



HAL
open science

Etudes des translocations chromosomiques en utilisant les méthodes d'édition du génome : des mécanismes moléculaires à l'oncogenèse

Loélia Babin

► **To cite this version:**

Loélia Babin. Etudes des translocations chromosomiques en utilisant les méthodes d'édition du génome : des mécanismes moléculaires à l'oncogenèse. Cancer. Université Paris Saclay (COMUE), 2019. Français. NNT : 2019SACLS240 . tel-03510173

HAL Id: tel-03510173

<https://theses.hal.science/tel-03510173v1>

Submitted on 4 Jan 2022

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.

Etudes des translocations chromosomiques en utilisant les méthodes d'édition du génome : des mécanismes moléculaires à l'oncogenèse

Thèse de doctorat de l'Université Paris-Saclay
préparée à L'institut Imagine

École doctorale n°582 Cancérologie : biologie - médecine - santé
(CBMS)
Spécialité de doctorat: Aspects moléculaires et cellulaires de la biologie

Thèse présentée et soutenue à Paris, le 27-09-2019, par

Mme Loélia BABIN

Composition du Jury :

Dr Gaëlle LEGUBE Directrice de recherche, CBI Toulouse (– UMR 5088)	Rapporteuse Présidente
Dr Didier SURDEZ Ingénieur de recherche, Institut Curie Paris (– U830) Dr Camille LOBRY Chargé de recherche, IGR Villejuif (– UMR 1170)	Rapporteur Examineur
Dr Patricia KANNOUCHE Directrice de recherche, IGR Villejuif (– UMR 8200) Dr Erika BRUNET Directrice de recherche, Institut Imagine Paris (– U1163)	Examinatrice Directrice de thèse

imagine
INSTITUT DES MALADIES GÉNÉTIQUES



Table des matières

INTRODUCTION	7
I. Les translocations chromosomiques	8
A. Mécanismes de formation.....	10
1. Mécanismes de surveillance du génome et de réparation	10
2. Des mécanismes différents chez l'Homme et la Souris.....	12
B. Mécanismes favorisant l'apparition des translocations.....	13
1. La localisation nucléaire	13
2. La chromoplexie	14
3. Translocations induites par RAG/AID	16
4. Translocations et sites fragiles	19
5. Translocations et traitement aux inhibiteurs de topoisomérase II.....	20
C. Le rôle des translocations dans l'oncogenèse.....	21
1. La dérégulation d'oncogènes ou gènes suppresseurs de tumeurs (Figure 4 A)	21
2. La création d'un oncogène de fusion (Figure 5B).....	22
D. Le paradoxe des translocations et de leur rôle dans l'oncogenèse	24
1. Incidence des translocations dans les cancers.....	24
2. Addiction aux oncogènes de fusion	25
3. Les translocations retrouvées chez des individus sains	25
E. La modélisation des translocations.....	27
1. Surexpression du gène de fusion : un modèle limité	28
2. Utilisation des nucléases artificielles afin de générer des translocations chromosomiques	29
a) Les ZFN (Nucléase à doigt de Zinc)	29
b) Les TALEN (Transcription Activator-Like Effector Nucleases)	30
c) Le système CRISPR/Cas9.....	31
d) La question de la spécificité	34
II. La translocation t(2;5)(p23,q35) comme modèle d'étude	37
A. Aspects moléculaires.....	37
1. Nucléophosmine 1 (NPM1)	37
2. Anaplastic Lymphoma Kinase (ALK).....	39
3. La protéine de fusion NPM1-ALK	40
4. Potentiel oncogénique de NPM1-ALK	43

B.	Lymphome Anaplasique à grandes cellules (ALCL)	43
1.	Présentation générale	43
2.	Histologie.....	44
3.	Réarrangements chromosomiques présents dans les ALCL ALK+.....	46
4.	Le statut p53 des ALCL ALK+.....	46
5.	Une origine cellulaire controversée	47
6.	Signatures moléculaires	49
a)	Profil transcriptionnel.....	49
b)	Les ARN non codants.....	50
7.	Pronostic et traitements.....	53
8.	Les modèles de l'ALCL <i>in vivo</i>	54
III.	Les ARN circulaires	56
A.	Synthèse des ARN circulaires	56
B.	Détection des ARN circulaires	59
C.	Les différentes formes d'ARN circulaires	60
D.	Rôles des ARN circulaires (Figure 15).....	61
E.	ARN circulaires et cancers	64
F.	ARN circulaires et translocations.....	67
G.	Rôles des ARN circulaires dans d'autres processus cellulaires	68
IV.	Le timing de réplication	70
A.	Timing de réplication, ADN et chromatine.....	70
1.	Régions AT/GC rich.....	71
2.	Marques histones.....	71
3.	Localisation nucléaire.....	72
4.	Organisation 3D.....	72
5.	Timing de réplication et expression génique	75
B.	Timing de réplication et stabilité du génome	78
C.	Timing de réplication et translocation	79
D.	Timing de réplication et pathologies associées.....	80
1.	Timing de réplication et cancer.....	80
2.	Autres pathologies	81
	PROJET DE THESE	83
	RESULTATS.....	85
I.	Etude des ARN circulaires de fusion (Article 1, publié)	86

II.	Etude du timing de réplication (Article 2, en cours de préparation)	118
A.	Introduction.....	119
B.	Results	121
C.	Conclusion/Discussion.....	128
D.	Material and methods.....	130
E.	Figures legends.....	134
F.	Supplementary Data.....	139
G.	Bibliography.....	146
III.	Résultats préliminaires : induction de la translocation dans des cellules T primaires	150
	DISCUSSION/PERSPECTIVES	158
I.	Induction de la translocation t(2,5) dans des lignées cellulaires humaines et murines	159
II.	Etude des ARN circulaires de fusion	160
A.	Quels sont les mécanismes d'action et le rôle de ces ARN circulaires de fusion ?	161
B.	Les ARN circulaires de fusion identifiés sont-ils présents chez les patients ?.....	162
C.	Identification des ARN circulaires et problèmes d'artefacts de PCR.....	162
III.	Analyse du timing de réplication et du profil transcriptionnel	163
A.	Une signature transcriptionnelle proche de celle des patients ALCL.....	164
B.	Comment expliquer les modifications de timing de réplication à la jonction ?.....	165
C.	La question de la synchronicité de réplication des chromosomes homologues.....	166
D.	Comment expliquer les modifications de timing de réplication observées à l'échelle du génome ?	167
E.	Lien entre timing de réplication, marques épigénétiques et organisation de la chromatine.	168
F.	Modification du timing de réplication et différenciation cellulaire	168
IV.	Vers un modèle cellulaire plus physiologique	169
A.	Caractérisation cellulaire.....	170
B.	Potentiel tumoral	170
C.	Existence d'un effet donneur spécifique.....	171
D.	Analyse du timing de réplication.....	172
E.	Vers l'origine cellulaire de l'ALCL	172
V.	Réflexion personnelle générale	174
	BIBLIOGRAPHIE	176
	ARTICLES EN COLLABORATION	211

REMERCIEMENTS

Je tiens à remercier les Docteurs Gaëlle Legube et Didier Surdez qui ont accepté d'être rapporteurs de ma thèse. J'espère que ce manuscrit sera à la hauteur de vos attentes. Je remercie également les Docteurs Patricia Kannouche et Camille Lobry qui ont accepté d'être membres de mon jury de thèse.

Je remercie la fondation ARC pour la Recherche contre le Cancer de m'avoir financé durant ma dernière année de thèse.

*Je tiens à remercier tout particulièrement le Docteur Erika Brunet, qui m'a accueillie dans son équipe et m'a encadré tout au long de ces quatre années de thèse. Un immense merci pour ton soutien, aussi bien scientifique qu'humain, ta bienveillance, ta patience et ta confiance. J'ai appris énormément à tes côtés durant ces quatre ans. Tu as toujours été disponible et à l'écoute, et tu as toujours su m'aider à trouver des solutions lorsque les choses ne fonctionnaient pas « comme prévu ». Merci également de m'avoir fait confiance sur ce projet tout en me guidant dans mes travaux de recherche. Tu es une encadrante extraordinaire !
Merci beaucoup !*

Je souhaite également remercier les membres du laboratoire « Structure et instabilité des génomes » au Muséum Nationale d'Histoire Naturelle de Paris, où j'ai débuté ma thèse. Un grand merci au Docteur Carine Giovannangeli qui m'a accueillie dans son laboratoire et m'a apporté de précieux conseils. Merci également à Anne De Cian pour son aide et ses conseils. Merci à Loïc Ponger et Evelyne Duvernois pour leur aide sur les analyse bio-informatiques. Je tiens à remercier particulièrement Marion Piganeau qui m'a fait découvrir le laboratoire lors de mon arrivée en M2 et avec qui j'ai appris énormément de choses. Merci pour tous les bons moments passés ensemble.

Un grand merci à l'équipe « Dynamique du génome et système immunitaire » qui m'a accueillie durant mes deux dernières années de thèse. Merci aux Docteurs Jean-Pierre de Villartay et Patrick Revy pour leur accueil, leur disponibilité et tous leurs conseils. Un merci particulier à

mes collègues de bureau, Anna, Marie et Laëtitia grâce à qui venir au labo est un vrai plaisir. Un grand merci également à Alice, qui m'a énormément aidée sur les dernières manip. Merci pour ta motivation, ton efficacité et ta gentillesse.

Un grand merci également à nos collaborateurs, les Docteurs Fabienne Meggetto, Thomas Mercher et Ludovic Deriano qui m'ont apporté de nombreux conseils, et m'ont aidée à avancer dans mon projet. Merci également au Dr Maria Jasin avec qui j'ai eu la chance de collaborer.

Je souhaite également remercier l'ensemble des personnes avec qui j'ai eu la chance de travailler durant ces quatre ans. Merci Khadija pour ta précieuse aide sur les ARN circulaires, merci à Armel et Sonia qui m'ont fait découvrir les joies de la mitochondrie. Merci aux doctorants et post-doctorants que j'ai eu l'occasion de croiser, Hind, Astrid, Laurianne, Marine, Benjamin, Sheldom, Gabriel, Lorena, Stefania, Marta, Maname, Aurélie, Benoit F et Benoit R, Satish... Travailler à vos côtés a été un réel plaisir.

Pour finir je souhaite dire un immense merci à ma famille. Merci à Clément qui m'a épaulée durant toute ma thèse. Merci à mes deux sœurs ainsi qu'à mes grands-parents pour leur soutien. Un merci tout particulier à mes parents qui m'ont toujours accompagnée et soutenue, sans qui je n'en serais jamais arrivée là aujourd'hui. Merci d'être à mes côtés depuis toujours.

ABBREVIATIONS

<i>4sU</i>	<i>4-thiouridine</i>
<i>ADN</i>	<i>Acide désoxyribonucléique</i>
<i>ALCL</i>	<i>Anaplastic large cell lymphoma</i>
<i>ALL</i>	<i>Acute lymphoblastic leukemia</i>
<i>AML</i>	<i>Acute myeloïd leukemia</i>
<i>ARN</i>	<i>Acide ribonucléique</i>
<i>ARNg</i>	<i>ARN guide</i>
<i>ARNm</i>	<i>ARN messenger</i>
<i>CAR-T</i>	<i>Chimeric antigen receptor (car) t-cell</i>
<i>CAS9</i>	<i>CRISPR associated protein 9</i>
<i>CNV</i>	<i>Copy Number Variant</i>
<i>CRISPR</i>	<i>Clustered Regularly Interspaced Short Repeats</i>
<i>DSB</i>	<i>Double-strand break</i>
<i>EBV</i>	<i>Epstein Barr Virus</i>
<i>ERCE</i>	<i>Early Replication Control Elements</i>
<i>FISH</i>	<i>Fluorescent in situ hybridization</i>
<i>HR</i>	<i>Homologous recombination</i>
<i>kDa</i>	<i>Kilo dalton</i>
<i>KO</i>	<i>Knock-out</i>
<i>LAD</i>	<i>Lamin Associated Domain</i>
<i>NHEJ</i>	<i>Non-homologous end-joining</i>
<i>PAM</i>	<i>Protospacer Adjacent Motif</i>
<i>PCA</i>	<i>Principal Component Analysis</i>
<i>PCR</i>	<i>Polymerase chain reaction</i>
<i>PHA</i>	<i>Phytohemmaglutinine</i>
<i>ROS</i>	<i>Reactive oxygen species</i>
<i>RSS</i>	<i>Recombination Signal Sequences</i>
<i>RT-PCR</i>	<i>Reverse Transcriptase PCR</i>
<i>RVD</i>	<i>Repeat Variable di-Residues</i>
<i>snoRNA</i>	<i>Small nucleolar RNA</i>
<i>snRNP</i>	<i>Small nuclear ribonucleoprotein</i>
<i>TAD</i>	<i>Topologically Associated Domains</i>
<i>TALE</i>	<i>Transcription Activator-Like Effector</i>
<i>TALEN</i>	<i>Transcription Activator-Like Effector Nucleases</i>
<i>TSS</i>	<i>Transcription Starting Sites</i>
<i>WHO</i>	<i>World Health Organization</i>
<i>WT</i>	<i>Wild-type</i>
<i>ZFN</i>	<i>Zinc finger nucleases</i>

INTRODUCTION

I. Les translocations chromosomiques

Les translocations chromosomiques sont caractérisées par un échange de matériel génétique entre deux chromosomes hétérologues. Historiquement le premier cas de translocation, observé chez un patient atteint du syndrome de Down, a été identifié en 1960. Il s'agissait d'une translocation faisant intervenir les chromosomes 15 et 21 (Penrose et al., 1960). La même année Peter Nowell et David Hungerford ont identifié la première translocation chromosomique récurrente dans le cancer : le chromosome de Philadelphie impliqué dans les leucémies myéloïdes chroniques (Nowell and Hungerford, 1960). Il s'agit d'une translocation entre les chromosomes 9 et 22.

Les translocations chromosomiques peuvent être classées en deux catégories (**Figure 1**) :

- Les translocations réciproques, qui entraînent l'apparition de deux chromosomes dérivatifs mais conservent généralement l'ensemble de l'information génétique.
- Les translocations Robertsoniennes, qui mènent à une perte de matériel génétique. Ce type de translocation apparaît entre deux chromosomes acrocentriques (13, 14, 15, 21 et 22). La cassure a lieu dans les régions juxta-centromériques et entraîne une fusion centrique. Le dérivatif formé des bras courts des chromosomes est généralement perdu au cours des divisions cellulaires. Les individus porteurs de ce type de translocation présentent donc un caryotype à 45 chromosomes.

Ces translocations peuvent parfois être formées au cours de la gamétogenèse. Il a été montré que la plupart des translocations réciproques apparaissent sur les chromosomes paternels au cours de la spermatogenèse (Thomas et al., 2010). A l'inverse les translocations Robertsoniennes ont une origine majoritairement maternelle au cours de l'ovogenèse (Page and Shaffer, 1997) et sont transmissibles à la descendance, augmentant notamment les risques de trisomie (Boue and Gallano, 1984).

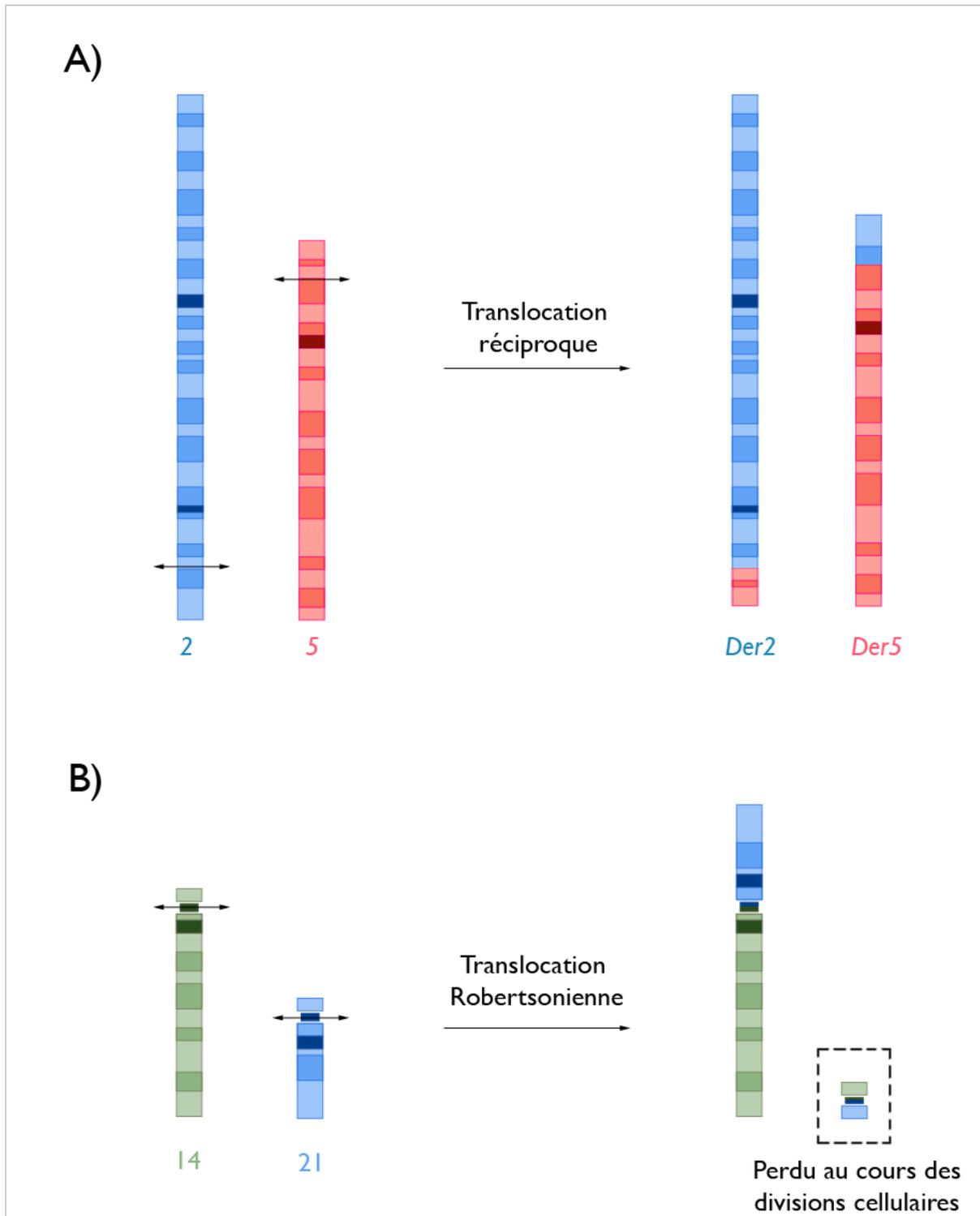


Figure 1 : Translocations chromosomiques. A) exemple d'une translocation réciproque entre les chromosomes 2 et 5 menant à la formation de deux chromosomes dérivés Der2 et Der5, sans perte de matériel génétique. B) exemple d'une translocation Robertsonienne entre les chromosomes acrocentriques 14 et 21. Le chromosome formé des bras courts des chromosomes est généralement perdu au cours des divisions cellulaires.

A. Mécanismes de formation

1. Mécanismes de surveillance du génome et de réparation

L'organisme est constamment soumis à des stress cellulaires d'origine endogène et exogène. Ces stress peuvent entraîner des altérations de l'ADN et générer une instabilité génétique. La synthèse de ROS est par exemple à l'origine de cassures simple et double-brin de l'ADN dans la cellule (Karanjawala et al., 2002). Il a également été montré que des conflits réplication/transcription pouvaient générer de l'instabilité génétique (Hoffman et al., 2015). Des facteurs exogènes tels que les rayonnements ionisants, ou certains agents chimiques peuvent également favoriser les dommages à l'ADN. La présence de plusieurs cassures double-brin de l'ADN (DSB) simultanées peut mener à l'apparition de translocations chromosomiques lorsqu'elles ne sont pas réparées correctement (Richardson and Jasin, 2000).

Dans la cellule, il existe des mécanismes de surveillance de l'intégrité du génome permettant de détecter les anomalies présentes sur l'ADN. L'apparition de cassures double-brin entraîne une activation des kinases ATM et ATR permettant d'initier le processus de réparation (Lee and Paull, 2005). L'activation de ces kinases s'accompagne d'une cascade de phosphorylations de plus de 900 protéines cibles (Matsuoka et al., 2007). Dans les cellules mammifères, la réparation des cassures double-brin est effectuée par deux voies principales : la recombinaison homologue (HR) et le « *Non Homologous End-Joining* » (NHEJ) (**Figure 2**). Le choix entre ces deux mécanismes est notamment lié à la phase du cycle cellulaire. En effet, la recombinaison homologue n'a lieu qu'en fin de phase S et en G2, alors que le NHEJ est actif tout au long du cycle cellulaire (Hustedt and Durocher, 2016). La recombinaison homologue nécessite une première étape de résection des extrémités, suivie d'une étape de réparation utilisant la chromatide sœur comme matrice. Ce mécanisme est extrêmement fidèle (Johnson and Jasin, 2000). Le NHEJ consiste quant à lui en une re-ligation simple des extrémités avec ou sans utilisation de micro-homologies (NHEJ classique ou alternatif) (Lieber, 2010). Ce mécanisme a été décrit comme moins fidèle, bien que cette observation reste controversée (Betermier et al., 2014; Waters et al., 2014). Généralement, la réparation via le NHEJ entraîne une ligation correcte des deux extrémités d'ADN. Cependant, dans de rares cas, la ligation a

lieu entre deux extrémités appartenant à des chromosomes hétérologues, ce qui aboutit à la formation de translocations chromosomiques (Ghezraoui et al., 2014; Weinstock et al., 2007).

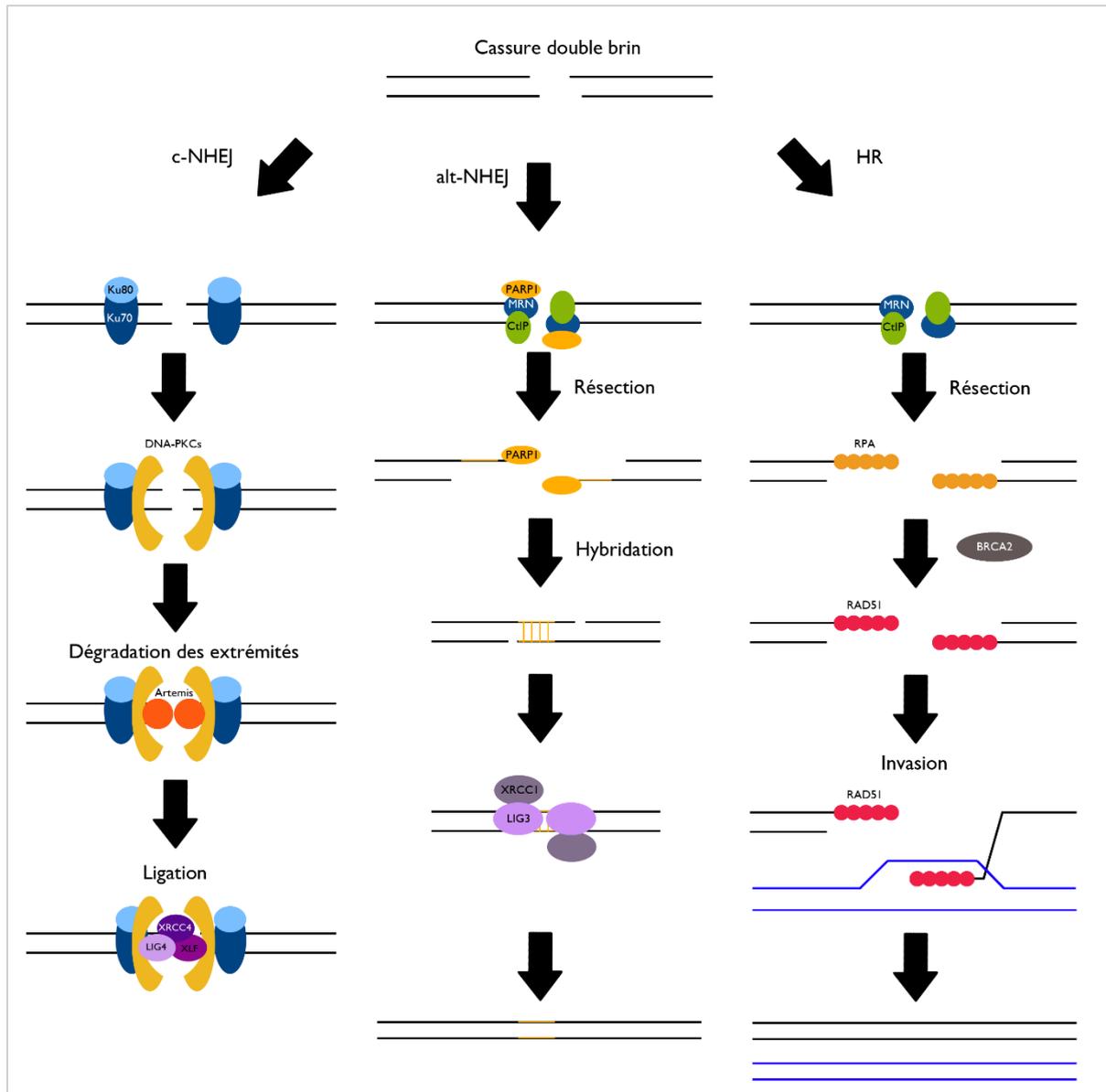


Figure 2 : Mécanismes de réparation des cassures double-brin de l'ADN. A gauche le NHEJ classique entraîne une re-ligation simple des extrémités. Au centre le NHEJ alternatif utilise les micro-homologies à distance pour réparer. A droite la recombinaison homologue entraîne une résection des extrémités suivie d'une réparation par copie de la chromatide sœur.

Les systèmes de surveillance et de réparation de l'ADN assurent le maintien de l'intégrité génomique. Une altération de ces mécanismes peut être à l'origine d'une augmentation des dommages de l'ADN, pouvant mener à la formation de translocations chromosomiques. Des modèles murins déficients en ATM ont par exemple mis en évidence une augmentation de la

fréquence de translocations chromosomiques et de lymphomes associés (Callen et al., 2007; Camacho et al., 2002; Vaites et al., 2014). De plus, chez les patients atteints d'Ataxie Téléngectiasique, caractérisée par un défaut en ATM, le nombre de translocations chromosomiques est fortement augmenté (Kojis et al., 1991; Kojis et al., 1989).

2. Des mécanismes différents chez l'Homme et la Souris

Des études sur les mécanismes de formation des translocations chromosomiques ont mis en évidence des différences majeures entre l'Homme et la Souris. En effet il a été montré que la formation de translocations chez la souris était dépendante du NHEJ alternatif (Alt-NHEJ), et nécessitait la présence de la ligase 3 (Simsek et al., 2011), de la polymérase Théta (Mateos-Gomez et al., 2015) et de de l'enzyme de résection CtIP (Zhang and Jasin, 2011). A l'inverse chez l'Homme l'analyse des jonctions révèle une faible maturation des extrémités, avec peu de micro-homologies, indiquant une faible implication du Alt-NHEJ (Brunet et al., 2009; Ghezraoui et al., 2014). La perte de la ligase 3 et de CtIP n'affecte pas le taux de translocations dans les cellules humaines, à l'inverse de la perte de la ligase 4 et de XRCC4, deux acteurs majeurs du c-NHEJ (Ghezraoui et al., 2014). En absence des facteurs du c-NHEJ la formation de translocations est alors médiée par le Alt-NHEJ avec une augmentation des grandes délétions à la jonction, et de l'utilisation de micro-homologies.

Plus récemment, une équipe a étudié le rôle potentiel d'autres protéines du système de réparation dans la formation de translocations chromosomiques. En utilisant un système rapporteur qui permet de suivre la formation de translocation dans des lignées cellulaires humaines, ils ont montré que la protéine RAD50 limitait les réarrangements chromosomiques, alors que les protéines 53BP1, PARP3 et DDB1 les favorisaient (Day et al., 2017).

B. Mécanismes favorisant l'apparition des translocations

Les translocations chromosomiques sont des réarrangements rares de l'ADN. Cependant il est aujourd'hui démontré que certaines conditions permettent d'augmenter la fréquence de ces événements.

1. La localisation nucléaire

Il est connu depuis de nombreuses années que l'organisation de la chromatine dans le noyau n'est pas aléatoire et forme des territoires chromosomiques (Cremer et al., 1982). Chez l'Homme il est globalement admis qu'il existe une organisation radiale dépendante de la densité en gènes. Les régions riches en gènes sont majoritairement centrales, alors que les régions pauvres en gènes sont localisées en périphérie (Croft et al., 1999). Cette organisation nucléaire présente une certaine plasticité. En effet une étude sur six types cellulaires différents chez la souris a montré une organisation nucléaire spécifique à chaque type cellulaire (Mayer et al., 2005). Des observations similaires ont été observées lors de la comparaison de fibroblastes humains quiescents ou proliférant (Bridger et al., 2000). Enfin, au cours de la différenciation, des modifications de l'organisation nucléaire ont été constatées. Ces changements ont notamment été observés lors de la différenciation des adipocytes humains (Kuroda et al., 2004) ou encore des lymphocytes T (Kim et al., 2004).

L'observation de translocations récurrentes chez les patients atteints de cancers a mené à l'hypothèse d'un rôle de la localisation spatiale dans le choix des partenaires impliqués dans les translocations. Des études utilisant un marquage fluorescent des loci d'intérêt ont mis en évidence la proximité des partenaires pour différentes translocations. Par exemple, il a été montré que les loci *BCR* et *ABL*, fréquemment transloqués dans les leucémies myéloïdes chroniques, étaient initialement proches dans les cellules non tumorales de la moelle osseuse (Lukasova et al., 1997). Des résultats similaires ont été obtenus pour les gènes *MYC* et *IGH* fréquemment transloqués dans les lymphomes B (Roix et al., 2003). De manière intéressante cette localisation pourrait influencer le timing d'apparition des translocations lors de la différenciation cellulaire. En effet une étude a montré que des loci fréquemment transloqués

dans les liposarcomes étaient proches dans les adipocytes et non dans les pré-adipocytes (Kuroda et al., 2004). De plus, une étude sur les cancers de la prostate a montré que l'activation du récepteur aux androgènes entraînait un rapprochement des loci *TMPRSS2*, *ERG*, et *ETV1* favorisant leur translocation lors d'un stress génotoxique (Lin et al., 2009). Ces résultats pourraient apporter une explication à l'association de translocations à certains types cellulaires particuliers. Les études citées précédemment ont utilisées des approches de microscopie ciblée, permettant l'analyse de seulement quelques loci. Les technologies récentes de Hi-C, ont permis de réaliser des analyses plus globales en déterminant les contacts entre l'ensemble des régions du génome (Lieberman-Aiden et al., 2009). Une étude a par exemple apporté de nouvelles preuves de l'importance de la proximité initiale des partenaires transloqués (Zhang et al., 2012). Les auteurs induisent des cassures double-brin dans les cellules par irradiation. Ils déterminent ensuite l'ensemble des loci transloqués. En comparant les translocations et la carte de contact avant irradiation, ils observent une corrélation entre proximité et fréquence de translocation.

Dans le cas de la translocation t(2,5) menant à la fusion *NPM1-ALK*, qui est utilisée comme modèle dans cette thèse, la proximité entre les loci *NPM1* et *ALK* n'est pas totalement établie. Une étude a montré qu'un rapprochement *NPM1* et *ALK* pourrait être induit par une dérégulation de l'expression de gènes situés à proximité (Mathas et al., 2009). En effet les auteurs observent une dérégulation des gènes *FRA2* et *HLH* sur le chromosome 2 et du gène *CSF1-R* sur le chromosome 5 dans les ALCL portant ou non la translocation t(2,5). Dans les cellules tumorales, en absence de translocation ils observent une proximité spatiale de ces gènes qui n'est pas retrouvée dans des cellules T non tumorales. Cette proximité pourrait ensuite favoriser l'apparition de la translocation qui serait alors un événement secondaire.

2. La chromoplexie

La chromoplexie consiste en un réarrangement en chaîne de plusieurs portions de chromosomes. Ce phénomène a été montré pour la première fois en 2013 dans le cas de cancers de la prostate (Baca et al., 2013). Une analyse de 57 tumeurs par séquençage du génome entier a révélé une forte abondance de translocations (5596 réarrangements

déTECTÉS). Ces translocations apparaissent lors de réarrangements en chaîne en inter et/ou en intra chromosomique, avec ou sans délétions à la jonction (**Figure 3**) et qui sont le plus souvent équilibrés. Lors de l'analyse de la clonalité des réarrangements, les auteurs ont montré que plusieurs phases de réarrangements en chaîne pouvaient intervenir au cours de l'oncogenèse. Ces réarrangements entraînent notamment la dérégulation de gènes pouvant jouer un rôle dans l'oncogenèse, tels que *PTEN*, *NKX3-1*, *CDKN1B*, *TP53* ou encore *RB* dans le cas du cancer de la prostate.

Par la suite, des phénomènes similaires ont pu être décrits dans le cas du carcinome NUT-midline (Lee et al., 2017), du sarcome d'Ewing (Anderson Science 2018), et dans l'adénocarcinome du poumon (Lee et al., 2019). Quelques cas ont également été observés dans le myélome avec une apparition plus tardive de ces événements au cours de l'oncogenèse (Maura et al., 2019). Dans le sarcome d'Ewing, la fusion *EWSR1-ETS*, oncogène principal de ce cancer, est formée dans 42% des cas par un événement complexe de chromoplexie, et non une simple translocation chromosomique. La présence de chromoplexie est associée à un mauvais pronostic dans ce cancer. Tout comme dans le cas du cancer de la prostate, dans le sarcome d'Ewing, la chromoplexie entraîne la dérégulation de l'expression de certains gènes pouvant jouer un rôle dans l'oncogenèse, tels que *AR1D1B* ou *AP1B1*.

La nécessité de chromoplexie dans la formation de remaniements oncogènes pourrait être en partie liée à l'orientation des gènes sur les chromosomes. Par exemple dans le cas du sarcome d'Ewing, les auteurs ont montré que la fusion *EWSR1-ERG* était toujours formée par chromoplexie alors que la fusion *EWSR1-FLI1* pouvait être formée en absence de chromoplexie (Anderson et al., 2018). L'orientation des gènes pourrait expliquer cette différence, car une simple translocation ne permet pas de placer les gènes *EWSR1* et *ERG* dans une bonne orientation transcriptionnelle. Néanmoins, la question de la nécessité de la formation de plusieurs réarrangements dans les autres cas reste ouverte. Ces réarrangements sont, dans la plupart des cas, équilibrés, ce qui est en accord avec le fait que la chromoplexie est un événement initial de l'oncogenèse et se produit dans des cellules saines dans lesquels une perte du nombre de copies induirait un arrêt du cycle.

Les mécanismes sous-jacents aux événements de chromoplexie restent aujourd'hui peu connus. Cependant, les réarrangements observés dans le cas du cancer de la prostate sont

associés à des régions fortement transcrites présentant une chromatine ouverte (Baca et al., 2013). Il semblerait donc que la chromoplexie puisse être liée à la transcription. De plus, les travaux de Anderson et al. ont mis en évidence une corrélation entre chromoplexie, réplication précoce de l'ADN et chromatine ouverte (Anderson et al., 2018). Or les régions fortement exprimées sont globalement répliquées précocement au cours de la phase S du cycle cellulaire. Ces résultats apportent des arguments supplémentaires en faveur d'un lien entre transcription et chromoplexie.

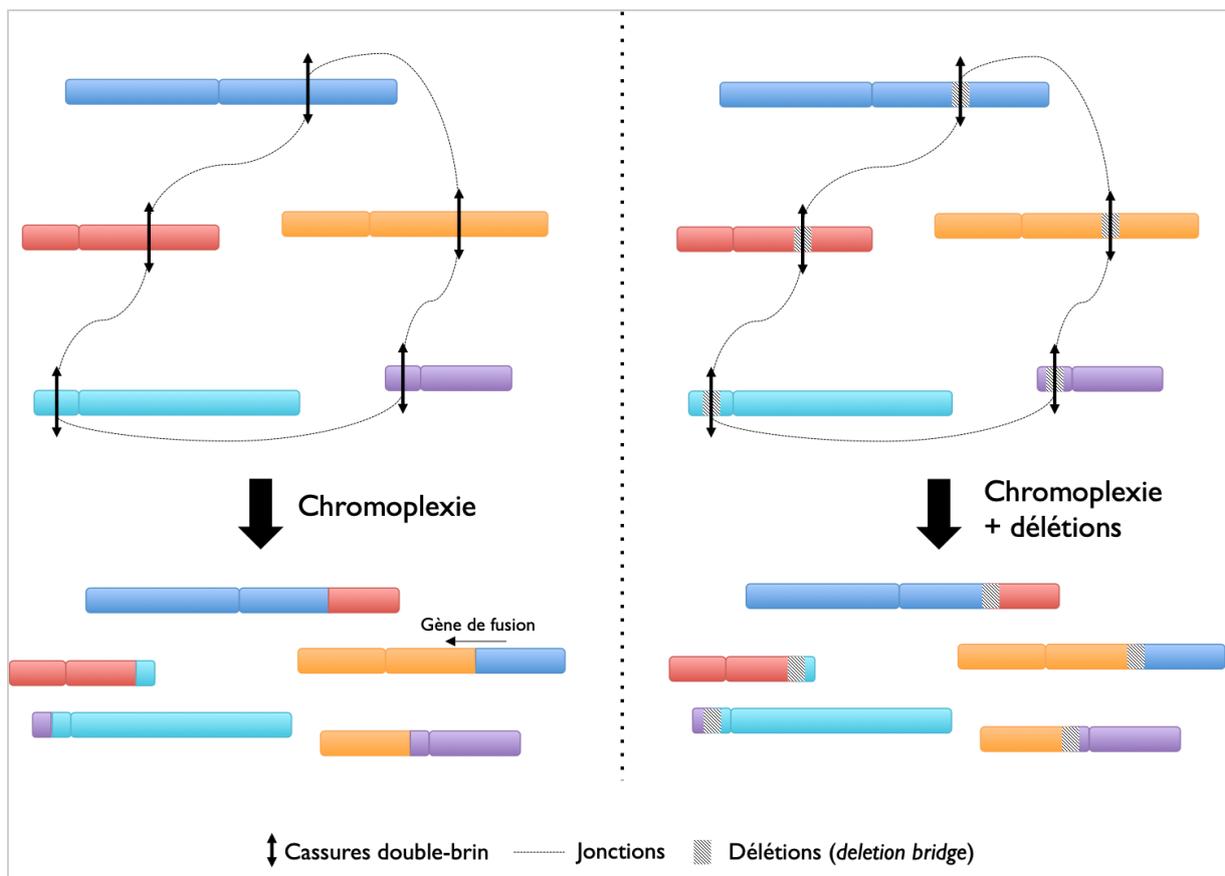


Figure 3 : Schéma du mécanisme de chromoplexie sans (à gauche) ou avec (à droite) délétions au niveau des points de cassure. Des réarrangements en chaîne ont lieu, pouvant entraîner la formation de gènes de fusion.

3. Translocations induites par RAG/AID

Au cours de la maturation des lymphocytes B et T, a lieu une étape de diversification du répertoire immunitaire faisant intervenir des translocations programmées à des sites particuliers (Pour revue (Prochazkova and Loizou, 2016)). Ce processus fait intervenir la recombinaison V(D)J ainsi que l'hypermutation somatique. La recombinaison V(D)J a lieu au

niveau de loci spécifiques : le gène codant la chaîne lourde des immunoglobulines pour les lymphocytes B et le TCR beta pour les lymphocytes T. Lors de ce mécanisme un réarrangement des fragments V (*Variable*), D (*Diversity*) et J (*Joining*) a lieu (**Figure 4**). L'initiation de ce mécanisme est induite par les endonucléases RAG1 et RAG2 capables de créer des cassures double-brin au niveau de leurs sites de reconnaissance RSS (*Recombination Signal Sequences*) (Bassing et al., 2000; Oettinger et al., 1990; van Gent et al., 1995). Les cassures générées par le complexe RAG sont ensuite réparées par le NHEJ classique (c-NHEJ). (Taccioli et al., 1993). La recombinaison V(D)J est initiée au niveau de la moelle osseuse ainsi que dans le foie fœtal pour les progéniteurs-B (pro-B) et dans le thymus pour les progéniteurs-T (pro-T). La première étape consiste en un assemblage (D)J, puis le segment V est accolé à ce fragment. Ceci résulte en un exon V(D)J fonctionnel qui est assemblé avec la région constante C μ lors de la transcription de l'ARN (**Figure 4**). Au cours de la maturation des lymphocytes B a ensuite lieu une étape de modification de la région V, appelée hypermutation somatique, qui permet d'augmenter la diversité du répertoire immunitaire. Une étape de commutation de classe est également possible, entraînant la modification de la région constante C μ de manière à modifier son isotype (en IgE, IgA ou IgG) tout en conservant la spécificité de reconnaissance. Ces deux étapes nécessitent l'intervention de la protéine AID (*Activation-Induced Cytidine Deaminase*) (Muramatsu et al., 2000). Cette enzyme déamine les cytosines en uraciles sur l'ADN simple-brin (lors de la transcription), entraînant l'apparition d'un mésappariement réparé par cassure et mutation de l'ADN. La fenêtre d'expression d'AID au cours de la différenciation des lymphocytes B est extrêmement réduite (Muramatsu et al., 1999). Ceci participe au contrôle de l'hypermutation et de la commutation de classe.

