresponding to labeling during the first pulse only (pattern 4: green only) is disputed in the literature. These signals are recorded either as termination events occurring during the first pulse<sup>9,13,32</sup> or as initiation events occurring during the first pulse without further elongation during the second pulse, an outcome of hypothetical stalling of both nascent forks.<sup>16,31,33–35</sup> We reasoned that termination events, or initiation events, might occur at the same frequency during either pulse since they last for the same time and, for the same reason, that termination might occur at the same frequency during the first pulse than initiation during the second one. We analyzed the results of three independent experiments performed with each JEFF lymphoblastoid cells and mouse embryonic fibroblasts (MEFs) (Table 1). We found that (i) green-only tracks (pattern 4) account for  $42.7 \pm 6.4\%$  and  $44.8 \pm 2.5\%$  of terminations (patterns 4 + 5) occurring during the two pulse periods in JEFF and MEF cells, respectively. Consistently, red-only tracks (pattern 3) represent  $42.4 \pm 6.2\%$  and  $45.2 \pm 1.6\%$

**Fig. 1.** Principles of replication dynamics analyses. (a) Scheme of the protocol used for double pulse labeling of replication forks in asynchronous cell cultures. Successive pulses with IdU and CldU enable recognition of ongoing forks, initiation and termination events. Green and red arrows represent neo-synthesized DNA, respectively, labeled with IdU and CldU. The labeling pattern depends on whether the considered event occurs before the first pulse, during the first pulse, during the second pulse or after the second pulse. Six different patterns of initiations and terminations are expected, numbered 1–6. (b) Examples of DNA fibers harboring initiation and termination events. Blue, DNA; green, IdU; red, CldU. (c) Examples of replicon clusters. The IOD is the distance separating two adjacent initiations (ori). The ITD is the distance separating two adjacent terminations (ter). A scheme is shown below each fiber. DNA is shown in blue. (d and e) IOD and ITD distributions in MEF (d) and JEFF (e) cells. Medians are represented by black horizontal lines and are indicated above each distribution. Medians rather than means are used as they are less sensitive to outliers in the data.  $n$  is the number of individual measures. Distributions were compared by the mean of the Mann–Whitney test (see [Materials and Methods](#)). ns indicates that  $p$ -value is  $> 0.5$ . (d) As indicated, we excluded, or not, fibers containing IdU-only tracks (green only).



**Table 1.** Frequencies of green-only and red-only tracks

		Red only/initiations [3/(2 + 3)] <sup>a</sup>	Green only/terminations [4/(4 + 5)] <sup>a</sup>	Red only/total [3/(2 + 3 + 4 + 5)] <sup>a</sup>	Green only/total [4/(2 + 3 + 4 + 5)] <sup>a</sup>
JEFF	Experiment 1	71/144 (49.3%)	105/210 (50%)	71/354 (20%)	105/354 (29.6%)
	Experiment 2	67/165 (40.6%)	72/180 (40%)	67/345 (19.4%)	72/345 (20.9%)
	Experiment 3	115/309 (37.2%)	117/307 (38.1%)	115/616 (18.7%)	117/616 (19%)
	Mean ± SD <sup>b</sup>	42.4 ± 6.2	42.7 ± 6.4	19.4 ± 0.7	23.2 ± 5.7
MEF	Experiment 1	102/235 (43.4%)	105/248 (42.3%)	102/483 (21.1%)	105/483 (21.7%)
	Experiment 2	115/252 (45.6%)	97/216 (44.9%)	115/468 (24.6%)	97/468 (20.7%)
	Experiment 3	99/213 (46.5%)	124/262 (47.3%)	99/475 (20.8%)	124/475 (26.1%)
	Mean ± SD <sup>b</sup>	45.2 ± 1.6	44.8 ± 2.5	22.2 ± 2.1	22.8 ± 2.8

<sup>a</sup> Numbers refer to the patterns of replication described in Fig. 1.

<sup>b</sup> Mean ± SD (%) of the three independent experiments presented above.

of initiations (patterns 2 + 3) occurring during the two pulse periods in these cells. The slight deficit in single-color tracks we observed in both cases can be explained by the fact that short single-color tracks are not distinguishable from the background and are therefore not counted. (ii) Green-only and red-only tracks are observed at similar frequencies in both cell lines. These results strongly support the idea that the vast majority of green-only tracks are true termination events. Finally, we used DNA fibers from MEFs grown under normal conditions to measure ITDs, taking into account either all fibers displaying termination events, including those with green-only tracks, or only fibers displaying two-color terminations. Typical replicon clusters with green-only tracks present or absent are shown in Fig. 1c. We found similar ITDs in both cases (Fig. 1d). We also measured the IOD in the same samples and found that both the medians and global distributions of ITDs precisely resemble those of IOD (Fig. 1d). This latter result was confirmed from an independent data set obtained from human lymphoblastoid JEFF cells (Fig. 1e). Together, our data show, as expected, that initiations and terminations occur at the same frequency and unambiguously indicate that the vast majority of green tracks represent genuine terminations. Therefore, conclusions from studies using green-only tracks to estimate fork stalling need to be re-evaluated.

#### Impact of sample size on IOD and fork speed estimation

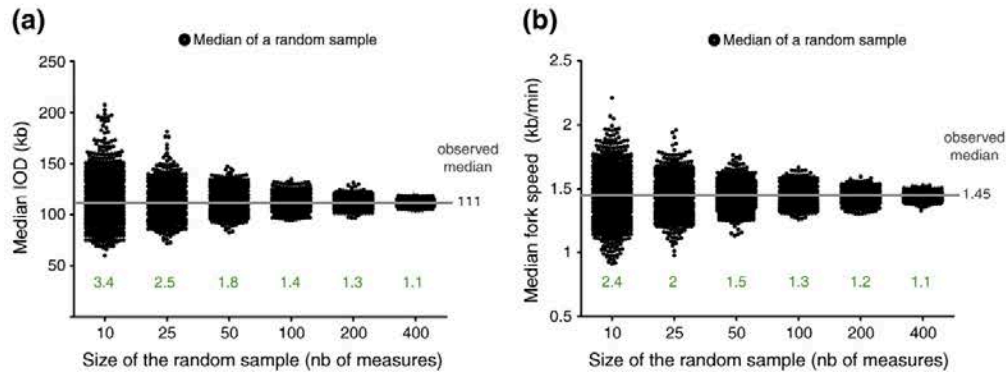
As shown in Fig. 1d and e, measured values for IOD or ITD of a given sample are widely scattered. It is thus mandatory to analyze distributions, not only means or medians, to perform correct comparisons of different samples and/or replication parameters (see *Materials and Methods*). Moreover, because of this dispersion, the number of fibers analyzed in each experiment might be critical to obtain reliable estimations of replication parameters. To assess sampling effect on IOD, we simulated 1000 data sets of bootstrap samples of various sizes (namely, 10, 25, 50, 100,

200 and 400) from a pool of over 700 IODs measured in MEF cells grown under standard conditions. This process mimics measurement of 10, 25, 50 and so on IODs per experiment. The distribution of the median IOD values for each bootstrap sample is presented in Fig. 2a. We found that, if one evaluates the median IOD of a sample from 25 to 50 individual measures, the estimated value could fluctuate by 2- to 3-fold due to sampling effects (i.e., the median IOD inferred from 10 individual measures can range from 60 kb to 208 kb). Increasing the number of measures limits this variability, as illustrated by the decreasing ratio between the maximal and minimal median IOD values obtained from 10 to 400 measures (Fig. 2a). Similar results were obtained when assessing sampling effects on fork speed (Fig. 2b). Thus, a large number of measures, at least 100–200, are required for a reliable estimation of IOD or fork speed in a given sample and, *a fortiori*, to compare two samples.

#### Impact of fiber length on IOD and fork speed estimation

DNA fiber analysis relies primarily on measurement of signal length or of the length spacing two signals. Therefore, the length of DNA fibers could impact the estimation of replication parameters. Determination of the length of DNA fibers is commonly achieved through YOYO-1 counterstaining<sup>36</sup> or by immunodetection of single-stranded DNA.<sup>29</sup> Alternatively, dig-dUTP labeling of DNA for the entire length of S phase<sup>37,38</sup> or extremely long IdU/CldU pulses can be used.<sup>39</sup> Of note, at a magnification of 63x, one microscope field corresponds to ~250 kb. Observation of longer fibers thus requires the capture of adjacent fields or the use of a fluorescence microscope equipped with a motorized stage that enables scanning of slides with high precision (see *Materials and Methods*).

To assess the impact of fiber length on IOD and fork speed, we choose three samples (named samples 1–3) of JEFF cells cultured under standard conditions in which the distribution of fiber lengths varies importantly, with median lengths ranging from



**Fig. 2.** Impact of sample size on measurement of fork speed and IOD. (a and b) From a large data set obtained in MEFs, we computed random samples of various sizes, namely, of 10, 25, 50, 100, 200 and 400 individual measures. Distributions of median IOD (a) and median fork speed (b) of random draws bootstrapped a thousand times are shown. The ratio of the maximal/minimal median is indicated below each distribution in green. The gray horizontal lines represent the observed median IOD (a) and fork speed (b).

104 to 378 kb (Fig. 3a). We observed that IOD estimation is very sensitive to fiber length (Fig. 3b). For example, IOD estimation was almost 5-fold lower in sample 3 than in sample 1. A significant difference was still observed between samples 1 and 2 (Fig. 3b). Similar IODs were estimated from DNA fiber sets with median lengths from 350 kb to 650 kb (data not shown), indicating that robust assessment of IOD requires fiber sets with a median size of at least 350 kb. Fork speed estimated from samples 1 and 2 was similar, but it was considerably lower in sample 3 (Fig. 3c). Although fork speed seems less sensitive to fiber length than IOD, fibers of at least 200 kb are required for accurate speed estimation.