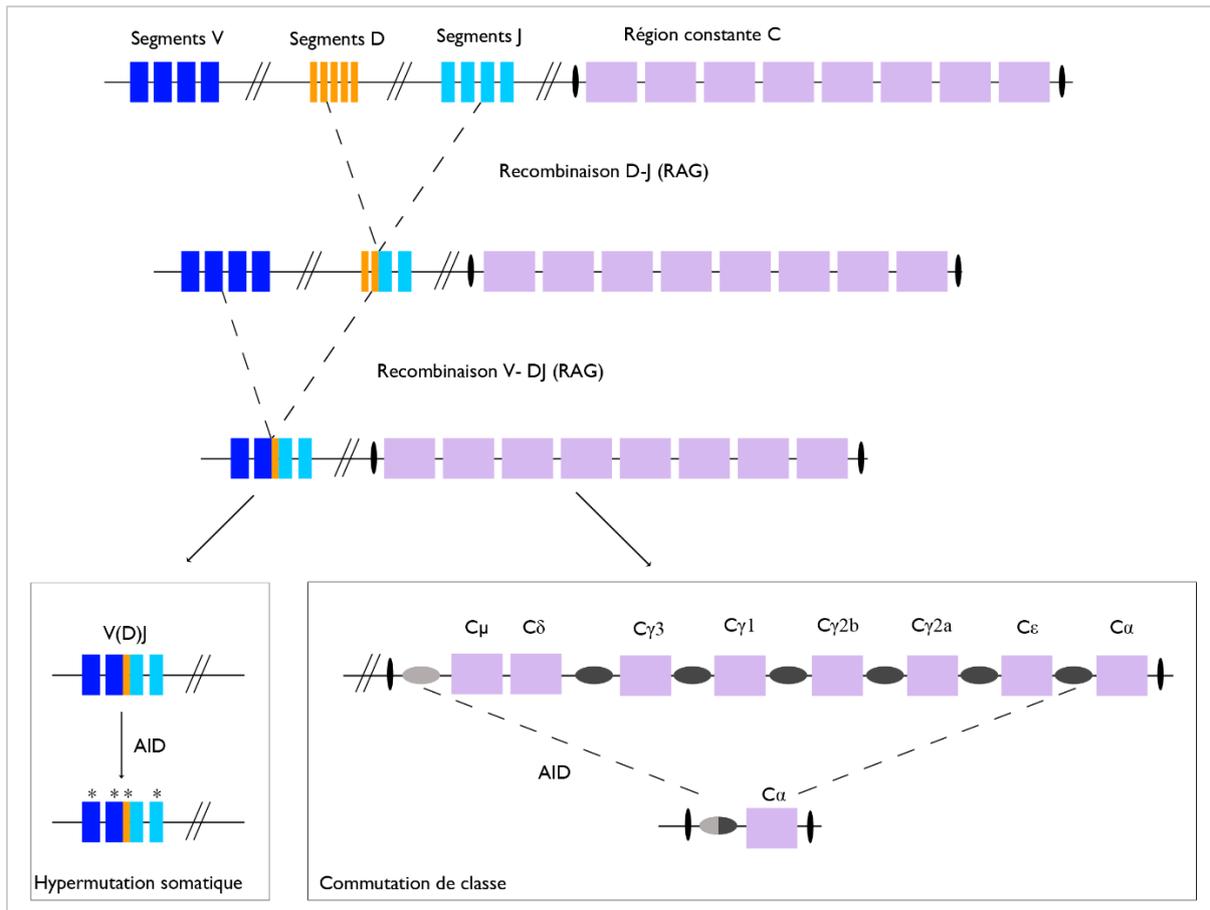


Figure 4 : Schéma simplifié de la diversification du répertoire immunitaire. La recombinaison V(D)J permet en premier lieu d'assembler les segments V, D et J à l'aide des protéines RAG1 et RAG2 qui introduisent des cassures de l'ADN, réparées par le c-NHEJ. La région V(D)J est ensuite soumise à une hypermutation somatique induite par la protéine AID. Cette étape permet d'induire des mutations dans les segments V, D et J. La région constante C est quant à elle soumise à la commutation de classe qui permet de modifier la classe de l'immunoglobuline, à l'aide de la protéine AID

Bien que ce système soit extrêmement contrôlé, la présence du signal RSS peut être suffisante à l'induction de coupure par le complexe RAG même en dehors des loci normalement concernés (Hesse et al., 1987). Ces erreurs de clivages conduisent à la formation anormale de translocations entre un segment appartenant à une région V, D ou J et un locus appartenant à un autre chromosome. Il a été observé que le locus anormalement remanié présentait un signal similaire au RSS (Tycko et al., 1989). De nombreuses translocations ont ainsi été décrites, dont l'un des partenaires est localisé au niveau des loci *IGH* ou *TCR*. Notamment les translocations $t(8;14)(q24;q32)$, $t(2;8)(p12;q24)$, $t(8;22)(q24;q11)$, $t(8;14)(q24;q11)$, et $t(8;12)(q24;q22)$ impliquent une relocalisation du gène *MYC* au locus *IGH* ou *TCR*, et sont associées aux leucémies lymphoïdes aiguës (Rabbitts and Boehm, 1991; Rimokh et al., 1991; Shima et al., 1986). Ou encore la translocation $t(14;18)(q32;q21)$ qui relocalise le gène *BCL2*

au locus *IGH* dans les lymphomes folliculaires (Bakhshi et al., 1987). Des translocations peuvent également être observées en absence de sites RSS cryptiques sur le 2^e partenaire, pouvant s'expliquer par la présence simultanée d'une cassure double-brin dans le génome. De manière similaire des remaniements utilisant des sites cryptiques de reconnaissance pour AID ont été décrits. Notamment il a été montré que des translocations entre le locus *MYC* et le locus *IGH* pouvaient avoir lieu et nécessitaient la présence d'AID (Robbiani et al., 2008). Une activité d'hypermutation aberrante a également été décrite au niveau des proto-oncogènes *PIM1*, *RHOH/TTF* et *PAX5* à l'origine de translocations chromosomiques présentes dans les lymphomes diffus à grandes cellules (Pasqualucci et al., 2001).

Un rôle de la protéine RAG2 a également été démontré dans le maintien de la stabilité du génome chez la souris (Deriano et al., 2011). En effet les auteurs ont montré que dans un contexte déficient en P53 chez la souris, la suppression du domaine c-terminal de RAG2 entraînait une augmentation du nombre de lymphomes ainsi que du nombre de translocations présentes dans ces tumeurs. Une seconde étude a montré que le domaine "*acidic hinge*" de RAG2 était également indispensable au maintien de l'intégrité génomique (Coussens et al., 2013). Les auteurs ont observé une augmentation du nombre d'aberrations chromosomiques sur les métaphases.

4. Translocations et sites fragiles

Dans le génome, il existe des régions d'ADN qui présentent un fort taux de cassures double-brin. Ces régions ont été qualifiées de sites fragiles. Il a été montré que le taux de cassures au niveau de ces sites était lié à la réplication de l'ADN. L'inhibition de la réplication par l'Aphidicholine entraîne par exemple une augmentation du nombre de cassures (Glover et al., 1984). Ce phénomène favorise l'apparition de délétions et de translocations au niveau de ces sites fragiles (Glover and Stein, 1988). Dans le cas de myélomes, il a par exemple été montré que des translocations avaient lieu au niveau du site fragile *FRA16D* (Ried et al., 2000). D'autres translocations ont été observées aux sites *FRA3B* dans l'adénocarcinome de l'œsophage (Fang et al., 2001), *FRA2G* dans le carcinome rénal (Druck et al., 2001), *FRA8C* dans

le lymphome de Burkitt (Ferber et al., 2004), et *FRA6E* dans les leucémies lymphoblastiques aiguës (Sinclair et al., 2005).

5. Translocations et traitement aux inhibiteurs de topoisomérase II

Les inhibiteurs de topoisomérase II, utilisés en clinique, présentent deux mécanismes d'action différents : certains stabilisent le complexe ADN/topo II, entraînant une fixation covalente de l'enzyme au niveau de l'ADN (Osheroff, 1989), alors que d'autres empêchent la religation de l'ADN après clivage (Robinson and Osheroff, 1991). Dans les deux cas on observe une stabilisation de la cassure qui entraîne normalement la mort cellulaire. Cependant cette stabilisation peut également être à l'origine de remaniements chromosomiques.

L'utilisation d'inhibiteurs de topoisomérase II tels que l'étoposide, la doxorubicine, ou encore le mitoxantrone en chimiothérapie aboutit parfois à des leucémies secondaires, caractérisées par des réarrangements chromosomiques. Historiquement, le 1^e cas de translocation secondaire à l'utilisation d'inhibiteur de topoisomérase II, a été décrit en 1985 dans le cas d'un neuroblastome traité par doxorubicine (Secker-Walker et al., 1985). Le patient avait alors développé une leucémie lymphoblastique aiguë portant une translocation t(4,11). La majorité des leucémies secondaires aux traitements inhibiteurs de topoisomérase II portent des remaniements faisant intervenir les gènes *MLL* ou *PML* (Pour revue (Pendleton et al., 2014)). La région BCR (*Breakpoint Cluster Region*) du gène *MLL* est localisée à proximité d'un site de fixation des topoisomérase II ce qui pourrait expliquer les remaniements observés. Plus de 64 partenaires ont été identifiés pour le gène *MLL* (Meyer et al., 2009). Parmi ces partenaires, les plus fréquents sont les gènes *AF4*, *AF9* et *ENL* (Wright and Vaughan, 2014). Le gène *PML* est quant à lui fréquemment associé avec le gène *RAR α* après traitement au mitoxantrone avec des cassures observées au niveau de sites TOPO II (Mistry et al., 2005).

C. Le rôle des translocations dans l'oncogenèse

La présence de translocations chromosomiques dans des cellules tumorales est une observation fréquente. Ces translocations peuvent être retrouvées dans de nombreux types de cancers. Deux mécanismes d'action principaux ont été décrits afin d'expliquer leurs rôles dans la tumorigenèse.

1. La dérégulation d'oncogènes ou gènes suppresseurs de tumeurs (Figure 5A)

Le premier mécanisme connu entraîne la relocalisation d'un oncogène à proximité d'un « *enhancer* » ou d'un promoteur actif. L'oncogène est alors surexprimé dans les cellules transloquées. Ce phénomène a été largement décrit dans le cas d'oncogènes relocalisés à proximité de la région régulatrice des *IGH* et du *TCR*, notamment pour le gène *MYC* (Rabbitts and Boehm, 1991; Rimokh et al., 1991; Shima et al., 1986) ou encore *BCL2* (Bakhshi et al., 1987). Ce mécanisme est très présent dans les cancers touchant les lymphocytes B et T en raison du rôle de la recombinaison V(D)J décrit précédemment.

Des évènements similaires ont pu être décrits dans des tumeurs solides. Par exemple, dans le cancer de la prostate, des réarrangements chromosomiques entraînant la fusion du promoteur de *TMPRSS2* aux gènes *ETS* ont été décrits comme oncogènes (Tomlins et al., 2007). Dans des cas de fibrosarcomes, le gène *PDGFB* peut être placé sous le contrôle du promoteur du gène *COL1A1*. Le 1^e exon de *PDGFB* est délété lors de cette translocation, entraînant une perte de sa régulation (Simon et al., 1997). Il existe d'autres exemples de dérégulation d'oncogènes dans des tumeurs solides : la translocation impliquant le gène *CTNNB1* et l'oncogène *PLAG1* dans les adénomes pléomorphiques (Kas et al., 1997), la translocation impliquant le gène *ACTB* et l'oncogène *GLI1* dans le périicytome (Dahlen et al., 2004) ou encore la translocation impliquant le gène *COL1A1* et l'oncogène *USP6* dans le kyste osseux anévrisimal (Oliveira et al., 2005).

Quelques études ont également décrit un mécanisme d'inactivation de gènes suppresseurs de tumeurs par troncation lors de translocations chromosomiques. Ce phénomène a notamment été montré avec l'inactivation de *P16* dans la leucémie aigue lymphoblastique lors de la

translocation t(9,14), ou encore avec l'inactivation de *NF1* dans le neurofibrome (Duro et al., 1996; Storlazzi et al., 2005).

2. La création d'un oncogène de fusion (Figure 5B)

Le second mécanisme entraîne la création d'un nouvel oncogène de fusion. Ce phénomène peut avoir lieu lorsque les cassures apparaissent dans les introns de deux gènes différents. Ceci entraîne la formation d'une protéine chimérique pouvant avoir une activité aberrante. La majorité des protéines concernées impliquent des kinases ou des facteurs de transcription. Dans le cas des kinases on observe généralement une activation constitutive de la protéine. Un très grand nombre de kinases de fusion ont été décrites (Stransky et al., 2014).

BCR-ABL est la première kinase de fusion découverte en 1986 (Ben-Neriah et al., 1986). Elle est impliquée dans les leucémies myéloïdes chroniques et localisée sur le chromosome de Philadelphie, formé lors de la translocation t(2,9). Une translocation similaire est retrouvée dans 20% des leucémies aiguës lymphoblastiques (ALL) adultes, 5% des ALL pédiatriques et dans quelques cas de leucémies myéloïdes aiguës (AML) (Laurent et al., 2001). Lorsqu'ABL est fusionnée à BCR, son activité tyrosine-kinase reste constitutivement active ce qui entraîne l'activation de différentes voies de signalisation cellulaire dont MAPK, PI3K/AKT, STAT1/5 (Cilloni and Saglio, 2012). De plus, la protéine de fusion BCR-ABL interfère avec l'interaction BCR/XPB, menant à des anomalies de réparation et une instabilité génétique (Takeda et al., 1999).

Il existe généralement une bonne association entre une fusion particulière et un type de cancer. Cependant, certains gènes sont impliqués dans différentes fusions et différents cancers. C'est notamment le cas de *MLL* qui, comme nous l'avons vu précédemment, peut avoir plus de 64 partenaires différents (Meyer et al., 2009). Il est également possible de citer le gène *ETV6* qui présente plus de 20 partenaires (Bohlander, 2005). De plus, il a été montré que la même protéine de fusion ETV6-NTRK3 pouvait être impliquée dans le fibrosarcome, le néphrome mésoblastique, le carcinome du sein ainsi que les leucémies aiguës myéloïdes (Lannon and Sorensen, 2005).

Au laboratoire, deux protéines de fusion sont principalement étudiées : NPM1-ALK (qui sera décrite en détail dans la suite de cette thèse), et EWS-FLI1. Cette seconde protéine de fusion est issue de la translocation $t(11;22)(p24;q12)$ présente dans le sarcome d'Ewing. Elle entraîne la fusion de du domaine de liaison à l'ADN de FLI1 et du domaine trans-activateur de EWS. Ceci induit l'activation constitutive des cibles d'EWS dans la cellule ainsi qu'une modification de l'épigénome (Pour revue (Grünwald et al., 2018)).

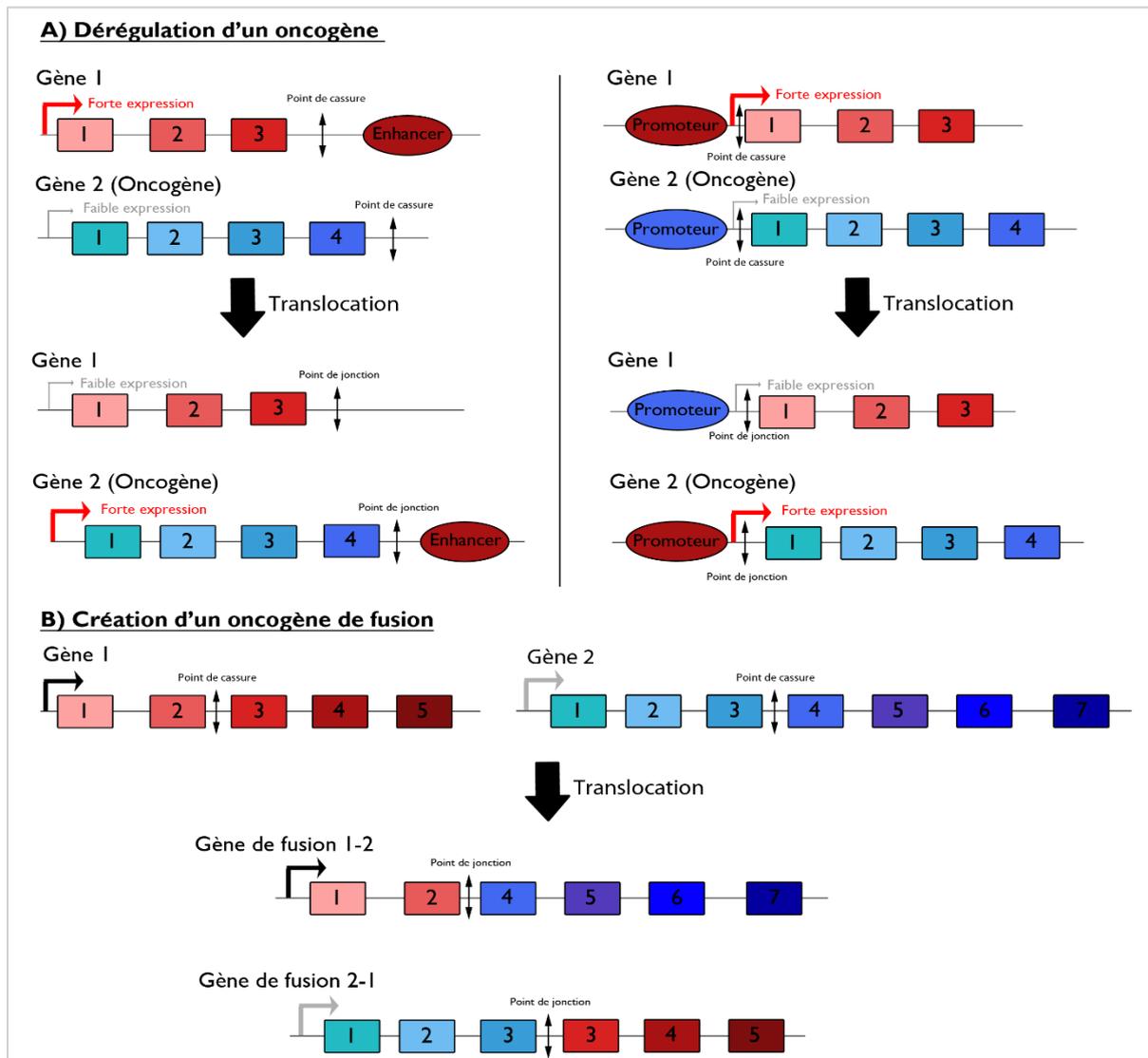


Figure 5 : Mécanismes d'action des translocations au cours de l'oncogenèse. A) L'oncogène 2 est transloqué à proximité d'une séquence *enhancer* (gauche) ou d'un promoteur fort (droite), ce qui entraîne sa surexpression. B) Les cassures double-brin ont lieu dans les introns de deux gènes, la translocation entraîne la formation de deux gènes de fusion plus ou moins exprimés.

D. Le paradoxe des translocations et de leur rôle dans l'oncogénèse

Bien que de nombreuses translocations soient présentes dans une large variété de cancers, il est parfois difficile d'établir un lien direct entre certaines translocations et les mécanismes d'oncogénèse.

1. Incidence des translocations dans les cancers

Il est aujourd'hui établi que la fréquence de translocations chromosomiques dans les cancers est très élevée (**Figure 6**). Dans les cancers hématologiques elle varie de 50% dans les leucémies à 90% dans les lymphomes. Dans le cas des sarcomes, un quart des cancers présentent des translocations. Dans les carcinomes, qui représentent 80% des cancers, le pourcentage de translocation n'est pas clairement établi mais des translocations sont retrouvées par exemple dans les carcinomes rénaux ou pulmonaires (Nambiar and Raghavan, 2011).

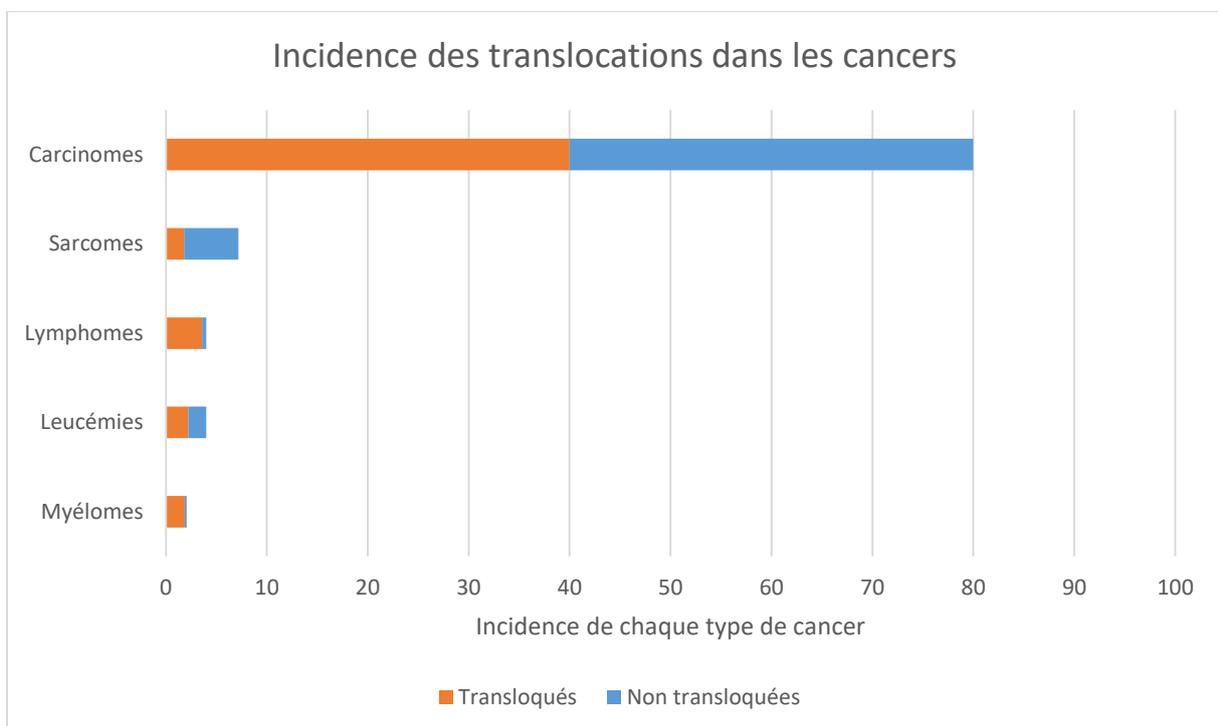


Figure 6: Incidence des translocations dans les cancers (Adapté de (Nambiar and Raghavan, 2011)). En abscisses l'incidence de chaque type de cancers parmi l'ensemble des cancers. En orange et en bleu le pourcentage de cancers avec/sans translocations.

2. Addiction aux oncogènes de fusion

Un argument en faveur de l'importance des translocations dans l'oncogenèse est l'existence d'une addiction aux oncogènes, et notamment aux protéines de fusion. Ce terme d'addiction a été utilisé pour la première fois par le Dr. Weinstein, lors d'une étude dans laquelle il montre que l'inhibition de la protéine kinase C dans des cellules cancéreuses peut entraîner la mort de ces cellules (Weinstein et al., 1997). Par la suite, ce phénomène a été décrit pour de nombreux oncogènes : *MYC* (Felsher and Bishop, 1999; Pelengaris et al., 1999), *HRAS* (Chin et al., 1999), *KRAS* (Aoki et al., 1997).

Dans le cas des protéines de fusion, ce phénomène d'addiction a été bien décrit pour la protéine BCR-ABL. Par exemple, la protéine de fusion BCR-ABL a été introduite dans une lignée cellulaire murine dépendante à l'IL3. En présence de BCR-ABL les cellules sont capables de survivre sans IL3, à l'inverse l'inhibition de cette kinase de fusion entraîne une mort massive des cellules (Klucher et al., 1998). Une étude plus récente a montré que cette addiction à BCR-ABL est dû à l'activation des voies STAT3/5, AKT et ERK (Sharma et al., 2006). Ces observations ont orienté la recherche vers des molécules thérapeutiques ciblant spécifiquement les oncogènes de fusion. Une équipe a mis au point un analogue d'ATP permettant l'inhibition des tyrosine-kinase (dont BCR-ABL), utilisé en clinique depuis 2001 sous le nom d'Imatinib (Druker et al., 1996). Par la suite cet inhibiteur a été utilisé dans le cas d'autres fusions chromosomiques (*COL1A1-PDGFRB*, *FIP1L1-PDGFR*, autres fusions incluant *PDGFR*). D'autres inhibiteurs de kinases sont également utilisés en clinique (Crizotinib et Ceritinib pour les fusions incluant *ALK*, Dasatinib, Nilotinib, Bosutinib, Ponatinib pour BCR-ABL) (Mertens et al., 2015).

3. Les translocations retrouvées chez des individus sains

Il a été montré que certaines translocations étaient suffisantes à la transformation cellulaire. C'est par exemple le cas de la $t(9,11)(p22,q23)$ menant à la fusion *MLL-AF9* (Horton et al., 2013). A l'inverse d'autres translocations ne permettent pas, à priori, d'induire l'oncogenèse à elles seules. C'est le cas de la translocation $t(12,21)(p13,q22)$ menant à la fusion *ETV6-*

RUNX1. Cette translocation est retrouvée chez des jumeaux monozygotes, ce qui indique une origine prénatale (Wiemels et al., 1999). Cependant les jumeaux ne développent pas de leucémies lymphoïdes aigues au même âge, ce qui semble indiquer la nécessité d'évènements secondaires à la translocation permettant l'oncogenèse.

De nombreuses translocations chromosomiques, fréquemment associées à des cancers, peuvent être détectées dans le sang d'individus sains. En effet, cette observation a été faite en 1991, lors d'une étude sur la translocation t(14 ;18) menant à la fusion IGH-BCL2 (Limpens et al., 1991). Le transcrite de fusion a été détecté par RT-PCR dans les leucocytes de personnes saines à une fréquence de $1/10^5$ cellules. De nombreuses études ultérieures ont confirmé ces résultats (Pour Revue (Basecke et al., 2002)).

En 1995, des analyses similaires ont été réalisées pour BCR-ABL (Biernaux et al., 1995). Lors de ces travaux, les auteurs ont réussi à détecter le transcrite codant pour la protéine de fusion BCR-ABL par RT-PCR, chez des individus sains. Ils ont ainsi détecté l'ARN dans le sang de 22/73 adultes, et 1/22 enfants. Ils n'ont cependant détecté aucune translocation dans 22 échantillons de sang de cordon. Ce résultat a par la suite été confirmé par Bose et al. (Bose et al., 1998). Ces observations ont également été faites pour de nombreuses autres translocations impliquées dans les leucémies et lymphomes (Janz et al., 2003). Notamment une étude de 2002 a montré que la translocation t(2 ;5) menant à la fusion NPM1-ALK pouvait être détectée chez 14/29 individus sains (Trumper et al., 1998).

Une étude plus approfondie, sur 15 ans, a été réalisée pour la translocation t(14 ;18) (Roulland et al., 2014). Les auteurs ont utilisé une cohorte de 520 000 individus, chez qui des prélèvements ont été effectués à t0. Parmi ces participants, 100 ont développé un lymphome folliculaire dans les 3 mois – 15 ans suivant le début de l'essai. Les auteurs ont alors analysé la prévalence de la translocation t(14 ;18) à t0 entre ces 100 individus malades et 218 individus contrôles. Les résultats révèlent une fréquence de détection de la translocation à t0 plus élevée chez les individus malades (fréquence $> 10^{-4}$). De plus, en comparant les points de cassures détectés à t0 et au diagnostic, les auteurs n'observent aucune variation. Ceci suggère une expansion clonale à partir d'une cellule transloquée préexistante. L'ensemble de ces résultats semble indiquer que la fréquence de détection de la translocation t(14 ;18) dans le sang d'un individu sain pourrait être un bon indicateur du risque de développer un lymphome

folliculaire. Cependant cette étude a été réalisée uniquement sur des individus d'origine Européenne, or une étude de 2017 a montré que la prévalence de la translocation $t(14;18)$ était dépendante de l'origine ethnique (Levy et al., 2017). En effet, dans cette étude les auteurs ont recherché la présence de la translocation dans le sang d'individus Brésiliens d'origines ethniques différentes. L'étude inclus 85 « Blancs », 72 « Noirs » et 70 « Descendants Japonais ». Les résultats montrent de très fortes variations de la prévalence de la $t(14;18)$ entre les trois groupes : elle est détectée chez 52% des « Blancs », 79% des « Noirs » et 97% des « Japonais ». Une seconde étude réalisée sur une population Américaine montre que la prévalence de la $t(14;18)$ est la même entre la population « Blanche » et « Noire », alors que l'incidence du lymphome folliculaire est plus élevée au sein de la population « Blanche » (Colon-Otero et al., 2017). Les résultats de ces différentes études semblent indiquer que la $t(14;18)$ n'est pas un bon indicateur de risque pour toutes les origines ethniques et qu'il est important de tenir compte du fond génétique lors de la recherche de biomarqueurs.

De même, dans le cas du sarcome d'Ewing, présentant une translocation entre les loci *EWSR1* et *FLI1*, il a été observé que l'incidence était dix fois plus faible dans la population Africaine comparé à la population Européenne (Beck et al., 2012). Cette différence peut être en partie expliquée par l'existence de polymorphisme au niveau des sites de reconnaissance microsatellites de la protéine de fusion EWS-FLI1.

L'ensemble des observations précédentes semblent indiquer que les translocations ne sont pas toujours suffisantes à l'oncogénèse. Il se pourrait que des évènements secondaires, ou un contexte cellulaire particulier soient nécessaires en plus de la translocation.

E. La modélisation des translocations

L'étude des translocations chez les patients ne permet pas toujours de disséquer les mécanismes de formation et/ou d'action des translocations. Pour cela il est parfois nécessaire d'utiliser des modèles simplifiés dont les paramètres sont davantage contrôlés. Ainsi au cours des dernières années, de nombreuses équipes ont tenté de reproduire certaines

translocations chromosomiques dans des lignées cellulaires, des cellules primaires, mais également *in vivo*.

1. Surexpression du gène de fusion : un modèle limité

La plupart des modèles sont basés sur une surexpression du gène de fusion. De tels modèles ont été générés dans des lignées cellulaires humaines et murines (Amarante-Mendes et al., 1998; Canitrot et al., 1999; Jin et al., 2007) afin d'étudier les conséquences sur la prolifération, la survie et une éventuelle transformation des cellules.

La surexpression de gènes de fusion peut également être réalisée dans des cellules primaires. Il est par exemple possible d'induire une surexpression stable des fusions BCR-ABL et MLL-AF9 dans des CD34+ humains à l'aide de systèmes rétroviraux ou lentiviraux (Horton et al., 2013; Rizo et al., 2010; Sontakke et al., 2016; Wei et al., 2008). Cependant dans ces études, les xénogreffes des cellules surexprimant le gène de fusion peinent à récapituler les maladies observées chez l'Homme. En effet dans le cas de la fusion MLL-AF9, les auteurs obtiennent majoritairement des leucémies lymphoïdes et non myéloïdes, montrant une des limites de ces modèles.

Des modèles murins ont également pu être générés, permettant une expression contrôlée du gène de fusion au cours du développement, ou encore une expression tissu-spécifique (Guryanova and Levine, 2013; Hauer et al., 2014; Rabbitts et al., 2001). Ce contrôle de l'expression des gènes de fusion est un paramètre essentiel. En effet une expression ubiquitaire peut s'avérer létal. C'est par exemple le cas de la fusion BCR-ABL qui induit une létalité embryonnaire chez la souris (Heisterkamp et al., 1991).

Bien que les modèles de surexpression aient permis de nombreuses avancées, ils présentent un certain nombre de limitations. Tout d'abord, le niveau d'expression n'est pas soumis au contrôle du promoteur endogène, empêchant ainsi l'analyse d'un éventuel effet dose. De plus, le contexte chromosomique est totalement différent : la translocation est absente, de même que le gène réciproque, qui peut parfois être exprimé chez les patients et pourrait jouer un rôle dans l'oncogenèse. Notamment, il a été montré que la protéine de fusion ABL-BCR était

exprimée après translocation, et qu'elle favorisait la prolifération cellulaire (Zheng et al., 2009). Les éventuels effets d'haploinsuffisance liés à la perte d'un des allèles sont également absents dans les modèles de surexpression (Groschel et al., 2014). Enfin, les translocations chromosomiques impliquent un remaniement de bras entiers de chromosomes pouvant générer des modifications de localisation nucléaire, de marques épigénétiques et d'organisation de la chromatine. L'ensemble de ces éléments ne peut pas être étudié dans les modèles de surexpression. Il est donc nécessaire de générer de nouveaux modèles de translocations plus proches de la physiologie des patients, afin de pouvoir étudier les conséquences globales des translocations chromosomiques.

Des modèles utilisant le système Cre-lox chez la souris avaient été proposés afin de générer les translocations *Mll-Af9* et *Ewsr1-Fli1* (Rabbitts et al., 2001; Tanaka et al., 2015). Ces modèles reproduisent la translocation de manière fidèle, cependant la technique est longue et complexe, et peut difficilement être appliquée aux cellules humaines.

2. Utilisation des nucléases artificielles afin de générer des translocations chromosomiques

L'émergence de nouveaux outils d'édition du génome a ouvert la porte à la création de nouveaux modèles d'étude des translocations, plus pertinents.

a) Les ZFN (Nucléase à doigt de Zinc)

Les nucléases à doigt de zinc (ZFN) ont été les 1^{er} outils d'édition du génome utilisés (Kim et al., 1996). Ils sont composés d'un domaine de reconnaissance de l'ADN à doigt de zinc et d'un domaine FokI capable de cliver l'ADN (**Figure 7**). Les doigts de zinc peuvent s'insérer dans le grand sillon de l'ADN par reconnaissance de trois nucléotides (Pavletich and Pabo, 1991). Le domaine FokI possède une activité endonucléase lorsqu'il est présent sous forme de dimère (Bitinaite et al., 1998). Ainsi pour générer une cassure double-brin de l'ADN il est nécessaire d'utiliser un couple de ZFN reconnaissant chacune un brin d'ADN. Les ZFN ont été les 1^{er} outils d'édition

du génome utilisés pour générer des translocations chromosomiques dans des cellules humaines (Brunet et al., 2009).

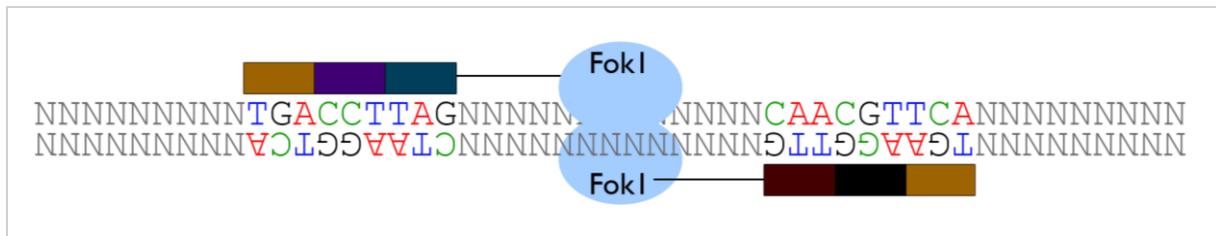


Figure 7 : Mécanisme d'action des ZFN. La reconnaissance de l'ADN est assurée par des domaines à doigts de zinc qui se fixent sur une succession de 3 nucléotides. La coupure est réalisée par le domaine FokI qui fonctionne en dimère.

b) Les TALEN (*Transcription Activator-Like Effector Nucleases*)

Les TALEN (*Transcription Activator-Like Effector Nucleases*) sont composées de deux domaines : un domaine TALE (*Transcription Activator-Like Effector*) assurant la spécificité de reconnaissance de la séquence d'ADN, et un domaine FokI capable de cliver l'ADN de manière simple brin (**Figure 8**) (Christian et al., 2010; Miller et al., 2011; Mussolino et al., 2011). La spécificité de reconnaissance est effectuée par une suite de répétitions RVD (*Repeat Variable di-Residues*). Chaque RVD reconnaît un acide aminé particulier : NI s'associe avec A, HD avec C, NG et HG avec T, et NN avec G. L'association de ces RVD entre eux permet de cibler n'importe quelle séquence d'ADN. De même que pour les ZFN il est nécessaire d'utiliser un couple de TALEN afin de générer des cassures double-brin. L'un des inconvénients de l'utilisation des TALEN, est le temps et la complexité de l'assemblage des répétitions entre elles.

Ce système a été utilisé au laboratoire afin de générer deux cassures double-brin ciblées de l'ADN, menant à la formation de translocations chromosomiques (Piganeau et al., 2013; Piganeau et al., 2016). Cette technique a permis de générer les translocations t(11;22)(q24;q12) menant à la fusion EWS-FLI1 et t(2;5)(p23;q35) et menant à la fusion NPM1-ALK avec une fréquence d'environ 1 cellule transloquée sur 1000. Par la suite ce système a permis d'étudier plus en détail les mécanismes de formation des translocations dans les cellules humaines (Ghezraoui et al., 2014).

cellulaires humaines. Une étude similaire a montré l'efficacité de cette technique dans des cellules de souris, permettant notamment de générer la translocation *Pax3-Foxo1* associée au rhabdomyosarcome alvéolaire (Lagutina et al., 2015). A l'inverse, une équipe a montré qu'il était possible de réverter la translocation *BCR-ABL*, dans une lignée cellulaire humaine, à l'aide de CRISPR/Cas9 (Lekomtsev et al., 2016). La technologie CRISPR/Cas9 peut également permettre de détecter l'ensemble des réarrangements présents dans le génome suite à l'induction d'une seule cassure (Frock et al., 2015). Des cassures endogènes présentes dans le génome peuvent mener à différentes translocations détectées par séquençage.

Ce système a ensuite été rapidement adapté à la génération de translocations *in vivo*. En 2014 et 2015 deux équipes sont parvenues à induire l'inversion *Eml4-Alk*, associée au cancer du poumon non à petites cellules. Les expériences ont été réalisées chez des souris adultes par instillation intratrachéale d'un adénovirus contenant la Cas9 et les guides ARN (Maddalo et al., 2014) ou par instillation intratrachéale ou injection intra-pulmonaire directe d'un lentivirus (Blasco et al., 2014). Dans les deux études, 6 à 8 semaines après infection, les auteurs ont pu détecter la présence de tumeurs dans les poumons. L'analyse biochimique des tumeurs a révélé la présence de marqueurs spécifiques des adénocarcinomes tels que *Nkx-2.1/TTF1* et *SpC*. De plus, une analyse par PCR ou par FISH a permis de montrer la présence de l'inversion *Eml4-Alk* dans l'ensemble des tumeurs analysées (n=4 et n=10). Les auteurs ont également montré que l'utilisation de Crizotinib (inhibiteur d'ALK utilisé en clinique) permettait la régression tumorale. Ces études mettent en avant la pertinence des modèles de réarrangements chromosomiques utilisant la technologie CRISPR/Cas9, qui permettent de reproduire de manière fidèle les événements qui ont lieu chez les patients.

La génération de translocations via le système CRISPR/Cas9 a également été effectuée dans le modèle animal *C. elegans* (Chen et al., 2015). La micro-injection des guides et de la Cas9 a permis de générer une population F1 fertile porteuse de translocations.

Par la suite, cette technique a été appliquée aux cellules souches. En 2016, une étude a montré qu'il était possible de générer des translocations dans des cellules souches de souris (mESC) à l'aide de CRISPR/Cas9 (Jiang et al., 2016). Les auteurs ont établi une lignée mESC exprimant la protéine Cas9 de manière constitutive. Les cellules ont ensuite été infectées avec un lentivirus porteur de guides ARN dirigés contre *Cdx2* et *Gsk3*, ce qui permet l'induction d'une

translocation t(5;7). Les cellules transloquées conservent leur pluripotence et sont capables de participer à la formation d'embryons chimériques lorsqu'elles sont injectées dans des blastocystes. En 2017, une étude a montré qu'il était également possible d'induire des translocations par CRISPR/Cas9 dans des cellules souches humaines (hMSC et hiPSC) (Torres-Ruiz et al., 2017). En 2017, un second groupe a induit la translocation *MLL-ENL* dans des cellules souche hématopoïétiques humaines CD34+ (Reimer et al., 2017). Les auteurs ont obtenu des cellules transloquées en utilisant un lentivirus portant la Cas9 et les deux guides ARN. Par la suite ils ont étudié le potentiel transformant de cette translocation. *In vitro* ils n'ont pas constaté de transformation complète des cellules en méthylcellulose ou en culture liquide. Ils ont également réalisé des xénogreffes chez la souris. Les premières greffes n'ont pas entraîné l'apparition de leucémies bien qu'elles présentent un phénotype proche des leucémies monocytaires. Une seconde transplantation des cellules récupérées sur les premiers receveurs permet cependant d'induire des leucémies B lymphoblastiques aigues. Ces observations montrent que l'induction de la translocation n'est pas toujours suffisante à l'oncogenèse, et qu'un environnement particulier pourrait permettre l'apparition secondaire de cancers. Ces résultats sont différents de ceux obtenus précédemment avec des modèles de surexpression (Barabe et al., 2007). En effet la surexpression avait été montrée suffisante à l'induction de l'oncogenèse. L'ensemble de ces résultats montre l'importance de l'utilisation de modèles les plus proches possibles de la pathologie afin de limiter les biais dans les résultats.