#### Impact of DNA counterstaining on estimation of IOD, ITD, fork speed and fork asymmetry

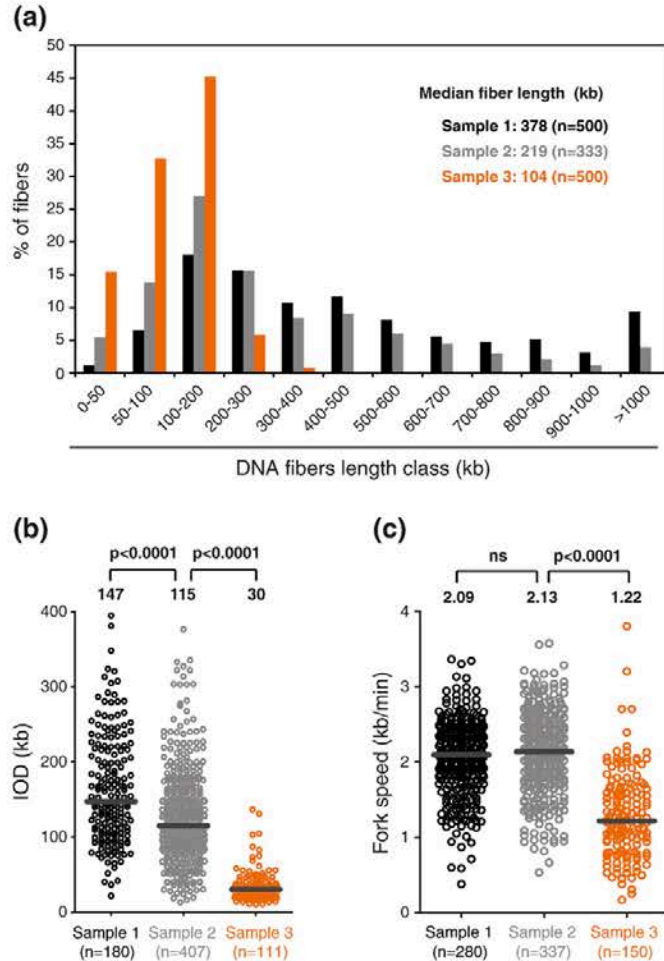
DNA staining is crucial to determine if a DNA fiber is broken and if two replication signals belong to the same DNA molecule, a prerequisite to IOD or ITD measurement.<sup>11</sup> Broken fibers can lead to the underestimation of fork speed and create false asymmetrical forks (Fig. 4a). Still, DNA staining has often been omitted in DNA fiber analysis. Strikingly, we found that DNA fibers are broken at the level of replication signals, in either IdU or CldU track, in almost one-third of all replication signals recorded in samples from JEFF cells cultured under standard conditions (Fig. 4b). The distribution of fiber lengths of the three independent samples used to assess the frequency of broken signals is shown in Fig. 4c.

To test the influence of DNA staining on IOD, ITD, fork asymmetry and fork speed, we analyzed the same set of images with or without DNA counterstaining. The

absence of counterstaining was simulated by omitting to display the signal corresponding to DNA, namely, that of the cyanine 5.5 filter (see [Materials and Methods](#)). These fibers correspond to sample 1 in Fig. 3. We observed that omission of DNA counterstaining altered the estimation of both IOD and ITD (Fig. 5a and b), most likely because some initiation or termination events on the same DNA fiber, but separated by long distances, were considered as belonging to different fibers. In contrast, fork speed was not impacted by the absence of DNA counterstaining (Fig. 5c). This might be explained by the fact that sample 1 is constituted of fairly long DNA fibers (Fig. 3a). One can indeed expect that analysis of a sample with shorter DNA fibers (i.e., more sheared) will be more sensitive to the absence of DNA staining. This was confirmed by analysis of sample 3 (from Fig. 3a), for which the omission of DNA counterstaining led to an estimation of median fork speed of 1.05 kb/min, which is significantly different from 1.22 kb/min estimated with DNA counterstaining (data not shown).

We next assessed the impact of DNA staining on the estimation of fork asymmetry, a parameter used to estimate the occurrence of fork stalling. Abnormal fork progression during one of the two labeling pulses can be estimated from the calculation of the IdU/CldU ratio<sup>32</sup> (Fig. 5d; see [Materials and Methods](#)). Alternatively, it is possible to estimate the synchronized progression of forks emanating from a given origin through the ratio between the length of left and right traveling forks<sup>12,30</sup> (Fig. 5d; see [Materials and Methods](#)). For both methods, an unbalanced ratio is indicative of fork stalling. We observed that both the IdU/CldU and left fork/right fork ratios differ significantly when measured with or without DNA counterstaining



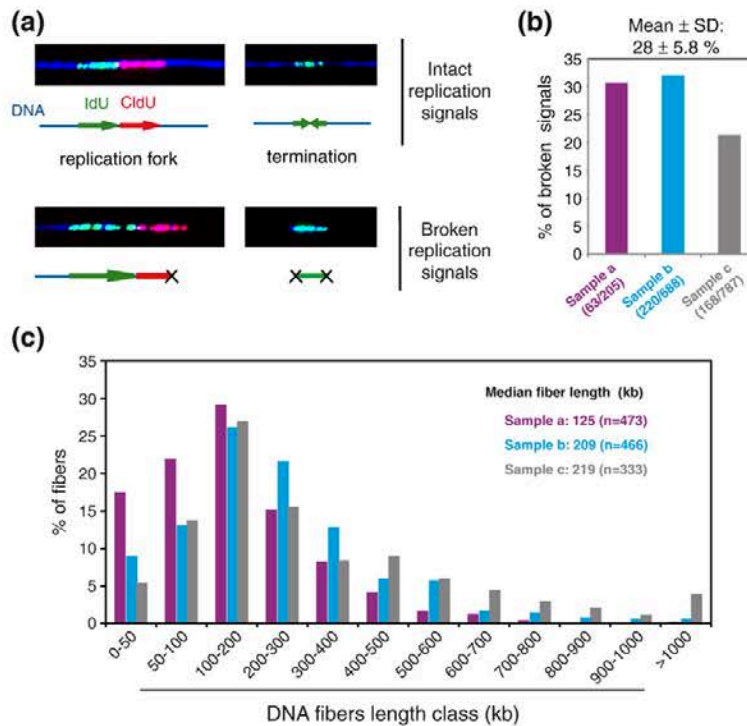


**Fig. 3.** Impact of fiber length on measurement of fork speed and IOD. (a) The distribution of fiber length is shown for three independent preparations of JEFF cells, named samples 1, 2 and 3. The medians and the number of measures ( $n$ ) are indicated for each sample. Dot plots of IOD (b) and fork speed (c) are shown. The number of individual measures ( $n$ ) and the medians are indicated for each distribution. The  $p$ -values are indicated. (b) IOD measurements take into account the patterns 1, 2 and 3 described in Fig. 1.

(Fig. 5e and f). Omission of DNA counterstaining leads to an overestimation of fork asymmetry regardless of the type of ratio considered. We indeed observed that ratios measured without DNA staining are enriched in high asymmetry values, a bias that most likely originates from broken fibers (Fig. 4a and b). An independent experiment in MEFs also showed a strong impact of counterstaining on measurement of fork asymmetry and showed that this increase is in the same order of magnitude as that resulting from treatment of the cells with 250  $\mu$ M hydroxyurea (Fig. 5g). Collectively, our results demonstrate that DNA staining, which is compulsory to discard replication signals arising from broken DNA fibers, is essential for accurate estimation of all replication parameters. Of note, all analyses presented above were performed with DNA counterstaining.

### Biological variability

Finally, it remains to be determined what the level of biological variability of replication parameters is. To address this question, we compared independent experiments analyzed according to the abovementioned standards, namely, median length of DNA fibers above 200 kb (350 kb for IOD), minimum of 100–200 events recorded and DNA counterstaining. Over the course of many years of experiments, we have accumulated 49 independent experiments in JEFF cells and, respectively, 12 and 10 in primary mouse and Chinese hamster fibroblasts. In either cell model, we found a 10–15% fluctuation of median fork speed independently of cell types or species (Fig. 6). These data indicate that fork speed measurement and, presumably, any other replication parameter display



**Fig. 4.** Analysis of broken replication signals. (a) Examples of intact and broken replication signals are shown. Schemes of signals are presented as in Fig. 1, except that here a black cross represents the site of breakage on DNA fibers. (b) We estimated the frequency of replication signals broken at the level of either IdU tracks or CldU tracks. Frequency of broken replication signals is shown as a percentage of total signals from three independent experiments of JEFF cells, named samples a, b and c. Mean  $\pm$  SD is indicated. (c) The distribution of fiber length is shown for samples a, b and c.

inherent variability of  $\sim 15\%$ . Differences of less than 15% between samples must therefore be reproduced several times before being considered as significant.