Afin de faciliter la sélection des cellules transloquées par le système CRISPR/Cas9, et de contrôler l'expression du gène de fusion, deux équipes ont proposé l'utilisation d'un gène rapporteur intégré à la jonction (Spraggon et al., 2017; Vanoli and Jasin, 2017). Les travaux de Vanoli et al. ont été réalisés en collaboration avec notre laboratoire au cours de ma thèse. Ces études ont été réalisées sur les translocations *EWSR1-FLI1* et *EWSR1-WT1*. Les auteurs proposent de combiner le système CRISPR/Cas9 avec la recombinaison homologue et le système Cre/Lox. Les deux guides ARN, la Cas9 et une matrice de réparation sont co-transfectés dans les cellules. La matrice de réparation est composée de deux régions homologues présentes aux extrémités (une pour chaque gène impliqué dans la translocation), de deux sites LoxP et d'un gène de résistance (Hygromycine ou Puromycine). Si la réparation des cassures générées par la Cas9 utilise la recombinaison homologue et intègre la matrice au

site de réparation, alors les cellules transloquées sont résistantes à l'Hygromycine ou à la Puromycine et peuvent ainsi être sélectionnées. Le gène de fusion n'est pas exprimé en raison de la présence du gène de résistance. Par la suite, il est possible d'exciser le gène de résistance par expression de la Cre et ainsi permettre l'expression du gène de fusion. Cette approche permet à la fois de sélectionner les cellules transloquées, et de contrôler l'expression du gène de fusion. Cependant les événements de recombinaison étant plus rares que la réparation par le NHEJ, Vanoli et al. ont montré que dans leurs cellules mésenchymateuses humaines non immortalisées, il était nécessaire d'utiliser un inhibiteur des DNA-PKcs afin de limiter le NHEJ, pour obtenir des clones transloqués porteurs de l'insert. Ce traitement n'est pas nécessaire avec la lignée cellulaire HEK293T utilisée par Spraggon et al.

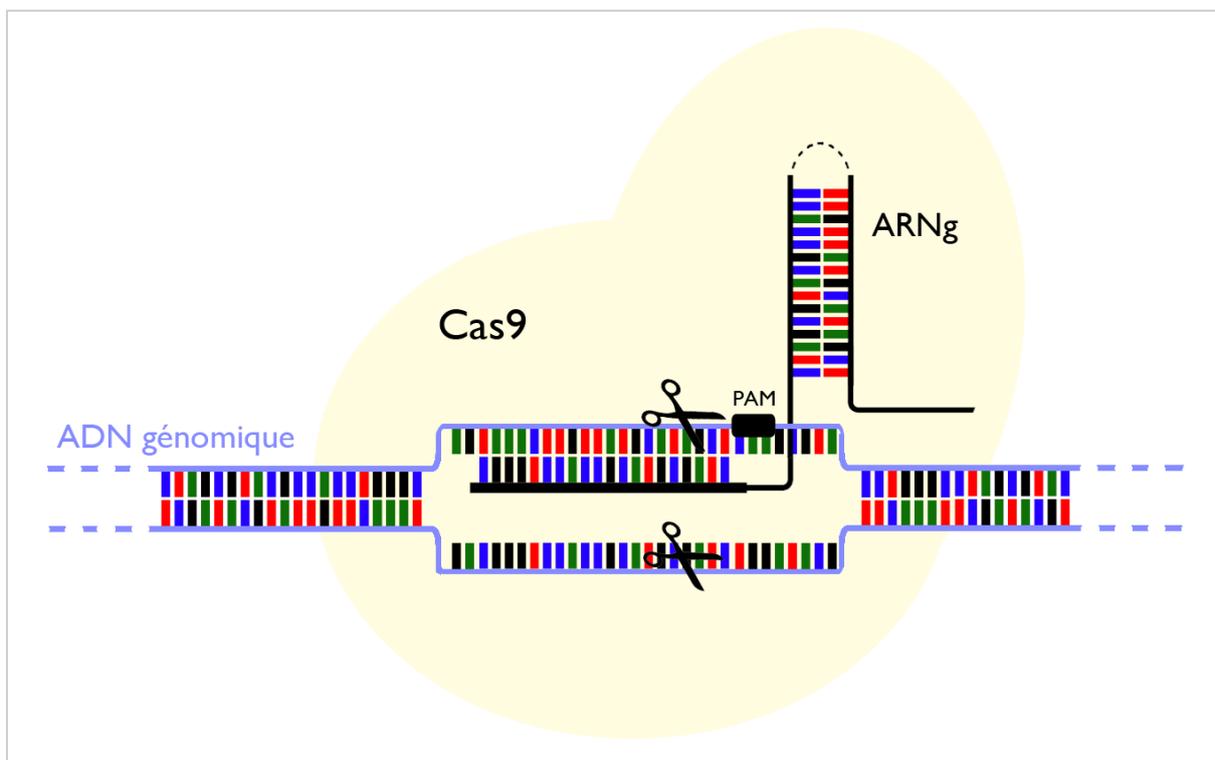


Figure 9 : Mécanisme d'action du système CRISPR/Cas9. L'ARN guide est composé d'une séquence complémentaire à la séquence ADN cible, ainsi que d'une structure appelée ARNtrac permettant la liaison à la protéine Cas9. La Cas9 est capable de couper l'ADN sur les deux brins quelques bases en amont de la séquence PAM.

d) *La question de la spécificité*

Le système CRISPR/Cas9 a révolutionné la modification ciblée du génome, de par sa simplicité de synthèse et d'utilisation. Cependant la question de la spécificité de coupure fait l'objet de

nombreux débats. Bien que plusieurs études aient révélé une bonne spécificité du système CRISPR/Cas9 (Smith et al., 2014; Veres et al., 2014) il a été montré que ce système pouvait tolérer jusqu'à 5 mésappariements et ainsi entraîner des coupures non spécifiques (Fu et al., 2013).

De nombreux logiciels de prédiction bio-informatique ont été mis au point afin de prédire et d'évaluer la spécificité et l'efficacité des guides ARN (Bae et al., 2014; Haeussler et al., 2016; Heigwer et al., 2014; Lei et al., 2014; MacPherson and Scherf, 2015; Montague et al., 2014; Naito et al., 2015; Prykhodzhiy et al., 2015; Xiao et al., 2014; Xie et al., 2014; Zhu et al., 2014). Il est ainsi possible d'aller vérifier la présence/absence de mutations aux sites de coupure non spécifique prédits. De plus, il existe des méthodes d'analyse permettant de détecter l'ensemble des mutations non spécifiques générées sur le génome (Crosetto et al., 2013; Gabriel et al., 2011; Tsai et al., 2017).

Afin de pallier ce manque de spécificité, différentes solutions ont été proposées. Il a notamment été montré qu'il était possible d'utiliser une Cas9 dite « nickase » qui génère des cassures simple-brin plutôt que double-brin (Mali et al., 2013; Ran et al., 2013). Il est alors possible d'utiliser un couple de « nickases » ciblant deux régions très proches afin de générer deux cassures simple-brin qui seront traitées comme une cassure double-brin. D'autres équipes ont généré des mutants de la protéine Cas9 permettant d'augmenter la spécificité (Cho et al., 2014; Kleinstiver et al., 2016; Slaymaker et al., 2016). Notamment l'utilisation de la Cas9-HF1 (*Cas9-High Fidelity*) semble supprimer totalement les coupures non spécifiques (en dessous du seuil de détection). Plus récemment, une équipe a mis au point une « Sniper-Cas9 » obtenu par sélection de mutants naturels chez E.coli présentant une efficacité semblable à la Cas9-WT et une augmentation de la spécificité de 2 à 16 fois (Lee et al., 2018). D'autres équipes de recherche ont proposé de modifier le guide ARN afin d'augmenter la spécificité (Cho et al., 2014; Fu et al., 2014; Kocak et al., 2019). Notamment, l'utilisation de guides plus courts semble donner de bons résultats.

L'ensemble de ces éléments montrent qu'il existe bel et bien un manque de spécificité du système CRISPR/Cas9. Dans le cas d'études sur des lignées cellulaires, ou des modèles animaux, les effets de coupure non spécifique prédits peuvent être vérifiés, notamment par séquençage. Cependant dans l'idée d'une utilisation en clinique, il est essentiel de contrôler

et limiter les effets non spécifiques afin de ne pas perturber de manière involontaire d'autres régions du génome. Des modifications de la protéine Cas9 ou des guides ARN permettent de limiter les effets de coupure non spécifique. Ces résultats ouvrent de nouvelles possibilités pour une utilisation prochaine en thérapeutique. Dans le cadre de l'induction des translocations chromosomiques, il est important de noter que l'utilisation de deux ARN guides simultanément peut augmenter les effets non spécifiques.

II. La translocation t(2;5)(p23,q35) comme modèle d'étude

Dans le cadre de cette thèse nous avons choisi la translocation t(2,5)(p23,q35) comme modèle d'étude. Cette translocation est fréquemment associée au lymphome anaplasique à grandes cellules (ALCL) et entraîne la formation d'une protéine de fusion NPM1-ALK (**Figure 10**).

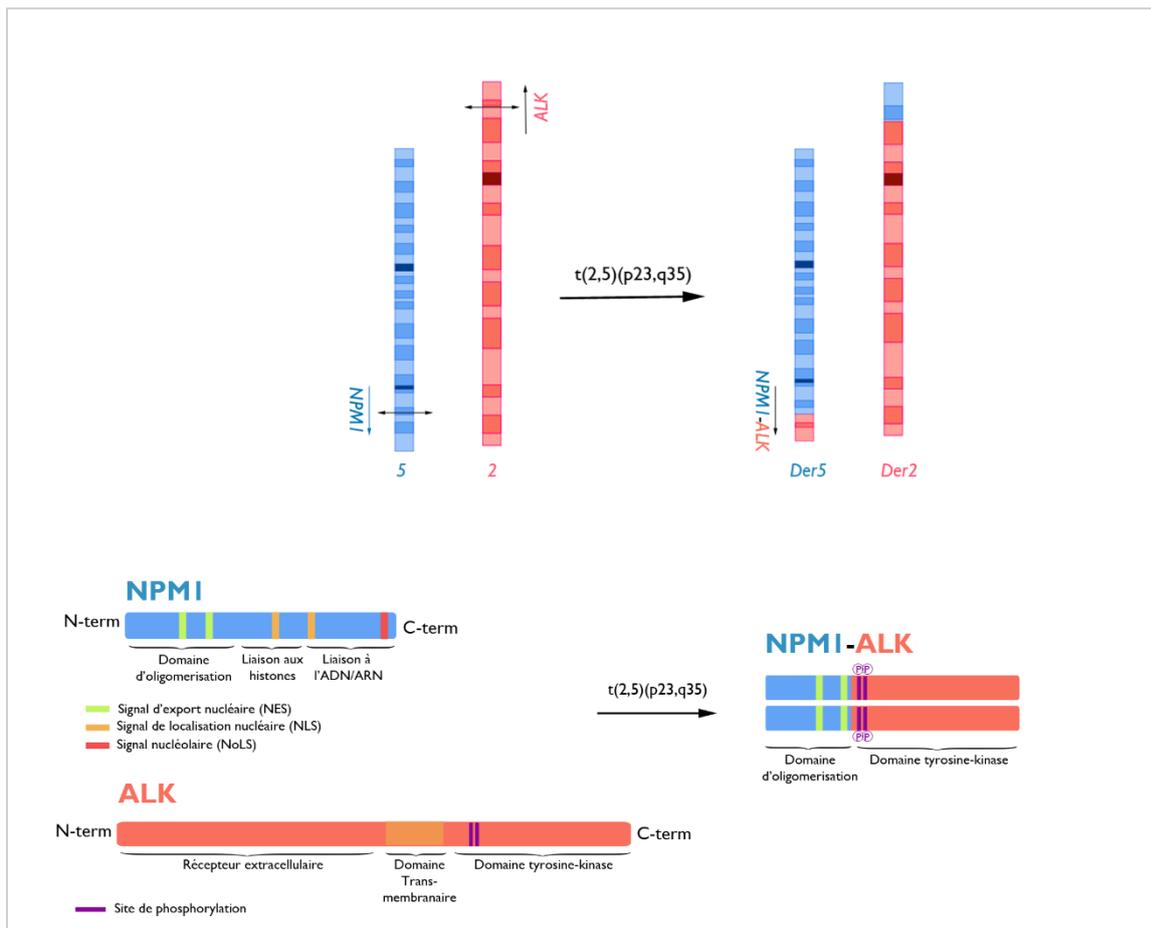


Figure 10: Schéma de la translocation t(2,5)(p23,q35) et de la protéine de fusion NPM1-ALK formée.

A. Aspects moléculaires

1. Nucléophosmine 1 (NPM1)

Le gène *NPM1*, porté par le chromosome 5, code pour une protéine ubiquitaire de 32 kDa : la nucléophosmine 1 (également connue sous le nom de B23). Il s'agit d'une protéine navette, capable de transiter entre le noyau et le cytoplasme (Borer et al., 1989), grâce à la présence de signaux d'import et d'export nucléaire tous deux présents sur la protéine (**Figure 10**). Elle

possède également un signal d'import nucléolaire dans sa partie C-terminale (Nishimura et al., 2002). En conditions physiologiques normales l'import nucléaire est favorisé par rapport à l'export (Bolli et al., 2007).

Différentes fonctions ont été décrites pour la protéine NPM1. Elle est notamment impliquée dans la maturation et le transport des ribosomes (Sipos and Olson, 1991), et a été montrée capable d'interagir avec l'ADN (Dumbar et al., 1989). Elle joue également un rôle dans la duplication des centrioles lors du cycle cellulaire (Tarapore et al., 2002) ainsi que dans la réparation de l'ADN via le mécanisme d'excision de base (BER) (Poletto et al., 2014). De plus, il a été montré que NPM1 était capable de se lier à différentes protéines de la voie P53 : P53 elle-même, entraînant une stabilisation de la forme active de P53 (Colombo et al., 2002), mais également les protéines RB (Takemura et al., 2002), ARF (Bertwistle et al., 2004) et HDM2 (Kurki et al., 2004). Via ces différentes interactions, NPM1 joue un rôle majeur dans la réponse aux stress tels que l'hypoxie, les agents cytotoxiques ou encore l'irradiation UV (Colombo et al., 2002; Li et al., 2004; Wu and Yung, 2002).

NPM1 est fréquemment mutée dans différents cancers. De nombreuses mutations ont notamment été décrites dans les leucémies aiguës myéloïdes (AML) (Falini et al., 2005). La plupart des mutations observées entraînent une modification du ratio import/export nucléaire et une accumulation de NPM1 dans le cytoplasme. D'autres travaux ont également montré qu'une haploinsuffisance du gène *NPM1* pouvait participer à l'oncogenèse (Grisendi et al., 2005; Sportoletti et al., 2008). Dans le cas des AML, *NPM1* est donc majoritairement considéré comme un suppresseur de tumeur. Cependant, une étude a montré que les patients porteurs de mutations inactivatrices dans le gène *NPM1* présentaient une meilleure réponse au traitement ainsi qu'une diminution du risque de rechute (Thiede et al., 2006). De plus, dans les tumeurs solides il a été décrit comme oncogène. En effet une surexpression de *NPM1* a par exemple été observée dans les cancers gastriques, du colon, de l'ovaire, de la prostate et des glandes salivaires (Li et al., 2017b; Nozawa et al., 1996; Shields et al., 1997; Subong et al., 1999; Tanaka et al., 1992). Cette surexpression est associée avec un mauvais pronostic dans les tumeurs solides (Chen et al., 2018b). Enfin la protéine NPM1 peut également être impliquée dans des oncogènes de fusion. A ce jour, trois fusions impliquant NPM1 ont été décrites : NPM1-MLF1 impliquée dans les syndromes myélo-dysplasiques et les leucémies

myéloïdes aigues (Yoneda-Kato et al., 1996), NPM1-RAR impliquée dans les leucémies promyélocytiques aigues (Redner et al., 1996) et NPM1-ALK impliquée dans les lymphomes anaplasiques à grandes cellules (Morris et al., 1994). Dans l'ensemble de ces translocations, NPM1 apporte au minimum son domaine d'oligomérisation, avec parfois son domaine de liaison aux histones (dans le cas de NPM1-MLF1). Ceci entraîne l'activation constitutive des protéines avec lesquelles NPM1 est fusionnée.

2. Anaplastic Lymphoma Kinase (ALK)

Le gène *ALK*, localisé sur le chromosome 2, code pour une récepteur tyrosine kinase transmembranaire de 220 kDa (**Figure 10**), exprimé exclusivement dans le système nerveux et principalement au cours du développement embryonnaire (Iwahara et al., 1997; Morris et al., 1997; Shiota et al., 1994). ALK appartient à la super famille des récepteurs à l'insuline. L'activation de la protéine ALK par phosphorylation entraîne une cascade de signalisation incluant notamment les facteurs CBL, FRS2, IRS1, SHC1 et MAPK (Motegi et al., 2004; Souttou et al., 2001). Cette activation est dépendante de la présence d'un ligand extracellulaire qui entraîne une dimérisation du récepteur ainsi qu'une autophosphorylation. ALK est capable de se lier aux ligands pleiotrophine, midkine et FAM150A/FAM150B (Stoica et al., 2001; Stoica et al., 2002). Les cascades de signalisation ainsi activées ont un effet anti-apoptotique et favorisent la prolifération cellulaire. Cependant, le rôle physiologique de la protéine ALK n'est aujourd'hui pas totalement élucidé. Par exemple, un KO de la protéine ALK chez la souris n'entraîne que quelques défauts mineurs des fonctions cérébrales, avec par exemple un retard de la reconnaissance d'objets ou encore un léger défaut de la mémoire spatiale (Bilsland et al., 2008; Weiss et al., 2012).

La protéine ALK est fréquemment associée à différents cancers. Il existe trois types de dérégulations possibles : les mutations ponctuelles, la surexpression et les réarrangements chromosomiques (Pour revue (Hallberg and Palmer, 2013)). Dans les trois cas on observe une activation constitutive de la kinase, indépendante de la présence d'un ligand, en dehors du système nerveux. Dans le cas de translocations chromosomiques, plus de 20 partenaires à ALK

ont été décrits. Dans le cadre de cette thèse nous nous sommes intéressés à la fusion NPM1-ALK.

3. La protéine de fusion NPM1-ALK

La protéine de fusion NPM1-ALK a été décrite pour la 1^e fois en 1994 chez un patient atteint d'un lymphome non-Hodgkinien (Morris et al., 1994). La fusion a lieu entre le domaine d'oligomérisation N-terminal de NPM1 et le domaine tyrosine kinase C-terminal d'ALK (**Figure 10**). Cette fusion entraîne une dimérisation ainsi qu'une activation constitutive de la kinase ALK en absence de ligand.

Il a été montré que la protéine NPM1-ALK était localisée aussi bien dans le noyau que dans le cytoplasme (Bischof et al., 1997). De plus, cette étude montre qu'il est possible de conserver l'effet oncogène de la kinase en remplaçant le domaine d'oligomérisation de NPM1 par un autre domaine de dimérisation de type fermeture éclair à leucine. La protéine est alors uniquement cytoplasmique dans cette construction. Cependant une étude plus récente a montré que la régulation de NPM1-ALK par le domaine NPM1 de cette dernière était essentielle à la survie cellulaire (Ceccon et al., 2016). En effet la kinase NPM1-ALK est capable de se dimériser avec elle-même, ou bien avec la protéine NPM1 endogène. La dimérisation avec NPM1 entraîne une localisation à la fois cytoplasmique et nucléaire de NPM1-ALK. En absence de NPM1 la protéine de fusion est uniquement cytoplasmique, ce qui s'accompagne d'une induction de l'apoptose cellulaire via ATM/CHK2. Lors de la surexpression de la protéine NPM1-ALK dans des lignées ALCL (ALK+) les auteurs observent également une accumulation de NPM1-ALK dans le cytoplasme et une apoptose cellulaire. Ce phénomène n'est pas observé lors de la surexpression d'une forme inactive de la kinase de fusion. Ces résultats semblent indiquer qu'il existe une régulation de NPM1-ALK par NPM1 capable de séquestrer la protéine de fusion dans le noyau sous forme inactive. Lorsque l'équilibre entre NPM1-ALK et NPM1 est rompu on observe une accumulation de la kinase de fusion active dans le cytoplasme, accompagnée de l'apoptose cellulaire. Ces observations peuvent expliquer la difficulté des modèles de surexpression de NPM1-ALK à transformer les cellules car une trop forte expression de NPM1-ALK devient toxique.

La protéine de fusion NPM1-ALK est capable d'activer un certain nombre de voies de signalisation cellulaire impliquées dans des processus tels que la survie et la prolifération cellulaires. Certaines cibles de la protéine de fusion sont aujourd'hui bien identifiées (**Figure 11**).

NPM1-ALK est capable d'activer la voie PI3K/AKT par phosphorylation (Bai et al., 2000; Slupianek et al., 2001). L'activation de la kinase de fusion entraîne la phosphorylation de PI3 et permet l'activation des protéines AKT1 et AKT2. Via l'activation de la voie mTOR et la séquestration cytoplasmique de FOXO3A, AKT permet l'inhibition de l'apoptose et la progression du cycle cellulaire (Gu et al., 2004). L'utilisation d'inhibiteurs de PI3-kinase ou de mutants d'AKT supprime les propriétés transformantes de NPM1-ALK (Slupianek et al., 2001).

La protéine de fusion peut également activer la voie MEK/ERK (Marzec et al., 2007), entraînant notamment l'activation du facteur anti-apoptotique BCL-XL et des facteurs de prolifération cellulaire CDK4 et P-RB ainsi que de m-TOR. L'activation de cette voie entraîne également l'activation de JUN-B qui permet l'expression du marqueur CD30. En retour CD30 est capable d'activer l'expression de JUN-B (Watanabe et al., 2005).

Enfin la voie la mieux décrite est la voie JAK3/STAT3 (Amin et al., 2004; Chiarle et al., 2005; Zamo et al., 2002). L'activation de STAT3 est possible par phosphorylation directe via NPM1-ALK, ou par l'intermédiaire de JAK3 (Zamo et al., 2002). L'activation de STAT3 entraîne l'activation de BCL-XL, de Cycline-D3 ainsi que de C-MYC alors qu'une inhibition de la voie STAT3 entraîne un arrêt du cycle cellulaire. STAT3 permet également l'activation de C/EBP β participant à la prolifération cellulaire en coopération avec la voie MAPK (Anastasov et al., 2010). De plus, STAT3 permet la transcription de DNMT1, une méthyltransférase responsable du maintien de la méthylation des promoteurs au cours du cycle cellulaire (Zhang et al., 2006). L'activation de DNMT1 pourrait entraîner des modifications épigénétiques participant à l'oncogenèse. Il existe également une boucle de rétro-contrôle positif de DNMT1 sur l'expression de STAT3. L'activation de la voie STAT3 serait également responsable de la perte de certains marqueurs T au cours de l'oncogenèse (Ambrogio et al., 2009). Enfin, STAT3 permet l'inhibition de la sénescence cellulaire par inhibition de P16INK4A (Martinelli et al., 2011).

D'autres voies pourraient également être activées par NPM1-ALK comme le montrent deux études de spectrométrie de masse qui ont identifié plus de 46 protéines interagissant avec NPM1-ALK (Crockett et al., 2004; Wu et al., 2009). Notamment ces travaux ont montré une interaction de NPM1-ALK avec des protéines impliquées dans l'organisation des centrosomes (CDC14), le transport nucléaire (NUP98) et le remodelage de la chromatine (SNF2). Wu et al. ont également mis en évidence l'interaction de NPM1-ALK avec plusieurs protéines impliquées dans la réparation de l'ADN (en particulier MCM6 et MSH2). MSH2 est une protéine importante pour la réparation des mésappariements. Deux études ultérieures ont confirmé le lien entre NPM1-ALK et MSH2 (Bone et al., 2015; Young et al., 2011). La dérégulation de MSH2 entraîne une perte des mécanismes de réparation des mésappariement, pouvant participer à l'oncogenèse.

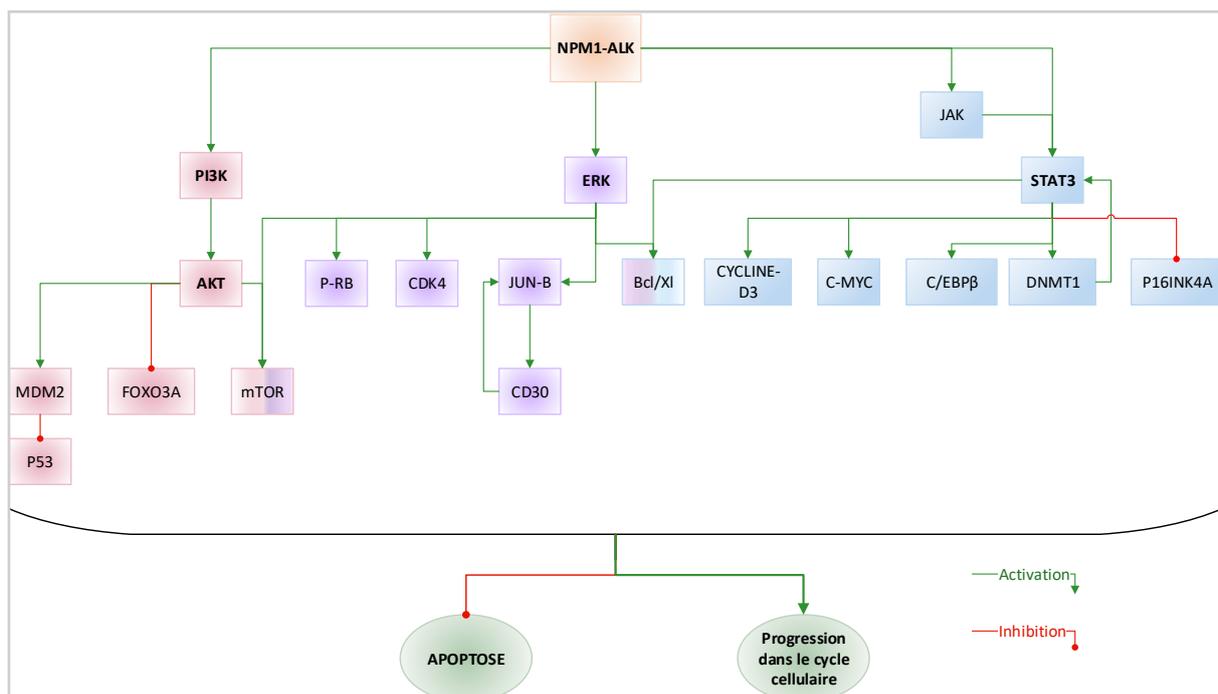


Figure 11 : Principales voies de signalisation cellulaire activées par NPM1-ALK. La protéine de fusion NPM1-ALK permet l'activation des voies PI3K/AKT, ERK et STAT3 qui permettent la prolifération cellulaire et inhibent l'apoptose. (Données non exhaustives).

4. Potentiel oncogénique de NPM1-ALK

Des études réalisées dans des lignées cellulaires murines BA/F3 et NIH3T3 ont permis de mettre en évidence le potentiel oncogénique de NPM1-ALK (Bai et al., 1998; Fujimoto et al., 1996). En effet, la surexpression de la protéine de fusion humaine dans ces cellules induit la capacité des NIH3T3 à pousser en agar mou, démontrant une perte de l'inhibition de contact, et la capacité des BA/F3 à pousser en absence d'IL3 (normalement nécessaire à leur survie). L'injection de ces cellules chez la souris permet d'induire des tumeurs (Bischof et al., 1997; Slupianek et al., 2001).

En 2013, une équipe est parvenue à transformer des cellules CD4 humaines par surexpression de la protéine de fusion, en utilisant une stratégie lentivirale (Zhang et al., 2013a). Les cellules transformées expriment faiblement le marqueur CD3, et fortement les marqueurs CD30, CD4 et CD25. Les cellules expriment également les molécules immunosuppressives IL-10 et PD-L1. La greffe de ces cellules chez la souris a permis d'induire la formation de tumeurs possédant une morphologie proche des ALCL de patients.

Ces différentes études mettent en avant le potentiel oncogénique de la protéine de fusion, cependant, comme nous l'avons vu précédemment, le gène de fusion *NPM1-ALK* peut être détecté dans le sang d'individus sains (Trumper et al., 1998), la fusion n'est donc pas toujours suffisante pour induire l'oncogenèse et des évènements secondaires pourraient être nécessaires.

B. Lymphome Anaplasique à grandes cellules (ALCL)

1. Présentation générale

Le lymphome anaplasique à grandes cellules (ALCL) a été décrit pour la première fois en 1985 par Stein et al. (Stein et al., 1985). Les auteurs ont montré un marquage des cellules par l'anticorps Ki-1 (CD30) utilisé initialement dans le marquage des lymphomes Hodgkiniens. Cependant, aucune cellule de Reed-Sternberg n'est observée, le lymphome anaplasique à grandes cellules appartient ainsi aux lymphomes non-Hodgkiniens. Il est généralement de

phénotype T ou null. Il s'agit d'un lymphome rare qui représente environ 2-5% des lymphomes non Hodgkiniens. Cependant, sa prévalence est supérieure chez l'enfant et l'adolescent, avec une fréquence qui atteint 10-15% (1997; Mora et al., 2000). Dès 1989, la présence d'une translocation entre les chromosomes 2 et 5 a été découverte dans certains ALCL (Kaneko et al., 1989; Le Beau et al., 1989; Rimokh et al., 1989). En 1994, Morris et al. montrent que cette translocation est à l'origine d'une protéine de fusion NPM1-ALK (Morris et al., 1994). D'autres translocations ont également été décrites, permettant l'activation de la protéine ALK (Boi et al., 2015). Aujourd'hui les ALCL sont ainsi divisés en deux catégories par la classification WHO (*World Health Organization*): les ALCL (ALK+) et les ALCL (ALK-) (Matutes, 2018).

2. Histologie

Bien que les ALCL présentent une morphologie cellulaire relativement homogène, il a été montré que de nombreux sous-types pouvaient exister (Benharroch et al., 1998). En 2008, la classification WHO a décrit cinq profils différents parmi les ALCL : les ALCL communs, les ALCL lymphohistiocytiques, les ALCL à petites cellules, les ALCL « Hodgkin-like », et les ALCL composites (**Figure 12**) (Sabattini et al., 2010). Un type cellulaire particulier (cellules marqueurs) présent dans l'ensemble des ALCL a cependant pu être décrit : il s'agit de grandes cellules lymphoïdes, possédant un cytoplasme abondant ainsi qu'un noyau excentré en forme de fer à cheval ou de rein, comprenant généralement une région éosinophile proche du noyau.

Le type commun représente 60% des ALCL et présente une majorité de « cellules marqueurs » décrites précédemment. Les cellules poussent généralement de manière cohésive et peuvent ainsi être confondues avec des tumeurs métastatiques. Le type lymphohistiocytique représente 10 % des ALCL. Il est caractérisé par des tumeurs agrégées autour de vaisseaux sanguins. Les « cellules marqueurs » sont associées à des histiocytes (Pileri et al., 1990). Le diagnostic repose sur la détection du marqueur CD30 permettant de distinguer les cellules tumorales des histiocytes. Le type à petites cellules représente 5-10% des ALCL. Il est caractérisé par la présence d'une population de petites cellules avec des noyaux irréguliers et un cytoplasme abondant (Kinney et al., 1993). Le diagnostic est difficile car les cellules marqueurs sont masquées par la présence des petites cellules. De plus, il a été montré qu'elles

exprimaient peu voire pas le marqueur CD30 (Falini et al., 1999a). Le type « Hodgkin-like » représente 3% des ALCL. Il présente une morphologie similaire à celle des lymphomes Hodgkiniens avec la présence de nodules (Leoncini et al., 1990). Le type composite représente 10-30% des ALCL. Il s'agit d'un mélange de plusieurs types morphologiques présentés ci-dessus.

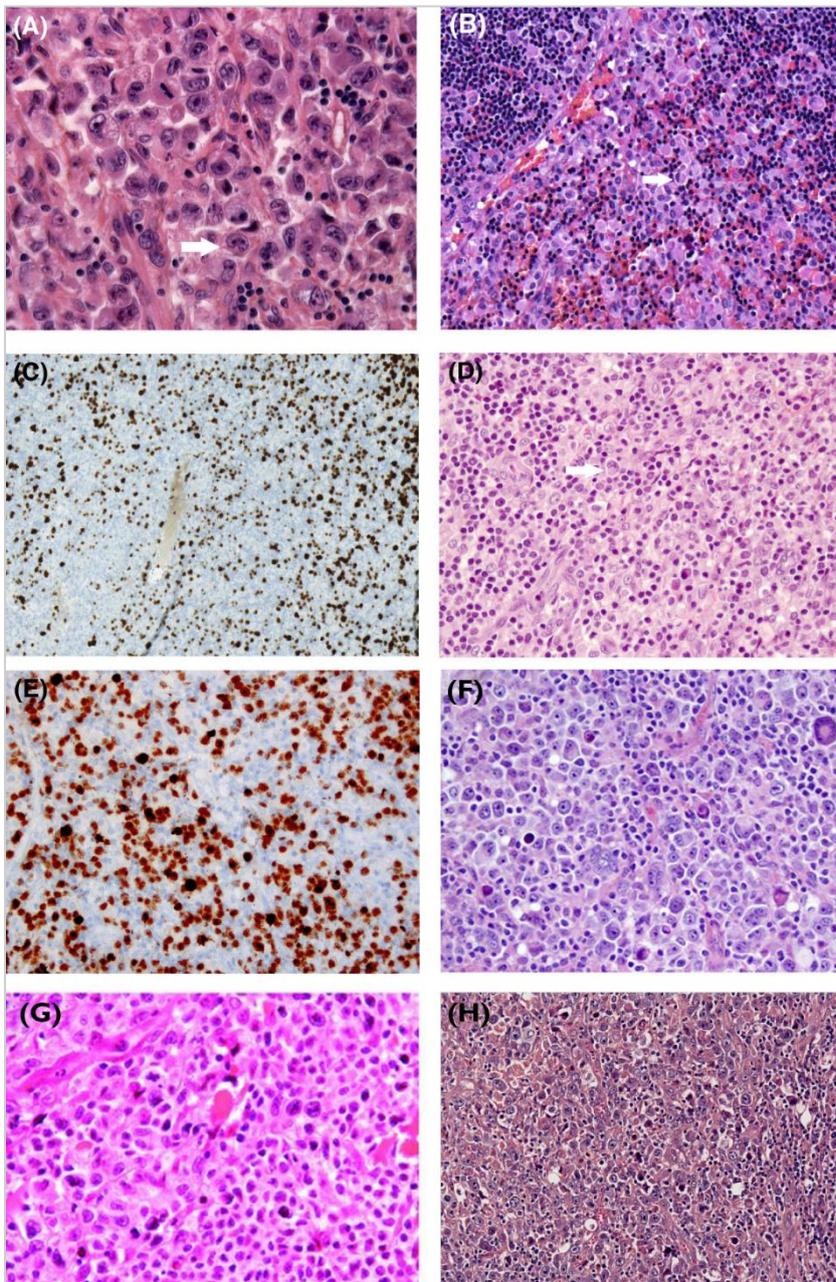


Figure 12: Coupes histologiques des différents types d'ALCL (Adapté de (Lamant et al., 2011; Turner et al., 2016; Wang and Jaffe, 2018)). A) ALCL commun caractérisé par la présence des « cellules marqueurs » de grande taille. B)C) ALCL lymphohistiocytique contenant des « cellules marqueurs » mélangées aux histiocytes. En C un marquage CD30 permet de mettre en évidence les cellules tumorales. D)E) ALCL à petites cellules. En D un marquage ALK permet de différencier les cellules tumorales des petites cellules avoisinantes. F) ALCL « Hodgkin-Like » avec présence de nodules. G) ALCL composite. (H) Exemple d'ALCL (ALK-).

3. Réarrangements chromosomiques présents dans les ALCL ALK+

La translocation t(2,5) est présente dans environ 80 % des ALCL (ALK+) (Falini et al., 1999b). Cependant, il existe d'autres translocations permettant l'activation constitutive de la kinase ALK dans les ALCL. La 2^e translocation la plus fréquente est la translocation t(1,2) menant à une fusion TPM3-ALK (Lamant et al., 1999). D'autres fusions, plus rares, ont également été décrites : TFG-ALK, CLTC-ALK, ATIC-ALK et MSN-ALK (Bridge et al., 2001; Colleoni et al., 2000; Hernandez et al., 1999; Tort et al., 2004; Touriol et al., 2000). Une étude de 2004 s'est intéressée aux différences entre ces fusions (Armstrong et al., 2004). Les auteurs ont surexprimé chacune d'entre elle dans une lignée cellulaire murine NIH3T3 et ont étudié quelques conséquences cellulaires. Une phosphorylation constitutive de la protéine ALK a été observée dans tous les cas. Concernant la localisation cellulaire, seule NPM1-ALK montre une localisation à la fois cytoplasmique, nucléaire et nucléolaire. Les autres fusions sont uniquement présentes dans le cytoplasme. L'ensemble des fusions permet une croissance en trois dimensions (agar-mou), ainsi qu'une augmentation des capacités migratoires. Par greffe sous-cutané chez la souris, les auteurs ont montré que l'ensemble des fusions permettaient d'induire des tumeurs *in vivo*, cependant les fusions NPM1-ALK et TGF-ALK semblent plus agressives. Au niveau mécanistique, les auteurs observent des différences dans les niveaux d'activation des voies PI3K et STAT3 entre les différentes fusions.

4. Le statut p53 des ALCL ALK+

La protéine P53 joue un rôle essentiel dans la surveillance du génome, et se trouve fréquemment modulée dans de nombreux cancers. Dans le cas des ALCL, le gène *TP53* est rarement directement muté (< 10%) (Cesarman et al., 1993; Rassidakis et al., 2005). Rassidakis et al. ont même montré que le gène *TP53* était surexprimé dans 65% des tumeurs ALCL étudiées. Cependant, des études plus récentes ont mis en évidence une régulation de P53 par NPM1-ALK. En effet, les auteurs ont montré que NPM1-ALK était capable d'inhiber l'activité de P53 via l'induction de JNK et MDM2 (Cui et al., 2009; McDuff and Turner, 2011). NPM1-ALK permet l'activation de MDM2 par la voie PI3K. MDM2 se lie à P53 et la séquestre dans le cytoplasme, empêchant son activité. Dans le cas de transformation cellulaire *in vitro*, McDuff

et al. ont montré que l'absence de P53 dans des fibroblastes embryonnaires de souris était indispensable à la transformation cellulaire par NPM1-ALK.

5. Une origine cellulaire controversée

L'origine cellulaire des ALCL (ALK+) est une question qui fait débat depuis de nombreuses années. En effet, bien que l'ALCL soit classifié comme un lymphome T, les marqueurs exprimés en surface des cellules tumorales ne permettent pas de déterminer de manière fiable le précurseur tumoral.

Historiquement les ALCL ont été caractérisés par la présence du marqueur CD30 (Stein et al., 1985). Cependant ce marqueur peut parfois être absent dans certains types d'ALCL (Falini et al., 1999a). De plus, CD30 n'est pas un marqueur spécifique des ALCL, il est par exemple présent à la surface des lymphocytes B et T activés (Stein et al., 1982). Il est également exprimé lors d'infection par EBV (Epstein Barr Virus) (Abbondanzo et al., 1990) ou lors de la stimulation PHA (Phytohemmaglutinine) (Stein et al., 1985).

L'étude immunologique des ALCL a mis en évidence la présence d'au moins un marqueur T (CD2, CD4, CD3, CD7, CD4 ou CD8) dans les tumeurs (Juco et al., 2003). 65% des tumeurs expriment le marqueur CD4 et 20% le marqueur CD8. Cependant, seulement 75% des tumeurs expriment le marqueur commun CD3 et seuls 4% expriment le TCR (Bonzheim et al., 2004). Malgré cette observation des analyses génétiques ont permis de mettre en évidence l'existence de réarrangements du locus *TCR* dans les ALCL (Malcolm et al., 2016; O'Connor et al., 1987). Certains de ces réarrangements sont aberrants et ne devrait pas pouvoir passer la sélection lymphocytaire β en conditions normales. Dans cette étude les auteurs ont utilisé un modèle murin surexprimant NPM1-ALK dans le lignage T en présence ou en absence de la protéine RAG2 responsable des réarrangements du *TCR*. En absence de RAG2 les cellules NPM1-ALK sont capables de passer la sélection β à l'inverse des cellules WT. Dans les souris (RAG2^{-/-}) et (RAG2^{+/+}) la présence de NPM1-ALK permet le développement de tumeurs T. La présence du TCR n'est donc pas nécessaire à l'oncogénèse. Cependant la tumorigénèse est plus rapide dans les souris (RAG2^{+/+}) ce qui pourrait indiquer un rôle du TCR favorisant

l'apparition de tumeurs, son expression serait ensuite perdue. Enfin, différentes études ont mis en évidence l'expression de marqueurs cytotoxiques tels que TIA-1 et granzyme-B (Foss et al., 1996; Krenacs et al., 1997).

L'ensemble de ces résultats semble indiquer une origine T potentiellement cytotoxique dans l'ALCL, les marqueurs seraient ensuite partiellement perdus par les cellules ALCL. Une étude a notamment montré que NPM1-ALK était capable d'inhiber l'expression de facteurs T par la voie STAT3 (Ambrogio et al., 2009). De plus, NPM1-ALK active des voies similaires à celles activées par le TCR.