### Concluding remarks

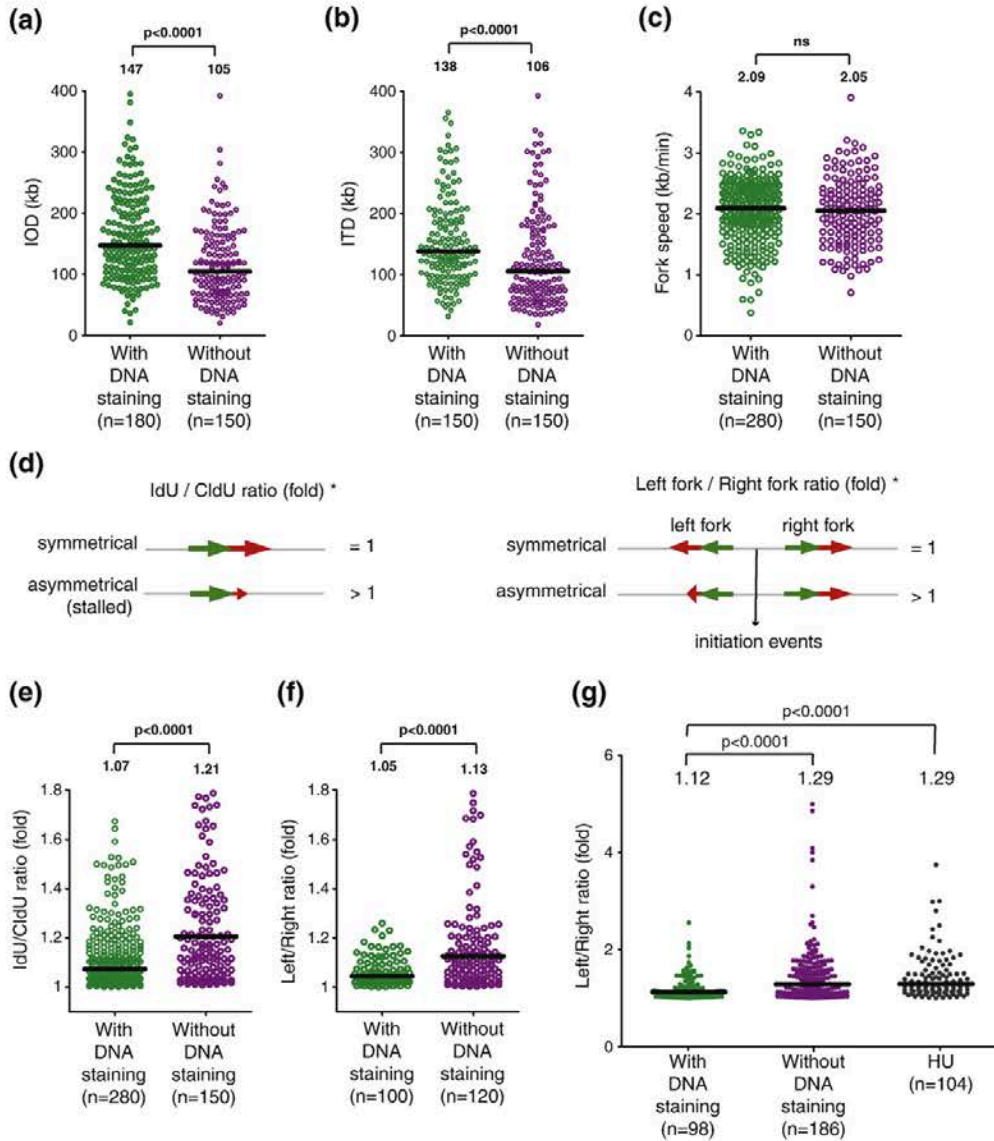
DNA fiber analysis is a powerful technology to monitor replication dynamics. Here we tested the most common biases affecting DNA fiber analysis, providing experimentally determined standards that should raise the technique to a high level of robustness, which is presently not the case. In addition, besides the methodological biases we have identified and for which we have established conditions allowing these biases to be overcome, it is important to keep in mind that stretching of DNA molecules itself is subjected to some variation, estimated at  $\sim 2\%$  for DNA combing<sup>5</sup> and  $\sim 10\%$  for DNA spreading.<sup>4,7</sup> Moreover, endogenous physiological variations may influence the replication dynamics. It is well known that replication dynamics is exquisitely sensitive to the pool of dNTPs.<sup>9,13,40–42</sup> The metabolic state of the cells is thus of extreme

importance. The pool of dNTPs is regulated during the cell cycle,<sup>43–47</sup> and it has been reported that the density of active forks increases during S phase in HeLa cells.<sup>17</sup> When questioning if a factor regulates DNA replication, it is thus critical to establish if uncontrolled alterations of S phase progression might interfere with the estimation of replication dynamics.

## Materials and Methods

### Cell culture

Epstein-Barr virus-immortalized human B-lymphocytes (JEFF cells) were grown in RPMI 1640+ GlutaMAX-I medium (Gibco) and 10% fetal calf serum (PAN Biotech GmbH). Chinese hamster embryonic fibroblasts (CHEFs) were grown in MEM (minimal essential medium) supplemented with Earle's salts (Gibco), 1% MEM non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), 2 mM L-glutamine (Gibco) and 10% fetal calf serum (PAN Biotech GmbH). MEFs were grown for a maximum of 4 passages in DMEM (Dulbecco's modified Eagle's medium) + GlutaMAX medium (Gibco), 15% fetal calf serum



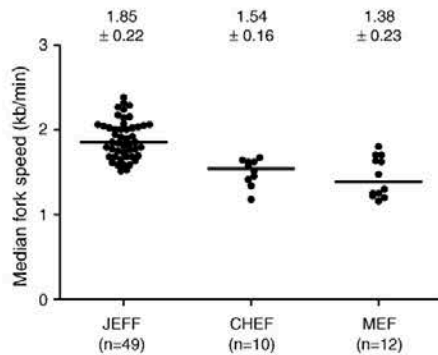
**Fig. 5.** Impact of DNA counterstaining on the measurement of replication parameters. Distribution of IOD (a), ITD (b) and fork speed (c) in JEFF cells. The same set of images was analyzed with or without visualizing DNA counterstaining. Data are presented as in the previous figures. The sample used here is the sample 1 of Fig. 3 (median length of fibers: 378 kb). (d) IdU/CldU and left fork/right fork ratios are depicted. An asterisk indicates that the IdU/CldU or left fork/right fork ratio corresponds to the ratio of the longest signal over the shortest signal. A ratio of  $> 1$  indicates fork asymmetry (see *Materials and Methods*). (e–g) Distribution of IdU/CldU and left fork/right fork ratios with or without DNA counterstaining in JEFF (e and f) and MEF cells (g). In (g), MEFs treated with 250  $\mu$ M hydroxyurea (HU) are used as an internal control of the left fork/right fork ratio.

(Biowest), 0.1 mM  $\beta$  mercaptoethanol (Gibco) and 0.01 mM non-essential amino acids (Gibco). All media were supplemented with 100 IU/mL penicillin and 100  $\mu$ g/mL streptomycin (Gibco). JEFF cells were grown at 37  $^{\circ}$ C, 20%  $O_2$  and 5%  $CO_2$ ; MEF cells, at 37  $^{\circ}$ C, 3%  $O_2$  and 5%  $CO_2$ ; CHEF cells, at 37  $^{\circ}$ C, 20%  $O_2$  and 8%  $CO_2$ .

#### Molecular combing and immunofluorescent detection

Molecular combing and immunodetection were performed as previously described.<sup>9,29</sup> Briefly, cells were pulse-labeled with IdU, then CldU. Pulses lasted 20 min for CHEF and JEFF cells and 30 min for MEFs. Cells were





**Fig. 6.** Determination of fork speed variability in different cell types. Dot plots of the median fork speed in JEFF, CHEF and MEF cells obtained from independent experiments. Black horizontal lines represent the median of the median distribution, and  $n$  is the number of independent medians obtained from cells cultured independently under standard conditions. Median  $\pm$  SD is indicated above each distribution.

harvested, and DNA fibers were purified by digestion of proteins in agarose plugs and subsequently stretched at a rate of 2 kb/ $\mu$ m on silanized coverslips prepared as previously described.<sup>48</sup> Immunodetection of neo-synthesized DNA and DNA fibers was performed by successive incubations with the following antibodies: (1) mouse anti-bromodeoxyuridine (BrdU) (BD Biosciences, #347583) and rat anti-BrdU (AbD Serotec, OBT 0030), (2) A488-conjugated goat anti-mouse (Invitrogen, A11029) and A555-conjugated goat anti-rat (Abcam, A21434), (3) mouse anti-single-stranded DNA (Millipore, MAB 3034), (4) cy5.5-conjugated goat anti-mouse (Abcam, ab6947) and (5) cy5.5-conjugated donkey anti-goat (Abcam, ab6951). We used an epifluorescence microscope (Axio Imager Z2; Carl Zeiss) equipped with a 63 $\times$  objective lens (PL APO, NA 1.4 Oil DIC M27) and a motorized stage (M-689 XY Microscope Stage with PILine Motor; Physik Instrumente), connected to a charge-coupled device camera (Cool-SNAP HQ2; Roper Scientific). MetaMorph software (Roper Scientific) was used for image acquisition and for building large-scale mosaic comprising up to 100 images. This technology theoretically enables us to recover long fibers, up to approximately 3 Mb (10 adjacent fields at a magnification of 63 $\times$ ). Measurements of fork speed and asymmetry, fiber length and IOD/ITD were performed either with MetaMorph software (Roper Scientific) or with Photoshop software (Adobe). We systematically used DNA counterstaining to ensure that replication signals belong to the same fiber, that two fibers are not overlapping and that signals are not broken.

#### Statistical analysis

Replication fork speed was estimated on individual forks displaying an IdU track flanked by a CldU track. Only intact forks were recorded as ascertain by DNA counterstaining. For each fork, the ratio  $(LI + LCI)/(tI + tCI)$  was calculated

[LI is the measured track (kb) and tI is the labeling time (min) for IdU; LCI and tCI are the corresponding parameters for CldU]. Note that the so-called IdU/CldU ratio that estimates fork asymmetry is calculated as the ratio of the longest track over the shortest. The left fork/right fork ratio corresponds to the ratio of the longest fork over the shortest fork. Thus, the IdU/CldU or left fork/right fork ratios are  $\geq 1$ . IOD and ITD correspond to the distance (kb) between the centers of two adjacent initiation or termination signals located on the same DNA fiber. We used GraphPad Prism (GraphPad Software) to generate graphs and perform statistical analysis. DNA replication parameters do not generally display a Gaussian distribution.<sup>29</sup> Statistical comparisons of the distributions were therefore assessed using the nonparametric Mann-Whitney rank sum test. Statistical significance was set to  $p \leq 0.05$ . Two-tailed tests were systematically used. The R environment was used to compute random samples in Fig. 2 [R Development Core Team (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.]. R code will be sent upon request.