A l'inverse, des travaux réalisés dans des lignées cellulaires dérivées d'ALCL ont mis en évidence l'existence d'une sous-population présentant des caractéristiques de progéniteurs immatures lors d'une analyse en RNAseq (Moti et al., 2015). Ces résultats pourraient laisser imaginer une origine plus immature des cellules d'origine de l'ALCL. Cependant, il s'agit d'une analyse sur lignées cellulaires, les résultats obtenus pourraient être différents avec l'utilisation d'échantillons de patients. Une seconde étude, réalisée sur des cellules de patients ALCL (ALK+) et (ALK-), a tenté d'identifier l'origine cellulaire des ALCL en utilisant le profil de méthylation de l'ADN (Hassler et al., 2016). Hassler et al. ont montré que le méthylome des ALCL (ALK+) était très proche de celui des progéniteurs T multipotents (CD34+/CD1a-) alors que les ALCL (ALK-) étaient plus proches des cellules T pre-TCR ou double positives. Ces résultats semblent indiquer une origine T immature des ALCL (ALK+). Cependant il n'est pas exclu que l'ALCL (ALK+) se forme dans des cellules T matures qui acquièrent secondairement un profil plus immature lors de l'oncogenèse.

6. Signatures moléculaires

a) Profil transcriptionnel

Différentes études ont tenté d'identifier une signature transcriptionnelle spécifique aux ALCL, et de mettre en évidence des gènes dont l'expression varie entre les ALCL (ALK+) et (ALK-). Deux études en 2005 et 2007 ont utilisé la technologie des puces à ADN afin de comparer le profil d'expression des ALCL (ALK+) et (ALK-) à partir d'échantillons de patients (Lamant et al., 2007; Thompson et al., 2005). Dans ces études les auteurs ont découvert des profils d'expression différents en fonction du statut ALK. Lamant et al. ont notamment mis en évidence 303 gènes différentiellement exprimés entre les (ALK+) et les (ALK-), avec une très forte surexpression des gènes *BCL6*, *PTPN12*, *CEBPB* et *SERPINA1* dans les ALCL (ALK+). Globalement les deux études ont montré une suractivation de voies de transduction de signal telles que SYK, LIN, CDC37, NF- κ B et une répression de facteurs de transcriptions tels que HOXC6, HOXA3, TTF-1, ADN(cytosine-5-)methyltransférase, ATF-5. La transition G1/S pourrait également être régulée différemment entre les (ALK+) et (ALK-), avec la surexpression de la cycline D3 dans les (ALK+) et la répression de l'inhibiteur de cycle cellulaire P19INK4D dans les (ALK-). Malgré des profils très différents, certaines kinases sont activées dans les deux types d'ALCL. Notamment LCK, protéine kinase C, VAV2, et NKIAMRE, ainsi que des molécules anti-apoptotiques (NIP-1, E1B 19K/BCL-2, PAR-3) pouvant révéler certains mécanismes communs.

Des études de séquençage nouvelle génération ont par la suite été réalisées, permettant d'obtenir des profils transcriptomiques plus précis (Eckerle et al., 2009; Piva et al., 2010). Lors de ces analyses, les auteurs ont utilisé des cellules T normales comme contrôle. Ils ont retrouvé une différence de profil entre ALCL (ALK+) et (ALK-), avec cependant une trentaine de gènes différentiellement exprimés dans les deux types comparés aux cellules T contrôle. Ces résultats montrent l'existence d'une signature ALCL commune. Les gènes les plus surexprimés dans les ALCL (ALK+) sont *IL1RAP*, *GAS1*, *PRF1*, *TMEM158*, et *IL2RA* dans l'étude de Piva et al, et *PRF1*, *IL1RAP*, *CCND3*, *BCL3*, et *GAS1* dans l'étude d'Eckerle et al. De plus, Piva et al ont montré que les deux tiers des gènes surexprimés dans les ALCL (ALK+) étaient liés à la voie STAT3. En effet, l'utilisation d'un shSTAT3 entraîne une importante diminution de la signature transcriptionnelle.

b) Les ARN non codants

(1) Les micro-ARN

Les micro-ARN sont de petits ARN non codants (de 21-24 nucléotides) permettant l'inhibition de la traduction ou la dégradation de leurs ARNm cibles. Ils participent ainsi à la régulation de l'expression génique.

En 2010, une première étude a permis de déterminer une signature micro-ARN pour les ALCL (ALK+) et (ALK-) à l'aide de puce micro-ARN (Merkel et al., 2010). Cette étude regroupe des données obtenues chez des patients, dans des modèles murins ainsi que dans des lignées cellulaires ALCL. Les auteurs ont ainsi pu identifier différents micro-ARN modulés dans les ALCL. Par exemple, l'expression du mir-101 est fortement diminuée dans l'ensemble des ALCL ((ALK+) et (ALK-)), cependant en forçant son expression les auteurs ont observé une diminution de la prolifération cellulaire uniquement dans les ALCL (ALK+). L'action du mir-101 pourrait se faire par la voie m-TOR. Une augmentation des mir-17-92 est observée dans les ALCL (ALK+) alors que le mir-155 est augmenté dans les ALCL (ALK-). Quatre micro-ARN sont modulés dans l'ensemble des échantillons utilisés : miR-29a, miR-29c, miR-101, and miR-150. En comparant les ALCL (ALK+)/(ALK-) les auteurs identifient notamment les micro-ARN miR-17, miR-20a, miR-20b, miR-106a, et miR-886-3p comme spécifiques des (ALK+). Une étude de 2015, a par la suite montré un mécanisme de régulation original pour le mir-150 dans les ALCL (ALK+) (Hoareau-Aveilla et al., 2015). En effet, les auteurs ont montré que la méthylation du gène MIR150 par DNMT1 (activé par STAT3) entraînait une diminution de l'expression du miR-150. Une augmentation de l'expression de mir-150 entraîne un arrêt du cycle cellulaire *in vitro* ainsi qu'une meilleure réponse au traitement des cellules initialement résistantes à la Crizotinib. Les auteurs proposent l'utilisation en clinique d'agents hypométhylants lors de résistance aux thérapies classiques. Une seconde étude en 2015 a établi une nouvelle signature micro-ARN dans les ALCL (ALK+)/(ALK-) par séquençage de nouvelle génération (Steinhilber et al., 2015). Les auteurs ont identifié 56 micro-ARN significativement modulés entre les ALCL (ALK+) et (ALK-). Les meilleurs candidats sont miR-340, miR-203, miR-135b, miR-182, miR-183 surexprimés dans les ALCL (ALK+) et miR-196b, miR-155, miR-146a, miR-424, miR-503, miR-424, miR-542-3p réprimés dans les ALCL (ALK+). De même que dans l'étude de

Merkel et al. les auteurs ont retrouvé une surexpression du miR-17-92 dans les ALCL (ALK+). En s'intéressant plus particulièrement aux ALCL (ALK+) surexprimant C/EBP β les auteurs ont montré un rôle du miR-181 dans la différenciation T et la régulation du TCR.

D'autres travaux ont montré le rôle du micro-ARN miR-135b (Matsuyama et al., 2011). Ce micro-ARN est surexprimé dans les ALCL. Il entraîne l'inhibition de FOXO1 et induit une résistance aux agents thérapeutiques tels que la cytosine β -d-arabinofuranoside.

Une étude récente a décrit le rôle d'un nouveau micro-ARN dans les ALCL (ALK+) : le cluster miR-497-195 (Hoareau-Aveilla et al., 2019). Dans les cellules ALCL (ALK+) ces deux micro-ARN présentent une expression diminuée. Les auteurs ont montré que cette diminution était directement liée à l'expression de NPM1-ALK ainsi qu'à la méthylation des promoteurs des gènes MIR195 et MIR497. Lors de la réexpression de miR-497 dans les cellules tumorales, une diminution de la viabilité cellulaire a pu être observée. Cet effet est médié par l'inhibition de la progression dans le cycle cellulaire via la répression de CCNE1, CDC25A, CDK6 et E2F3 impliqués dans le passage G1/S.

L'ensemble de ces différentes études montre le rôle des micro-ARN dans les ALCL et ouvre la voie à l'utilisation de nouvelles cibles thérapeutiques ou de nouveaux biomarqueurs. L'ensemble des micro-ARN dont la fonction est clairement démontrée dans l'ALCL est présenté **Tableau 1**.

mir	Voies de signalisation impliquées	Effet Oncogène/Suppresseur de tumeur
mir-101	PI3K/AKT/mTOR	Suppresseur de tumeur
mir-155	TLR	Suppresseur de tumeur
mir-17-92	STAT3	Oncogène
mir-29a	Apoptose (MCL-1)	Suppresseur de tumeur
mir-16	VEGF	Suppresseur de tumeur
mir-135b	Th2 polarisation (STAT6, GATA3)	Oncogène
mir-181a	différenciation T, TCR	Suppresseur de tumeur
mir-146a	NF-kappaB	Suppresseur de tumeur
mir-150	différenciation T, TCR	Suppresseur de tumeur
mir-497	Cycle cellulaire (CDK6)	Suppresseur de tumeur
mir-125b	ADN hyperméthylation	Suppresseur de tumeur

Tableau 1 : Liste des micro-ARN dont le rôle a été démontré dans les ALCL (Liste non exhaustive)

(2) Les longs ARN non codants et les petits ARN nucléolaires (snoRNA)

Les longs ARN non codants ont été décrits comme des ARN de plus de 200 nucléotides, généralement intergéniques. Il a notamment été montré qu'ils étaient capables de recruter des remodeleurs de chromatine et ainsi moduler l'expression de gènes (Derrien et al., 2012).

Une 1^e étude a comparé l'expression des longs ARN non codants entre des cellules saines de patients et des cellules ALCL (ALK+) correspondantes (Chung et al., 2017). Ces travaux ont été effectués à l'aide de puces spécifiques des longs ARN non codants. Les auteurs ont mis en évidence la surexpression de 20 et la répression de 31 longs ARN non codant dans les cellules (ALK+). Ils se sont intéressés en particulier à LINC01013. La déplétion de cet ARN entraîne une suppression des capacités d'invasion des cellules. Une seconde étude s'est intéressée à MIR503HG surexprimé dans les ALCL (ALK-) (Huang et al., 2018). Les auteurs ont montré que cet ARN non codant jouait un rôle dans la prolifération cellulaire. Ce rôle pourrait être médié par la voie miR-503/SMURF2/TGFBR.

Les snoARN sont formés à partir de régions introniques. Ils jouent un rôle dans la modification des autres classes d'ARN et notamment des ARN ribosomiaux (Dupuis-Sandoval et al., 2015). Cependant, leur fonction exacte n'est aujourd'hui pas totalement élucidée. Une étude sur les snoARN présents dans les lymphomes T a montré une diminution globale des snoARN dans le cancer. 30 snoARN ont été identifiés comme spécifiques de l'ALCL, cependant aucune analyse fonctionnelle n'a été réalisée.

Ces trois études sont les seules existantes à ce jour sur un rôle des longs ARN non codants et des snoARN dans les ALCL. Avec l'émergence d'outils de séquençage de plus en plus performants on peut imaginer des études futures plus nombreuses permettant d'identifier des cibles thérapeutiques et/ou des biomarqueurs.

7. Pronostic et traitements

La première ligne de traitement des ALCL consiste en une chimiothérapie cumulant quatre agents chimiques : cyclophosphamide, doxorubicine, vincristine et prednisone (CHOP) (Querfeld et al., 2010). Mais des résistances sont parfois développées et la radiothérapie est alors utilisée en combinaison à la chimiothérapie.

Les patients atteints de tumeurs (ALK+) présentent un meilleur pronostic que les patients (ALK-) avec un taux moyen de survie à long terme de 70-90 % contre 40-60 % dans les ALCL (ALK-) (Falini et al., 1999a; Gascoyne et al., 1999; Schmitz et al., 2010; Sibon et al., 2012).

Des traitements plus ciblés existent également. Notamment le Brentuximab Vebotin ciblant le marqueur CD30 a démontré son efficacité (Donato et al., 2018). Il s'agit d'un IgG anti-CD30 couplé à l'auristatine E, une molécule perturbant le réseau de microtubules, dont l'efficacité a été démontrée dans différents cancers. La réponse au traitement est indépendante du statut ALK de la tumeur ainsi que du nombre de thérapies effectuées précédemment.

Dans le cas des ALCL (ALK+) des molécules ciblant spécifiquement la protéine ALK ont été développées. La Crizotinib est la première molécule utilisée en clinique, dans les cancers du poumon à petites cellules (ALK+) dans un premier temps (Gerber and Minna, 2010). Les résultats très prometteurs obtenus dans ce type de cancer ont mené à l'utilisation de cet inhibiteur dans le cas des ALCL (Gambacorti-Passerini et al., 2011; Gambacorti Passerini et al., 2014; Mosse et al., 2013). Dans ces différents essais cliniques une rémission complète a été observée chez 9/11 patients et 8/9 patients résistants à la chimiothérapie. Ces thérapies ciblées présentent donc une bonne efficacité. Cependant le développement de résistances aux inhibiteurs de ALK a également été observé dans différents types de cancers (ALK+) (Choi et al., 2010; Sasaki et al., 2010; Shaw et al., 2013) dont les ALCL (Gambacorti Passerini et al., 2014). Deux mécanismes de résistances acquises principaux ont été décrits : des mutations dans le domaine tyrosine kinase de ALK empêchant la fixation de l'inhibiteur, ou une activation des voies K-RAS, EGFR ou C-KIT compensant l'inactivation de la protéine ALK (Vijayvergia and Mehra, 2014). Des inhibiteurs de seconde génération ont par la suite été développés afin de contrer l'apparition de résistances dues aux mutations du domaine tyrosine kinase tels que le Brigatinib ou l'Alactinib (Kodama et al., 2014; Sabari et al., 2017). Des inhibiteurs de troisième

génération ont également été générés, notamment Lorlatinib qui cible à la fois ALK et ROS1 (fréquemment impliqué dans les résistances aux inhibiteurs). Des essais pré-cliniques et clinique ont montré des résultats très encourageants (Shaw et al., 2017; Zou et al., 2015).

Des mécanismes de résistance intrinsèques peuvent également exister, notamment lors de la présence de cellules souches cancéreuses. Il a notamment été montré dans des lignées cellulaires NPM1-ALK que l'utilisation d'un inhibiteur de mTOR, permettant de cibler les cellules souche cancéreuses, avait un effet synergique avec la Crizotinib (Redaelli et al., 2016). De plus, cet inhibiteur permet de ré-induire une sensibilité à la Crizotinib dans des cellules résistantes (Moore et al., 2014).

Récemment il a été proposé d'utiliser des cellules T porteuses d'un récepteur chimérique (CAR-T) ciblant CD30 dans les ALCL. Les cellules CAR-T sont des cellules lymphocytaires de type T chimériques. Elles sont modifiées génétiquement afin d'exprimer un récepteur spécifique à leur surface (ici un récepteur anti-CD30) et d'activer une réponse immunitaire en aval. Les premières cellules CAR-T ont été générées en 1993 (Eshhar et al., 1993). L'utilisation de ces cellules a notamment montré une très bonne efficacité dans les leucémies myéloïdes aiguës, en utilisant un marqueur anti-CD19 (Pour revue (Hay and Turtle, 2017)). Deux essais cliniques ont déjà été réalisés pour le marqueur CD30 (Ramos et al., 2017; Wang et al., 2017). Cependant les résultats obtenus ne sont pas aussi concluant qu'attendus, aucune rémission complète n'a pu être observée. D'autres essais sont actuellement en cours.

8. Les modèles de l'ALCL *in vivo*

La modélisation *in vivo* des ALCL exprimant la fusion NPM1-ALK est un véritable challenge. En effet de nombreuses équipes ont tenté de créer des modèles murins de la pathologie, mais la plupart ne sont pas parvenus à récapituler la maladie humaine.

Deux études ont tenté d'infecter des cellules murines de la moelle osseuse avec un rétrovirus permettant l'expression de NPM1-ALK (Kuefer et al., 1997; Miething et al., 2003). Les cellules ont ensuite été ré-injecter dans la moelle de souris irradiées. Bien que des tumeurs se développent, elles présentent un phénotype très différent de celui des tumeurs de patients,

avec notamment l'absence du marqueur CD30 et l'expression d'un grand nombre de marqueurs B. Aucun lymphome T n'a pu être observé dans ces conditions.

D'autres équipes ont généré des souris transgéniques, exprimant le gène de fusion *NPM1-ALK* sous contrôle de promoteurs particuliers. L'expression sous contrôle du promoteur *Vav*, qui entraîne une expression dans l'ensemble de la lignée hématopoïétique a mené à la formation de lymphomes B et non T (Turner et al., 2003). De même, en forçant une expression majoritairement T à l'aide des promoteurs *Cd4* ou *Cd2* (Cellules T et NK), les auteurs ont observé la formation de lymphome T mais également B (Chiarle et al., 2003; Turner et al., 2006). De plus, les lymphomes T obtenus sont très différents des ALCL. Cependant, Jager et al sont parvenus à induire la formation d'un lymphome T similaire à l'ALCL en exprimant *NPM1-ALK* sous contrôle du promoteur *Lck* (spécifique des cellules T précoces) (Jager et al., 2005). Les auteurs ont observé la formation de tumeurs au niveau du thymus, mais également dans les ganglions lymphoïdes. L'analyse histochimique des tumeurs a révélé l'expression du marqueur CD30 dans une grande fraction de cellules.

Les résultats obtenus dans ces différentes études ont montré la difficulté à reproduire la pathologie humaine chez la souris. Il est possible que des mécanismes espèce spécifique participent au choix de la lignée cellulaire à l'origine des tumeurs. De plus, comme nous l'avons montré précédemment, l'origine cellulaire de l'ALCL n'est pas clairement établie, et le choix du promoteur est donc difficile. Les modèles de surexpression ne permettent pas de reproduire l'ensemble des conséquences de la translocation, ce qui pourrait expliquer la difficulté à reproduire la pathologie.

III. Les ARN circulaires

Les ARN circulaires forment une classe d'ARN particuliers sous forme circulaire et composés de seulement quelques exons de l'ARNm épissés en ordre inverse. L'existence des ARN circulaires a été découverte il y a plus de 30 ans (Hsu and Coca-Prados, 1979), mais ils ont tout d'abord été considérés comme des erreurs d'épissage et par conséquent peu étudiés (Cocquerelle et al., 1993). Des avancées récentes en séquençage de nouvelle génération ont permis de mettre en évidence l'abondance et la conservation de ces ARN non codants (Jeck et al., 2013; Salzman et al., 2012).

A. Synthèse des ARN circulaires

L'étude des ARN circulaires étant récente, les mécanismes de synthèse de ces derniers ne sont pas totalement connus à l'heure actuelle. Il a cependant été montré que la formation des ARN circulaires était dépendante du spliceosome. La mutation des sites d'épissage empêche par exemple la synthèse des ARN circulaires (Ashwal-Fluss et al., 2014). Différentes recherches ont abouti à deux modèles possibles de production des ARN circulaires (Jeck and Sharpless, 2014; Jeck et al., 2013; Lasda and Parker, 2014; Salzman et al., 2012) (**Figure 13**). Les deux modèles sont basés sur la notion de « épissage inverse ». Il s'agit d'un épissage particulier, permettant de relier un site donneur situé en aval d'un site accepteur, à l'inverse de l'épissage canonique qui relie un site donneur en 5' avec un site accepteur en 3'.

Le premier modèle est appelé modèle du saut d'exon. Il implique une première étape d'épissage classique avec la formation d'un lasso intermédiaire suite à un saut d'exon (**Figure 13**). Une seconde étape d'« épissage inverse » a ensuite lieu, menant à la formation d'une structure circulaire contenant les exons présents dans le lasso initial. L'existence de ce mécanisme a notamment été montrée chez *S. pombe* pour le transcrit *mrps16* (Barrett et al., 2015). Cette étude montre cependant que la formation d'un lasso intermédiaire n'est pas suffisante à la synthèse d'ARN circulaire. Une autre étude, réalisée dans des cellules humaines, apporte également des éléments en faveur de ce modèle (Zhang et al., 2016). Dans cet article, les auteurs s'intéressent aux ARN naissants, par blocage de la transcription puis relargage en présence de 4sU. Les ARN naissants sont ensuite isolés à l'aide du système HPDP-Biotine/Streptavidine. Cette étude met en évidence un retard de synthèse des ARN circulaires

par rapport aux ARN linéaires, ceci allant dans le sens d'une formation ultérieure à la transcription.

Le second modèle, ou modèle d' « épissage inverse » direct, fait intervenir l'hybridation de séquences introniques de part et d'autres des exons circularisés (**Figure 13**). L' « épissage inverse » intervient alors avant l'épissage canonique. En effet, chez l'humain, 70% des introns comportent des séquences capables de s'apparier. La mutation de ces séquences par CRISPR entraîne une perte des ARN circulaires correspondants (Ivanov et al., 2015; Zhang et al., 2016). Les séquences impliquées peuvent être des séquences répétées (type ALU) mais également des séquences complémentaires non répétées (Ivanov et al., 2015; Jeck et al., 2013; Zhang et al., 2014). Notamment, l'analyse comparée de gènes homologues Homme/Souris montre la présence d'ARN circulaires uniquement chez l'espèce porteuse de séquences complémentaires dans les introns. De plus, Zhang et al ont montré qu'il était possible d'observer une circularisation lors de la présence d'une demi-répétition ALU (≈ 23 nt). Il est également possible de prédire la formation d'ARN circulaires en se basant sur la présence de séquences homologues dans les introns (Zhang et al., 2014).

La synthèse des ARN circulaires peut être modulée par différents mécanismes. Il est possible de limiter ou favoriser l'appariement des introns. Cela permet la formation des ARN circulaires selon le modèle d' « épissage inverse » direct. Il existe par exemple une compétition entre l'appariement intra-intronique et inter-intronique, limitant ainsi la formation d'ARN circulaires (Zhang et al., 2014). L'appariement des introns peut également être régulé par différentes protéines. Par exemple, les protéines ADAR, responsables de la transformation des Adénosines en Inosines (A \rightarrow I) dans l'ARN, notamment au niveau des séquences ALU, limitent la synthèse des ARN circulaires (Ivanov et al., 2015). A l'inverse la présence de site de fixation pour certaines protéines de liaison à l'ARN dans les introns peut permettre le rapprochement spatial de ces derniers et ainsi favoriser la formation des ARN circulaires. Ce phénomène a été montré pour la protéine Quaking (QKI) lors de la transition épithélio-mésenchymateuse (Conn et al., 2015), ainsi que pour le facteur d'épissage ESPR1 dans les cellules ES (Yu et al., 2017), ou encore pour certaines isoformes de la protéine MBL chez la drosophile, capables de se fixer sur leur transcrit parental et d'augmenter la formation d'ARN circulaire au détriment de l'ARN linéaire (Ashwal-Fluss et al., 2014). La vitesse de transcription, notamment affectée par le

statut chromatinien ou encore les dommages à l'ADN, joue également un rôle dans la synthèse des ARN circulaires : une transcription rapide va favoriser leur production (Zhang et al., 2016). De plus, il a été montré que la taille des exons participant à l'ARN circulaire jouait également un rôle. En effet, la plupart des ARN circulaires contiennent plusieurs exons, et lorsqu'ils n'en contiennent qu'un seul, celui-ci a une taille supérieure à la moyenne (Barrett et al., 2015; Zhang et al., 2014). Ceci pourrait s'expliquer par des contraintes physiques liées à la circularisation. Enfin, il a été montré qu'il était possible d'obtenir des ARN circulaires incluant différents exons à partir d'un même transcrite linéaire (Jeck et al., 2013; Zhang et al., 2014). Il pourrait ainsi exister un mécanisme de compétition entre les ARN circulaires formés à partir d'un même transcrite, en plus de la compétition entre transcrits linéaires et circulaires.

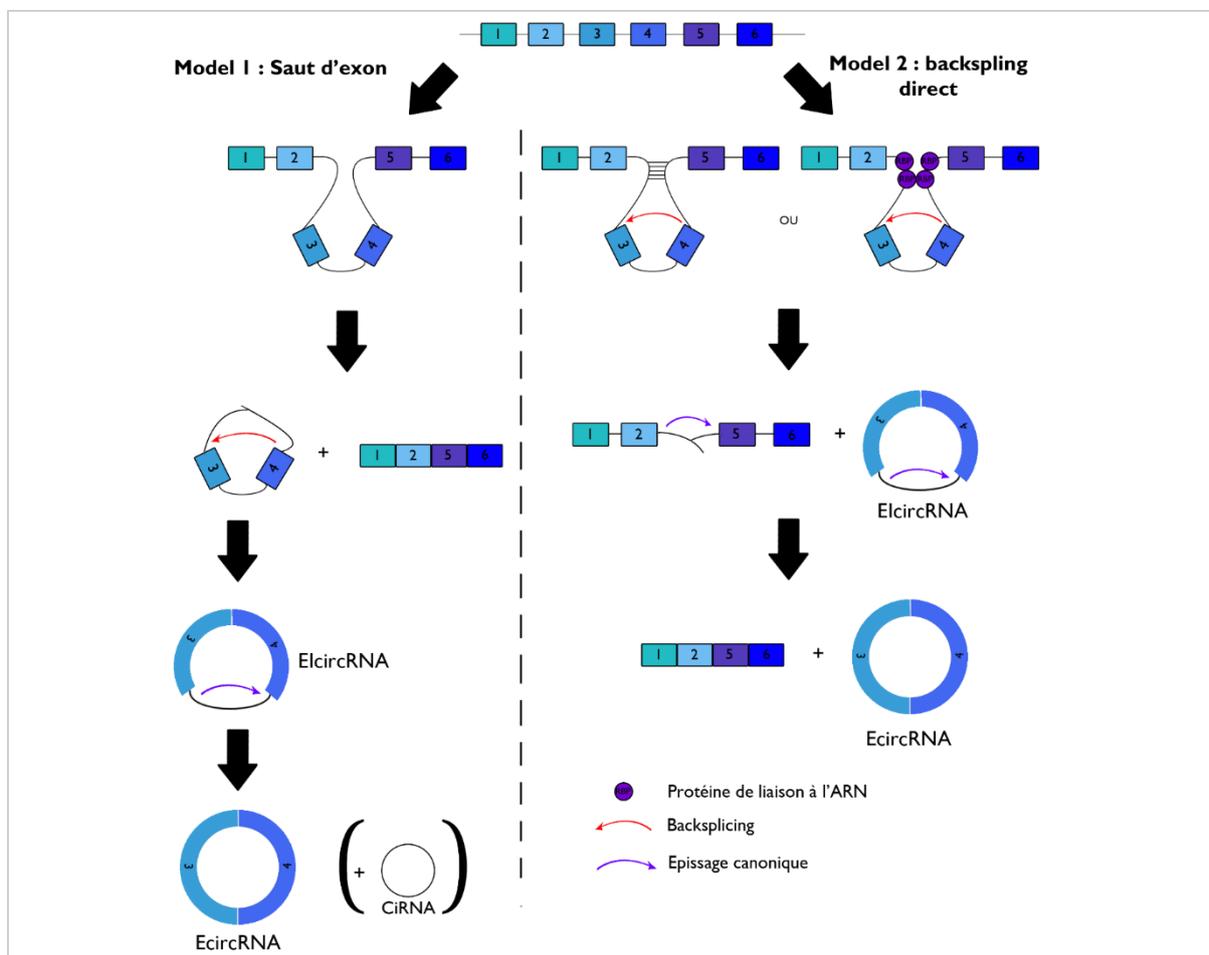


Figure 13: Modèles de synthèse des ARN circulaires. A gauche le modèle de saut d'exon, faisant intervenir l'épissage canonique en 1^{er} suivi d'un épissage inverse. A droite le modèle d'épissage inverse direct faisant intervenir un appariement des séquences introniques par homologies ou par fixation de protéines de liaison à l'ARN. L'étape d'épissage inverse a lieu en premier suivi de l'épissage canonique. Ces 2 modèles permettent de générer 3 types d'ARN circulaires : les ElcircRNAs contenant des séquences exoniques et introniques, les EcircRNA ne contenant que des exons, et les CirRNA ne contenant qu'un intron.

B. Détection des ARN circulaires

Il est possible d'enrichir et de détecter spécifiquement les ARN circulaires à partir d'ARN total en utilisant une 3' → 5' exoribonucléase spécifique : la Rnase R. Cette enzyme permet de dégrader spécifiquement les ARN linéaires et ainsi d'enrichir en ARN circulaires (Salzman et al., 2013; Salzman et al., 2012; Suzuki et al., 2006). Cependant cette méthode ne permet pas la dégradation complète des ARN linéaires. Il a été proposé de coupler la RNaseR avec une technique de déplétion des ARN polyadénylés pour obtenir des banques d'ARN circulaires (Panda et al., 2017). Cette étape supplémentaire permet d'obtenir un meilleur enrichissement avec une élimination presque complète des ARN linéaires. Une fois la purification réalisée il est possible de réaliser une analyse ciblée par RT-PCR ou bien une analyse globale par séquençage de nouvelle génération.

L'analyse par RT-PCR consiste à utiliser des amorces divergentes au niveau de la région d'intérêt (Panda and Gorospe, 2018). Cette méthode permet d'amplifier spécifiquement les ARN circulaires d'intérêt sans amplification des ARN linéaires. On peut ainsi obtenir un profil en barreaux d'échelle correspondant aux ARN circulaires incluant différents exons/introns **(Figure 14)**.

L'analyse globale consiste quant à elle, à réaliser un séquençage à partir de l'ARN extrait et enrichi ou non par traitement RNaseR. La difficulté de cette analyse réside dans le fait qu'il n'existe pour l'instant aucun programme consensus permettant la détection systématique des ARN circulaires. En effet de nombreux programmes existent, adaptés d'algorithmes utilisés pour des analyses de séquençage d'ARN classiques dans lesquels les jonctions d'épissage inverse sont spécifiquement détectées. De nombreux biais peuvent apparaître à chaque étape (préparation des banques d'ARN, alignement...) rendant l'analyse des données complexe.

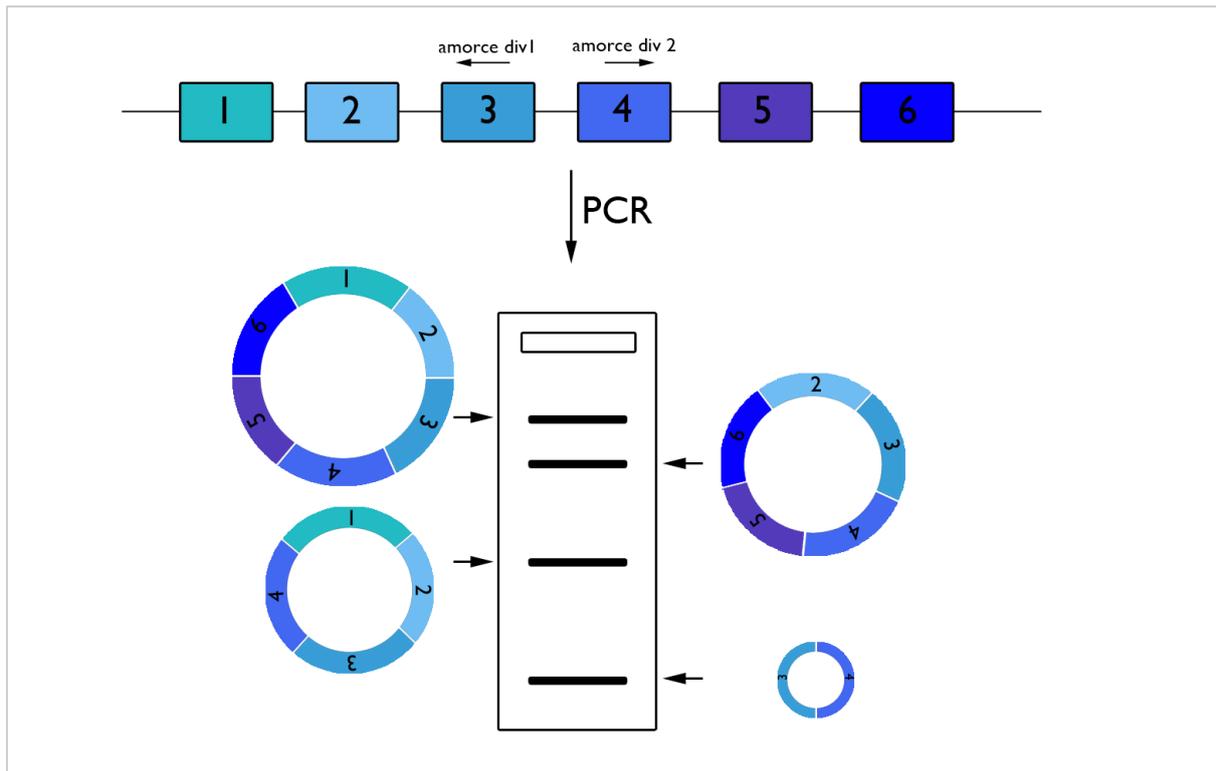


Figure 14: Détection des ARN circulaires par PCR. Les amorces sont choisies dans les exons d'intérêt et sont orientées de manière divergente. Ceci permet l'amplification d'ARN circulaires contenant différents exons.

C. Les différentes formes d'ARN circulaires

Différentes classes d'ARN circulaires ont pu être décrites, basées sur leur composition en introns/exons (**Figure 13**):

- Les ecircRNAs (ARN circulaires exoniques) contiennent uniquement des séquences exoniques. Il a été montré que ce type d'ARN circulaire pouvait être transporté dans le cytoplasme et jouer un rôle post-transcriptionnel d'éponge à micro-ARN (voir détail ci-dessous) (Memczak et al., 2013).
- Les CiRNAs (ARN circulaires introniques) contiennent uniquement des séquences introniques et auraient une localisation principalement nucléaire et pourraient moduler la transcription (Zhang et al., 2013b) (voir détail ci-dessous).

- Les ElciRNA (ARN circulaires exoniques et introniques) contiennent des séquences introniques retenues entre des séquences exoniques. Il a été montré que ces ARN circulaires pouvaient être localisés dans le noyau et jouer un rôle dans la transcription. (Li et al., 2015) (voir détail ci-dessous).

D. Rôles des ARN circulaires (Figure 15)

Tout comme les mécanismes qui régulent la synthèse des ARN circulaires, leurs rôles dans la cellule ne sont aujourd'hui que partiellement décrits. Il a été montré que les ARN circulaires étaient capables d'agir comme éponges à micro ARN par complémentarité de séquences (Hansen et al., 2013; Memczak et al., 2013). Les éponges à micro ARN ont été décrites comme des séquences capables d'entrer en compétition avec les cibles des micro-ARN et ainsi d'entraîner l'augmentation d'expression de ces cibles (Ebert et al., 2007). Hansen et al. ont notamment étudié l'interaction entre l'ARN circulaire ciRS-7 et le micro ARN miR-7 dans des cellules humaines HEK293 ainsi que dans le cerveau de souris. Les auteurs montrent une conservation entre les espèces des séquences cibles de ciRS-7 pour miR-7. De plus, ils constatent une co-localisation de ces deux ARN dans les cellules et dans des coupes de cerveau. Enfin en présence de ciRS-7 ils mettent en évidence une perte de l'inhibition des cibles de miR-7, entraînant par exemple la surexpression de SNCA (Synuclein Alpha) et IRS2 (Insulin Receptor Substrate 2), notamment impliquées dans la maladie de Parkinson et le diabète respectivement. Des résultats similaires ont pu être obtenus à partir d'un autre ARN circulaire : SRY, interagissant avec miR-138 (Hansen et al., 2013).

Un rôle de régulateur transcriptionnel a également été décrit pour les ARN circulaires. Il a notamment été montré dans des lignées cellulaires humaines (Hela et HEK 293) que des ARN circulaires de type ElciRNA étaient capables de réguler la transcription de leurs gènes parentaux par interaction avec RNA pol II et U1 snRNP (Li et al., 2015). Les auteurs observent une co-localisation des ARN circulaires et du promoteur de leurs gènes parentaux.

Un rôle similaire a été démontré pour des ARN circulaires introniques. Zhang et al, ont montré que les CiRNAs étaient principalement localisés dans le noyau à l'inverse de leurs ARNm

parentaux qui sont présents dans le cytoplasme. Notamment, l'ARN circulaire ci-ankrd52 est co-localisé avec le promoteur de son gène parental. De la même manière que les ElciRNA, ce CiRNA peut interagir avec RNA pol II et ainsi moduler la transcription du gène parental (Zhang et al., 2013b).

Ce rôle de régulateur transcriptionnel a également été décrit dans le cas d'ecircRNA pour le gène FLI1. Les auteurs ont mis en évidence un ARN circulaire contenant les exons 4-2-3 du gène FLI1 capable d'augmenter l'expression du gène FLI1 par hypométhylation de son promoteur (Chen et al., 2018a). Ce mécanisme fait intervenir le recrutement de TET au promoteur de FLI1 en *cis*, entraînant la déméthylation des îlots CpG, et la répression de DNMT1 en *trans*, par fixation au niveau de son promoteur. On observe donc ici un double rôle pour un même ARN circulaire : répresseur au niveau de DNMT1 et activateur au niveau de FLI1.

Il a également été démontré dans les années 90 que les ARN circulaires pouvaient être transcrits en peptides ou petites protéines, lorsqu'ils présentaient des séquences de reconnaissance pour les ribosomes ainsi qu'une phase ouverte de lecture (ORF) (Chen and Sarnow, 1995; Perriman and Ares, 1998). Plus récemment, Legnini et al. ont mis en évidence la traduction de l'ARN circulaire Circ-ZNF609 chez l'Homme et la souris (Legnini et al., 2017). La circularisation de cet ARN entraîne l'apparition d'un nouveau codon stop donnant naissance à une protéine tronquée. De plus, cet ARN circulaire contient deux ATG permettant la formation de deux protéines de tailles différentes, détectables par Western-Blot. Par analyse bio-informatique ils ont montré que cet ARN circulaire était enrichi au cours de la différenciation musculaire. Le rôle des peptides formés restent cependant non élucidé.

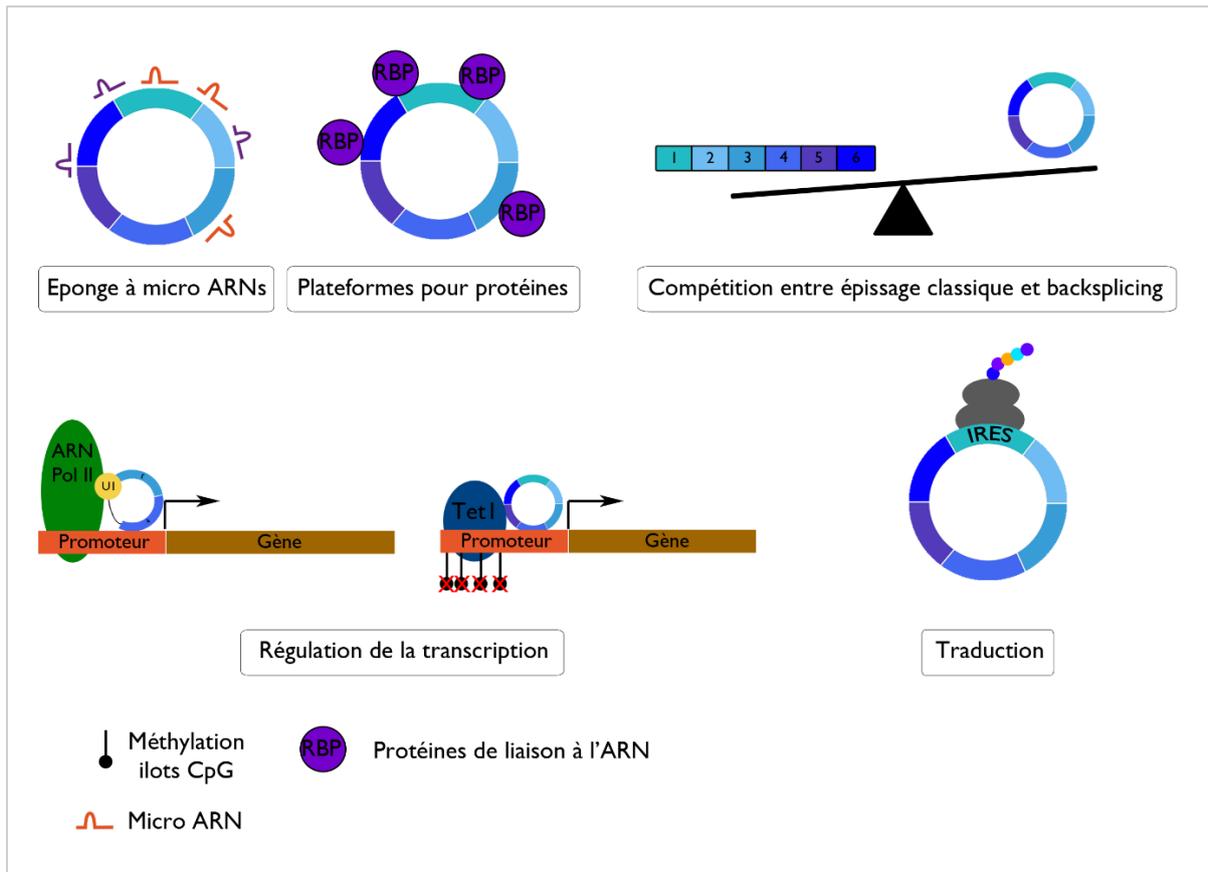


Figure 15: Rôles des ARN circulaires dans la cellule. Les ARN circulaires sont capables de capter des micro-ARN ou des protéines capables de se fixer sur les ARN. Leur synthèse entre en compétition avec la synthèse des ARN linéaires correspondants, permettant ainsi une régulation de l'expression des gènes parentaux. Les ARN circulaires sont également capables de réguler le promoteur des gènes par interaction avec l'ARN polymérase 2 et/ou U1snRNP, ou encore par modification de la méthylation du promoteur via le recrutement de Tet1. Enfin, lorsqu'ils portent des séquences de reconnaissance pour les ribosomes, les ARN circulaires peuvent être traduits en petits peptides ou protéines.

Une étude de 2017 a montré que les ARN circulaires pouvaient avoir un rôle de plateforme pour certaines protéines (Du et al., 2017). L'ARN circulaire circ-Foxo3 est capable d'induire l'apoptose en empêchant la dégradation de FOXO3 par MDM2. En absence de circ-Foxo3, MDM2 peut induire à la fois la dégradation de P53 et celle de FOXO3, empêchant ainsi la répression en aval de PUMA (cible de P53 et de FOXO3). Circ-Foxo3 présente des sites de liaison à la fois à P53 et à MDM2, favorisant ainsi un rapprochement des deux protéines et facilitant l'ubiquitination de P53 par MDM2. Ceci a pour effet d'entraîner la dégradation de P53. En présence de circ-Foxo3, la protéine FOXO3 est protégée de la dégradation grâce à la séquestration de MDM2 avec P53. FOXO3 peut alors activer ses cibles en aval, en particulier via la répression de PUMA qui entraîne l'apoptose cellulaire (**Figure 16**).

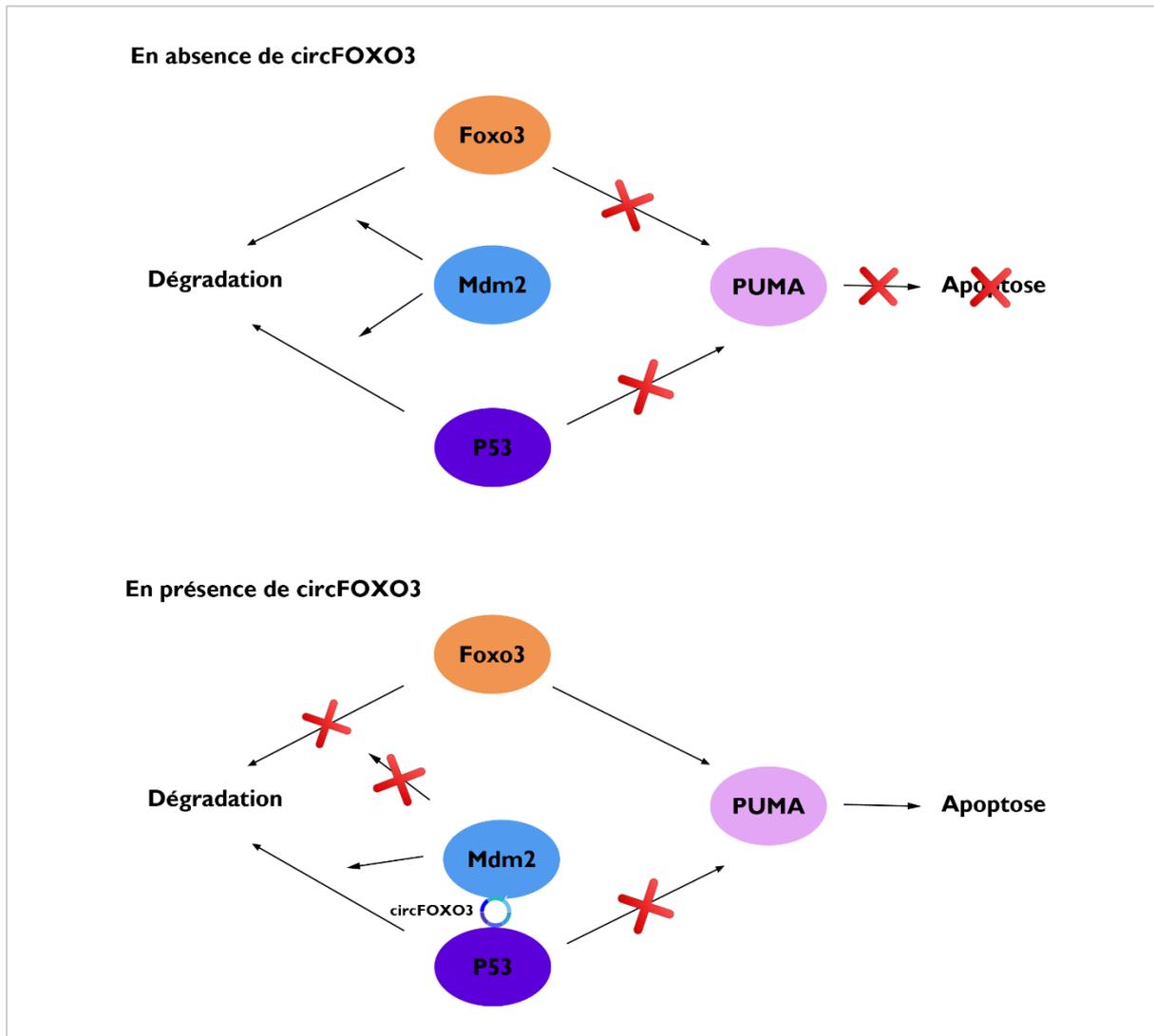


Figure 16 : Régulation de l'apoptose par circFOXO3. En absence de l'ARN circulaire MDM2 entraîne la dégradation de FOXO3 et P53 par ubiquitination. Ceci empêche l'activation de PUMA et réduit l'apoptose. En présence de circFOXO3, l'ARN circulaire se fixe à la fois sur P53 et MDM2 ce qui entraîne un rapprochement des deux protéines et une dégradation de P53. MDM2 n'est alors plus en mesure de se fixer sur FOXO3, ce qui empêche sa dégradation et permet l'activation de PUMA et l'apoptose. Adapté de (Du et al., 2017)

E. ARN circulaires et cancers

De nombreuses études sur différents types de cancers se sont intéressées au rôle des ARN circulaires dans l'oncogénèse. Les analyses de RNAseq ont permis de mettre en évidence la surexpression/répression de nombreux ARN circulaires dans divers cancers (**Tableau 2** (Kristensen et al., 2018)). Le rôle de ces ARN circulaires dans la progression tumorale est variable.

Différentes études ont pu mettre en évidence un rôle des ARN circulaires dans la transition épithélio-mésenchymateuse. Au cours de ce processus les ARN circulaires peuvent jouer un rôle bivalent selon les voies de signalisation affectées. Certains ARN circulaires entraînent une inhibition de la transition épithélio-mésenchymateuse en inhibant par exemple les voies NOTCH (Zhang et al., 2019a), TGF- β (Wang et al., 2018), ou N-cadhérine (Li et al., 2017a). D'autres ARN circulaires favorisent cette transition par activation des voies WNT (Qu et al., 2019), TWIST1 (Meng et al., 2018), ou AKT/mTOR (Zhang et al., 2019b).

Il a également été montré que certains ARN circulaires étaient capables de favoriser la croissance cellulaire. Par exemple circHIPK3 a été retrouvé surexprimé dans 6 cancers différents (prostate, foie, sein, colorectal, vessie et estomac) (Zheng et al., 2016). Les auteurs ont montré que l'inhibition de cet ARN circulaire dans trois lignées cellulaires humaines permettait une réduction significative de la prolifération cellulaire. Dans le cancer de la prostate, une étude de séquençage à forte profondeur a permis d'identifier un grand nombre d'ARN circulaires associés à la maladie (7232) (Chen et al., 2019a). Parmi eux, 11,3% sont essentiels à la prolifération cellulaire. De plus, l'abondance des transcrits circulaires semble liée au grade tumoral, avec une augmentation pour les grades avancés.

Dans le cas de cancer de la prostate, la surexpression de l'ARN circulaire circ-MYLK permet l'activation des voies VEGFA/VEGFR2 favorisant l'angiogenèse (Zhong et al., 2017). Il agit par liaison avec le micro ARN miR-29a. Les auteurs ont montré que le niveau d'expression de circ-MYLK était lié au grade du cancer.

Une étude sur les leucémies myéloïdes aiguës a montré qu'il était possible de distinguer différents sous-groupes de patients en fonction de l'expression de l'ARN circulaire hsa_circ_0075001 issu du gène *NPM1* (Hirsch et al., 2017). L'expression de cet ARN circulaire présente un lien avec le niveau de maturation des cellules tumorales. Chez les patients présentant des cellules très immatures, l'expression de cet ARN circulaire est très élevée. L'inverse est observé chez les patients présentant des cellules plus matures. L'expression de cet ARN circulaire est corrélée au niveau d'expression du gène parental.

Une étude récente a étudié le potentiel diagnostique et pronostique des ARN circulaires dans le cancer du poumon (Huang et al., 2019). Cette étude a rassemblé les résultats de 23 études antérieures. Les résultats indiquent une utilisation possible en diagnostique avec une bonne sensibilité et une bonne spécificité. Ces travaux ouvrent la voie à une utilisation médicale future, cependant ils restent à confirmer à plus grande échelle et sur différents types de cancers.

circBase ID (Alias)	Fonction putative	Gène parental	Type de cancer	Surexprimé/Réprimé
hsa_circ_0004277	Eponge à micro ARN	WDR37	Leucémie Myéloïde aigue	Surexprimé
hsa_circ_0075828	Eponge à micro ARN	LINC00340	Carcinome des cellules basales	Surexprimé
hsa_circ_0075825	Eponge à micro ARN	LINC00340	Carcinome des cellules basales	Surexprimé
hsa_circ_0022383	Eponge à micro ARN	FADS2	Carcinome des cellules basales	Reprimé
hsa_circ_0022392	Eponge à micro ARN	FADS2	Carcinome cutané des cellules squameuses	Reprimé
			Carcinome des cellules basales	Reprimé
hsa_circ_0041103	Eponge à micro ARN	TCF25	Cancer de la prostate	Surexprimé
hsa_circ_0002768	Eponge à micro ARN	MYLK	Cancer de la prostate	Surexprimé
(circ-Foxo3)	Plateforme à protéine	FOXO3	Cancer du sein	Reprimé
hsa_circ_0008717	Eponge à micro ARN	ABCB10	Cancer du sein	Surexprimé
hsa_circ_0001946 (ciRS-7)	Eponge à micro ARN	CDR1AS	Carcinome colorectal	Surexprimé
			Carcinome colorectal	Surexprimé
			Gliome	Reprimé
			Hépto-carcinome	Surexprimé
hsa_circRNA_104700	Eponge à micro ARN	PTK2	Carcinome colorectal	Reprimé
hsa_circ_0000069	Non étudié	STIL	Carcinome colorectal	Surexprimé
(circCCDC66)	Eponge à micro ARN	CCDC66	Carcinome colorectal	Surexprimé
(cir-ITCH)	Eponge à micro ARN	ITCH	Carcinome colorectal	Reprimé
			Cancer du poumon	Reprimé
			Carcinome des cellules squameuses de l'oesophage	Reprimé
			Hépto-carcinome	Reprimé
hsa_circ_0000523 (circ_006229)	Non étudié	METTL3	Carcinome colorectal	Reprimé
hsa_circ_0001346	Non étudié	RNF13	Carcinome colorectal	Reprimé
hsa_circ_0001793	Non étudié	IKKB	Carcinome colorectal	Surexprimé
hsa_circ_0070933	Eponge à micro ARN	LARP1B	Carcinome cutané des cellules squameuses	Surexprimé
hsa_circ_0070934	Eponge à micro ARN	LARP1B	Carcinome cutané des cellules squameuses	Surexprimé
hsa_circ_0067934	Non étudié	PRKCI	Carcinome des cellules squameuses de l'oesophage	Surexprimé
(circPVT1)	Eponge à micro ARN	PVT1	Cancer gastrique	Surexprimé
hsa_circ_0000190	Non étudié	CNIH4	Cancer Gastrique	Reprimé
hsa_circ_0000096	Non étudié	HIAT1	Cancer Gastrique	Reprimé
hsa_circ_0000140	Non étudié	KIAA0907	Cancer Gastrique	Reprimé
(circ_VCAN)	Non étudié	VCAN	Gliome	Surexprimé
(cZNF292)	Non étudié	ZNF292	Gliome	Reprimé
hsa_circ_0000594	Eponge à micro ARN	TTBK2	Gliome	Surexprimé
hsa_circ_0001649	Eponge à micro ARN	SHPRH	Hépto-carcinome	Reprimé
hsa_circ_0001727	Eponge à micro ARN	ZKSCAN1	Hépto-carcinome	Reprimé
hsa_circ_0005075	Eponge à micro ARN	EIF4G3	Hépto-carcinome	Surexprimé
hsa_circ_0000284 (circHIPK3)	Eponge à micro ARN	HIPK3	Hépto-carcinome	Surexprimé
hsa_circ_0007874	Eponge à micro ARN	MTO1	Hépto-carcinome	Reprimé
hsa_circ_0004018	Eponge à micro ARN	SMYD4	Hépto-carcinome	Reprimé
hsa_circ_0003570	Non étudié	FAM53B	Hépto-carcinome	Reprimé
hsa_circ_0013958	Eponge à micro ARN	ACP6	Cancer des poumons	Surexprimé
hsa_circ_0016347	Eponge à micro ARN	KCNH1	Osteosarcome	Surexprimé
(circUBAP2)	Non étudié	UBAP2	Osteosarcome	Surexprimé
hsa_circ_0013339	Eponge à micro ARN	SLC30A7	Carcinome oral des cellules squameuses	Surexprimé

Tableau 2 : ARN circulaires potentiellement impliqués dans différents cancers, et leur fonction putative (adapté de (Kristensen et al., 2018)).

F. ARN circulaires et translocations

Une étude, réalisée par le groupe de du Dr. Pandolfi en 2016, a mis en évidence l'existence d'ARN circulaires spécifiques des translocations chromosomiques, appelés ARN circulaires de fusion (Guarnerio et al., 2016). Ces ARN circulaires sont composés d'exons provenant des 2 gènes à la jonction (**Figure 17**) et ne peuvent donc pas être retrouvés dans des cellules non transloquées. Ils ont montré la présence de ces ARN dans des cellules transloquées de patients atteints de différents cancers, et ont notamment étudié les translocations *PML-RAR α* et *MLL-AF9*. Pour chaque translocation ils ont trouvé un ou deux ARN circulaires présents dans l'ensemble des tumeurs étudiées. Dans le cas de la translocation *MLL-AF9* ils ont étudié plus en détails le rôle de ces ARN circulaires de fusion (f-circRNA). Ils ont ainsi mis en évidence un effet additif de la protéine de fusion et des f-circRNA par surexpression de ces deux éléments dans des cellules WT. Les f-circRNA participent à la prolifération cellulaire, à la formation tumorale *in vivo* ainsi qu'à la résistance aux thérapies médicamenteuses. Les auteurs ont émis l'hypothèse qu'en présence de translocation il pourrait y avoir rapprochement de séquences répétées type ALU (**Figure 17**), ce qui faciliterait la formation de ces ARN circulaires de fusion. Des travaux plus récents ont également pu mettre en évidence l'existence d'autres isoformes d'ARN circulaires de fusion dans les translocations impliquant *MLL* et ses partenaires (Dal Molin et al., 2019).

Les travaux de Chen et al sur le cancer de la prostate ont également montré la présence d'ARN circulaires de fusion principalement dans les tumeurs de grade avancé (Chen et al., 2019a).

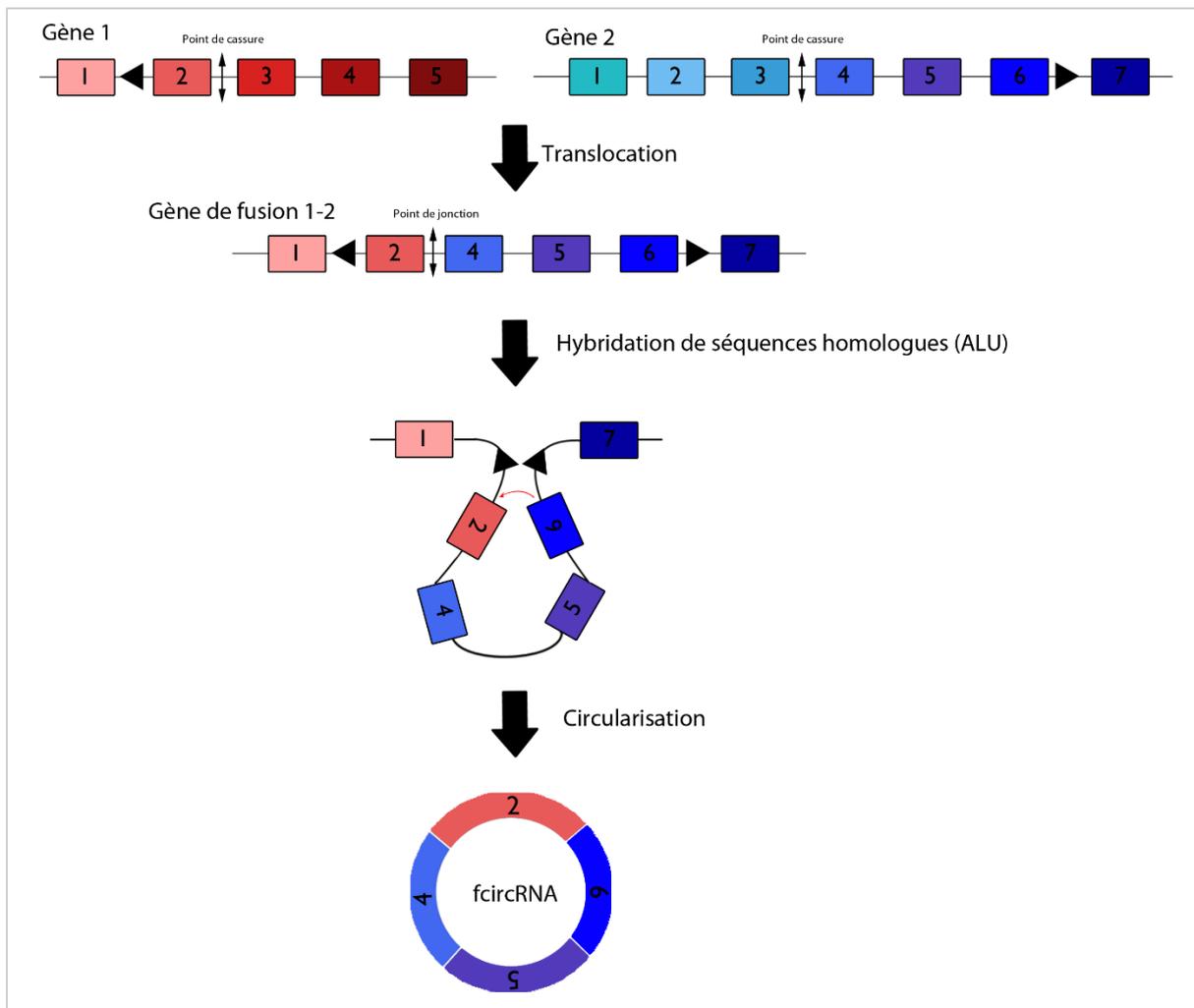


Figure 17: Formation des ARN circulaires de fusion à partir d'une translocation chromosomique. La translocation est formée entre les gènes 1 et 2 suite à l'apparition de 2 points de cassures. Un gène de fusion 1-2 est généré. On observe une hybridation de séquences introniques homologues (ALU) permettant la circularisation par épissage inverse.

G. Rôles des ARN circulaires dans d'autres processus cellulaires

Une étude chez la souris a montré un rôle potentiel des ARN circulaires au cours du vieillissement (Gruner et al., 2016). Les auteurs ont réalisé des analyses de RNAseq à partir de trois tissus différents chez la souris jeune et âgée : le cœur, le cortex cérébral et l'hippocampe. Dans les deux tissus cérébraux ils ont pu observer une très forte augmentation globale de l'expression des ARN circulaires. Ce résultat n'est pas retrouvé dans le cœur. De plus, cette augmentation n'est pas retrouvée au niveau des transcrits linéaires correspondants, ce qui semble indiquer que ce phénomène est spécifique des ARN circulaires. Ces résultats concordent avec une étude antérieure réalisée chez la drosophile qui avait également montré

une accumulation d'ARN circulaires dans le système nerveux au cours du vieillissement (Westholm et al., 2014).

Une étude a également montré que les ARN circulaires pouvaient jouer un rôle dans le maintien de la pluripotence cellulaire (Yu et al., 2017). Les auteurs ont identifié une vingtaine d'ARN circulaires surexprimés dans les cellules souches embryonnaires humaines. Parmi ces ARN circulaires, ils ont montré que circBIRC6 et circCORO1C étaient nécessaires au maintien du statut souche des cellules. En effet l'inhibition de ces ARN circulaires entraîne une diminution de l'expression des facteurs souches PATF, NANOG, OCT4, KLF4, et MYC et une augmentation des facteurs de différenciation GSC, MIXL1, SOX17, et SOX1. De plus, la surexpression de circBIRC6 et circCORO1C, bien qu'insuffisante à la reprogrammation de fibroblastes en cellules souche, permet d'augmenter l'efficacité de cette reprogrammation lors de l'utilisation du vecteur Sendai exprimant OCT4, SOX2, KLF4, et CMYC.

D'autres travaux ont pu mettre en évidence le rôle potentiel de l'ARN circulaire ciRS-7 dans la maladie d'Alzheimer (Zhao et al., 2016). En effet l'expression de cet ARN circulaire est fortement diminuée chez les patients atteints d'Alzheimer. Ceci aurait pour conséquence l'augmentation du micro ARN miR-7 capable de réprimer l'expression d'UBE2A (protéine capable d'induire la dégradation des peptides d'amyloïdes). Or dans le cas présent la diminution d'UBE2A pourrait entraîner une accumulation de ces peptides dans le cerveau et ainsi participer à la progression de la maladie. Cependant cet article ne montre aucun lien direct entre ciRS-7 et la diminution d'UBE2A, il est donc impossible d'affirmer le rôle de ciRS-7 dans la maladie d'Alzheimer. Il serait intéressant de réaliser des analyses de surexpression/répression de ciRS-7 afin de confirmer son rôle sur miR-7 et UBE2A.

L'ensemble de ces différents travaux ouvre de nouvelles perspectives sur le potentiel des ARN circulaires dans différentes pathologies, et notamment en oncologie. Ces ARN pourraient également jouer un rôle dans des processus physiologiques normaux, tels que le vieillissement. Une meilleure compréhension de leur synthèse/dégradation et de leur activité dans la cellule pourrait permettre de trouver de nouvelles cibles thérapeutiques ou de nouveaux biomarqueurs diagnostiques et/ou pronostiques.

IV. Le timing de réplication

Au cours de la phase S du cycle cellulaire les différentes régions de l'ADN sont répliquées selon un ordre spécifique : on appelle cela le timing de réplication. On peut ainsi délimiter des domaines d'ADN plus ou moins précoces/tardifs. Ce pattern de réplication est propre à chaque type cellulaire et relativement bien conservé entre les espèces pour un même type cellulaire (Ryba et al., 2010; Yaffe et al., 2010).

A. Timing de réplication, ADN et chromatine

Il semble exister un lien entre timing de réplication, organisation spatiale du noyau, état chromatinien, et composition des séquences d'ADN (pour revue (Rhind and Gilbert, 2013), **Figure 18**), cependant la nature exacte de cette relation n'est pas toujours clairement établie.

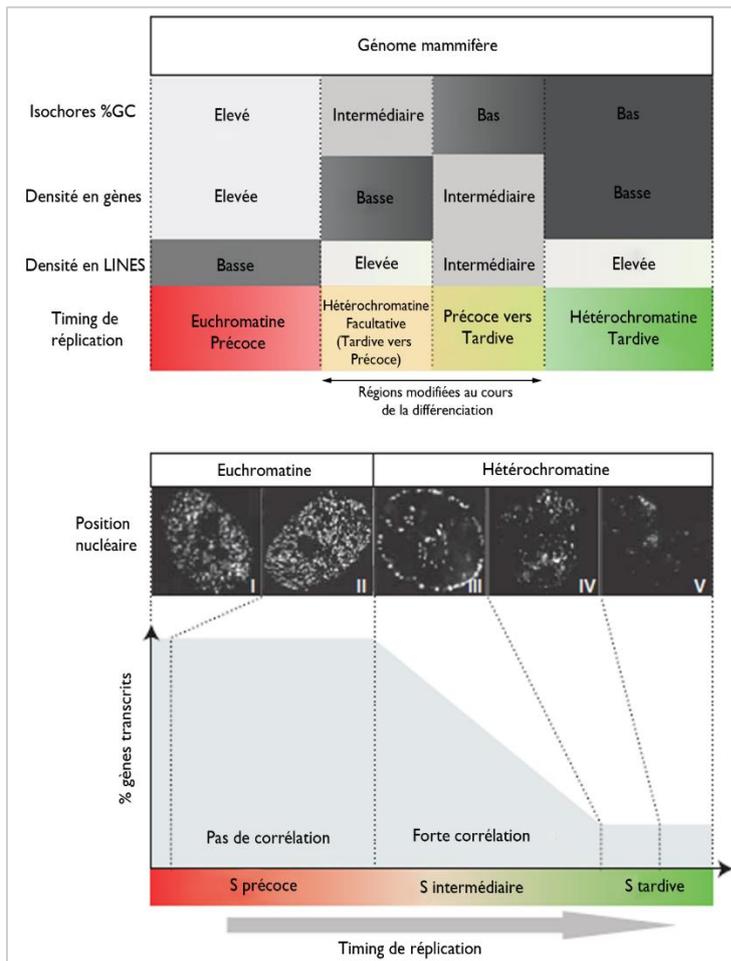


Figure 18: Relation entre propriétés des isochores, timing de réplication, organisation nucléaire et transcription. (Haut) Relation entre timing de réplication, %GC, densité en gènes et en éléments répétés LINES. L'euchromatine répliquée précocement de manière constitutive présente un %GC élevé ainsi qu'une forte densité en gènes et une faible densité en LINES. L'opposé est observé pour les régions d'hétérochromatine constitutive. Un profil intermédiaire est observé pour les régions modifiées au cours de la différenciation cellulaire. (Bas) Corrélation entre localisation nucléaire, transcription et timing de réplication. L'euchromatine localisée au centre du noyau est répliquée en début de phase S, elle contient de nombreux gènes transcrits, mais la corrélation entre transcription et timing est faible, autant de gènes transcrits que ne gènes non transcrits sont répliqués dans cette phase. En milieu et fin de phase S la réplication a lieu en périphérie du noyau. En milieu de phase S la corrélation entre transcription et réplication est très forte. En phase S tardive l'hétérochromatine constitutive contenant très peu de gènes est

répliquée. Adapté de (Rhind and Gilbert, 2013).

1. Régions AT/GC riches

Une 1^{ère} étude a montré, à l'aide de techniques d'immunofluorescence, une bonne corrélation entre l'organisation en isochore (région sur laquelle le taux de GC est homogène), le %GC et le timing de réplication (Schmegner et al., 2007). Cependant cette étude se limite à l'analyse d'une seule région du génome, et non pas à une analyse globale. De plus, deux autres articles montrent qu'il existe une très forte conservation du timing de réplication entre l'Homme et la souris, alors qu'ils n'observent pas de conservation du %GC (Ryba et al., 2010; Yaffe et al., 2010). Néanmoins, les régions riches en AT sont plus aptes à modifier leur profil de réplication lors de la différenciation cellulaire. C'est en effet ce que met en avant l'étude de Hiratani et al. dans laquelle ils induisent la différenciation neurale de cellules ES murines. Ce modèle leur permet d'étudier le timing de réplication d'une cinquantaine de gènes (AT ou GC riches) dont l'expression est modifiée entre les 2 stades de différenciation (Hiratani et al., 2004).

2. Marques histones

L'association entre marques d'histones et timing de réplication a également fait l'objet de plusieurs études qui semblent toutes s'accorder sur la corrélation entre réplication précoce et marques histones activatrices telles que H3K4me1,2,3, H3K9ac, H3K27ac, H3K36me3, H4K20me1 (Jullienne et al., 2013; Thurman et al., 2007; Yaffe et al., 2010). L'analyse de Jullienne et al., réalisée dans une lignée cellulaire humaine (K562), met notamment en évidence l'existence de 4 clusters en PCA (Analyse en Composante Principale), déterminés par une association « timing de réplication/marques histones ». A l'inverse le lien entre marques histones inhibitrices et réplication tardive est controversé. Une seule étude montre une association réplication tardive/H3K27me3 (Thurman et al., 2007) alors que cette corrélation n'est pas retrouvée dans les analyses citées précédemment. Cette étude a été réalisée sur seulement 1% du génome humain ce qui pourrait induire un biais et expliquer les différences observées avec les analyses sur génome entier.

3. Localisation nucléaire

Des études sur la localisation nucléaire de la chromatine ont mis en évidence une compartimentalisation de la réplication. En effet lorsque l'on marque les cellules au BrdU au cours de la phase S, on peut observer des patterns de réplication spécifiques qui varient lors de la progression du cycle cellulaire (Panning and Gilbert, 2005). Les régions périphériques sont globalement répliquées plus tardivement. De plus, au cours de la différenciation cellulaire, des relocalisations nucléaires ont lieu, qui s'accompagnent d'une modification du timing de réplication (Hiratani et al., 2008). Cette relocalisation peut également être observée au niveau du chromosome X chez la femelle : lors de l'inactivation du X, on constate un déplacement du X inactif vers la périphérie accompagné d'une réplication tardive (Hiratani and Gilbert, 2010; Koren and McCarroll, 2014).

4. Organisation 3D

Des techniques récentes de Hi-C ont pu mettre en évidence l'existence de deux niveaux d'organisation de la chromatine au sein du noyau :

- Les compartiments A/B (Lieberman-Aiden et al., 2009) présentent de fortes interactions au sein de chaque domaine mais une absence d'interaction entre les domaines A et B. Le compartiment A est associé à des régions riches en gènes exprimés et en marques histones activatrices telles que H3K36me3. Le compartiment B est quant à lui caractérisé par un fort contact avec la lamine B1 dans le noyau (Guelen et al., 2008). En raison de cette association il est parfois nommé LAD (*Lamin Associated Domain*). Les auteurs observent une forte diminution de l'expression des gènes dans ce compartiment ainsi que de la marque H3K4me2 associée à une chromatine ouverte. De plus, ils constatent un enrichissement des marques H3K9me2 et H3K27me3 associées à une chromatine fermée. La frontière entre les 2 compartiments est caractérisée par un changement abrupt de ces différentes caractéristiques. De plus, cette bordure est fortement enrichie en CTCF et en îlots CpG présents dans les 10kb entourant chaque LAD.

- Les domaines TAD (*Topologically Associated Domains*) sont d'une taille d'environ 100kb (Dixon et al., 2012). La frontière de ces domaines est également marquée par un enrichissement en CTCF, en marques histones associées aux promoteurs, ainsi qu'en TSS (*Transcription Starting Sites*). A l'inverse les auteurs observent une déplétion en marques histones non associées aux promoteurs H3K4me1 et H3K9me3. Les bordures des TAD semblent relativement bien conservées entre les différents types cellulaires, mais également entre l'Homme et la Souris.

Différentes études ont pu mettre en évidence un lien étroit entre timing de réplication, compartiments A/B et TAD (**Figure 19**) (Pope et al., 2014; Ryba et al., 2010; Yaffe et al., 2010). Notamment il a été montré que les zones de transition précoces/tardives des domaines de réplication correspondaient aux bordures des TAD (Pope et al., 2014) bien que le niveau de corrélation entre les deux soit variable d'un TAD à un autre. De plus, l'établissement de ces deux éléments a lieu au même moment du cycle cellulaire (en début de G1) (Dileep et al., 2015). Un modèle de domaine de réplication a été proposé par Dileep et al. 2015. Ce modèle suggère qu'au cours de la différenciation cellulaire certains TAD changent de compartiment A/B entraînant une modification du timing de réplication. Ces domaines sont appelés domaines développementaux et s'opposent aux domaines dits constitutifs, pour lesquels aucun changement de compartiment n'est observé au cours de la différenciation.

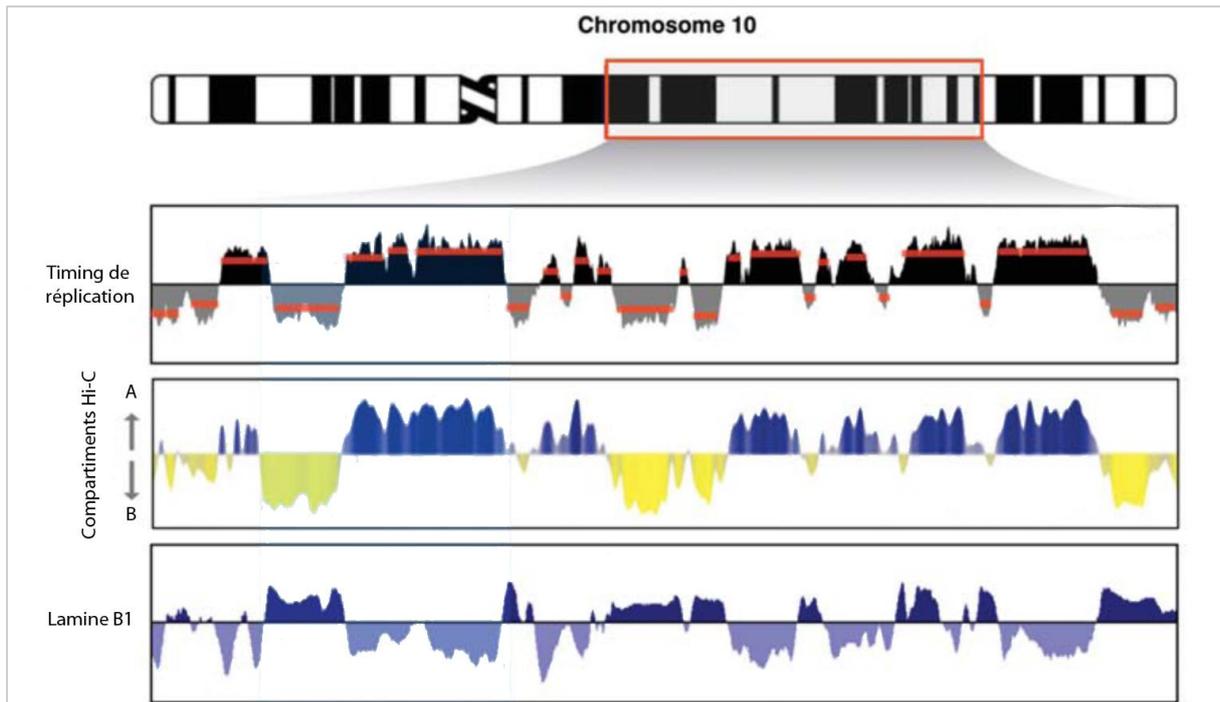


Figure 19: Relation entre le timing de réplication et l'organisation de la chromatine. Les domaines de réplication précoce sont localisés dans le compartiment A et dépourvus d'interaction avec la lamine B1. L'opposé est observé pour les domaines de réplication tardive. Adapté de (Dileep et al., 2015).

Une étude récente s'est intéressée aux éléments *cis* régulateurs pouvant entraîner une modification du timing de réplication (Sima et al., 2019). Les auteurs se sont focalisés sur locus *Dppa2/4* dans des cellules embryonnaires de souris. En déplaçant la région particulière, associée aux CTCF à la bordure du TAD contenant le locus *Dppa2/4*, ils n'ont observé aucun effet sur le timing de réplication. A l'inverse, en déplaçant des séquences internes au domaine, conservant une forte interaction entre elles et avec *Dppa2/4* en absence de CTCF, ils ont pu observer une modification du timing de réplication de précoce vers tardif. Ils ont nommé ces séquences ERCE (*Early Replication Control Elements*). Cette modification s'accompagne également d'un passage du compartiment A vers B ainsi que d'une diminution de la transcription du gène *Dppa2/4*. Les auteurs ont également pu prédire et vérifier un certain nombre d'ERCE dans tout le génome en utilisant les critères suivants afin de déterminer les ERCE : indépendants des CTCF, enrichis en marques épigénétiques activatrices (DNase1 HS, P300, H3K27ac, H3K4m1, H3K4m3, MED1, and OCT4, SOX2, NANOG) et présents dans le compartiment A. Parmi les ERCE prédits ils observent 30% d'*enhancers* et 20% de *super-enhancers*. Leur analyse aboutit au modèle présenté **Figure 20**.

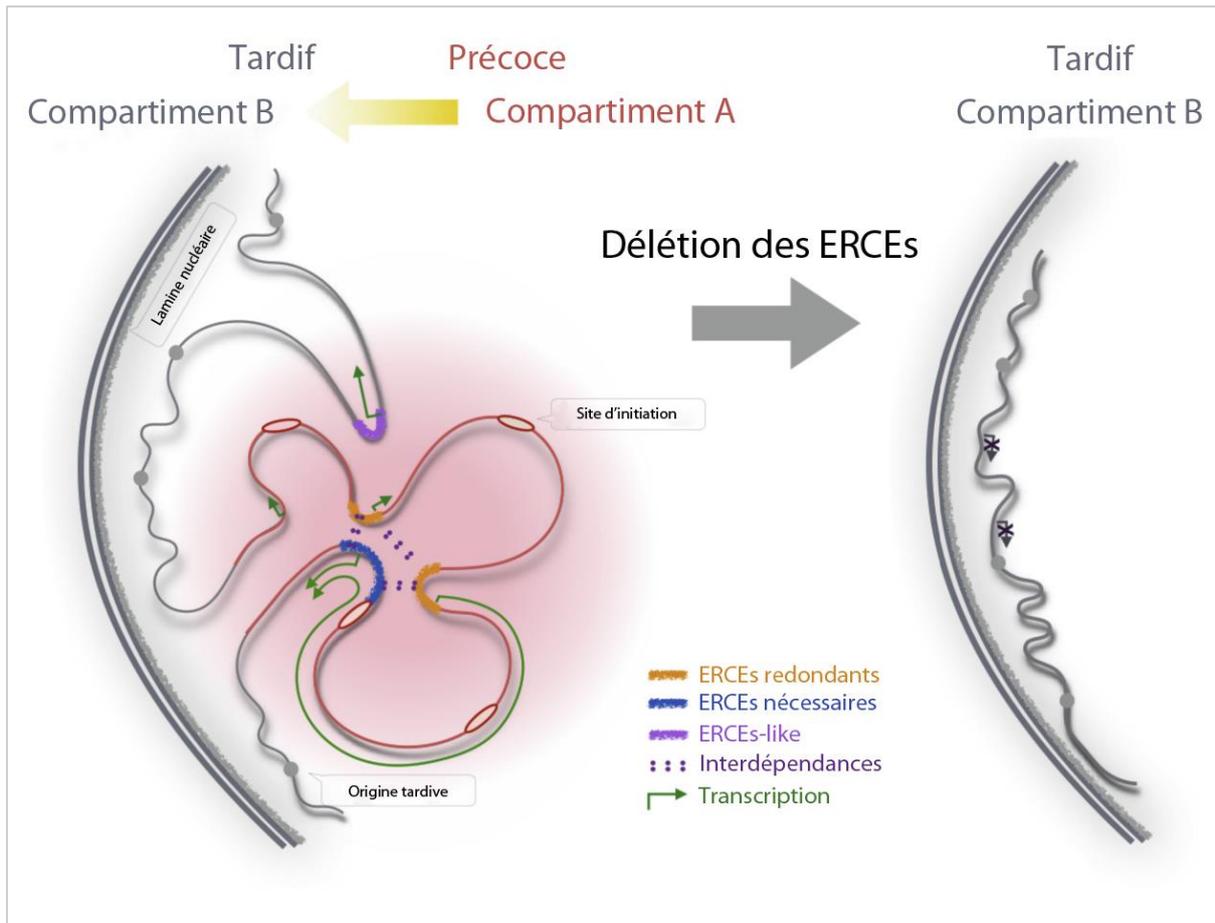


Figure 20: Modèle des ERCE (*Early Replication Control Elements*). Il existe différentes catégories d'ERCE : ceux qui sont nécessaires à l'induction d'un timing précoce, ceux qui participent à la réplication précoce mais sont redondants, ceux qui ont des caractéristiques d'ERCE mais n'en sont pas (ERCE-like). La délétion des ERCE entraîne le déplacement vers une réplication tardive. Adaptée de (Sima et al., 2019).

5. Timing de réplication et expression génique

La relation entre timing de réplication et expression génique fait également débat. En effet il est souvent admis que les régions riches en gènes transcrits sont répliquées précocement au cours de la phase S. Cependant cette règle ne semble pas applicable dans tous les cas (**Figure 21**). Par exemple l'étude de Hiratani et al. 2004 montre que lors de la différenciation des cellules neurales de souris, il n'y a pas de corrélation entre le timing de réplication et la transcription des régions riches en GC. A l'inverse, les régions riches en AT voient leur timing de réplication modifié en même temps que l'expression génique (Hiratani et al., 2004).