#### Acknowledgements

We thank Prof. Franck Toledo's laboratory for providing MEFs. We thank Dr. Gaël A. Millot for advices on statistical analysis. We thank Genomic Vision for providing molecular combing technology. The M.D. team is supported by the Institut National du Cancer (2009-1-PLBIO-10-IC-1), by the Agence Nationale de la Recherche (ANR-09-GENO-000/repinsCFS) and by the Association pour la Recherche sur le Cancer (ARC Subvention Libre no. SL220100601348 and Equipements mi-lourds no. 8514). B.L.T. is supported by a fellowship from Institut National du Cancer; H.T. is supported by a MENRT graduate student fellowship and a fellowship from the Association pour la Recherche sur le Cancer.

**Conflict of Interest.** The authors declare no competing financial interests.

Received 28 February 2013;

Received in revised form 22 March 2013;

Accepted 25 March 2013

Available online 1 April 2013

#### Keywords:

DNA combing;  
replication fork speed;  
inter-origin distance;  
inter-termination distance;  
fork asymmetry

† H.T. and S.K. contributed equally to this work.

‡ <http://www.R-project.org/>

**Abbreviations used:**

CHEF, Chinese hamster embryonic fibroblast; IOD, inter-origin distance; ITD, inter-termination distance; MEF, mouse embryonic fibroblast.

**References**

- Hyrien, O. & Goldar, A. (2010). Mathematical modeling of eukaryotic DNA replication. *Chromosome Res.* **18**, 147–161.
- Gilbert, D. M. (2010). Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat. Rev., Genet.* **11**, 673–684.
- Mechali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev., Mol. Cell Biol.* **11**, 728–738.
- Parra, J. & Windle, B. (1993). High resolution visual mapping of stretched DNA by fluorescent hybridization. *Nat. Genet.* **5**, 17–21.
- Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F. & Bensimon, D. (1994). Alignment and sensitive detection of DNA by a moving interface. *Science*, **265**, 2096–2098.
- Michalet, X., Ekong, R., Fougereuse, F., Rousseaux, S., Schurra, C., Hornigold, N. *et al.* (1997). Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science*, **277**, 1518–1523.
- Jackson, D. A. & Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* **140**, 1285–1295.
- Pasero, P., Bensimon, A. & Schwob, E. (2002). Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. *Genes Dev.* **16**, 2479–2484.
- Anglana, M., Apiou, F., Bensimon, A. & Debatisse, M. (2003). Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell*, **114**, 385–394.
- Herrick, J. & Bensimon, A. (2008). Global regulation of genome duplication in eukaryotes: an overview from the epifluorescence microscope. *Chromosoma*, **117**, 243–260.
- Tuduri, S., Tourriere, H. & Pasero, P. (2010). Defining replication origin efficiency using DNA fiber assays. *Chromosome Res.* **18**, 91–102.
- Conti, C., Sacca, B., Herrick, J., Lalou, C., Pommier, Y. & Bensimon, A. (2007). Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Mol. Biol. Cell*, **18**, 3059–3067.
- Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O. & Debatisse, M. (2008). Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature*, **455**, 557–560.
- Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J. & Bensimon, A. (2006). DNA replication origin interference increases the spacing between initiation events in human cells. *Mol. Biol. Cell*, **17**, 5337–5345.
- Maya-Mendoza, A., Petermann, E., Gillespie, D. A., Caldecott, K. W. & Jackson, D. A. (2007). Chk1 regulates the density of active replication origins during the vertebrate S phase. *EMBO J.* **26**, 2719–2731.
- Scorah, J. & McGowan, C. H. (2009). Claspin and Chk1 regulate replication fork stability by different mechanisms. *Cell Cycle*, **8**, 1036–1043.
- Guilbaud, G., Rappailles, A., Baker, A., Chen, C. L., Arneodo, A., Goldar, A. *et al.* (2011). Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput. Biol.* **7**, e1002322.
- Herrick, J., Jun, S., Bechhoefer, J. & Bensimon, A. (2002). Kinetic model of DNA replication in eukaryotic organisms. *J. Mol. Biol.* **320**, 741–750.
- Loveland, A. B., Habuchi, S., Walter, J. C. & van Oijen, A. M. (2012). A general approach to break the concentration barrier in single-molecule imaging. *Nat. Methods*, **9**, 987–992.
- Ge, X. Q., Jackson, D. A. & Blow, J. J. (2007). Dormant origins licensed by excess Mcm2–7 are required for human cells to survive replicative stress. *Genes Dev.* **21**, 3331–3341.
- Katsuno, Y., Suzuki, A., Sugimura, K., Okumura, K., Zineldeen, D. H., Shimada, M. *et al.* (2009). Cyclin A-Cdk1 regulates the origin firing program in mammalian cells. *Proc. Natl Acad. Sci. USA*, **106**, 3184–3189.
- Conti, C., Seiler, J. A. & Pommier, Y. (2007). The mammalian DNA replication elongation checkpoint: implication of Chk1 and relationship with origin firing as determined by single DNA molecule and single cell analyses. *Cell Cycle*, **6**, 2760–2767.
- Dominguez-Sanchez, M. S., Barroso, S., Gomez-Gonzalez, B., Luna, R. & Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet.* **7**, e1002386.
- Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I. *et al.* (2010). Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev.* **24**, 2812–2822.
- Ibarra, A., Schwob, E. & Mendez, J. (2008). Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc. Natl Acad. Sci. USA*, **105**, 8956–8961.
- Petermann, E., Helleday, T. & Caldecott, K. W. (2008). Claspin promotes normal replication fork rates in human cells. *Mol. Biol. Cell*, **19**, 2373–2378.
- Dominguez-Kelly, R., Martin, Y., Koundrioukoff, S., Tanenbaum, M. E., Smits, V. A., Medema, R. H. *et al.* (2011). Wee1 controls genomic stability during replication by regulating the Mus81-Eme1 endonuclease. *J. Cell Biol.* **194**, 567–579.
- Petermann, E., Woodcock, M. & Helleday, T. (2010). Chk1 promotes replication fork progression by controlling replication initiation. *Proc. Natl Acad. Sci. USA*, **107**, 16090–16095.
- Bianco, J. N., Poli, J., Saksouk, J., Bacal, J., Silva, M. J., Yoshida, K. *et al.* (2012). Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods*, **57**, 149–157.
- Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A. *et al.* (2009). Topoisomerase I

- suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **11**, 1315–1324.
31. Wijsker, D., Petermann, E., Helleday, T. & Bunz, F. (2008). Essential function of Chk1 can be uncoupled from DNA damage checkpoint and replication control. *Proc. Natl Acad. Sci. USA*, **105**, 20752–20757.
  32. Letessier, A., Millot, G. A., Koundrioukoff, S., Lachages, A. M., Vogt, N., Hansen, R. S. *et al.* (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature*, **470**, 120–123.
  33. Merrick, C. J., Jackson, D. & Diffley, J. F. (2004). Visualization of altered replication dynamics after DNA damage in human cells. *J. Biol. Chem.* **279**, 20067–20075.
  34. Conti, C., Leo, E., Eichler, G. S., Sordet, O., Martin, M. M., Fan, A. *et al.* (2010). Inhibition of histone deacetylase in cancer cells slows down replication forks, activates dormant origins, and induces DNA damage. *Cancer Res.* **70**, 4470–4480.
  35. Rao, V. A., Conti, C., Guirouilh-Barbat, J., Nakamura, A., Miao, Z. H., Davies, S. L. *et al.* (2007). Endogenous  $\gamma$ -H2AX-ATM-Chk2 checkpoint activation in Bloom's syndrome helicase deficient cells is related to DNA replication arrested forks. *Mol. Cancer Res.* **5**, 713–724.
  36. Marheineke, K., Hyrien, O. & Krude, T. (2005). Visualization of bidirectional initiation of chromosomal DNA replication in a human cell free system. *Nucleic Acids Res.* **33**, 6931–6941.
  37. Herrick, J., Stanislawski, P., Hyrien, O. & Bensimon, A. (2000). Replication fork density increases during DNA synthesis in *X. laevis* egg extracts. *J. Mol. Biol.* **300**, 1133–1142.
  38. Marheineke, K. & Hyrien, O. (2004). Control of replication origin density and firing time in *Xenopus* egg extracts: role of a caffeine-sensitive, ATR-dependent checkpoint. *J. Biol. Chem.* **279**, 28071–28081.
  39. Norio, P. & Schildkraut, C. L. (2001). Visualization of DNA replication on individual Epstein-Barr virus episomes. *Science*, **294**, 2361–2364.
  40. Taylor, J. H. (1977). Increase in DNA replication sites in cells held at the beginning of S phase. *Chromosoma*, **62**, 291–300.
  41. Nicander, B. & Reichard, P. (1985). Relations between synthesis of deoxyribonucleotides and DNA replication in 3T6 fibroblasts. *J. Biol. Chem.* **260**, 5376–5381.
  42. Poli, J., Tsaponina, O., Crabbe, L., Keszthelyi, A., Pantescio, V., Chabes, A. *et al.* (2012). dNTP pools determine fork progression and origin usage under replication stress. *EMBO J.* **31**, 883–894.
  43. Eledge, S. J., Zhou, Z., Allen, J. B. & Navas, T. A. (1993). DNA damage and cell cycle regulation of ribonucleotide reductase. *BioEssays*, **15**, 333–339.
  44. Eriksson, S., Graslund, A., Skog, S., Thelander, L. & Tribukait, B. (1984). Cell cycle-dependent regulation of mammalian ribonucleotide reductase. The S phase-correlated increase in subunit M2 is regulated by de novo protein synthesis. *J. Biol. Chem.* **259**, 11695–11700.
  45. Nordlund, P. & Reichard, P. (2006). Ribonucleotide reductases. *Annu. Rev. Biochem.* **75**, 681–706.
  46. Thelander, L. (2007). Ribonucleotide reductase and mitochondrial DNA synthesis. *Nat. Genet.* **39**, 703–704.
  47. Hakansson, P., Hofer, A. & Thelander, L. (2006). Regulation of mammalian ribonucleotide reduction and dNTP pools after DNA damage and in resting cells. *J. Biol. Chem.* **281**, 7834–7841.
  48. Labit, H., Goldar, A., Guilbaud, G., Douarche, C., Hyrien, O. & Marheineke, K. (2008). A simple and optimized method of producing silanized surfaces for FISH and replication mapping on combed DNA fibers. *Biotechniques* **45**, 649–652, 654, 656–658.