De plus, certaines équipes de recherche ont tenté d'induire une modification du timing de réplication local en forçant l'expression génique dans une région particulière. Une de ces

études a été réalisée dans des cellules de poulet, avec introduction d'un transgène portant différents promoteurs et/ou insulateurs. Une modification du timing de réplication de tardif vers précoce est observée uniquement avec transgène porteur d'un insulateur protégé de la méthylation, qui induit l'ouverture de la chromatine. L'effet est observé avec ou sans transcription en aval bien que la transcription augmente légèrement cet effet. Les résultats obtenus sur différentes lignées cellulaires et différentes régions du génome sont similaires. Cette étude semble donc indiquer que la modification du timing de réplication est liée à l'état chromatinien et relativement indépendante de la transcription (Hassan-Zadeh et al., 2012). Une étude similaire, menée chez la souris, aboutit aux mêmes conclusions (Simon et al., 2001). De plus, de nombreux gènes transcrits sont présents dans des régions répliquées tardivement (Rivera-Mulia et al., 2015; Rivera-Mulia and Gilbert, 2016) indiquant qu'une réplication précoce n'est pas nécessaire à la transcription.

Néanmoins une autre étude, réalisée dans des cellules souches de souris, et utilisant le système TALE en fusion avec VP64 afin d'activer la transcription mène à des conclusions différentes. En effet, l'activation de la transcription entraîne une relocalisation spatiale vers le centre du noyau ainsi qu'un passage de réplication tardive à précoce. Cependant lors de l'utilisation d'un transgène entraînant une ouverture de la chromatine sans activation de la transcription, on observe une relocalisation mais aucune modification du timing de réplication (Therizols et al., 2014). Ces résultats vont donc à l'encontre des études précédentes et semblent indiquer un rôle de la transcription dans la modification du timing de réplication. De plus, une étude de 2019 réalisée par l'équipe du Dr. Debatisse a mis en évidence une corrélation entre niveau de transcription et timing de réplication pour les longs gènes dans des cellules humaines et de poulet (Blin et al., 2019). Les auteurs ont montré qu'il était possible d'induire une réplication plus précoce des gènes *DMD*, *CCSER1* et *FHIT* en augmentant fortement leur niveau de transcription. Ils ont également observé une forte corrélation entre le niveau de transcription et la précocité du Timing de réplication. Ces modifications ne sont plus observées lorsque la terminaison de la transcription est induite à proximité du site d'initiation (~1kb en aval) via l'introduction d'un site de polyadénylation SV40 artificiel. Ces différentes observations apportent des arguments en faveur d'un rôle de la transcription sur le timing de réplication, en particulier pour les gènes longs. De plus, chez la drosophile, le poisson zèbre et l'embryon de xénope il a été montré que la mise en place du programme de

timing de réplication précédait l'activation de la transcription, indiquant ainsi un lien entre timing de réplication et transcription (Kaaij et al., 2018).

Il a été montré chez la levure *S. cerevisiae* que la réplication précoce des gènes codant pour les histones permettait de maximiser leur expression lors de l'entrée en phase S. L'induction d'un délai de réplication de ces gènes entraîne une diminution de leur transcription 45 min après l'entrée en phase S. Ceci montre ainsi l'importance du timing de réplication pour l'expression de certains gènes au cours de la phase S du cycle cellulaire (Muller and Nieduszynski, 2017). Cette augmentation transitoire du niveau d'expression génique semble limitée aux gènes associés au cycle cellulaire ainsi qu'aux histones. Pour l'ensemble des autres gènes répliqués précocement, il existe un système de compensation permettant d'inhiber l'expression des brins néo-synthétisés et ainsi de conserver un niveau d'expression similaire tout au long du cycle cellulaire (Voichek et al., 2016).

Une étude portant sur les sites d'initiation et de terminaison de la réplication dans des cellules humaines a également montré que les sites d'initiation étaient enrichis au niveau des TSS (*Transcription Starting Sites*), en particulier pour les gènes fortement exprimés (Chen et al., 2019b). De la même manière, les sites de terminaison sont enrichis au niveau des sites TTS (*Transcription Terminating Sites*). Ces résultats semblent indiquer qu'il existe également un lien entre initiation/terminaison de la réplication et transcription.

L'ensemble de ces études aboutit donc à des résultats pouvant être assez contradictoires. On peut supposer qu'il existe différents niveaux de régulation, spécifiques pour chaque organisme, types cellulaires ou encore régions du génome. De plus, la corrélation entre timing de réplication et transcription est controversée. L'augmentation de l'expression génique permet l'avancée du timing de réplication des gènes longs, mais dans d'autres cas l'ouverture de la chromatine en absence de transcription est suffisante à l'induction d'un timing de réplication plus précoce. De plus, il semblerait qu'il existe des sites particuliers ERCE au sein des domaines de réplication capables de réguler le timing en *cis*. Il est donc important de tenir compte de l'état chromatinien, de la taille des gènes, ainsi que de la composition nucléotidique lors de l'analyse du timing de réplication.

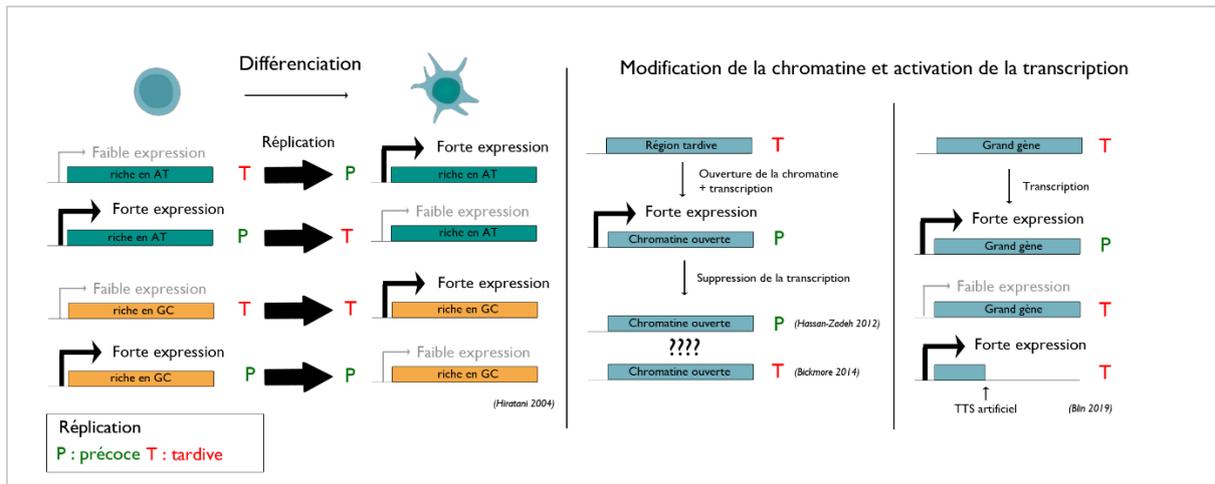


Figure 21: Lien entre timing de réplication et expression génique. Gauche : Au cours de la différenciation cellulaire seule les régions riches en AT changent de Timing lorsque l'expression des gènes qu'elles portent est modifiée. Centre : l'avancement du timing de réplication est possible lorsque l'on induit l'ouverture de la chromatine et la transcription. Selon les études la suppression de la transcription empêche ou non l'avancement du timing de réplication. Droite : Dans le cas des grands gènes, une forte expression permet l'avancement du timing. L'ajout d'un site de terminaison de la transcription (TTS) artificiel empêche cet avancement.

B. Timing de réplication et stabilité du génome

Il a été montré qu'il existait une corrélation entre le timing de réplication et différentes altérations du génome.

Des études comparant les mutations ponctuelles au sein d'un génome et entre les génomes de différentes espèces ont pu mettre en évidence une augmentation du nombre de mutations dans les régions répliquées tardivement au cours du cycle cellulaire (Chen et al., 2010; Stamatoyannopoulos et al., 2009; Weber et al., 2012). De plus, une étude chez la levure *S. Cerevisiae* a montré qu'il était possible de modifier le taux de mutation d'un même gène en fonction de sa localisation sur le chromosome. Ce taux est corrélé au timing de réplication, avec une augmentation lorsque la réplication est tardive. Les auteurs ont montré que cette hausse pourrait être dû à l'utilisation des polymérase translésionnelles lors de la réplication tardive, moins fidèles que les polymérase classiques (Lang and Murray, 2011). Cette hypothèse est confortée par une étude chez l'humain montrant une signature mutationnelle de la polymérase translésionnelle Pol ζ dans les régions répliquées tardivement (Seplyarskiy et al., 2015). Une seconde hypothèse a été proposée par Stamatoyannopoulos et al. En effet lors de la réplication tardive, il a été observé une augmentation du blocage des fourches de

réplication entraînant une accumulation d'ADN simple brin plus susceptible d'accumuler des mutations (Lindahl, 1993; Sogo et al., 2002). Ce phénomène pourrait expliquer en partie l'augmentation du taux de mutations observée lors de la réplication tardive. De plus, Blin et al ont montré un lien étroit entre transcription des grands gènes et timing de réplication (Blin et al., 2019). Leurs résultats indiquent que la transcription serait un moyen de pallier à l'instabilité des grands gènes, dont un grand nombre sont porteurs de site fragiles. Un avancement de leur timing de réplication permettrait de compléter leur réplication avant l'entrée en mitose.

Les observations ne sont pas aussi claires en ce qui concerne les variations du nombre de copies (CNV). En effet des études chez la drosophile ont trouvé une association délétion/réplication précoce et duplication/réplication tardive (Cardoso-Moreira et al., 2011) alors qu'une étude réalisée dans des cellules cancéreuses aboutit au résultat inverse (De and Michor, 2011). Des travaux plus récents, analysant le profil des CNV lors de la reprogrammation de fibroblastes en iPS, ont montré que les duplications étaient majoritairement retrouvées dans les régions passant d'une réplication tardive à précoce. A l'inverse les délétions sont majoritaires dans les régions conservant une réplication tardive (Lu et al., 2014). Les différences observées entre ces travaux pourraient être dû à des mécanismes espèce/type cellulaire spécifiques.

C. Timing de réplication et translocation

Au cours de l'évolution, de nombreux réarrangements chromosomiques ont eu lieu. L'étude comparée des génomes de différentes espèces permet de mieux comprendre l'apparition de ces réarrangements. Une étude comparant les génomes Souris/Homme/Chien/Rat/Singe a notamment montré que les réarrangements chromosomiques ont souvent lieu entre des régions présentant le même timing de réplication. Lorsque ce n'est pas le cas on observe le plus souvent une invasion précoce vers tardif. Dans la plupart des cas, les réarrangements distants ont lieu entre des régions répliquées précocement et présentant des interactions physiques détectées en Hi-C (Yaffe et al., 2010). Une autre étude a montré que les régions de

transition entre réplication précoce et tardive étaient davantage sujettes aux remaniements (Watanabe et al., 2002).

Une étude sur des cellules leucémiques a également mis en évidence l'existence d'anomalies du timing de réplication au niveau des réarrangements chromosomiques et notamment des translocations. Cependant ces anomalies peuvent également être retrouvées dans des cellules leucémiques non transloquées, ce qui semble indiquer que ces modifications interviendraient en premier et faciliteraient l'apparition de translocations (Ryba et al., 2012). Une étude plus récente a montré que les remaniements chromosomiques en *cis*, observés dans des cellules cancéreuses, ont généralement lieu entre des régions répliquées tardivement, alors que les remaniements en *trans* ont lieu entre des régions au timing précoce (Du et al., 2019).

Ces résultats mettent en évidence un impact du timing de réplication sur la probabilité de translocation. Cependant on peut se demander si l'inverse est également vrai. Comme nous l'avons vu précédemment, le timing de réplication semble étroitement lié à l'organisation spatiale et chromatinienne. Ainsi l'apparition d'une translocation pourrait venir perturber cette organisation et entraîner des modifications du timing de réplication des chromosomes dérivatifs, et éventuellement de l'ensemble des chromosomes. Il a par exemple été montré que l'induction de translocations dans des cellules humaines entraînait un retard de réplication des chromosomes dérivatifs ainsi qu'un retard de condensation mitotique (Breger et al., 2005). De même une étude de Hi-C, dans des cellules cancéreuses présentant des remaniements chromosomiques, a mis en évidence des modifications locales du timing de réplication au niveau des réarrangements chromosomiques (Dixon et al., 2018).

D. Timing de réplication et pathologies associées

1. Timing de réplication et cancer

Des anomalies de synchronisation de réplication entre les 2 chromosomes homologues, ont été trouvées dans des cellules cancéreuses. Notamment dans le cancer du col de l'utérus, il a été observé une augmentation du nombre de loci asynchrones pour les cancers de haut grade (Amiel et al., 1998). De plus, cette asynchronisme peut être retrouvé dans les lymphocytes

circulants chez des patients atteints de cancers hématologiques mais également de tumeurs solides (Dotan et al., 2000; Grinberg-Rashi et al., 2010; Korenstein-Ilan et al., 2002).

L'analyse de cellules issues de patients atteints de leucémies aiguës lymphoblastiques a également révélé la présence d'altérations locales du timing de réplication (Ryba et al., 2012). Une modification du timing est observée pour 9-18% du génome. De plus, le clustering hiérarchique met en évidence de fortes similitudes entre les échantillons provenant de différents patients. Les modifications observées concernent des régions de petite taille (500kb environ) pouvant correspondre à des domaines de réplication développementaux. De manière plus ciblée, une étude a pu montrer un changement dans le timing de réplication de gènes suppresseurs de tumeurs et d'oncogènes dans des cellules de cancer du sein (Fritz et al., 2013). La modification de la réplication s'accompagne d'un changement d'expression pour 3 des 9 gènes étudiés (*TP53*, *RAD51* et *MYC*).

2. Autres pathologies

Des anomalies du timing de réplication ont pu être montrées dans des cas de syndromes progéroïdes, caractérisés par un vieillissement prématuré (Rivera-Mulia et al., 2017). Dans cette étude, grâce à l'utilisation de cellules iPS, les auteurs ont notamment montré qu'une altération du timing de réplication du gène *TP63* été présente très tôt au cours de la différenciation cellulaire. Ce gène code pour une protéine tumorale. Dans les cellules de patient il est répliqué plus précocement et présente une augmentation de sa transcription.

D'autres anomalies ont pu être observées dans le cas de cholangite sclérosante primitive et maladie intestinale inflammatoire. L'analyse de 2 loci (13q et 15q) chez les patients a montré une augmentation de l'asynchronisme de réplication.

Dans le syndrome de DiGeorges, caractérisé par une perte de la région 22q11.21, un retard de réplication a été observé pour le gène *HIRA/TUPLE1* porté par l'allèle non délété. Ce retard s'accompagne d'un déplacement du locus vers la périphérie du noyau (D'Antoni et al., 2004).

Ces différents travaux semblent donc indiquer un rôle important du timing de réplication pour le maintien de l'intégrité cellulaire. Des anomalies de ce timing sont associées à différentes pathologies, et notamment à divers cancers.

PROJET DE THESE

Mon projet de thèse, intitulé « Etude des translocations chromosomiques : des mécanismes moléculaires à l'oncogénèse », avait pour but l'étude des conséquences globales des translocations chromosomiques et leur rôle dans l'oncogénèse. En effet, comme nous l'avons montré précédemment, la plupart des modèles se limitent à l'étude de la protéine de fusion en dehors de tout contexte chromosomique et souvent, ne permettent pas de récapituler l'ensemble de la pathologie. L'expression du gène de fusion n'a pas lieu sous le contrôle du promoteur endogène, la translocation est absente, et les éventuels effets d'haploinsuffisance sont perdus. Il est donc nécessaire d'utiliser des modèles plus proches de la pathologie. Dans le cadre de cette étude j'ai choisi d'utiliser la translocation t(2,5), associée au lymphome anaplasique à grandes cellules, comme modèle d'étude. Comme décrit dans l'introduction générale de ce manuscrit, dans cette pathologie le lien entre translocation et oncogénèse n'est pas totalement établi. De plus, l'origine cellulaire des tumeurs reste un sujet de débat. Il est donc important de créer un modèle cellulaire reproduisant une translocation identique à celle observée chez les patients afin d'apporter des éléments de réponse à ces différentes questions.

Le premier objectif de ma thèse a consisté à générer de nouveaux modèles d'étude cellulaire de la translocation t(2,5) en utilisant le système CRISPR/Cas9, dont la capacité à induire des translocations avait été démontrée au laboratoire (Ghezraoui et al., 2014; Renouf et al., 2014). Pour cela, j'ai utilisé une lignée cellulaire humaine immortalisée, non transformée, qui présente un caryotype peu remanié (la lignée hTert-RPE-1).

La seconde partie de ma thèse a été dédiée à l'étude des conséquences globales des translocations. Je me suis intéressée en particulier à deux aspects : la formation d'ARN circulaires de fusion, ainsi que l'étude du timing de réplication. L'objectif de cette seconde partie était de mieux comprendre les différents mécanismes pouvant participer à la transformation cellulaire.

Enfin, dans une dernière partie, je me suis attachée à reproduire la translocation t(2,5) dans un modèle cellulaire plus pertinent afin d'étudier l'oncogénèse. J'ai pour cela utilisé des cellules primaires humaines obtenues à partir de sang de donneur sain.

RESULTATS

I. Etude des ARN circulaires de fusion (Article 1, publié)

La première partie de mon projet de thèse s'est concentré sur l'étude des ARN circulaires de fusion spécifiques de la translocation *NPM1-ALK*. Ce travail a été réalisé en collaboration avec Marion Piganeau, une autre doctorante du laboratoire.

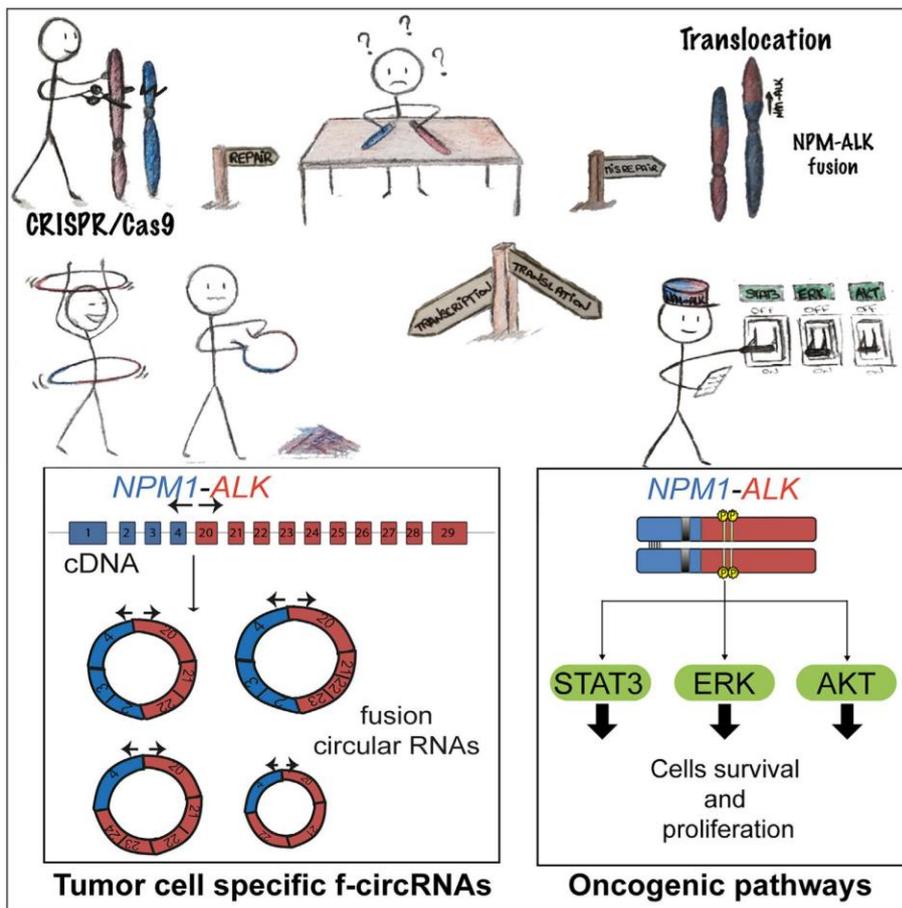
Dans un premier temps, nous avons généré des modèles cellulaires humains et murins pour la translocation *NPM1-ALK* en utilisant le système CRISPR/Cas9. Chez la souris, les gènes *Npm1* et *Alk* sont portés par les chromosomes 11 et 17 respectivement. Le modèle murin nous a permis de mettre en évidence le potentiel oncogénique de *NPM1-ALK* par la levée de la dépendance à l'IL3 dans la lignée BA/F3 via l'activation des voies ERK, AKT et STAT3. Le modèle humain, dans la lignée RPE-1, nous a également permis de montrer l'activation de la voie STAT3 (ERK et AKT étant déjà activées à l'état basal), ainsi que le potentiel oncogénique de la fusion en absence de facteurs de croissance (milieu de culture faible en sérum).

Nous avons ensuite mis en évidence la formation d'ARN circulaires de fusion spécifiques de la translocation dans les deux modèles. Nous avons découvert un ARN circulaire commun à l'Homme et la souris. Quatre ARN circulaires ont été identifiés dans l'ensemble des clones humains, et ont également été retrouvés dans les lignées ALCL SUPM2 et SUDHL1.

Enfin, nous avons étendu notre analyse à un autre modèle de translocation chez la souris. Nous avons induit une translocation entre les loci *Mll* et *Af9*, localisés sur les chromosomes 9 et 4 respectivement. Nous avons également pu mettre en évidence la formation d'ARN circulaires de fusion dans ce modèle.

Article

Chromosomal Translocation Formation Is Sufficient to Produce Fusion Circular RNAs Specific to Patient Tumor Cells



Loelia Babin,
Marion Piganeau,
Benjamin
Renouf, ...,
Thomas Mercher,
Carine
Giovannangeli,
Erika C. Brunet

erika.brunet@inserm.fr

HIGHLIGHTS

CRISPR/Cas9 model of ALCL translocation leads to oncogene pathway activation

CRISPR/Cas9 translocations generate *de novo* fusion circular RNAs

Shared fusion circular RNAs are found in CRISPR/Cas9 models and ALCL tumor cells

Babin et al., iScience 5, 19–29
July 27, 2018 © 2018 The Authors.
<https://doi.org/10.1016/j.isci.2018.06.007>



Article

Chromosomal Translocation Formation Is Sufficient to Produce Fusion Circular RNAs Specific to Patient Tumor Cells

Loelia Babin,^{1,2,6} Marion Piganeau,^{2,5,6} Benjamin Renouf,² Khadija Lamribet,² Cecile Thirant,³ Ludovic Deriano,⁴ Thomas Mercher,³ Carine Giovannangeli,² and Erika C. Brunet^{1,2,7,*}

SUMMARY

Circular RNAs constitute a unique class of RNAs whose precise functions remain to be elucidated. In particular, cancer-associated chromosomal translocations can give rise to fusion circular RNAs that play a role in leukemia progression. However, how and when fusion circular RNAs are formed and whether they are being selected in cancer cells remains unknown. Here, we used CRISPR/Cas9 to generate physiological translocation models of NPM1-ALK fusion gene. We showed that, in addition to generating fusion proteins and activating specific oncogenic pathways, chromosomal translocation induced by CRISPR/Cas9 led to the formation of *de novo* fusion circular RNAs. Specifically, we could recover different classes of circular RNAs composed of different circularization junctions, mainly back-spliced species. In addition, we identified fusion circular RNAs identical to those found in related patient tumor cells providing evidence that fusion circular RNAs arise early after chromosomal formation and are not just a consequence of the oncogenesis process.

INTRODUCTION

Circular RNAs (circRNAs) were discovered in eukaryotic cells more than 30 years ago (Hsu and Coca-Prados, 1979) but were first classified as splice errors (Cocquerelle et al., 1993). Only recently new reports showing the emerging role of circRNAs in multiple biological functions drew renewed attention to this specific repertoire of non-coding RNAs (Hansen et al., 2013; Jeck et al., 2013). In particular, recent advances in high throughput sequencing showed high abundance and conservation of circRNAs among species. Biogenesis of circRNAs remains unclear, but different mechanisms have been described that are associated with canonical and non-canonical splicing, and with the presence of repeat sequences in the introns surrounding the junction of circularization (Conn et al., 2015; Errichelli et al., 2017; Ivanov et al., 2015; Zhang et al., 2013, 2016). CircRNAs can act as transcription regulators (Li et al., 2015), microRNA sponges (Hansen et al., 2013; Memczak et al., 2013; Chen et al., 2017), or platforms for RNA binding proteins (Ashwal-Fluss et al., 2014). In addition, a small subset of circRNAs is translated (Pamudurti et al., 2017). Recently, circRNAs have been shown to be implicated in cell differentiation (Legnini et al., 2017) and to be associated with human diseases (Chen et al., 2017; Hsiao et al., 2017; Shao and Chen, 2016) and could be used as diagnostic biomarkers (Lu et al., 2017). A study from Pandolfi's laboratory demonstrated that chromosomal translocations lead to the expression of fusion-specific circRNAs in various types of tumor cells (Guarnerio et al., 2016).

Recurrent cancer translocations disrupt tumor suppressors, activate oncogenes, or generate aberrant fusion proteins. Expression of the resulting fusion cDNA has been the main strategy to study translocation-driven oncogenesis. Although this strategy provides valuable insights and understanding regarding the role of several specific translocations, the outcome does not always phenocopy the respective disease and may not accurately recapitulate the multistep process of translocation-driven oncogenesis. For instance, TEL-AML1 knock-in mice remain without evidence of disease, and treatment of these animals with mutagenic compounds mainly induces T cell malignancies without characteristics of B-ALL as observed in humans (Schindler et al., 2009). Particularly, overexpression of fusion cDNA hardly reproduces physiological expression of the fusion gene (dosage effects) and does not take into account other potential oncogenic consequences of translocations including reciprocal fusion gene formation, disruption of endogenous fused genes, or transcriptional and epigenetic changes resulting from repositioning of translocated chromosomes in the nucleus.

¹Laboratory "Genome Dynamics in the Immune System", Equipe Labellisée Ligue Contre le Cancer, INSERM UMR1163; Université Paris Descartes Sorbonne Paris Cité, Institut Imagine, Paris 75015, France

²Museum National d'Histoire Naturelle, INSERM U1154, CNRS UMR7196, Sorbonne Université, Paris 75005, France

³INSERM U1170, Equipe Labellisée Ligue Contre le Cancer, Gustave Roussy Institute, Université Paris Diderot, Université Paris-Sud, Villejuif 94800, France

⁴Genome Integrity, Immunity and Cancer Unit, Department of Immunology, Department of Genomes and Genetics, Institut Pasteur, Paris 75015, France

⁵Present Address: Haematopoietic Stem Cell Lab, The Francis Crick Institute, London NW1 1AT, UK

⁶These authors contributed equally

⁷Lead Contact

*Correspondence: erika.brunet@inserm.fr
<https://doi.org/10.1016/j.isci.2018.06.007>



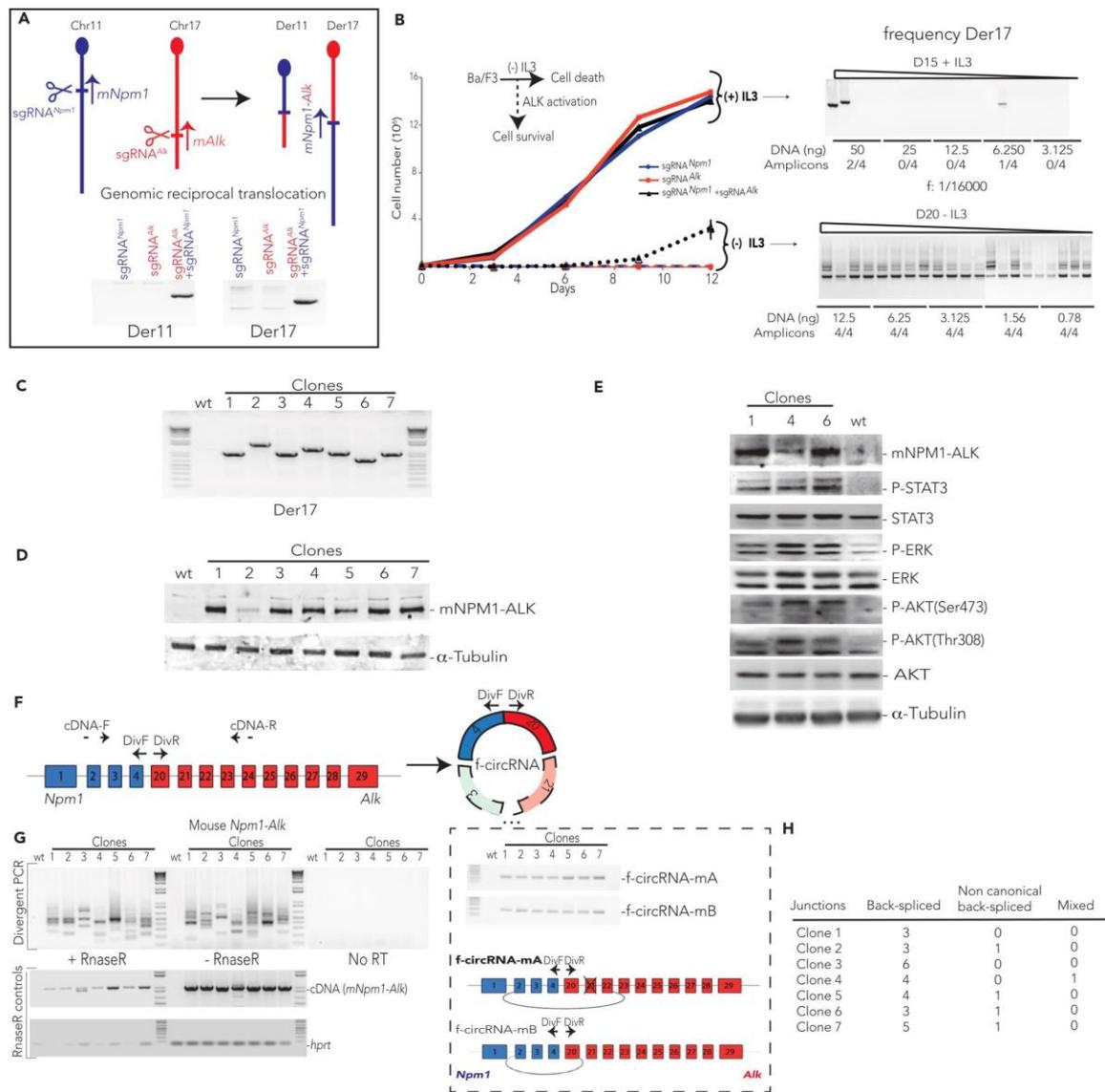


Figure 1. NPM1-ALK Translocation Induces Ba/F3 Cells Transformation and Leads to f-circRNA Formation

(A) In mouse cells, *Npm1* and *Alk* genes are located on chromosomes 11 and 17, respectively. To induce t(11; 17) translocation, CRISPR/Cas9 system is used to create specific *Npm1* (using *sgRNA^{Npm1}*) and *Alk* (using *sgRNA^{Alk}*) DSBs. The derivative chromosome 17 (Der17) will lead to *Npm1-ALK* fusion gene expression. Both derivative chromosomes, Der11 and Der17, are recovered only when *Npm1* and *Alk* DSBs are concomitantly induced (detected by nested PCR).

(B) Proliferation curve of CRISPR/Cas9-treated cells after IL-3 removal. Left panel: cytokine-independent growth was observed only from cells originating from the pool treated with *sgRNA^{Npm1}* and *sgRNA^{Alk}* (mean of three experiments \pm SD). Right panels: estimate of translocation frequency using PCR on serial dilutions of genomic DNA from Ba/F3 cells. The number of times the PCR was positive for each dilution is indicated (amplicons) (from four independent experiments). At day 15 after transfection (upper panel), translocation junction from cells cultured with IL-3 was detected in two wells of 50 ng DNA, reflecting a frequency around 6.25×10^{-5} . Instead, serial dilutions of DNA from cells growing in the absence of IL-3 (lower panel) showed PCR amplicons for all DNA dilutions (the gel showing all the dilutions is the result of the association of two consecutive gel lines).

(C) PCR amplification of Der17 from seven clones growing without IL-3.

(D) Western Blot analysis of NPM1-ALK expression. Each of the seven clones expressed NPM1-ALK fusion protein.

(E) Western Blot detection of STAT3, AKT, and ERK1/2 phosphorylation for clones 1, 4, and 6. STAT3, AKT, and ERK1/2 activation was detected in all clones.

Figure 1. Continued

(F) PCR strategy to detect fusion circular RNAs (f-circRNAs) using divergent primers (DivF and R). Convergent primers (cDNA-F and R) are used to detect linear fusion mRNA (*Npm1* exons in blue and *Alk* exons in red).

(G) PCR-amplified products obtained using divergent primers to detect f-circRNAs corresponding to *Npm1-Alk* gene in Ba/F3 cells (wild-type cells and seven translocated clones). Representative gel from three independent experiments. Schemes of f-circRNA-mA and f-circRNA-mB (right panel) detected in the clones. f-circRNA-mA and f-circRNA-mB were amplified with specific primers.

(H) Number of circular junctions sequenced for each different type of f-circRNA (back-spliced, non-canonical back-spliced, and mixed). Sequencing was done from three independent experiments.

See also Figure S1 and Table S1.

The consistent advances in programmable nuclease technology, including ZFNs (zinc-finger nucleases), TALENs (transcription activator-like effector nucleases), and CRISPR/Cas9 RNA-guided nucleases, facilitates easier and more precise genome editing. Particularly, we and others have shown that introduction of DNA double-strand breaks (DSBs) at patient translocation breakpoints enables specific oncogenic translocation formation (Brunet et al., 2009; Piganeau et al., 2013; Choi and Meyerson, 2014; Ghezraoui et al., 2014; Torres et al., 2014).

In this study, we used CRISPR/Cas9 technology to induce specific chromosomal translocations. We managed to assess immediate cellular consequences, including newly expressed fusion circular RNAs (f-circRNAs). Focusing on *NPM1-ALK* fusion found in anaplastic large cell lymphoma (ALCL) (Morris et al., 1994), we confirmed that CRISPR-induced *NPM1-ALK* translocation was not only sufficient to trigger murine Ba/F3 pro-B cell line transformation but also induced the activation of STAT3, MEK/ERK, and AKT pathways, which are upregulated in ALCL tumors. We demonstrated that newly formed translocations lead to direct expression of specific f-circRNAs transcribed from the *NPM1-ALK* breakpoint junction. Sequencing of f-circRNAs reveals different types of circularization junctions. Strikingly, f-circRNAs found in tumor cell lines of patients with ALCL were also identified in our different translocation models. Thus, our study provides strong evidences that different f-circRNAs, including specific f-circRNAs, found in patient tumor cells are expressed directly after *de novo* translocation induction. These results further support the use of CRISPR/Cas9 to induce *in situ* translocations to reach more relevant cancer models including expression of *de novo* f-circRNAs.

RESULTS**CRISPR/Cas9-Induced NPM1-ALK Fusion Leads to STAT3, AKT, and ERK Pathway Activation in Mouse Cells**

NPM1-ALK is an oncogenic fusion protein that is capable of transforming multiple rodent cell lines (Bai et al., 1998; Fujimoto et al., 1996). Particularly, human *NPM1-ALK* overexpression has been shown to confer interleukin (IL)-3-independent survival and proliferation of Ba/F3 murine pro-B lymphocytes (Bai et al., 1998). In the mouse genome, *Npm1* and *Alk* genes are located on chromosomes 11 and 17, respectively. *NPM1-ALK* is expressed from the derivative chromosome 17 (Der17) (Figure 1A). We designed single guide (sgRNAs) targeting murine *Npm1* intron 4 and *Alk* intron 19 to induce concomitant DSBs at loci found as breakpoints in human ALCL. Transient co-expression of Cas9 with sgRNA^{*Npm1*} and sgRNA^{*Alk*} led to the formation of the two derivative chromosomes (Der11 and Der17). In contrast, a single break on *Npm1* or *Alk* was not sufficient to induce translocation (Figure 1A). The frequency of translocation at day 5 after transfection was $\geq 2.5 \times 10^{-4}$ (Figure S1A) (as calculated in [Renouf et al., 2014]). However, the translocation frequency of cells growing in the presence of IL-3 dropped to 6.25×10^{-5} at day 15 after transfection, indicating that *Npm1-Alk* does not provide a growth advantage in these conditions (Figure 1B). To investigate the ability of CRISPR/Cas9-induced *Npm1-Alk* translocation to transform Ba/F3 cells, we removed IL-3 from the medium of transfected cells. IL-3 withdrawal led to major growth arrest and cell death when cells were treated with a single sgRNA. In contrast, after 6 days without IL-3, we observed proliferation of cellular clones from the pool treated with both sgRNA^{*Npm1*} and sgRNA^{*Alk*}. After 20 days, we were able to recover multiple translocation junctions. We assessed translocation frequency in proliferating cells without IL-3 and detected Der17 junctions starting from as little as 0.8 ng of DNA, suggesting that most cells carried translocation (Figure 1B). Thus, we confirmed that CRISPR/Cas9-induced mouse *Npm1-Alk* translocation leads to the transformation of Ba/F3 cells as recently shown (van de Krogt et al., 2017). In addition, to validate that IL-3-independent proliferation was driven by constitutive activation of *NPM1-ALK*, we treated the cells with crizotinib, an ALK phosphorylation inhibitor. We found that crizotinib led to complete proliferation arrest of selected cells in the absence of IL-3, indicative of active *NPM1-ALK* fusion protein in the whole population (Figure S1B).

We then isolated seven random Ba/F3 clones from cells growing without IL-3. We could detect Der17 in all of them, further confirming the observation that translocations had occurred in IL-3-independent cells (Figure 1C). Importantly, NPM1-ALK fusion protein was expressed in all the clones (Figure 1D). Insertions and deletions were detected at the translocation junctions but restricted to the intronic region between *Npm1* and *Alk* genes (Figure S1C). We further analyzed three clones (clones 1, 4, and 6). Each of them showed clear co-localization between one *Npm1* and one *Alk* signal as observed by fluorescence *in situ* hybridization (FISH) (Figure S1D). *In vitro* and *in vivo* studies have shown that NPM1-ALK kinase promotes cell growth, cell cycle progression, and survival through transphosphorylation of cytosolic targets, including effector proteins of STAT3, MEK/ERK, and AKT pathways (Marzec et al., 2007a, 2007b; Zamo et al., 2002; Zhang et al., 2002). We therefore wondered if *de novo* mouse t(11; 17) translocation induced by CRISPR activates the same signaling cascades. We showed that STAT3 was expressed both in surviving clones and wild-type Ba/F3 cells. However, phosphorylated STAT3 was exclusively detected in the three NPM1-ALK-positive clones, indicating that newly formed NPM1-ALK activates the STAT3 pathway. Similarly, the phosphorylation of ERK1/2 protein was substantially enhanced in the three translocated clones compared with wild-type cells and we could observe greater phosphorylation of specific phospho-sites of AKT (Ser473 and Thr308) (Figure 1E). Our results clearly demonstrate that CRISPR-induced *Npm1-Alk* translocation is sufficient to trigger transformation of murine Ba/F3 pro-B cell line through the activation of ALCL-specific oncogenic pathways.

De Novo *Npm1-Alk* Translocation Leads to Expression of Circular RNAs Specific to Breakpoint Junctions (f-circRNAs) in Mouse Cells

Recently, f-circRNAs expressed from translocation junctions have been identified in various tumor cells associated with *PML-RAR α* , *MLL-AF9*, *EWSR1-FLI1*, or *EML4-ALK1* translocation (Guarnerio et al., 2016). This work demonstrated that the presence of *MLL-AF9*-specific f-circRNAs contributes to cell survival upon therapy and plays an active role in leukemia progression *in vivo*. Thus, we wondered if *de novo* translocation formation leads to the expression of f-circRNAs specific to the breakpoint junction. We amplified f-circRNAs using outward-facing PCR primers around the *Npm1-Alk* translocation junction, after RnaseR treatment of our samples to decrease the quantity of linear RNAs as previously described (Guarnerio et al., 2016) (Figures 1F and 1G). Importantly, no f-circRNAs could be amplified in wild-type Ba/F3 cells. We investigated the formation of f-circRNAs in the Ba/F3 clones growing in the absence of IL-3, positive for NPM1-ALK translocation. Several PCR amplicons could be detected in each of the seven clones, indicating that f-circRNAs are specific to *de novo* translocation (Figures 1G and 1H). To validate that our amplicons result from f-circRNA formation in cells and are not amplified from the linear *Npm1-Alk* counterpart, we transcribed *in vitro* the NPM1-ALK cDNA from human ALCL cells (homology between mouse and human of NPM1-ALK exons is >80%), and in parallel, an identified f-circRNA containing six exons as positive control (f-circRNA-hF described in the following section) (Figure S1E). After RT-PCR and PCR amplification using similar conditions as for f-circRNA amplification (in the presence of 1 μ g RNA from wild-type cells and with RnaseR treatment), we could detect some PCR amplicons using divergent primers and starting from as little as 1 ng of the linear NPM1-ALK *in vitro* transcript. Sequencing revealed that these unspecific amplicons (18 sequences) corresponded exclusively to sequences containing short direct repeats located at an intra-exonic junction between NPM1 and ALK (Table S1). These unspecific products have been considered as PCR artifacts and have subsequently been excluded from our datasets. More precisely, we excluded every sequence with at least 1 bp found in common at the breakpoint between NPM1 and ALK intra-exonic sequences.

We sequenced four to seven random PCR products from *Npm1-Alk* clones 1 to 7. After excluding PCR artifacts, we defined three different groups of f-circRNAs: (1) back-spliced f-circRNAs reflecting proper back-spliced events (circularized from two canonical splice sites), (2) non-canonical back-spliced f-circRNAs (circularized within two exons that probably arise from alternative splice sites), and (3) mixed f-circRNAs (circularized from one canonical splice site to a site located within an exon) (Figure 1H and Table S1). Of note, some f-circRNAs showed exon skipping (for example, f-circRNA-mA loses the exon 21, Figure 1G). This phenomenon could be similar to circRNAs described by Zhao's team and was called alternative splicing events (Gao et al., 2016). We could identify back-spliced f-circRNAs in every clone (three to six for each clone), and importantly, sequencing revealed that some back-spliced f-circRNAs could be found in several clones (Table S1). We could detect f-circRNA-mA (joining exon 2 of *Npm1* to exon 22 of *Alk*) and f-circRNA-mB (joining exon 2 of *Npm1* to exon 20 of *Alk*) in each of our seven clones by using specific PCR amplification (Figure 1G). In conclusion, *Npm1-Alk*-translocated mouse clones expressed different types of f-circRNAs. Two of them are commonly expressed in all clones, indicating a potential selection process of specific f-circRNAs upon transformation of Ba/F3 cells.

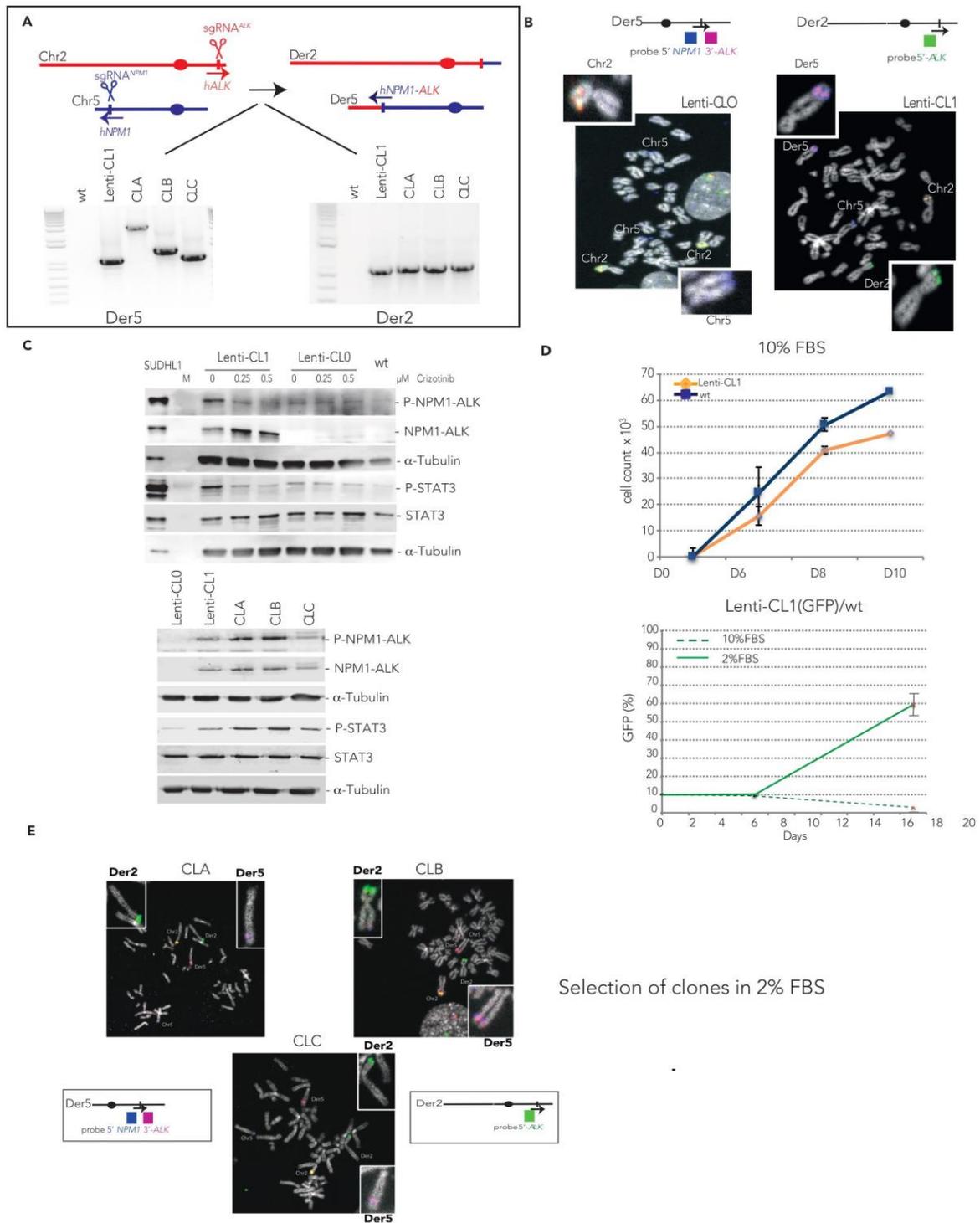


Figure 2. NPM1-ALK Activates STAT3 Pathway and Increases Cell Growth in Low-Serum Culture

(A) In human cells, *NPM1* and *ALK* genes are located on chromosomes 5 and 2, respectively. To induce t(2; 5) translocation, CRISPR/Cas9 system is used to induce specific *NPM1* (using sgRNA^{NPM1}) and *ALK* (using sgRNA^{ALK}) DSBs. The derivative chromosome (Der5) will lead to *NPM1*-*ALK* fusion gene expression. Der5 and Der2 junctions were detected in the four clones Lenti-CL1 (obtained using a lentivirus vector), CLA, CLB, and CLC isolated from sgRNA^{NPM1}- and sgRNA^{ALK}-treated cells.

(B) FISH for *NPM1* and *ALK* detection in Lenti-CL0 (control clone) and Lenti-CL1 metaphases. *ALK* locus is detected with a break-apart probe (green + purple). Blue signal corresponds to the 5' end *NPM1* locus: Der5 and Der2 formation leads to split purple and green signals. Colocalization of purple and blue signals reveals Der5 formation.

(C) Western blot detection of NPM1-ALK and STAT3 activation in the translocated clones. SUDHL1 is an *ALK*-positive ALCL cell line. Lenti-CL0 and Lenti-CL1 were also treated with different amounts of crizotinib.

(D) Upper panel: growth curve of wild-type RPE-1 cells and Lenti-CL1 in 10% FBS. Lower panel: competitive proliferation assay between wild-type RPE-1 cells and Lenti-CL1 in the presence of 10% or 2% FBS. Lenti-CL1 stably expressed GFP (mean of three experiments \pm SD).

(E) FISH analysis of RPE-1 clones growing in 2% FBS: CLA, CLB, and CLC. A break-apart probe (purple and green) was used for *ALK* and a BAC probe (blue) was used to probe *NPM1*. Der5 and Der2 formation leads to split purple and green signals. Colocalization of purple and blue signals reveals Der5 formation. See also Figure S2.

CRISPR/Cas9-Induced NPM1-ALK Fusion Leads to STAT3 Pathway Activation, Growth Advantage in Low-Serum Medium, and Expression of Specific f-circRNAs in Human Cells

We further decided to confirm the formation of f-circRNAs in a human cell line after *NPM1*-*ALK* formation. We first induced t(2; 5) (p23; q35) translocation in the human RPE-1 hTERT-immortalized retinal pigment epithelial cell line, which has a near-diploid karyotype, making it relevant to analyze chromosomal rearrangements. We previously set up the CRISPR/Cas9 transfection system to efficiently induce t(2; 5) (p23; q35) in RPE-1 cells using sgRNAs targeting *NPM1* intron 4 and *ALK* intron 19 (Ghezraoui et al., 2014; Renouf et al., 2014). However, in these studies, we failed to efficiently recover translocated clones. Here we adapted a lentiviral strategy (Heckl et al., 2014) to stably integrate the Cas9 nuclease and the two sgRNAs targeting *NPM1* and *ALK* (Figure S2A). After transduction and fluorescence-activated cell sorting (FACS) sorting of transduced cells, we isolated one *NPM1*-*ALK* positive clone (Lenti-CL1) and one *NPM1*-*ALK*-negative clone (Lenti-CL0) confirmed by FISH analysis (Figures 2A and 2B). Sequencing of derivative chromosome junctions confirmed the formation of fusion junctions in Lenti-CL1 (Figure S2A). *NPM1*-*ALK* fusion protein was stably expressed and active as shown by its auto-phosphorylation (Figure 2C). Importantly, the phosphorylation dropped back to background level upon crizotinib treatment. Similarly, STAT3 showed enhanced phosphorylation in Lenti-CL1 clone compared with negative Lenti-CL0 cells and phosphorylation went back to a basal level with crizotinib treatment, indicating a reliance on *NPM1*-*ALK* activity. Of note, RPE-1 cells have high endogenous ERK1/2 and AKT phosphorylation, preventing the assessment of *NPM1*-*ALK*-driven phosphorylation for those pathways. Intriguingly, crizotinib treatment had no discernable effect on cell growth or survival of Lenti-CL1 using normal culture conditions (data not shown). Instead, wild-type cells grew faster than Lenti-CL1 cells in 10% of fetal bovine serum (FBS) (Figure 2D), possibly explaining the difficulties we previously faced in isolating translocated cells by traditional cloning method. Thus, as for Ba/F3 cells, we also observed a dramatic decrease in the frequency of translocated cells over time for cells transfected with sgRNA^{NPM1} and sgRNA^{ALK}. This phenomenon was independent of the initial frequency of translocation (5×10^{-4} [high] or 1.25×10^{-4} [low] [Figure S2B]). These results prompted us to test the growth of *NPM1*-*ALK* clone in the presence of low serum concentration (2% FBS). In this condition, Lenti-CL1 (GFP-positive cells) showed then a strong proliferative advantage over wild-type cells (2% FBS) (Figure 2D) and the translocation frequency increased up to 16-fold after 16 days (Figure S2B). These experiments strongly indicate that *NPM1*-*ALK* translocation led to a marked growth advantage in culture conditions containing lower growth factor concentrations. Thus we envisioned that the recovery of translocated clones would be enhanced by culturing sgRNA^{NPM1}- and sgRNA^{ALK}-transfected cells in 2% FBS medium. As expected, we easily isolated three additional *NPM1*-*ALK*-positive clones by simple cloning using low-serum medium (CLA, CLB, and CLC; Figures 2A and S2A). These clones were translocated as shown by FISH (Figure 2E) and expressed an active form of *NPM1*-*ALK*, leading to STAT3 phosphorylation (Figure 2C).

We next looked for the formation of f-circRNAs in our different human RPE-1 clones carrying *NPM1*-*ALK* translocation (Figure 3A). As for mouse cells, no f-circRNA could be amplified from wild-type cells. In contrast, several PCR products were detected and sequenced from all four *NPM1*-*ALK* clones (Figures 3A and 3B and Table S2). *NPM1*-*ALK* translocation induced in human cells also led to the expression of the three different types of f-circRNAs. Importantly, four back-spliced f-circRNAs were found common in all clones (Figure 3A and Table S2). Here we also confirmed that our treatment with RnaseR specifically

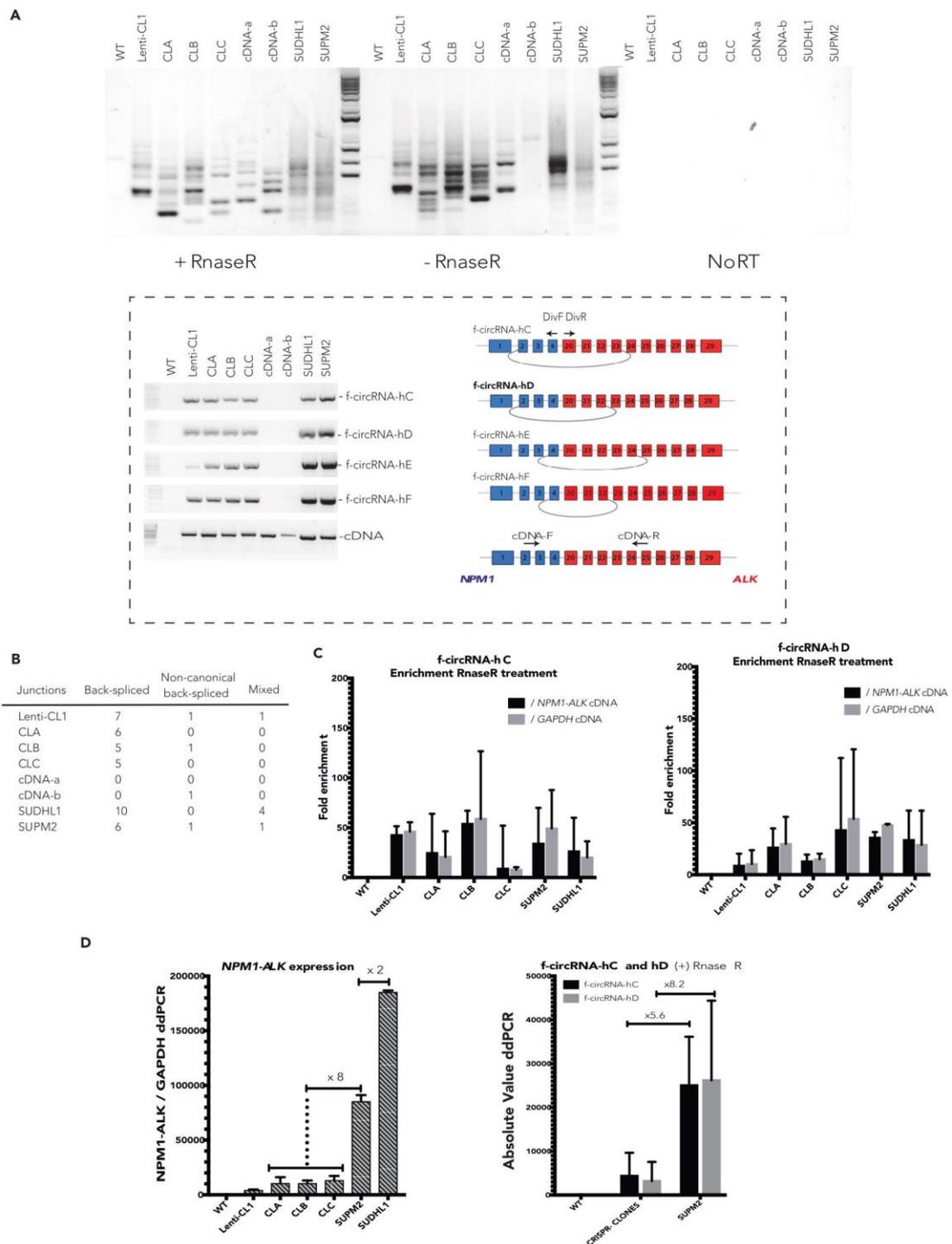


Figure 3. *NPM1-ALK* Induces the Expression of ALCL-Specific f-circRNAs

(A) PCR-amplified products obtained using divergent primers to detect f-circRNAs corresponding to *NPM1-ALK* gene in RPE-1 human cells: from wild-type cells, CRISPR/Cas9-translocated clones, cells overexpressing *NPM1-ALK* (labeled cDNA-a and cDNA-b), and SUDHL1 and SUPM2 ALCL cells. Representative gel from three independent experiments. Schemes of f-circRNA-hC, f-circRNA-hD, f-circRNA-hE, and f-circRNA-hF (middle panel) shared by Lenti-CL1, CLA, CLB, CLC, SUDHL1, and SUPM2 cells. The PCR product corresponding to f-circRNA-hD (bold) has similar circularization junction as f-circRNA-mA in Ba/F3 cells (Figure 1A). f-circRNA-hC, f-circRNA-hD, f-circRNA-hE, and f-circRNA-hF were also amplified with specific primers. Detection of *NPM1-ALK* fusion gene (cDNA) was done using the convergent primers.

(B) Number of circular junctions sequenced for each type of *NPM1-ALK* f-circRNA in RPE-1 and ALCL cells (back-spliced, non-canonical back-spliced, and mixed). Sequencing was done from three independent experiments.

(C) Fold enrichment for f-circRNA-hC/ f-circRNA-hD after RnaseR treatment normalized by *NPM1-ALK* linear cDNA and *GAPDH* cDNA (mean of three experiments \pm SEM).

(D) Left panel: *NPM1-ALK* linear cDNA expression level in CRISPR/Cas9 clones and SUPM2 and SUDHL1 cells. Right panel: f-circRNA-hC or f-circRNA-hD expression level in CRISPR clones compared with SUPM2 cells (average from all the clones) (mean of three experiments \pm SEM).

See also Figure S3 and Table S2.

diminished the quantity of linear RNAs compared with the identified f-circRNAs (from 8- to 50-fold as quantified in Figure 3C), decreasing the potential background of PCR due to unspecific amplification from linear cDNA (described in the previous section).

Furthermore, we generated RPE-1 clones stably expressing human *NPM1-ALK* cDNA (named cDNA-a and cDNA-b). *NPM1-ALK* expression was confirmed by PCR amplification of fusion cDNA (Figure 3A). In our conditions of amplification and after excluding PCR artifacts, only one non-canonical back-spliced f-circRNA has been detected in clones expressing *NPM1-ALK* cDNA (Figure 3B and Table S2), in contrast to cells carrying translocated chromosomes, which expressed more f-circRNAs and mostly canonical back-spliced f-circRNAs.

One particular f-circRNA from *MLL-AF9* fusion was extensively studied by Guarnerio et al. (2016) and found as tumor specific. Thus, to further investigate the circRNA repertoire formed after CRISPR-induced translocation, we designed a translocation system for *Mll-Af9* fusion (Drynan et al., 2005). We could reproduce the translocation in a mouse pro-B cell line and obtained three translocated clones for *Mll-Af9* genes as confirmed by PCR amplification of the derivative chromosomes, *Mll-Af9* cDNA amplification, and FISH analysis (Figures S3A and S3B). We could readily detect different types of f-circRNAs, and some back-spliced f-circRNAs were prevalent in all clones, as observed for *NPM1-ALK* model (Figure S3C). For instance, back-spliced f-circRNA-mG and f-circRNA-mH were found in each of the three clones (Figure S3C and Table S3). These preliminary results validate that common f-circRNAs are generated by the formation of endogenous *Mll-Af9* translocation, as observed for the *NPM1-ALK* model. Further validation by RNA-seq in human CRISPR/Cas9 clones would be needed.

In conclusion, the formation of derivative chromosomes using CRISPR/Cas9 technology leads to a specific landscape of f-circRNAs.

De Novo *NPM1-ALK* Translocation Leads to the Expression of f-circRNAs Specific to Human ALCL Tumor Cells

To gain insight into the relevance of the repertoire of f-circRNAs formed in our clones, we compared them with the ones found in ALCL tumor cells, SUDHL1 and SUPM2 cell lines. As for the RPE-1 clones, we recovered all types of f-circRNAs in the two ALCL cell lines. Notably, the four back-spliced f-circRNAs shared by all *NPM1-ALK* RPE-1 clones induced by CRISPR/Cas9 were also expressed in SUDHL1 and in SUPM2 patient cells: f-circRNA-hC joining exon 2 to exon 23, f-circRNA-hD joining exon 2 to exon 22, f-circRNA-hE joining exon 4 to exon 24, and f-circRNA-hF joining exon 4 to exon 22 (Figure 3A and Table S2). Remarkably, f-circRNA-hD, which circularized the exon 2 of *NPM1* to the exon 22 of *ALK*, was also expressed in all the mouse *Npm1-Alk*-positive clones (corresponding to f-circRNA-mA, Figure 1G). We then estimate the relative expression level of two of the four common f-circRNAs. The level of f-circRNA-hC and f-circRNA-hD in CRISPR/Cas9 clones was found to be 5- to 8-fold lower than in SUPM2 tumor cells. This level directly correlates to the difference of expression of *NPM1-ALK* cDNA between the clones and the tumor cells (of note SUDHL1 tumor cells contain two copies of the translocation, leading to 2-fold higher expression of *NPM1-ALK* compared with SUPM2 cells) (Figure 3D).

Thereby CRISPR/Cas9 approach to induce translocation leads to modifications of the non-coding RNA repertoire by expressing tumor-related f-circRNAs.



DISCUSSION

Here, we successfully used CRISPR/Cas9 nucleases to induce *de novo* chromosomal translocations leading to *NPM1-ALK* fusion oncogene expression in mouse and human cell lines. We confirmed that endogenous mouse *NPM1-ALK* expression is sufficient to confer cytokine-independent growth of mouse cells as recently shown by the group of Wlodarska (van de Krogt et al., 2017). Furthermore, we demonstrated that the cell survival depended on *NPM1-ALK* activation and was inhibited by crizotinib treatment. CRISPR/Cas9-mediated rearranged mouse and human cell lines displayed STAT3 (for our human cells), AKT, and ERK1/2 phosphorylation depending on *NPM1-ALK* kinase expression and activity. Activation of these three proteins was found essential for cell transformation and tumor maintenance in cancer cells (Marzec et al., 2007b; Zamo et al., 2002; Zhang et al., 2002). In addition, low concentrations of growth factors favored proliferation of translocated clones over wild-type cells. This observation is consistent with the fact that *NPM1-ALK* fusion protein expressed at physiological levels can substitute cytokine stimulation. In conditions in which wild-type cells grow easily (10% serum), translocated cells are probably diluted before acquiring any advantage phenotype.

Our results demonstrated that *de novo* translocation formation promptly leads to the expression of *de novo* circRNAs specific to the fusion breakpoint. Interestingly, f-circRNAs found in cancer-associated chromosomal translocation have already been shown to participate in cellular transformation (Guarnerio et al., 2016). In this study, expression of a specific f-circRNA in hematopoietic stem cells (containing the *MLL-AF9* fusion) led to a proliferative advantage and favored the progression of leukemia *in vivo*, demonstrating a physiological role for f-circRNAs. To gain further insight into f-circRNA formation, we generated CRISPR-mediated translocations. We observed the expression of several f-circRNAs upon translocation formation encompassing from two to nine exons (with a length up to 980 bp). The fact that several different f-circRNAs were expressed from the same fusion gene is quite surprising, and whether or not they all have a physiological role remains to be determined. The presence of specific exons may also have a particular biological relevance. Remarkably, we identified specific f-circRNAs expressed in translocated clones perfectly matching those found in tumor-related cells. We also recovered one specific f-circRNA (f-circRNA-mA and f-circRNA-hD) in our mouse and human *NPM1-ALK*-positive clones, respectively, that was also expressed in two ALCL cell lines, supporting the idea that circRNAs are well conserved between human and mouse cells (Legnini et al., 2017; Rybak-Wolf et al., 2015).

Several types of f-circRNAs were identified in this study. Importantly, CRISPR/Cas9 *NPM1-ALK*-translocated clones showed a different f-circRNA repertoire compared with clones that expressed ectopically *NPM1-ALK* cDNA. Several back-spliced species were expressed in CRISPR/Cas9 clones. This class of f-circRNAs properly joins a 5' splice site with a 3' splice site of an upstream intron. These events are thought to arise from the annealing of inverted repeat sequences inside flanking exons, enhancing the juxtaposition of the splice sites leading to back-splicing (Jeck et al., 2013). The f-circRNAs that were found common to several clones and tumor cells (but not in cells ectopically expressing *NPM1-ALK* cDNA) were exclusively canonical back-spliced f-circRNAs, in agreement with their identification in earlier RNA-seq studies.

Surprisingly, *in vitro* transcription revealed that unspecific PCR amplicons can be generated when amplifying linear cDNA using classical divergent PCR designed for circRNAs. All these events exhibit short direct repeats for at least 1 bp located at the circularized junction between *NPM1* and *ALK*. They are probably produced after PCR template switching, although we can wonder if 1 bp is enough to lead to efficient polymerase template switch. Efficient treatment with RnaseR is then essential for decreasing false-positives from these types of experiments. Instead, we kept mixed events where no repeat could be found, indicating that these types of PCR products were not found as amplified from linear cDNA. The relevance of these forms of f-circRNA (which probably reflect the presence of non-canonical back-spliced sites) remains to be elucidated in further RNA sequencing (RNA-seq) studies. In conclusion, the CRISPR/Cas9 strategy to induce chromosomal translocation provides experimental settings where *de novo* f-circRNAs are formed: this method will certainly help to precisely dissect the biogenesis and the functional roles of circRNAs since *de novo* circRNAs can be easily tracked.

To conclude, our study provides evidence that f-circRNAs found in tumor cells are produced directly from newly rearranged chromosomes. Therefore, precise genome rearrangements obtained with CRISPR/Cas9 provide faithful models phenocopying chromosomal translocation-associated diseases including activation of oncogenic pathways and also expression of tumor-specific f-circRNAs by bypassing limitations of models

expressing the fusion gene ectopically. Identification of specific f-circRNAs could provide new diagnostic markers related to cancer aggressiveness and might become therapeutic targets in the near future.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

DATA AND SOFTWARE AVAILABILITY

Original imaging data have been deposited to Mendeley Data and are available at <https://doi.org/10.17632/22mbdtz59s.1> (<https://doi.org/10.17632/22mbdtz59s.1>).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, three figures, and four tables and can be found with this article online at <https://doi.org/10.1016/j.isci.2018.06.007>.

ACKNOWLEDGMENTS

We are very grateful to Dr. S. Kabir for commenting on the manuscript, to Dr. J-P de Villartay/P. Revy lab and to Dr. A. De Cian for scientific feedback, to Dr. F. Megetto for providing Ba/F3 cells, and to Dr. C. Lescale and J. Bianchi for providing pro-B cell lines. This research work is supported by a joint grant (E.B., T.M., and L.D.) from the Institut National du Cancer and the Cancéropôle IdF (2016-1-PLBIO-07-INSERM11-1). E.B. research was supported by ANR-12-JSV6-0005. E.B.'s group and T.M.'s team belong to two "La ligue contre le Cancer" teams. M.P. was supported by fellowships from Le Cancéropôle IdF and La Ligue Contre le Cancer, and C.T., by a fellowship from la Fondation de France. L.D.'s research is supported by the Institut Pasteur as well as by the European Research Council under the ERC starting grant agreement # 310917, and T.M.'s research, by SIRIC SOCRATE and an Emergence grant from the Canceropole IdF.

AUTHOR CONTRIBUTIONS

E.C.B conceived the initial study. E.C.B, L.B., and M.P. designed research experiments with the help of L.D., T.M., and C.G. M.P. and L.B. performed most of the experiments, B.R. designed and tested the CRISPR/Cas9 system for NPM1-ALK, and K.L. provided technical support for FISH and protein analysis and for f-circRNA detection. C.T. established the lentiviral transduction and FACS sorting in RPE-1 cells. E.B., L.B., and M.P. wrote the manuscript with valuable feedback from all the authors.

DECLARATION OF INTERESTS

None declared.

Received: December 8, 2017

Revised: May 5, 2018

Accepted: June 14, 2018

Published: July 27, 2018

REFERENCES

- Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evtal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* **56**, 55–66.
- Bai, R.Y., Dieter, P., Peschel, C., Morris, S.W., and Duyster, J. (1998). Nucleophosmin-anaplastic lymphoma kinase of large-cell anaplastic lymphoma is a constitutively active tyrosine kinase that utilizes phospholipase C-gamma to mediate its mitogenicity. *Mol. Cell. Biol.* **18**, 6951–6961.
- Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V.M., Gregory, P., Urnov, F., Weinstock, D.M., and Jasin, M. (2009). Chromosomal translocations induced at specified loci in human stem cells. *Proc. Natl. Acad. Sci. USA* **106**, 10620–10625.
- Chen, L., Zhang, S., Wu, J., Cui, J., Zhong, L., Zeng, L., and Ge, S. (2017). circRNA_100290 plays a role in oral cancer by functioning as a sponge of the miR-29 family. *Oncogene* **36**, 4551–4561.
- Choi, P.S., and Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* **5**, 3728.
- Cocquerelle, C., Mascrez, B., Hetuin, D., and Bailleul, B. (1993). Mis-splicing yields circular RNA molecules. *FASEB J.* **7**, 155–160.
- Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salamanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125–1134.
- Drynan, L.F., Pannell, R., Forster, A., Chan, N.M., Cano, F., Daser, A., and Rabbitts, T.H. (2005). Mll fusions generated by Cre-loxP-mediated de novo translocations can induce lineage reassignment in tumorigenesis. *EMBO J.* **24**, 3136–3146.
- Errichelli, L., Dini Modigliani, S., Laneve, P., Colantoni, A., Legnini, I., Capauto, D., Rosa, A., De Santis, R., Scarfo, R., Peruzzi, G., et al. (2017). FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat. Commun.* **8**, 14741.

- Fujimoto, J., Shiota, M., Iwahara, T., Seki, N., Satoh, H., Mori, S., and Yamamoto, T. (1996). Characterization of the transforming activity of p80, a hyperphosphorylated protein in a Ki-1 lymphoma cell line with chromosomal translocation t(2;5). *Proc. Natl. Acad. Sci. USA* 93, 4181–4186.
- Gao, Y., Wang, J., Zheng, Y., Zhang, J., Chen, S., and Zhao, F. (2016). Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat. Commun.* 7, 12060.
- Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C., et al. (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol. Cell* 55, 829–842.
- Guarnerio, J., Bezzi, M., Jeong, J.C., Paffenholz, S.V., Berry, K., Naldini, M.M., Lo-Coco, F., Tay, Y., Beck, A.H., and Pandolfi, P.P. (2016). Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. *Cell* 165, 289–302.
- Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388.
- Heckl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R.V., McConkey, M.E., Thielke, A., Aster, J.C., Regev, A., and Ebert, B.L. (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* 32, 941–946.
- Hsiao, K.Y., Lin, Y.C., Gupta, S.K., Chang, N., Yen, L., Sun, H.S., and Tsai, S.J. (2017). Noncoding effects of circular RNA CCDC66 promote colon cancer growth and metastasis. *Cancer Res.* 77, 2339–2350.
- Hsu, M.T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* 280, 339–340.
- Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C., et al. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* 10, 170–177.
- Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19, 141–157.
- Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., et al. (2017). Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell* 66, 22–37 e29.
- Li, F., Zhang, L., Li, W., Deng, J., Zheng, J., An, M., Lu, J., and Zhou, Y. (2015). Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/beta-catenin pathway. *Oncotarget* 6, 6001–6013.
- Lu, L., Sun, J., Shi, P., Kong, W., Xu, K., He, B., Zhang, S., and Wang, J. (2017). Identification of circular RNAs as a promising new class of diagnostic biomarkers for human breast cancer. *Oncotarget* 8, 44096–44107.
- Marzec, M., Kasprzycka, M., Liu, X., El-Salem, M., Halasa, K., Raghunath, P.N., Bucki, R., Wlodarski, P., and Wasik, M.A. (2007a). Oncogenic tyrosine kinase NPM/ALK induces activation of the rapamycin-sensitive mTOR signaling pathway. *Oncogene* 26, 5606–5614.
- Marzec, M., Kasprzycka, M., Liu, X., Raghunath, P.N., Wlodarski, P., and Wasik, M.A. (2007b). Oncogenic tyrosine kinase NPM/ALK induces activation of the MEK/ERK signaling pathway independently of c-Raf. *Oncogene* 26, 813–821.
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338.
- Morris, S.W., Kirstein, M.N., Valentine, M.B., Dittmer, K.G., Shapiro, D.N., Saltman, D.L., and Look, A.T. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 263, 1281–1284.
- Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., et al. (2017). Translation of CircRNAs. *Mol. Cell* 66, 9–21 e27.
- Piganeau, M., Ghezraoui, H., De Cian, A., Guittat, L., Tomishima, M., Perrouault, L., Rene, O., Katibah, G.E., Zhang, L., Holmes, M.C., et al. (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Res.* 23, 1182–1193.
- Renouf, B., Piganeau, M., Ghezraoui, H., Jasin, M., and Brunet, E. (2014). Creating cancer translocations in human cells using Cas9 DSBs and nCas9 paired nicks. *Methods Enzymol.* 546, 251–271.
- Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., et al. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58, 870–885.
- Schindler, J.W., Van Buren, D., Foudi, A., Krejci, O., Qin, J., Orkin, S.H., and Hock, H. (2009). TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. *Cell Stem Cell* 5, 43–53.
- Shao, Y., and Chen, Y. (2016). Roles of circular RNAs in neurologic disease. *Front. Mol. Neurosci.* 9, 25.
- Torres, R., Martin, M.C., Garcia, A., Cigudosa, J.C., Ramirez, J.C., and Rodriguez-Perales, S. (2014). Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat. Commun.* 5, 3964.
- van de Krogt, J.A., Vanden Bempt, M., Finalet Ferreira, J., Mentens, N., Jacobs, K., Pluys, U., Doms, K., Geerdens, E., Uytendaele, A., Pierre, P., et al. (2017). ALK-positive anaplastic large cell lymphoma with the variant RNF213-, ATIC- and TPM3-ALK fusions is characterized by copy number gain of the rearranged ALK gene. *Haematologica* 102, 1605–1616.
- Zamo, A., Chiarle, R., Piva, R., Howes, J., Fan, Y., Chilosi, M., Levy, D.E., and Inghirami, G. (2002). Anaplastic lymphoma kinase (ALK) activates Stat3 and protects hematopoietic cells from cell death. *Oncogene* 21, 1038–1047.
- Zhang, Q., Raghunath, P.N., Xue, L., Majewski, M., Carpentieri, D.F., Odum, N., Morris, S., Skorski, T., and Wasik, M.A. (2002). Multilevel dysregulation of STAT3 activation in anaplastic lymphoma kinase-positive T/null-cell lymphoma. *J. Immunol.* 168, 466–474.
- Zhang, X.O., Dong, R., Zhang, Y., Zhang, J.L., Luo, Z., Zhang, J., Chen, L.L., and Yang, L. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 26, 1277–1287.
- Zhang, Y., Zhang, X.O., Chen, T., Xiang, J.F., Yin, Q.F., Xing, Y.H., Zhu, S., Yang, L., and Chen, L.L. (2013). Circular intronic long noncoding RNAs. *Mol. Cell* 51, 792–806.

ISCI, Volume 5

Supplemental Information

Chromosomal Translocation Formation Is Sufficient to Produce Fusion Circular RNAs Specific to Patient Tumor Cells

Loelia Babin, Marion Piganeau, Benjamin Renouf, Khadija Lamribet, Cecile Thirant, Ludovic Deriano, Thomas Mercher, Carine Giovannangeli, and Erika C. Brunet

SUPPLEMENTAL INFORMATION

SUPPLEMENTARY FIGURES

Supplementary Figure 1 (Related to Figure 1)

Characterization of *Npm1-Alk* translocation induced in mouse cells and *in vitro* study of circRNAs amplification.

(A) Estimate of translocation frequency (f) of Der17 formation from Ba/F3 cells, 5 days after transfection, using PCR on serial dilutions of genomic DNA. The number of times the PCR was positive (from 4 experiments) for each dilution is also indicated (Amplicons).

(B) Crizotinib treatment of IL-3 independent cells. 100 nM crizotinib treatment prevents growth of IL-3-independent cells with *Npm1-Alk* translocation selected in Figure 1-B (mean of 2 experiments +/-SD).

(C) Sequences of *Npm1-Alk* breakpoint junctions from Ba/F3 isolated clones. Clones 1, 3, 5, 6 and 7 harbor deletions (Δ), and clones 2 and 4 have insertions (+). All indels are restricted to intronic regions.

(D) FISH analysis of Ba/F3 *Npm1-Alk* clones. A red probe labeled the 5' end of the *Npm1* gene and a green probe labeled the 3' end of *Alk* gene. Colocalization of red and green revealed Der17 formation. It is worth noting that Ba/F3 cells are polyploid cells with in an average of 72 (+/- 2.5) chromosomes. They have 5 signals for *Npm1* locus and 4 signals for *Alk* locus.

(E) Schemes for *in vitro* transcription of linear and circular *NPMI-ALK* using the T7-promoter containing primers. After RnaseR treatment and retro-transcription, divergent PCR primers are used to detect circRNAs. Artefact amplicons from the linear *NPMI-ALK* cDNA are visibly detected. In contrast, f-circRNA-hF is properly amplified. As control for RnaseR digestion, we used convergent PCR primers to amplify *hNPMI-ALK* cDNA and *GAPDH*. Scheme of one PCR artefact (using divergent primers) obtained from the linear *NPMI-ALK* cDNA after sequencing (*direct repeats, in this example the repeat is "CCAACG") (see sequences on Table S1).

Supplementary Figure 2 (Related to Figure 2)

Characterization of *NPM1-ALK* translocation induced in human cells

(A) sgRNAs targeting *NPM1* and *ALK* genes were cloned in a lentiviral vector expressing a Cas9-2A-GFP (vector sgRNAs-Cas9-2A-GFP). Sequences of Der5 and Der2 breakpoint junctions from Lenti-CL1, CLA, CLB and CLC RPE-1 clones.

(B) Estimate of translocation frequency (f) using PCR on serial dilutions of genomic DNA (in duplicates) from RPE-1 cells at day 3 (D3) and day 16 (D16) after transfection for independent experiments (starting from a pool of cells with low and high frequency). Cells were either kept in 10% or 2% of FBS.

Supplementary Figure 3 (Related to Figure 3)

Characterization of *MLL-AF9* translocation induced in mouse cells

(A) Schematic representation of the CRISPR/Cas9-induced *MLL-Af9* translocation in mouse cells. PCR amplification shows Der4 and Der9 formation and *MLL-Af9* cDNA is detected in all 3 clones (we were unable to find a specific antibody for mouse MLL or AF9).

(B) FISH analysis of *MLL-Af9* positive clones in mouse B-cell lines. A green probe labelled the 5' end of *MLL* and a red probe the 3' end of *Af9*. Red and green colocalization signal shows the formation of Der9.

(C) PCR-amplified products obtained using divergent primers to detect f-circRNAs corresponding to *MLL-Af9* gene in mouse pro-B cell line (wild type cells and 3 clones). A representative gel from 3 independent experiments. Number of circular junctions sequenced for each type of *MLL-Af9* f-circRNAs in mouse pro-B cells (back-spliced, Non canonical back-spliced, and mixed). Sequencing was done from 3 independent experiments. Schemes of f-circRNA-mG and -mH (lower panel) found in the three clones.