# Références

---

Aguilera, A., and Garcia-Muse, T. (2013). Causes of genome instability. *Annual review of genetics* 47, 1-32.

Alver, R.C., Chadha, G.S., and Blow, J.J. (2014). The contribution of dormant origins to genome stability: from cell biology to human genetics. *DNA repair* 19, 182-189.

Anindya, R., Aygun, O., and Svejstrup, J.Q. (2007). Damage-induced ubiquitylation of human RNA polymerase II by the ubiquitin ligase Nedd4, but not Cockayne syndrome proteins or BRCA1. *Molecular cell* 28, 386-397.

Arlt, M.F., Miller, D.E., Beer, D.G., and Glover, T.W. (2002). Molecular characterization of FRAXB and comparative common fragile site instability in cancer cells. *Genes, chromosomes & cancer* 33, 82-92.

Arlt, M.F., Xu, B., Durkin, S.G., Casper, A.M., Kastan, M.B., and Glover, T.W. (2004). BRCA1 is required for common-fragile-site stability via its G2/M checkpoint function. *Molecular and cellular biology* 24, 6701-6709.

Ballabeni, A., Melixetian, M., Zamponi, R., Masiero, L., Marinoni, F., and Helin, K. (2004). Human geminin promotes pre-RC formation and DNA replication by stabilizing CDT1 in mitosis. *The EMBO journal* 23, 3122-3132.

Barlow, J.H., Faryabi, R.B., Callen, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.W., McKinnon, P., Wright, G., *et al.* (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell* 152, 620-632.

Beck, D.B., Burton, A., Oda, H., Ziegler-Birling, C., Torres-Padilla, M.E., and Reinberg, D. (2012). The role of PR-Set7 in replication licensing depends on Suv4-20h. *Genes & development* 26, 2580-2589.

Bell, S.P., Kobayashi, R., and Stillman, B. (1993). Yeast origin recognition complex functions in transcription silencing and DNA replication. *Science* 262, 1844-1849.

Bensaude, O. (2011). Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* 2, 103-108.

Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., *et al.* (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899-905.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.M., and Lemaitre, J.M. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature structural & molecular biology* *19*, 837-844.

Bignell, G.R., Greenman, C.D., Davies, H., Butler, A.P., Edkins, S., Andrews, J.M., Buck, G., Chen, L., Beare, D., Latimer, C., *et al.* (2010). Signatures of mutation and selection in the cancer genome. *Nature* *463*, 893-898.

Boldog, F., Gemmill, R.M., West, J., Robinson, M., Robinson, L., Li, E., Roche, J., Todd, S., Waggoner, B., Lundstrom, R., *et al.* (1997). Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B. *Human molecular genetics* *6*, 193-203.

Bozzoni, I., Baldari, C.T., Amaldi, F., and Buongiorno-Nardelli, M. (1981). Replication of ribosomal DNA in *Xenopus laevis*. *European journal of biochemistry / FEBS* *118*, 585-590.

Bramhill, D., and Kornberg, A. (1988). Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* *52*, 743-755.

Cadoret, J.C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 15837-15842.

Cairns, J. (1963). The bacterial chromosome and its manner of replication as seen by autoradiography. *Journal of molecular biology* *6*, 208-213.

Casper, A.M., Nghiem, P., Arlt, M.F., and Glover, T.W. (2002). ATR regulates fragile site stability. *Cell* *111*, 779-789.

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., and Mechali, M. (2012). New insights into replication origin characteristics in metazoans. *Cell Cycle* *11*, 658-667.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., *et al.* (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome research* *21*, 1438-1449.

Chabosseau, P., Buhagiar-Labarchede, G., Onclercq-Delic, R., Lambert, S., Debatisse, M., Brison, O., and Amor-Gueret, M. (2011). Pyrimidine pool imbalance induced by BLM helicase deficiency contributes to genetic instability in Bloom syndrome. *Nature communications* *2*, 368.

Chan, K.L., Palmai-Pallag, T., Ying, S., and Hickson, I.D. (2009). Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nature cell biology* *11*, 753-760.

Cicenas, J., Kalyan, K., Sorokinas, A., Jatulyte, A., Valiunas, D., Kaupinis, A., and Valius, M. (2014). Highlights of the Latest Advances in Research on CDK Inhibitors. *Cancers* *6*, 2224-2242.

Clijsters, L., and Wolthuis, R. (2014). PIP-box-mediated degradation prohibits re-accumulation of Cdc6 during S phase. *Journal of cell science* *127*, 1336-1345.

Conroy, R.S., Koretsky, A.P., and Moreland, J. (2010). Lambda exonuclease digestion of CGG trinucleotide repeats. *European biophysics journal : EBJ* *39*, 337-343.

Coquelle, A., Pipiras, E., Toledo, F., Buttin, G., and Debatisse, M. (1997). Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell* *89*, 215-225.

Corbin, S., Neilly, M.E., Espinosa, R., 3rd, Davis, E.M., McKeithan, T.W., and Le Beau, M.M. (2002). Identification of unstable sequences within the common fragile site at 3p14.2: implications for the mechanism of deletions within fragile histidine triad gene/common fragile site at 3p14.2 in tumors. *Cancer research* *62*, 3477-3484.

Cornacchia, D., Dileep, V., Quivy, J.P., Foti, R., Tili, F., Santarella-Mellwig, R., Antony, C., Almouzni, G., Gilbert, D.M., and Buonomo, S.B. (2012). Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells. *The EMBO journal* *31*, 3678-3690.

Cortez, D., Glick, G., and Elledge, S.J. (2004). Minichromosome maintenance proteins are direct targets of the ATM and ATR checkpoint kinases. *Proceedings of the National Academy of Sciences of the United States of America* *101*, 10078-10083.

Coulombe, P., Gregoire, D., Tsanov, N., and Mechali, M. (2013). A spontaneous Cdt1 mutation in 129 mouse strains reveals a regulatory domain restraining replication licensing. *Nature communications* *4*, 2065.

Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O., and Debatisse, M. (2008). Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature* *455*, 557-560.

Daboussi, F., Courbet, S., Benhamou, S., Kannouche, P., Zdzienicka, M.Z., Debatisse, M., and Lopez, B.S. (2008). A homologous recombination defect affects replication-fork progression in mammalian cells. *Journal of cell science* *121*, 162-166.

Debatisse, M., El Achkar, E., and Dutrillaux, B. (2006). Common fragile sites nested at the interfaces of early and late-replicating chromosome bands: cis acting components of the G2/M checkpoint? *Cell Cycle* 5, 578-581.

Debatisse, M., Le Tallec, B., Letessier, A., Dutrillaux, B., and Brison, O. (2012). Common fragile sites: mechanisms of instability revisited. *Trends Genet* 28, 22-32.

Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., and Pelicci, P.G. (2013). Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome research* 23, 1-11.

Denison, S.R., Callahan, G., Becker, N.A., Phillips, L.A., and Smith, D.I. (2003). Characterization of FRA6E and its potential role in autosomal recessive juvenile parkinsonism and ovarian cancer. *Genes, chromosomes & cancer* 38, 40-52.

DePamphilis, M.L. (2005). Cell cycle dependent regulation of the origin recognition complex. *Cell Cycle* 4, 70-79.

Dereli-Oz, A., Versini, G., and Halazonetis, T.D. (2011). Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. *Molecular oncology* 5, 308-314.

Diffley, J.F., Cocker, J.H., Dowell, S.J., and Rowley, A. (1994). Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell* 78, 303-316.

Dijkwel, P.A., and Hamlin, J.L. (1988). Matrix attachment regions are positioned near replication initiation sites, genes, and an interamplicon junction in the amplified dihydrofolate reductase domain of Chinese hamster ovary cells. *Molecular and cellular biology* 8, 5398-5409.

Drury, L.S., Perkins, G., and Diffley, J.F. (1997). The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *The EMBO journal* 16, 5966-5976.

Drusco, A., Pekarsky, Y., Costinean, S., Antenucci, A., Conti, L., Volinia, S., Aqeilan, R.I., Huebner, K., and Zanesi, N. (2011). Common fragile site tumor suppressor genes and corresponding mouse models of cancer. *Journal of biomedicine & biotechnology* 2011, 984505.

Durkin, S.G., Arlt, M.F., Howlett, N.G., and Glover, T.W. (2006). Depletion of CHK1, but not CHK2, induces chromosomal instability and breaks at common fragile sites. *Oncogene* 25, 4381-4388.

Durkin, S.G., and Glover, T.W. (2007). Chromosome fragile sites. *Annual review of genetics* 41, 169-192.

El Achkar, E., Gerbault-Seureau, M., Muleris, M., Dutrillaux, B., and Debatisse, M. (2005). Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 18069-18074.

Elder, F.F., and Robinson, T.J. (1989). Rodent common fragile sites: are they conserved? Evidence from mouse and rat. *Chromosoma* *97*, 459-464.

Fang, H.Y., Lin, T.S., and Cheng, C.Y. (2000). Pancreaticocolonic fistula after extensive corrosive injury from esophagus to jejunum. *Zhonghua yi xue za zhi = Chinese medical journal; Free China ed* *63*, 77-81.

Feng, D., Tu, Z., Wu, W., and Liang, C. (2003). Inhibiting the expression of DNA replication-initiation proteins induces apoptosis in human cancer cells. *Cancer research* *63*, 7356-7364.

Ferber, M.J., Eilers, P., Schuurin, E., Fenton, J.A., Fleuren, G.J., Kenter, G., Szuhai, K., Smith, D.I., Raap, A.K., and Brink, A.A. (2004). Positioning of cervical carcinoma and Burkitt lymphoma translocation breakpoints with respect to the human papillomavirus integration cluster in FRA8C at 8q24.13. *Cancer genetics and cytogenetics* *154*, 1-9.

Ferber, M.J., Thorland, E.C., Brink, A.A., Rapp, A.K., Phillips, L.A., McGovern, R., Gostout, B.S., Cheung, T.H., Chung, T.K., Fu, W.Y., *et al.* (2003). Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. *Oncogene* *22*, 7233-7242.

Fernandez-Cid, A., Riera, A., Tognetti, S., Herrera, M.C., Samel, S., Evrin, C., Winkler, C., Gardenal, E., Uhle, S., and Speck, C. (2013). An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. *Molecular cell* *50*, 577-588.

Finnis, M., Dayan, S., Hobson, L., Chenevix-Trench, G., Friend, K., Ried, K., Venter, D., Woollatt, E., Baker, E., and Richards, R.I. (2005). Common chromosomal fragile site FRA16D mutation in cancer cells. *Human molecular genetics* *14*, 1341-1349.

Focarelli, M.L., Soza, S., Mannini, L., Paulis, M., Montecuccio, A., and Musio, A. (2009). Claspin inhibition leads to fragile site expression. *Genes, chromosomes & cancer* *48*, 1083-1090.

Fu, H., Maunakea, A.K., Martin, M.M., Huang, L., Zhang, Y., Ryan, M., Kim, R., Lin, C.M., Zhao, K., and Aladjem, M.I. (2013). Methylation of histone H3 on lysine 79 associates with a group of replication origins and helps limit DNA replication once per cell cycle. *PLoS genetics* *9*, e1003542.

Fujita, M., Yamada, C., Tsurumi, T., Hanaoka, F., Matsuzawa, K., and Inagaki, M. (1998). Cell cycle- and chromatin binding state-dependent phosphorylation of human MCM heterohexameric complexes. A role for cdc2 kinase. *The Journal of biological chemistry* 273, 17095-17101.

Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., and Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nature cell biology* 8, 358-366.

Gay, S., Lachages, A.M., Millot, G.A., Courbet, S., Letessier, A., Debatisse, M., and Brison, O. (2010). Nucleotide supply, not local histone acetylation, sets replication origin usage in transcribed regions. *EMBO reports* 11, 698-704.

Ge, X.Q., and Blow, J.J. (2010). Chk1 inhibits replication factory activation but allows dormant origin firing in existing factories. *The Journal of cell biology* 191, 1285-1297.

Geraghty, D.S., Ding, M., Heintz, N.H., and Pederson, D.S. (2000). Premature structural changes at replication origins in a yeast minichromosome maintenance (MCM) mutant. *The Journal of biological chemistry* 275, 18011-18021.

Girard-Reydet, C., Gregoire, D., Vassetzky, Y., and Mechali, M. (2004). DNA replication initiates at domains overlapping with nuclear matrix attachment regions in the xenopus and mouse c-myc promoter. *Gene* 332, 129-138.

Glover, T.W., Hoge, A.W., Miller, D.E., Ascara-Wilke, J.E., Adam, A.N., Dagenais, S.L., Wilke, C.M., Dierick, H.A., and Beer, D.G. (1998). The murine Fhit gene is highly similar to its human orthologue and maps to a common fragile site region. *Cancer research* 58, 3409-3414.

Glover, T.W., and Stein, C.K. (1987). Induction of sister chromatid exchanges at common fragile sites. *American journal of human genetics* 41, 882-890.

Gomez, M. (2008). Controlled rereplication at DNA replication origins. *Cell Cycle* 7, 1313-1314.

Gros, J., Devbhandari, S., and Remus, D. (2014). Origin plasticity during budding yeast DNA replication in vitro. *The EMBO journal* 33, 621-636.

Guilbaud, G., Rappailles, A., Baker, A., Chen, C.L., Arneodo, A., Goldar, A., d'Aubenton-Carafa, Y., Thermes, C., Audit, B., and Hyrien, O. (2011). Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS computational biology* 7, e1002322.

Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., Schwob, E., Losada, A., and Mendez, J. (2010). Cohesin organizes chromatin loops at DNA replication factories. *Genes & development* 24, 2812-2822.

Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. *Science* 319, 1352-1355.

Halder, R., Halder, K., Sharma, P., Garg, G., Sengupta, S., and Chowdhury, S. (2010). Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. *Molecular bioSystems* 6, 2439-2447.

Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences of the United States of America* 107, 139-144.

Hayano, M., Kanoh, Y., Matsumoto, S., Renard-Guillet, C., Shirahige, K., and Masai, H. (2012). Rif1 is a global regulator of timing of replication origin firing in fission yeast. *Genes & development* 26, 137-150.

Hellman, A., Rahat, A., Scherer, S.W., Darvasi, A., Tsui, L.C., and Kerem, B. (2000). Replication delay along FRA7H, a common fragile site on human chromosome 7, leads to chromosomal instability. *Molecular and cellular biology* 20, 4420-4427.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular cell* 44, 966-977.

Helmrich, A., Stout-Weider, K., Hermann, K., Schrock, E., and Heiden, T. (2006). Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome research* 16, 1222-1230.

Hendrickson, M., Madine, M., Dalton, S., and Gautier, J. (1996). Phosphorylation of MCM4 by cdc2 protein kinase inhibits the activity of the minichromosome maintenance complex. *Proceedings of the National Academy of Sciences of the United States of America* 93, 12223-12228.

Hiratani, I., and Gilbert, D.M. (2010). Autosomal lyonization of replication domains during early Mammalian development. *Advances in experimental medicine and biology* 695, 41-58.

Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D.P., Plath, K., Dalton, S., *et al.* (2010). Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome research* 20, 155-169.

Holmes, W.F., Braastad, C.D., Mitra, P., Hampe, C., Doenecke, D., Albig, W., Stein, J.L., van Wijnen, A.J., and Stein, G.S. (2005). Coordinate control and selective expression of the full complement of replication-dependent histone H4 genes in normal and cancer cells. *The Journal of biological chemistry* 280, 37400-37407.

Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., Nakatsuka, A., Kunichika, T., Obuse, C., and Waga, S. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *The Journal of biological chemistry* 288, 30161-30171.

Howlett, N.G., Taniguchi, T., Durkin, S.G., D'Andrea, A.D., and Glover, T.W. (2005). The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Human molecular genetics* 14, 693-701.

Huberman, J.A. (1968). Visualization of replicating mammalian and T4 bacteriophage DNA. *Cold Spring Harbor symposia on quantitative biology* 33, 509-524.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic acids research* 33, 2908-2916.

Hyrien, O. (2015). Peaks cloaked in the mist: the landscape of mammalian replication origins. *The Journal of cell biology* 208, 147-160.

Hyrien, O., Maric, C., and Mechali, M. (1995). Transition in specification of embryonic metazoan DNA replication origins. *Science* 270, 994-997.

Iarovaia, O.V., Bystritskiy, A., Ravcheev, D., Hancock, R., and Razin, S.V. (2004). Visualization of individual DNA loops and a map of loop domains in the human dystrophin gene. *Nucleic acids research* 32, 2079-2086.

Iizuka, M., Matsui, T., Takisawa, H., and Smith, M.M. (2006). Regulation of replication licensing by acetyltransferase Hbo1. *Molecular and cellular biology* 26, 1098-1108.

Itzhaki, J.E., Gilbert, C.S., and Porter, A.C. (1997). Construction by gene targeting in human cells of a "conditional" CDC2 mutant that rereplicates its DNA. *Nature genetics* 15, 258-265.

Jacob, F., and Brenner, S. (1963). [On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences* 256, 298-300.

Jans, D.A., Moll, T., Nasmyth, K., and Jans, P. (1995). Cyclin-dependent kinase site-regulated signal-dependent nuclear localization of the SW15 yeast transcription factor in mammalian cells. *The Journal of biological chemistry* 270, 17064-17067.



Jiang, Y., Lucas, I., Young, D.J., Davis, E.M., Karrison, T., Rest, J.S., and Le Beau, M.M. (2009). Common fragile sites are characterized by histone hypoacetylation. *Human molecular genetics* 18, 4501-4512.

Kameoka, Y., Tagawa, H., Tsuzuki, S., Karnan, S., Ota, A., Suguro, M., Suzuki, R., Yamaguchi, M., Morishima, Y., Nakamura, S., *et al.* (2004). Contig array CGH at 3p14.2 points to the FRA3B/FHIT common fragile region as the target gene in diffuse large B-cell lymphoma. *Oncogene* 23, 9148-9154.

Karnani, N., Taylor, C.M., Malhotra, A., and Dutta, A. (2010). Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Molecular biology of the cell* 21, 393-404.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. *Nature reviews Genetics* 13, 204-214.

Koundrioukoff, S., Carignon, S., Techer, H., Letessier, A., Brison, O., and Debatisse, M. (2013). Stepwise activation of the ATR signaling pathway upon increasing replication stress impacts fragile site integrity. *PLoS genetics* 9, e1003643.

Krummel, K.A., Denison, S.R., Calhoun, E., Phillips, L.A., and Smith, D.I. (2002). The common fragile site FRA16D and its associated gene WWOX are highly conserved in the mouse at Fra8E1. *Genes, chromosomes & cancer* 34, 154-167.

Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 484, 115-119.

Labib, K. (2010). How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes & development* 24, 1208-1219.

Lagarkova, M.A., Svetlova, E., Giacca, M., Falaschi, A., and Razin, S.V. (1998). DNA loop anchorage region colocalizes with the replication origin located downstream to the human gene encoding lamin B2. *Journal of cellular biochemistry* 69, 13-18.

Le Beau, M.M., Rassool, F.V., Neilly, M.E., Espinosa, R., 3rd, Glover, T.W., Smith, D.I., and McKeithan, T.W. (1998). Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Human molecular genetics* 7, 755-761.

Le Tallec, B., Dutrillaux, B., Lachages, A.M., Millot, G.A., Brison, O., and Debatisse, M. (2011). Molecular profiling of common fragile sites in human fibroblasts. *Nature structural & molecular biology* 18, 1421-1423.

Le Tallec, B., Millot, G.A., Blin, M.E., Brison, O., Dutrillaux, B., and Debatisse, M. (2013). Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep* 4, 420-428.

Lebofsky, R., and Bensimon, A. (2005). DNA replication origin plasticity and perturbed fork progression in human inverted repeats. *Molecular and cellular biology* 25, 6789-6797.

Lee, S.H., Kim, W.H., Kim, H.K., Woo, K.M., Nam, H.S., Kim, H.S., Kim, J.G., and Cho, M.H. (2001). Altered expression of the fragile histidine triad gene in primary gastric adenocarcinomas. *Biochemical and biophysical research communications* 284, 850-855.

Lenz, G., Wright, G.W., Emre, N.C., Kohlhammer, H., Dave, S.S., Davis, R.E., Carty, S., Lam, L.T., Shaffer, A.L., Xiao, W., *et al.* (2008). Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proceedings of the National Academy of Sciences of the United States of America* 105, 13520-13525.

Letessier, A., Millot, G.A., Koundrioukoff, S., Lachages, A.M., Vogt, N., Hansen, R.S., Malfoy, B., Brison, O., and Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* 470, 120-123.

Li, C.J., Vassilev, A., and DePamphilis, M.L. (2004). Role for Cdk1 (Cdc2)/cyclin A in preventing the mammalian origin recognition complex's largest subunit (Orc1) from binding to chromatin during mitosis. *Molecular and cellular biology* 24, 5875-5886.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozcy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289-293.

Liku, M.E., Nguyen, V.Q., Rosales, A.W., Irie, K., and Li, J.J. (2005). CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. *Molecular biology of the cell* 16, 5026-5039.

Limongi, M.Z., Pelliccia, F., and Rocchi, A. (2003). Characterization of the human common fragile site FRA2G. *Genomics* 81, 93-97.

Lin, Y.L., and Pasero, P. (2012). Interference between DNA replication and transcription as a cause of genomic instability. *Current genomics* 13, 65-73.

Liu, P., Slater, D.M., Lenburg, M., Nevis, K., Cook, J.G., and Vaziri, C. (2009). Replication licensing promotes cyclin D1 expression and G1 progression in untransformed human cells. *Cell Cycle* 8, 125-136.

Lopez-estrano, C., Schwartzman, J.B., Krimer, D.B., and Hernandez, P. (1998). Co-localization of polar replication fork barriers and rRNA transcription terminators in mouse rDNA. *Journal of molecular biology* 277, 249-256.

Lucas, I., Palakodeti, A., Jiang, Y., Young, D.J., Jiang, N., Fernald, A.A., and Le Beau, M.M. (2007). High-throughput mapping of origins of replication in human cells. *EMBO reports* 8, 770-777.

Lutzmann, M., Maiorano, D., and Mechali, M. (2006). A Cdt1-geminin complex licenses chromatin for DNA replication and prevents rereplication during S phase in *Xenopus*. *The EMBO journal* 25, 5764-5774.

Lutzmann, M., and Mechali, M. (2008). MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. *Molecular cell* 31, 190-200.

Ly, T., Endo, A., and Lamond, A.I. (2015). Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *eLife* 4.

MacAlpine, H.K., Gordan, R., Powell, S.K., Hartemink, A.J., and MacAlpine, D.M. (2010). *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome research* 20, 201-211.

Macheret, M., and Halazonetis, T.D. (2015). DNA replication stress as a hallmark of cancer. *Annual review of pathology* 10, 425-448.

Machida, Y.J., Teer, J.K., and Dutta, A. (2005). Acute reduction of an origin recognition complex (ORC) subunit in human cells reveals a requirement of ORC for Cdk2 activation. *The Journal of biological chemistry* 280, 27624-27630.

Mantiero, D., Mackenzie, A., Donaldson, A., and Zegerman, P. (2011). Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *The EMBO journal* 30, 4805-4814.

Manzo, S.G., Zhou, Z.L., Wang, Y.Q., Marinello, J., He, J.X., Li, Y.C., Ding, J., Capranico, G., and Miao, Z.H. (2012). Natural product triptolide mediates cancer cell death by triggering CDK7-dependent degradation of RNA polymerase II. *Cancer research* 72, 5363-5373.

Martin, M.M., Ryan, M., Kim, R., Zakas, A.L., Fu, H., Lin, C.M., Reinhold, W.C., Davis, S.R., Bilke, S., Liu, H., *et al.* (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome research* 21, 1822-1832.

McConnell, K.H., Dixon, M., and Calvi, B.R. (2012). The histone acetyltransferases CBP and Chameau integrate developmental and DNA replication programs in *Drosophila* ovarian follicle cells. *Development* 139, 3880-3890.

McCormack, J.E., Pepe, V.H., Kent, R.B., Dean, M., Marshak-Rothstein, A., and Sonenshein, G.E. (1984). Specific regulation of c-myc oncogene expression in a murine B-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* *81*, 5546-5550.

McGarry, T.J., and Kirschner, M.W. (1998). Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* *93*, 1043-1053.

McKnight, S.L., and Miller, O.L., Jr. (1977). Electron microscopic analysis of chromatin replication in the cellular blastoderm *Drosophila melanogaster* embryo. *Cell* *12*, 795-804.

Mesner, L.D., and Hamlin, J.L. (2005). Specific signals at the 3' end of the DHFR gene define one boundary of the downstream origin of replication. *Genes & development* *19*, 1053-1066.

Mesner, L.D., Valsakumar, V., Cieslik, M., Pickin, R., Hamlin, J.L., and Bekiranov, S. (2013). Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome research* *23*, 1774-1788.

Mesner, L.D., Valsakumar, V., Karnani, N., Dutta, A., Hamlin, J.L., and Bekiranov, S. (2011). Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. *Genome research* *21*, 377-389.

Mimori, K., Druck, T., Inoue, H., Alder, H., Berk, L., Mori, M., Huebner, K., and Croce, C.M. (1999). Cancer-specific chromosome alterations in the constitutive fragile region FRA3B. *Proceedings of the National Academy of Sciences of the United States of America* *96*, 7456-7461.

Mimura, S., Seki, T., Tanaka, S., and Diffley, J.F. (2004). Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. *Nature* *431*, 1118-1123.

Minocherhomji, S., and Hickson, I.D. (2014). Structure-specific endonucleases: guardians of fragile site stability. *Trends in cell biology* *24*, 321-327.

Miotto, B., and Struhl, K. (2008). HBO1 histone acetylase is a coactivator of the replication licensing factor Cdt1. *Genes & development* *22*, 2633-2638.

Mishmar, D., Rahat, A., Scherer, S.W., Nyakatura, G., Hinzmann, B., Kohwi, Y., Mandel-Gutfroind, Y., Lee, J.R., Drescher, B., Sas, D.E., *et al.* (1998). Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. *Proceedings of the National Academy of Sciences of the United States of America* *95*, 8141-8146.

Mrasek, K., Schoder, C., Teichmann, A.C., Behr, K., Franze, B., Wilhelm, K., Blaurock, N., Claussen, U., Liehr, T., and Weise, A. (2010). Global screening and extended nomenclature for 230 aphidicolin-inducible fragile sites, including 61 yet unreported ones. *International journal of oncology* *36*, 929-940.

Mukhopadhyay, R., Lajugie, J., Fourel, N., Selzer, A., Schizas, M., Bartholdy, B., Mar, J., Lin, C.M., Martin, M.M., Ryan, M., *et al.* (2014). Allele-specific genome-wide profiling in human primary erythroblasts reveal replication program organization. *PLoS genetics* *10*, e1004319.

Muramatsu, S., Hirai, K., Tak, Y.S., Kamimura, Y., and Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon), and GINS in budding yeast. *Genes & development* *24*, 602-612.

Naim, V., and Rosselli, F. (2009). The FANC pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities. *Nature cell biology* *11*, 761-768.

Naim, V., Wilhelm, T., Debatisse, M., and Rosselli, F. (2013). ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nature cell biology* *15*, 1008-1015.

Negrini, M., Monaco, C., Vorechovsky, I., Ohta, M., Druck, T., Baffa, R., Huebner, K., and Croce, C.M. (1996). The FHIT gene at 3p14.2 is abnormal in breast carcinomas. *Cancer research* *56*, 3173-3179.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability--an evolving hallmark of cancer. *Nature reviews Molecular cell biology* *11*, 220-228.

Nevis, K.R., Cordeiro-Stone, M., and Cook, J.G. (2009). Origin licensing and p53 status regulate Cdk2 activity during G(1). *Cell Cycle* *8*, 1952-1963.

Oehlmann, M., Score, A.J., and Blow, J.J. (2004). The role of Cdc6 in ensuring complete genome licensing and S phase checkpoint activation. *The Journal of cell biology* *165*, 181-190.

Ohta, M., Inoue, H., Cotticelli, M.G., Kastury, K., Baffa, R., Palazzo, J., Siprashvili, Z., Mori, M., McCue, P., Druck, T., *et al.* (1996). The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell* *84*, 587-597.

Ozeri-Galai, E., Lebofsky, R., Rahat, A., Bester, A.C., Bensimon, A., and Kerem, B. (2011). Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Molecular cell* *43*, 122-131.

Palakodeti, A., Han, Y., Jiang, Y., and Le Beau, M.M. (2004). The role of late/slow replication of the FRA16D in common fragile site induction. *Genes, chromosomes & cancer* 39, 71-76.

Palumbo, E., Matricardi, L., Tosoni, E., Bensimon, A., and Russo, A. (2010). Replication dynamics at common fragile site FRA6E. *Chromosoma* 119, 575-587.

Pasic, I., Shlien, A., Durbin, A.D., Stavropoulos, D.J., Baskin, B., Ray, P.N., Novokmet, A., and Malkin, D. (2010). Recurrent focal copy-number changes and loss of heterozygosity implicate two noncoding RNAs and one tumor suppressor gene at chromosome 3q13.31 in osteosarcoma. *Cancer research* 70, 160-171.

Petermann, E., Helleday, T., and Caldecott, K.W. (2008). Claspin promotes normal replication fork rates in human cells. *Molecular biology of the cell* 19, 2373-2378.

Petermann, E., Maya-Mendoza, A., Zachos, G., Gillespie, D.A., Jackson, D.A., and Caldecott, K.W. (2006). Chk1 requirement for high global rates of replication fork progression during normal vertebrate S phase. *Molecular and cellular biology* 26, 3319-3326.

Petersen, B.O., Lukas, J., Sorensen, C.S., Bartek, J., and Helin, K. (1999). Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. *The EMBO journal* 18, 396-410.

Petersen, B.O., Wagener, C., Marinoni, F., Kramer, E.R., Melixetian, M., Lazzarini Denchi, E., Gieffers, C., Matteucci, C., Peters, J.M., and Helin, K. (2000). Cell cycle- and cell growth-regulated proteolysis of mammalian CDC6 is dependent on APC-CDH1. *Genes & development* 14, 2330-2343.

Picard, F., Cadoret, J.C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L., and Prioleau, M.N. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS genetics* 10, e1004282.

Ragland, R.L., Glynn, M.W., Arlt, M.F., and Glover, T.W. (2008). Stably transfected common fragile site sequences exhibit instability at ectopic sites. *Genes, chromosomes & cancer* 47, 860-872.

Ried, K., Finnis, M., Hobson, L., Mangelsdorf, M., Dayan, S., Nancarrow, J.K., Woollatt, E., Kremmidiotis, G., Gardner, A., Venter, D., *et al.* (2000). Common chromosomal fragile site FRA16D sequence: identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. *Human molecular genetics* 9, 1651-1663.

Rozier, L., El-Achkar, E., Apiou, F., and Debatisse, M. (2004). Characterization of a conserved aphidicolin-sensitive common fragile site at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. *Oncogene* 23, 6872-6880.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome research* 20, 761-770.

Saha, S., Shan, Y., Mesner, L.D., and Hamlin, J.L. (2004). The promoter of the Chinese hamster ovary dihydrofolate reductase gene regulates the activity of the local origin and helps define its boundaries. *Genes & development* 18, 397-410.

Savelyeva, L., Sagulenko, E., Schmitt, J.G., and Schwab, M. (2006). The neurobeachin gene spans the common fragile site FRA13A. *Human genetics* 118, 551-558.

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., *et al.* (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes & development* 22, 2048-2061.

Schwartz, M., Zlotorynski, E., and Kerem, B. (2006). The molecular basis of common and rare fragile sites. *Cancer letters* 232, 13-26.

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS genetics* 5, e1000446.

Shen, Z., Chakraborty, A., Jain, A., Giri, S., Ha, T., Prasanth, K.V., and Prasanth, S.G. (2012). Dynamic association of ORCA with prereplicative complex components regulates DNA replication initiation. *Molecular and cellular biology* 32, 3107-3120.

Shridhar, R., Shridhar, V., Wang, X., Paradee, W., Dugan, M., Sarkar, F., Wilke, C., Glover, T.W., Vaitkevicius, V.K., and Smith, D.I. (1996). Frequent breakpoints in the 3p14.2 fragile site, FRA3B, in pancreatic tumors. *Cancer research* 56, 4347-4350.

Siddiqui, K., On, K.F., and Diffley, J.F. (2013). Regulating DNA replication in eukarya. *Cold Spring Harbor perspectives in biology* 5.

Sozzi, G., Veronese, M.L., Negrini, M., Baffa, R., Cotticelli, M.G., Inoue, H., Tornielli, S., Pilotti, S., De Gregorio, L., Pastorino, U., *et al.* (1996). The FHIT gene 3p14.2 is abnormal in lung cancer. *Cell* 85, 17-26.

Speck, C., Chen, Z., Li, H., and Stillman, B. (2005). ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nature structural & molecular biology* 12, 965-971.

Stinchcomb, D.T., Struhl, K., and Davis, R.W. (1979). Isolation and characterisation of a yeast chromosomal replicator. *Nature* 282, 39-43.

Sugimoto, N., Tatsumi, Y., Tsurumi, T., Matsukage, A., Kiyono, T., Nishitani, H., and Fujita, M. (2004). Cdt1 phosphorylation by cyclin A-dependent kinases negatively regulates its function without affecting geminin binding. *The Journal of biological chemistry* 279, 19691-19697.

Swarnalatha, M., Singh, A.K., and Kumar, V. (2012). The epigenetic control of E-box and Myc-dependent chromatin modifications regulate the licensing of lamin B2 origin during cell cycle. *Nucleic acids research* 40, 9021-9035.

Takeda, D.Y., Shibata, Y., Parvin, J.D., and Dutta, A. (2005). Recruitment of ORC or CDC6 to DNA is sufficient to create an artificial origin of replication in mammalian cells. *Genes & development* 19, 2827-2836.

Tanaka, S., and Diffley, J.F. (2002). Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nature cell biology* 4, 198-207.

Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., and Araki, H. (2011). Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Current biology : CB* 21, 2055-2063.

Tardat, M., Brustel, J., Kirsh, O., Lefevbre, C., Callanan, M., Sardet, C., and Julien, E. (2010). The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nature cell biology* 12, 1086-1093.

Taylor, J.H. (1977). Increase in DNA replication sites in cells held at the beginning of S phase. *Chromosoma* 62, 291-300.

Techer, H., Koundrioukoff, S., Azar, D., Wilhelm, T., Carignon, S., Brison, O., Debatisse, M., and Le Tallec, B. (2013). Replication dynamics: biases and robustness of DNA fiber analysis. *Journal of molecular biology* 425, 4845-4855.

Thorland, E.C., Myers, S.L., Gostout, B.S., and Smith, D.I. (2003). Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene* 22, 1225-1237.

Toledo, L.I., Altmeyer, M., Rask, M.B., Lukas, C., Larsen, D.H., Povlsen, L.K., Bekker-Jensen, S., Mailand, N., Bartek, J., and Lukas, J. (2013). ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* 155, 1088-1103.

Torres-Rosell, J., De Piccoli, G., Cordon-Preciado, V., Farmer, S., Jarmuz, A., Machin, F., Pasero, P., Lisby, M., Haber, J.E., and Aragon, L. (2007). Anaphase onset before complete DNA replication with intact checkpoint responses. *Science* 315, 1411-1415.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., *et al.* (2009). Topoisomerase I suppresses genomic



instability by preventing interference between replication and transcription. *Nature cell biology* *11*, 1315-1324.

Valenzuela, M.S., Chen, Y., Davis, S., Yang, F., Walker, R.L., Bilke, S., Lueders, J., Martin, M.M., Aladjem, M.I., Massion, P.P., *et al.* (2011). Preferential localization of human origins of DNA replication at the 5'-ends of expressed genes and at evolutionarily conserved DNA sequences. *PLoS one* *6*, e17308.

Valton, A.L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintome, C., Riou, J.F., and Prioleau, M.N. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *The EMBO journal* *33*, 732-746.

Vassilev, L.T., Tovar, C., Chen, S., Knezevic, D., Zhao, X., Sun, H., Heimbrook, D.C., and Chen, L. (2006). Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 10660-10665.

Virgilio, L., Shuster, M., Gollin, S.M., Veronese, M.L., Ohta, M., Huebner, K., and Croce, C.M. (1996). FHIT gene alterations in head and neck squamous cell carcinomas. *Proceedings of the National Academy of Sciences of the United States of America* *93*, 9770-9775.

Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J., and Grunstein, M. (2002). Histone acetylation regulates the time of replication origin firing. *Molecular cell* *10*, 1223-1233.

Walter, J.C. (2000). Evidence for sequential action of cdc7 and cdk2 protein kinases during initiation of DNA replication in *Xenopus* egg extracts. *The Journal of biological chemistry* *275*, 39773-39778.

Wang, L., Darling, J., Zhang, J.S., Huang, H., Liu, W., and Smith, D.I. (1999). Allele-specific late replication and fragility of the most active common fragile site, FRA3B. *Human molecular genetics* *8*, 431-437.

Yamazaki, S., Ishii, A., Kanoh, Y., Oda, M., Nishito, Y., and Masai, H. (2012). Rif1 regulates the replication timing domains on the human genome. *The EMBO journal* *31*, 3667-3677.

Yang, S.C., Rhind, N., and Bechhoefer, J. (2010). Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Molecular systems biology* *6*, 404.

Ying, S., Minocherhomji, S., Chan, K.L., Palmai-Pallag, T., Chu, W.K., Wass, T., Mankouri, H.W., Liu, Y., and Hickson, I.D. (2013). MUS81 promotes common fragile site expression. *Nature cell biology* *15*, 1001-1007.

Yoshida, K., Bacal, J., Desmarais, D., Padioleau, I., Tsaponina, O., Chabes, A., Pantesco, V., Dubois, E., Parrinello, H., Skrzypczak, M., *et al.* (2014). The histone deacetylases sir2 and rpd3 act on ribosomal DNA to control the replication program in budding yeast. *Molecular cell* 54, 691-697.

Zegerman, P., and Diffley, J.F. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445, 281-285.

Zhang, H., and Freudenreich, C.H. (2007). An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Molecular cell* 27, 367-379.