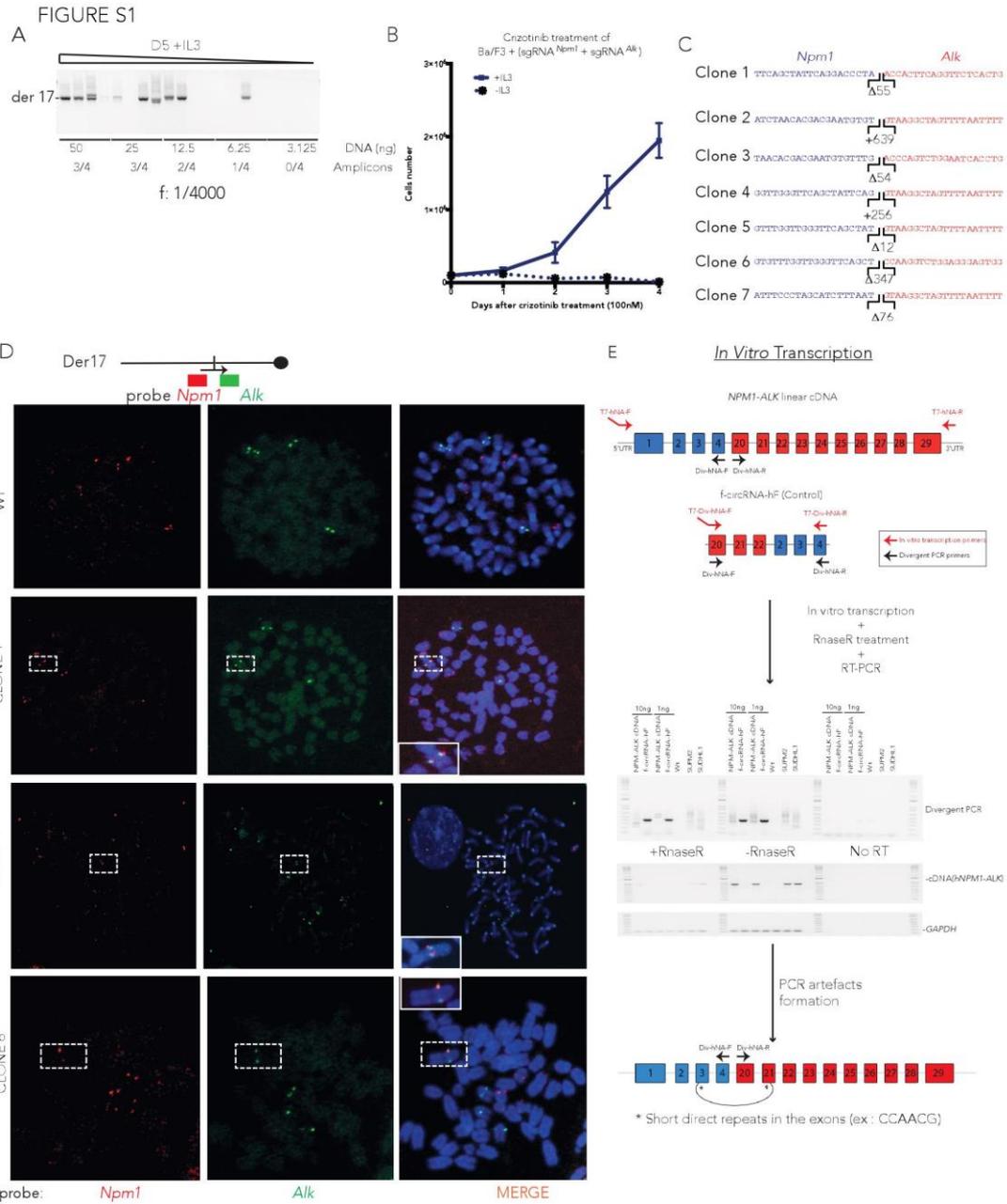


FIGURE S2

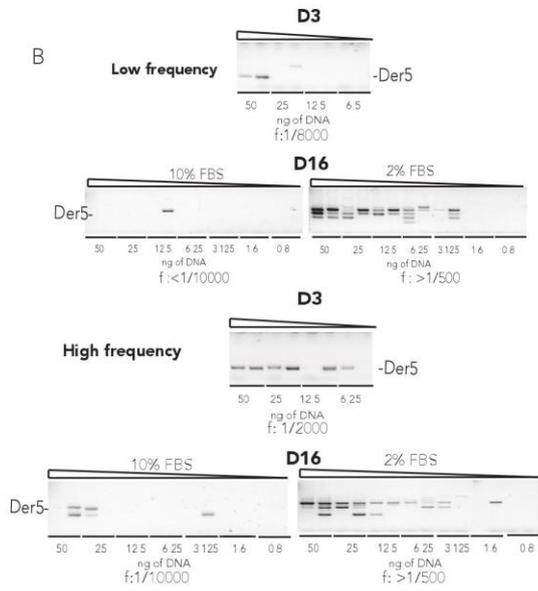
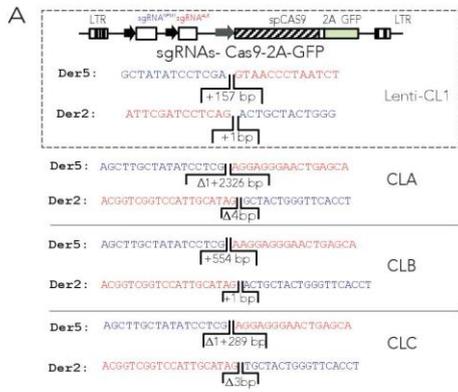
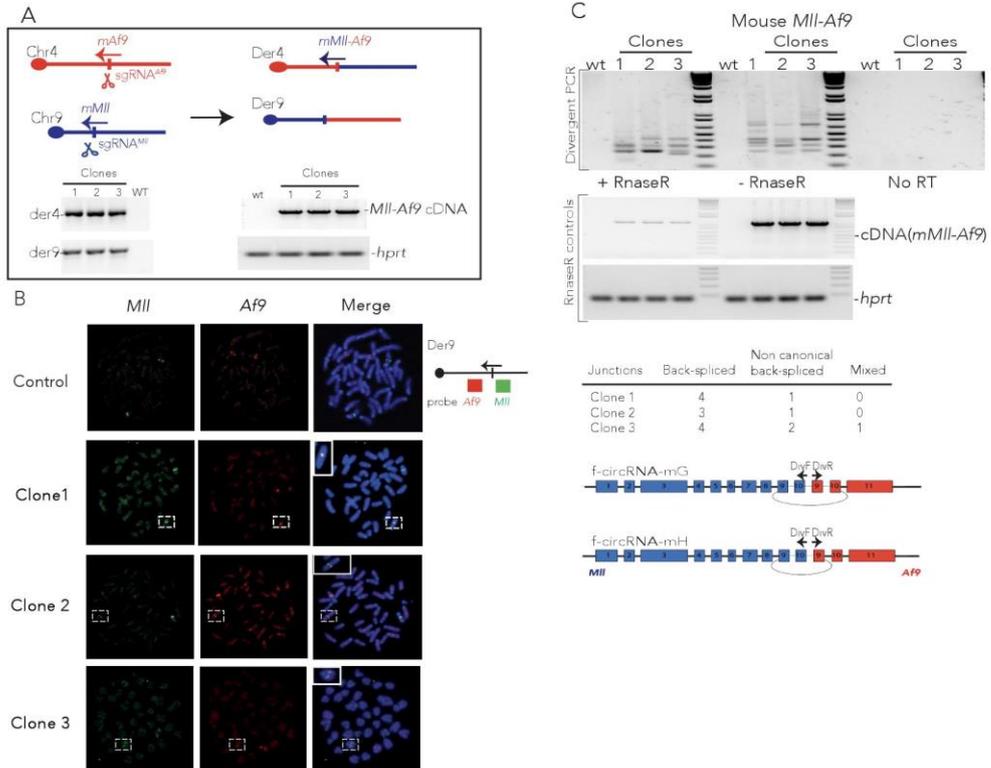


FIGURE S3



SUPPLEMENTARY TABLES

Table S1 (Related to Figure 1)

f-circRNAs expressed in mouse *Npm1-Alk* Ba/F3 clones

Schematic representation of f-circRNA junctions obtained from the sequencing of 3 independent experiments. Circularized junctions are linked by a line. Circularization junction sequences of Non-canonical back-spliced and Mixed f-circRNAs are described after each table.



* indicates short repeats.

Sequences mNpm1-Alk circular RNAs :

Clone 2 :

Non canonical back-spliced :

ALK Ex 24 AACCCACCTCCCTGGCCATG TGAAATTACACCACCTGTGG Ex 4 NPM

Clone 5 :

Non canonical back-spliced :

ALK Ex 23 TGAGTCTACAAGCCCTGCCCCGCTT TCTGTACAACCAACA Ex 3 NPM

Clone 6 :

Non canonical back-spliced :

ALK Ex 23 TCATGGCTGGCGGAGACCTC GCGTGATTCCGTCCTGCGCG Ex 1 NPM

Clone 7 :

Non canonical back-spliced :

ALK Ex 22 ACGTGCCAGA CTTTGAAATTACACCACCTGTGGTCTTAC Ex 4 NPM

Table S2 (Related to Figure 3)

f-circRNAs expressed in human *NPM1-ALK* RPE-1 clones, *SUDHLI* and *SUPM2* cells.

Schematic representation of f-circRNA junctions obtained from the sequencing of 3 independent experiments. Circularized junctions are linked by a line. Circularization junction sequences of Non-canonical back-spliced and Mixed f-circRNAs are described after each table.



circular RNAs with the junction in exon 1 corresponding to an identified alternative splice site (ENST00000517671.5)

Sequences *hNPM1-ALK* circular RNA :

Lenti-CL1, SUDHLI :

Mixed :

ALK Ex 21 CCCTGCAAGTGGCTGTGAAG CGTTC'TTTTATCTCCGTCCG Ex 1 NPM

Lenti-CL1 :

Non canonical back-spliced :

ALK Ex 20 CTGGCAAGACCTCCTCCATC TGCACATTGTTGAAGCAGAG Ex 3 NPM

CLB :

Non canonical back-spliced :

ALK Ex 20 GAGGTGCCGCGGAAAAACAT AGCAGAGGCAATGAATTACG Ex 3 NPM

SUDHLI :

Mixed :

ALK Ex 22 ATGGAAGCCCTGATCATCAG CCCCTGAGGCCCCAGAACT Ex 1 NPM

ALK Ex 20 AAAAACATCACCCTCATTCG CGTTCTTTTATCTCCGTCCG Ex 1 NPM

ALK Ex 20 GGCAAGACCTCCTCCATCAG GTTCCCTTGGGGGCTTTGA Ex 4 NPM

SUPM2 :

Non canonical back-spliced :

ALK Ex 22 GACGAACTGGATTTCCATCAT TTGAAGCAGAGGCAATGAAT Ex 3 NPM

Mixed :

ALK Ex 20 GACTACAACCCCACTACTG GTTGTGAACTAAAGGCCGAC Ex 2 NPM

CDNA2 :

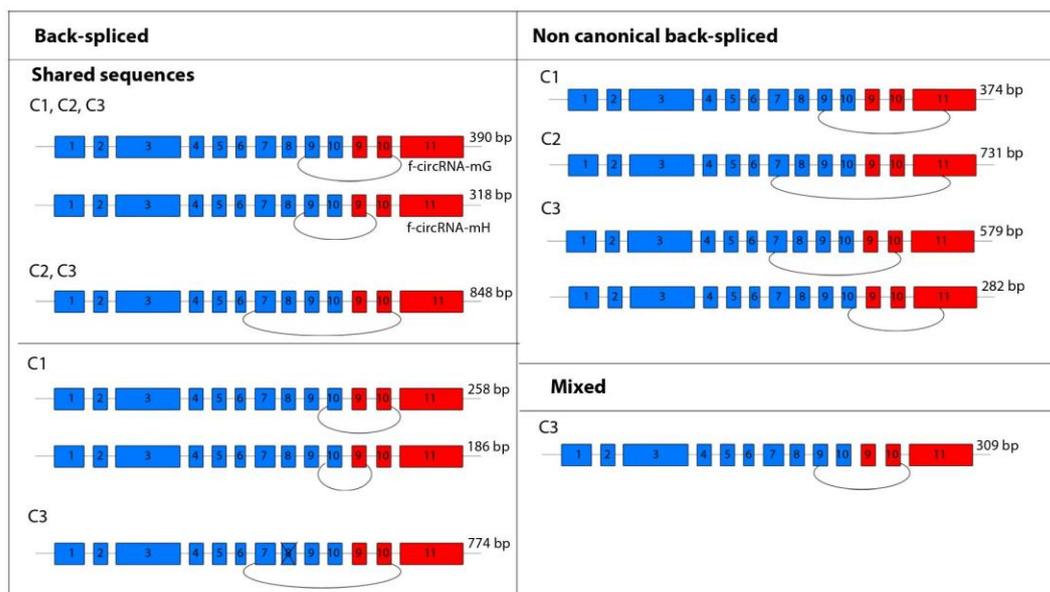
Non canonical back-spliced :

ALK Ex 24 CACGTGGCTCGGGACATTGC ACGAAGGCAGTCCAATTAAA Ex 3 NPM

Table S3 (Related to Figure S3)

f-circRNAs expressed in *Mll-Af9* mouse pro-B cells clones

Schematic representation of f-circRNA junctions obtained from the sequencing of 3 independent experiments. Circularized junctions are linked by a line. Circularization junction sequences of Non-canonical back-spliced and Mixed f-circRNAs are described after each table.



Sequences *mMll-Af9* circular RNA

Clone 1 :

Non canonical back-spliced :

AF9 Ex 11 AACCTTATAGAGGAAACTGG TTGAACATCCTCAACCCACT Ex 9 MLL

Clone 2 :

Non canonical back-spliced :

AF9 Ex 11 ACAACAGCTGGACGCATCAA CCAGCAGCCGTCGTCCCCC Ex 7 MLL

Clone 3 :

Non canonical back-spliced :

AF9 Ex 10 CTGCAGCAGATCGTGAACCT TCCCAGCCAGCAGCCGTCGT Ex 7 MLL

AF9 Ex 11 TTTGCTCGCTGGACAAAACCT CCCATAACACCAGAGTAGT Ex 10 MLL

Mixed :

AF9 Ex 10 GGAACGACACATTCTGCAGC AGTAAACAAGAGAATGCAGG Ex 9 MLL

Table S4 (Related to Transparent Methods)

sgRNA and primer sequences

sgRNA sequences:

Translocation induction

Human NPM1: **TATCCTCGAACTGCTAC** **G TTCACCTCAGC**
sgRNA^{NPM} 5' -GTGAACCCAGTAGCAGTTCG-3'
Human ALK: **GATCCTCAGGTAACCCTAATCTGATCACGGTC**
sgRNA^{ALK} 5' -GATCAGATTAGGGTTACCTG-3'
Mouse NPM1: **GGACCTATGACCAATGAGGCTATCTTTCAAT**
sgRNA^{NPM} 5' -AGATAGCCTCATTGGTCATA-3'
Mouse ALK: **ATAAAATCCTGGCATGTCTATCTGTAAAGGCTA**
sgRNA^{ALK} 5' -TCCTGGCATGTCTATCTGTA-3'
Mouse MLL : **CACCCACATCCGGTGGTTCACTGTA ACTCCA**
sgRNA^{MLL} 5' -TACAGTGAACCACCGGATGT-3'
Mouse AF9 : **TCTCCATACTCGTTGACTGGTCTATCAGTTTA**
sgRNA^{AF9} 5' -GATAGACCAGTCAACGAGTA -3'

Primer sequences:

Plasmid cloning

Cloning of sgRNA ^{NPM} :sgRNA ^{ALK} :Cas9-GFP	sgRNA_EE_F	AGGAACCAATgaattcGACTGGATCCGGTACCAAGG
	sgRNA_EE_R	TAGAGAGGTAgattcAGCGGCCCAAGCTTAAAAA
Cloning of NPM1-ALK cDNA	cDNA_NA_F	CTAAGTGCgctagcggcaccATGGAAGATTCGATGGACATGG
	cDNA_NA_R	ATAAAAACCTgaattcACCACATCTGGGCCTTGAT

PCR for translocation detection

hNPM1-ALK (PCR1)	hNA-Der5-F	CAGTTGCTTGGTCCAGTT
	hNA-Der5-R	AGGAATTGGCCTGCCTTAGT
hNPM1-ALK (PCR2)	hNA-Der5-NF	GGGAGAGGAAAATCTTGCTG
	hNA-Der5-NR	GCAGCTTCAGTGCAATCACA
hALK-NPM1 (PCR1)	hAN-Der2-F	TCCTTCAGTGTCATCACGA
	hAN-Der2-R	GAACCTTGCTACCACCTCCA
hALK-NPM1 (PCR2)	hAN-Der2-NF	CCCACCCCTAGACGTCCT
	hAN-Der2-NR	TTCACATCCTCCTCCTTCA
mNPM1-ALK (PCR1)	mNA-Der17-F	CATCGTAGAGGCAGAAGCAA

mNPM1-ALK (PCR2)	mNA-Der17-R	TGCAGTTCATCTGCATAGC
	mNA-Der17-NF	AGGTCTCTTGCGTCATTTGG
	mNA-Der17-NR	CAGATAGGTGTCCGCTGTGA
mALK-NPM1 (PCR1)	mAN-Der11-F	CTTCAGTGCCCAACACAATG
	mAN-Der11-R	CTTCTCTGGCACACACCTTT
mALK-NPM1 (PCR2)	mAN-Der11-NF	TGGCTCAAGCCATACTAGCC
	mAN-Der11-NR	AGAGGCAGGTGGATTTCTGA
mMLL-AF9 (PCR1)	mMA-Der9-F	TTTTGGGTGAGTTGTTTGGGA
	mMA-Der9-R	CAGCGGACAGTTCTGTGGTA
mMLL-AF9 (PCR2)	mMA-Der9-NF	ACACACCAGAAGAGGGCATC
	mMA-Der9-NR	ACCTGCAAAGGTTCTTGTGG
mAF9-MLL (PCR1)	mAM-Der4-F	CAGTGCACCCTTGAGCATAA
	mAM-Der4-R	TAGTCACCAGGCATTTGCAG
mAF9-MLL (PCR2)	mAM-Der4-NF	ACAACCTGGCATGTGCAACTC
	mAM-Der4-NR	TGCTAGGGATTGAAGCCAAG

RT-PCR

hNPM1-ALK cDNA	hRNA-NA-F	TTGTGAACTAAAGGCCGACAAA
	hRNA-NA-R	CAGTTTCTGGCAGCAATGTCTC
mNPM1-ALK cDNA	mRNA-NA-F	GTGGAACAGGAGGCAGTTGT
	mRNA-NA-R	GCAATATCCCGGTGGATAAA
mMLL-AF9 cDNA	mRNA-MA-F	CCTGAACCCAAACAGGTCAG
	mRNA-MA-R	TGGAATGAACCACCACAGAA
hGAPDH cDNA	hGAPDH-F	GCCATCAATGACCCCTTCAT
	hGAPDH-R	TGACAAGCTTCCCGTTCTCA
mHPRT cDNA	mHPRT-F	GCAGTACAGCCCCAAAATGG
	mHPRT-R	GGTCCTTTTACCAGCAAGCT

PCR for fusion circular RNA detection

hNPM1-ALK	hNA-div-F	AAGCACCAGGAGCTGCAAG
	hNA-div-R	AATATGCACTGGCCCTGAAC
mNPM1-ALK	mNA-div-F	ACCAGGAGTTGCAGGCTATG
	mNA-div-R	AGATGCTGTCCACTAATGTGC
mMLL-AF9	mMA-div-F	TCCAATAAAACAAAGCAAATCAGA
	mMA-div-R	CACTGCTGGCACAGAGAAAAG

PCR for specific fusion circular RNA detection

f-circRNA-mA	f-circ-mA-F	GCAGGCTATGCAGATGGAAC
	f-circ-mA-R	AGCCTTTAGTTCACAGCCTGA
f-circRNA-mB	f-circ-mB-F	GTGGAAGCACCAGGAGTT
	f-circ-mB-R	GCCTTTAGTTCACAGCCGG
f-circRNA-hC	f-circ-hC-F	GAGCTGCAGAGCCCTGAGTA
	f-circ-hC-R	GCCTTTAGTTCACAACCGGG
f-circRNA-hD	f-circ-hD-F	AGCACCAGGAGCTGCAAG
	f-circ-hD-R	GGCCTTTAGTTCACAACCTGA
f-circRNA-hE	f-circ-hE-F	GAGCTGCAGAGCCCTGAGTA

f-circRNA-hF	f-circ-hE-R	CCCCCAAGGGAAACCTGA
	f-circ-hF-F	GAGCTGCAGAGCCCTGAGTA
	f-circ-hF-R	CCCCCAAGGGAAACCGGT

PCR for digital droplet PCR

hGAPDH	GAPDH-F	GCCATCAATGACCCCTTCAT
	GAPDH-R	TGACAAGCTTCCCGTCTCA
hNPM-ALK	hNA-F	ACCACCAGTGGTCTTAAGGTTG
	hNA-R	CGGAGCTTGCTCAGCTTGTA

PCR for in vitro transcription

T7-hNPM-ALK	T7-hNA-F	gaaattaatacgactcactata GGGGT TCTCTGGAGCAGCGTTCT
	T7-hNA-R	CACATCAACAAGGCAAGGAA
T7-Div-hNA	T7-Div-hNA-F	gaaattaatacgactcactata GGGAAGC ACCAGGAGCTGCAAG
	T7-Div-hNA-R	AATATGCACTGGCCCTGAAC

TRANSPARENT METHODS

CRISPR/Cas9 plasmids

The sgRNAs vectors were cloned in MLM3636 plasmid (Addgene plasmid 43860). sgRNAs sequences are listed above. SgRNAs targeting human *NPM1* and *ALK* genes have been previously described (Ghezraoui et al., 2014; Renouf et al., 2014). All sgRNAs were designed using the CRISPOR designer tool (<http://crispor.tefor.net/crispor.py>) and are described in Table S4. Cas9-2A-GFP was expressed from pCas9-GFP vector (Addgene plasmid 44719). For construction of lentiviral plasmid expressing both sgRNAs, human sgRNA^{*NPM1*} was firstly cloned in the sgRNA:Cas9-GFP vector from Ebert's laboratory (Heckl et al., 2014). The U6-sgRNA^{*ALK*} cassette was PCR amplified from MLM3636 vector containing sgRNA^{*ALK*} and cloned in the EcoRI site of the sgRNA^{*NPM1*}:Cas9-GFP vector. Lentiviral construct was co-transfected with pCMV 8.74 packaging plasmid, and pMD2.G VSV envelope expression plasmid. Supernatants were collected at 48h and concentrated by ultra-centrifugation.

Cell culture and transfection

The Ba/F3 murine pro-B cell line (gift from Dr F.Meggetto) was cultured in RPMI-1640 medium supplemented with 10% FBS and 5 ng/mL murine recombinant IL-3 (Thermo Fisher Scientific). The SUDHL1 ALCL cell line was cultured in RPMI-1640 medium supplemented with 10% FBS. The SUPM2 ALCL cell line was cultured in RPMI-1640 medium supplemented with 20% FBS. The RPE-1 hTERT-immortalized retinal cell line was cultured in DMEM/F-12 medium supplemented with 10% or 2% FBS. The mouse v-Abl proB-cell line was cultured in RPMI-1640 medium supplemented with 10% FBS, beta-mercaptoethanol (50nM), penicillin and streptomycin (Lescale et al., 2016). (Mediums and FBS from Gibco) Typically, 1×10^6 cells were transfected using Nucleofection (Amaxa technology II (Lonza)) with 3 μ g of pCas9-GFP vector and 3 μ g of each sgRNA, using the Cell Line Nucleofector Kit V (programs D-032 for Ba/F3 and X-001 for RPE-1 and mouse v-Abl proB-cell line). For lentiviral transduction, 1×10^5 RPE-1 cells were seeded and transduced with concentrated sgRNA^{*NPM1*}:sgRNA^{*ALK*}:Cas9-GFP supernatant. Cells were FACS sorted based on GFP expression 72h post-transduction.

IL-3 removal and crizotinib treatment

5 days after transfection of pCas9-2A-GFP and sgRNA^{*NPM1*}, sgRNA^{*ALK*}, or both sgRNAs in Ba/F3 cell line, 10^5 cells were cultured without IL-3. Surviving cells were counted every 3

days. IL-3 starved and wild-type Ba/F3 cells were then seeded at 10^5 cells per well in a 6-well plate. Cells were cultured with addition of 100 nM crizotinib (SelleckChem) in medium and counted every days.

PCR-based detection of translocation junctions and translocation frequency

Genomic DNA from cells was extracted using E.Z.N.A Tissue DNA Kit (Omega Biotek). For detection of translocation junctions from pooled cells expressing CRISPR/Cas9, we performed a nested PCR on 50 ng of genomic DNA as previously described (Ghezraoui et al., 2014; Piganeau et al., 2016; Renouf et al., 2014). Primers sequences are listed in Table S4.

Translocation frequency was estimated by serial dilution as previously described (Ghezraoui et al., 2014; Renouf et al., 2014). Briefly one-round PCR was performed on four serial dilutions of DNA from transfected cells, using primers from PCR2.

RT-PCR and cDNA amplification of fusion transcript

5 days after transfection, RNA was extracted using Nucleospin RNA Kit (Macherey Nagel). RT-PCR was performed using ProtoScript® II First Strand cDNA Synthesis Kit (NEB) according to manufacturer's instructions. Primers sequences are listed in Table S4.

***NPM1-ALK* cDNA overexpression**

NPM1-ALK cDNA was amplified by PCR from SUDHLI cells using primers containing NheI and EcoRI sites. PCR product was cloned between NheI and EcoRI sites in pLJM1-EGFP plasmid (Addgene Plasmid 19319). pLJM1-NPM1-ALK vector was then transfected in RPE-1 cells. 24h after transfection 12 μ g/ml puromycin was added in the medium. *NPM1-ALK* cDNA expression was assessed by RT-PCR after 2 weeks of selection. Two positive clones were then isolated.

Cloning of translocated cells

Cloning of cytokine-independent Ba/F3 cells was performed by seeding 0.3 cell per well in a 96-well plate. Cloning of translocated RPE-1 cells was performed by sib selection. Pooled cells were seeded in a 96-well plate, at 10^3 cells per well. At confluence, plates were duplicated for translocation detection by PCR screen. One plate was maintained in culture. Cells from the other plate were incubated with lysis buffer (10 mM Tris-HCl pH 8, 0.45 % NP40, 0.45 % Tween20, 100 μ g/mL Proteinase K) for 1 hour at 55°C. Proteinase K was inactivated by heating at 95°C for 10 minutes. First-round PCR was performed on lysate with

primers used for translocation junction detection (listed below). Second-round PCR was performed with SYBR green in real-time PCR conditions (Mx3005p, Agilent) (Brunet et al., 2009; Ghezraoui et al., 2014; Renouf et al., 2014). Ba/F3 cells from the remaining plate corresponding to positive wells were amplified in cell culture. RPE-1 cells from remaining plate corresponding to positive wells were re-seeded for additional steps of selection to obtain translocated clones. Of note, cloning of RPE-1 clones A, B and C was performed by seeding 0.3 cell per well (in a 96-well plate) directly after 2 weeks of growth in 2% FBS medium.

Western-blot

Whole cell extracts were prepared with protein lysis buffer (50 mM Tris-HCl at pH 7.4, 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT prepared from a 1 M DTT stock), with addition of Complete cocktail protease inhibitor tablets (Roche) and Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific). Typically, 30 µg of protein extracts were run on an 8% (w/v) Tris-HCl SDS PAGE gel. For SUDHL1 control cell line, 10 µg of protein extract were used. Gel was blotted, and probed with Phospho-ALK (Tyr1604), ALK (31F12), Phospho-STAT3 (Tyr705), Stat3, Phospho-p44/42 MAPK (ERK1/2) (Thr202/Tyr204), p44/42 MAPK (ERK1/2), Phospho-AKT (Thr308), Phospho-AKT (Ser473), and AKT antibodies from Cell Signaling. Murine NPM1-ALK protein expression was assessed with ALK (M19) antibody (Santa Cruz), and Cas9 with Cas9 (7A9-3A3) antibody (Novus). Alpha-Tubulin mouse antibody (Sigma-Aldrich) was used as a loading control. Phospho-ALK and Phospho-STAT3 antibodies were used at 1/200 dilution. Other antibodies were used according to manufacturer's recommendations. Secondary antibodies were HRP-conjugated, and blot developed with ECL prime (GE Healthcare Life).

Fluorescence *in situ* hybridization

FISH was performed on metaphases from translocated clones. Human *ALK* and *NPM1* human genes were labeled using respectively ALK Breakapart LPS 019 probe (Cytocell) and homemade *NPM1* probe from BAC RP11-546B8. Murine *Alk* and *Npm1* genes were labeled using homemade probes from BAC RP23-100F5 and RP23-74P19. *Mll* and *Af9* genes were labeled using homemade probes from BAC RP23-23P9 and RP23-440D4. Metaphase spreading, probe labeling, hybridization, washing and fluorescence detection were performed according to standard procedures. Images were taken with the Zeiss Axio Observer.Z1 system.

Selective advantage in high/low serum medium

Cells were plated at day 0 with 10% Lenti-CL1 GFP-positives and 90% WT RPE-1 GFP-negatives. They were cultured in high serum condition (10% FBS) or low serum condition (2% FBS). After 6 and 14 days cells were analyzed on cytometer to detect GFP enrichment.

Translocation detection in high/low serum medium

72 hours after transfection with pCas9-GFP and sgRNA^{ALK} + sgRNA^{NPM1}, cells were cultured in 10% FBS or 2% FBS. DNA extraction was performed at day 3 and day 16 after transfection to estimate the frequency of translocation.

RT-PCR and cDNA amplification for detection of circular RNA

RNA was extracted using Nucleospin RNA Kit (Macherey Nagel). 1 µg RNA was treated by 2U RNaseR enzyme (Epicentre) for 20 minutes at 37°C. RT-PCR was performed using ProtoScript® II First Strand cDNA Synthesis Kit (NEB) according to manufacturer's instructions. cDNA was then PCR amplified. PCR products were cloned directly with TOPO TA cloning kit (ThermoFischer) according to manufacturer's instructions and sequenced by Sanger Sequencing using T7 primer.

***In vitro* transcription assay (Figure S1E)**

Linear *NPM1-ALK* cDNA was amplified from SUPM2 using primers located in 5' and 3' UTR of mRNA (see TABLE S4 for primers sequences). The 5' primer contained the T7 promoter sequence (GAAATTAATACGACTCACTATAGGG) upstream from the specific 5'UTR sequence. Similarly, f-circRNA-hF was amplified (from a TOPO TA clone) using an upstream primer containing the T7 promoter sequence and used as control for circular RNA amplification. *In vitro* transcription was performed using HiScribe T7 High Yield RNA Synthesis Kit (NEB) according to manufacturer's instructions. Purity and size of RNA were assessed by RNA ScreenTape (Agilent, data not shown). Then 10 ng or 1 ng of transcribed RNA was mixed with 1 µg of total RNA from wild type RPE-1 cell line. RNaseR treatment and RT-PCR were performed on the mix as described previously: PCR amplification was done with primers hNA-div-F/R, hRNA-NA-F/R, hGAPDH-F/R (as control).

cDNA amplification for detection of specific circular RNAs for *NPM1-ALK*

A nested PCR was performed on cDNA to specifically amplify back-spliced f-circRNA identified after TOPO TA sequencing. The first round PCR was performed with divergent

primers for 23 cycles. The second round was performed with a forward primer in exon 20 of *Alk* gene and a reverse primer overlapping the circularization junction specific for each f-circRNA (40 cycles, See Supplementary Material for primers sequences). For specific quantification of f-circRNA-hC and -hD, the second round was performed using digital droplet PCR (Biorad) with Evagreen (according to manufacturer's instructions). Primers sequences are listed in Table S4.

***NPM1-ALK* expression quantification using digital droplet PCR**

Quantification of *NPM1-ALK* expression was performed using digital droplet PCR (Biorad) with Evagreen (according to manufacturer's instructions). Normalization was done using *GAPDH* amplification (See below for primers sequences). Primers sequences are listed in Table S4.

RnaseR efficiency

NPM1-ALK fusion cDNA, *GAPDH* cDNA and f-circRNA-hC/-hD were quantified with or without RnaseR treatment using digital droplet PCR (Biorad, See Supplementary Material for primers sequences). RnaseR efficiency was assess as follow:

$$\text{Fold enrichment} = \frac{\text{circular RNA (+RnaseR)}}{\text{circular RNA (-RnaseR)}} \bigg/ \frac{\text{linear RNA(+RnaseR)}}{\text{linear RNA (-RnaseR)}}$$

II. Etude du timing de réplication (Article 2, en cours de préparation)

Au cours de la seconde partie de ma thèse je me suis intéressée aux modifications de timing de réplication induites par la translocation *NPM1-ALK*.

Au cours de ce travail, nous avons séparé le rôle de la protéine de fusion de celui de la translocation chromosomique. Pour cela nous avons généré des cellules porteuses de la translocation, dans lesquelles la protéine de fusion a été inactivée. Nous avons également généré des cellules double-transloquées afin de mieux comprendre le rôle de l'allèle transloqué.

Nous avons réalisé des expériences de Répli-Seq afin de déterminer le timing de réplication sur l'ensemble du génome pour chaque clone, ainsi que des expériences de RNAseq permettant d'analyser le profil transcriptionnel. Ces travaux nous ont permis de mettre en évidence des modifications de timing de réplication au niveau des deux jonctions de translocation, mais également à l'échelle du génome entier. Ces changements ont pu être corrélés avec les modifications d'expression génique observées en RNAseq. Nous avons également montré que l'induction de la translocation permettait de reproduire partiellement la signature transcriptionnelle observée chez les patients ALCL.

Des expériences sont actuellement en cours afin de terminer ce travail, notamment l'analyse en Répli-seq des cellules double-transloquées. De plus des expériences similaires sont réalisées dans des cellules T primaires humaines.

Ces résultats sont présentés comme une ébauche d'article :

t(2;5)(p23;q35) translocation formation leads to more dramatic changes in replication timing and gene expression than NPM1-ALK expression alone

Babin L.¹, Sole-Ferre A.¹, Darchen A.¹, Piganeau M.^{2,6}, Congras A.³, Meggetto F.³, Aid Z.⁴, Mercher T.⁴, Deriano L.⁵, Brunet E.¹.

¹Laboratory « Genome Dynamics in the Immune System », Equipe Labellisée Ligue Contre le Cancer, INSERM UMR1163; Université Paris Descartes Sorbonne Paris Cité, Institut Imagine, Paris 75015, France

²Museum National d'Histoire Naturelle, INSERM U1154, CNRS UMR7196, Sorbonne Université, Paris 75005, France

³Laboratory « Biologie des ARN dans les cancers hématologiques », Equipe Labellisée Ligue Contre le Cancer, Labex TOUCAN, CRCT, Toulouse 31037, France

⁴INSERM U1170, Equipe Labellisée Ligue Contre le Cancer, Gustave Roussy Institute, Université Paris Diderot, Université Paris-Sud, Villejuif 94800, France

⁵Genome Integrity, Immunity and Cancer Unit, Department of Immunology, Department of Genomes and Genetics, Institut Pasteur, Paris 75015, France

⁶Present Address: Haematopoietic Stem Cell Lab, The Francis Crick Institute, London NW1 1AT, UK

A. Introduction

During the cell cycle S phase, DNA is replicated following a precise pattern known as Replication Timing (RT) which defines early and late replicated regions throughout the genome. RT is specific to each cell type, developmental stage and is well conserved between species. Moreover, it is directly linked to chromatin organization in the nucleus (Ryba et al., 2010; Yaffe et al., 2010). Early replicated regions are generally located in the center of the nucleus whereas late replicated regions are more peripheral (Panning and Gilbert, 2005). During cell differentiation, modifications in nuclear localization are correlated to changes in RT (Hiratani et al., 2008; Williams et al., 2006). Furthermore a strong correlation was also observed between RT and A/B compartments/Topologically Associated Domains (TADs) (Pope et al., 2014; Ryba et al., 2010; Yaffe et al., 2010).

Although it is generally admitted that early replicated regions are transcribed gene richer compared to late replicated regions (Rhind and Gilbert, 2013), the precise link between RT and gene expression is not always clearly understood. Instead some studies showed that RT is relatively independent of transcription level (Hassan-Zadeh et al., 2012; Simon et al., 2001)

whereas others demonstrated that it is not possible to advance replication timing in absence of transcription (Blin et al., 2019; Therizols et al., 2014) and that RT program precedes transcription activation (Kaaij et al., 2018; Pourkarimi et al., 2016; Seller and O'Farrell, 2018; Siefert et al., 2017). Notably, Blin et al showed a strong correlation between RT and transcription level for the longest genes of the genome. It was further demonstrated that Transcription Starting Site are enriched in replication origins and Transcription Termination Site in replication termination (Chen et al., 2019). Eventually, Gilbert's team provided evidences of the existence of Early Replication Control Elements (ERCEs) controlling replication timing. These ERCEs features include enriched active epigenetic marks, association with compartment A but are not dependent of CTCF binding (Sima et al., 2019). Particularly ERCEs will directly drive earlier timing of surrounded genes.

In the context of cancer, modifications of replication timing were observed in oncogenes and tumor suppressor genes (Fritz et al., 2013), and more globally in leukemia cells 9-18% of the genome showed significant modifications in RT (Ryba et al., 2012). Likewise Dixon et al analyze cancer cells with chromosomal rearrangements and show local modification in TAD organization but also in RT (Dixon et al., 2018). Besides, it was previously described that replication is delayed for the derivative chromosomes (Breger et al., 2005). Some local changes in RT have been described as epigenetic because of the absence of genetic abnormalities. These modifications are stable through mitotic divisions (Adolph et al., 1992; Eul et al., 1988). These observations lead to the hypothesis that chromosomal rearrangements could disturb nuclear and chromatin organization and can induce modifications in Replication Timing. Whether or not these changes in RT are linked to gene expression profile modifications remains to be clarified.

In this study we focused on t(2,5)(p23,q35) translocation, involved in Anaplastic Large Cell Lymphoma (ALCL). This translocation leads to expression of NPM1-ALK fusion oncogene (Morris et al., 1994) which activates proliferation and survival pathways STAT3 (Amin et al., 2004; Chiarle et al., 2005; Zamo et al., 2002), AKT (Bai et al., 2000; Slupianek et al., 2001) and MAPK (Marzec et al., 2007).

Using a CRISPR/Cas9 approach to faithfully reproduce the NPM1-ALK translocation (Babin et al., 2018; Ghezraoui et al., 2014; Renouf et al., 2014), we hypothesized that translocation

formation could lead to overall cellular structural changes including in expression pattern and RT modifications resulting from the repositioning of chromosomal arms. Hence, we tried to decipher the consequences of translocation formation on replication modifications to gain insight into the RT and expression regulations. Importantly it was necessary to discriminate between the effects of the t(2,5)(p23,q35) translocation formation itself and the consequences of NPM1-ALK expression. To that end, we generated different t(2,5)(p23,q35) translocation cells expressing or not NPM1-ALK protein. Remarkably we found local modifications at translocation breakpoint junctions, but also genome wide. These modifications are mainly driven by expression of NPM1-ALK fusion gene and strongly correlate with modification in gene expression.

B. Results

Understanding the role of the translocation formation: generation of t(2;5) NPM1-ALK knock-out clones and t(2;5) double-translocated clones

We previously generated CRISPR/Cas9 NPM1-ALK expressing translocated RPE-1 clones that fully recapitulated activated pathways of ALCL patient cells and the formation of fusion circular RNA (Babin et al., 2018). This model enables us to compare cells before and after translocation formation whereas we have rarely access to the patient tumor cells of origin. Moreover, unlike for NPM1-ALK overexpression models, we preserve the entire chromosomal surroundings and expression level of the fusion protein. The formation of the two derivative chromosomes is a key point, since translocation appearance may disturb the overall chromatin organization and potentially the TAD structure (as shown in the study of Dr. Yue's Lab (Dixon et al., 2018)). Here we strove to separate the role of NPM1-ALK fusion protein on one hand from the role of the translocation itself on another hand. To tackle this question, we then generated NPM1-ALK knock-out translocated clones from the initial NPM1-ALK positive translocated clones thus enabling us to decipher the role of NPM1-ALK expression from the translocation itself. We used CRISPR/Cas9 system to delete a small part of NPM1-ALK fusion gene and generate a truncated nonfunctional protein (Figure 1A). We took advantage of the presence of insertions at the junctions between NPM1 and ALK genes for each of our clones. We designed gRNAs specific to the derivative 5, overlapping NPM1 extremity and the first part of the insertion in order to preserve the NPM1 wild-type allele. The second gRNA

was design in ALK exon 21 as ALK wild-type allele is not expressed in RPE-1 cells. Four different NPM1-ALK knock-out clones (further referred as Translocated-KO) were generated: 1 from CLA, 2 from CLB and 1 from CLC. PCR amplification and sequencing of the deletion were performed (Figure 1A and S1A). We then checked that NPM1-ALK protein is not expressed with no STAT3 activation. NPM1 endogenous protein level remains half of the wild type level as expected (Figure 1B). The t(2,5)(p23,q35) translocation stability was confirmed by FISH by the presence of both derivative chromosomes (Figure 1C).

In our CRISPR/Cas9 model, only one allele of Chr2 and Chr5 is translocated as found in patient cells. Our analysis relies on high throughput sequencing data making difficult the separation of allele specific reads thus leading to averaged results across the genome. To gain inside to the consequences of translocation formation only, we further modified our translocated clones in order to create a second translocation between the wild-type alleles of chr2 and chr5 (from Translocated-KO clones). Since NPM1 is essential to cell survival, we could not manage to directly induce the second translocation at the exact same breakpoint within NPM1 gene. Instead we targeted a breakpoint just downstream of NPM1 gene, in an intergenic region between NPM1 and FGF18 genes and such as we retained the endogenous expression of NPM1 gene. Importantly, we verified that this breakpoint was in the same TAD and RT domain than in wild-type cells (Figure 1D). ALK gene is not expressed in our cells so we kept the same breakpoint on chromosome 2. We generated double-translocated clones from CLA-KO, CLB-KO1 and CLC-KO clones. We were able to detect by PCR and sequence the second break junctions (Figure S1A). We could visualized the four derivative chromosomes by FISH experiment (Figure 1E).

t(2,5)(p23,q35) translocation formation leads to local Replication Timing modifications at breakpoint junctions

It was previously described that chromosomal rearrangements could lead to modification of Replication Timing (RT) at the breakpoint junctions (Dixon et al., 2018; Ryba et al., 2012). We investigated the presence of modifications in RT in our model cells. We performed Repli-seq analysis as described in (Marchal et al., 2018). RT was estimated on 50kb windows throughout the genome. We established RT at NPM1 and ALK loci and reconstituted the RT profile of the two derivative chromosomes (Figure 2A). By overlapping RT profiles of wild-type cells compared to translocated NPM1-ALK positive or KO cells, we could highlight local

modifications of RT at the two junctions. On derivative 5, the 5' part of NPM1 became earlier replicated, whereas the 3' end of ALK remained unchanged. Conversely in the derivative 2, the 5' part of ALK switched to later replicated when the 3' end NPM1 stayed late. The RT profiles of breakpoint junctions were strictly identical between Translocated and Translocated-KO clones, indicating that the formation of the translocation leads to a direct change on RT independently of NPM1-ALK fusion protein expression. Strikingly, modifications can spread over 200 kbp on the 5' side of NPM1 on der5 far beyond the NPM1 gene, including other genes (TLX3, RANBP17).

Ongoing work: We don't have yet the results of RepliSeq for the Double-Translocated clones. We did a preliminary experiment on two Double-Translocated clones to detect local modification of RT at the junction on der2 and der5 by droplet PCR. We used primers located on 5' ALK gene on derivative 2 (Figure S2A). In this region for Simple-Translocated clones (with or without NPM1-ALK expression), we observed a local modification in RT with half and half of early and late replicated DNA whereas 85% was replicated early in wild-type cells. This may reflect a mean between RT of the translocated allele and the wild-type allele. In contrast when we analyzed RT in the Double-Translocated clones we observed a major shift of the same region to late replication (90% late replicated DNA). So, at least at the breakpoint junctions, this result suggest that the translocated chromosome is responsible of the RT change at least on derivative 2 in 5'ALK gene. However we can still not exclude a direct RT synchronization of alleles when one is translocated and not the other.

Modifications in RT profiles are spread genome wide upon translocation formation

We next analyzed RT changes at the whole genome level. We first performed a hierarchical clustering analysis based on RT profiles for each clones (Figure 2B). Importantly, we were able to clearly separate wild-type cells from Translocated ones. Translocated-KO cells had an intermediate profile with 2 clones clustering with Translocated cells and 1 clone clustering with wild-type cells. As the two Translocated-KO clones obtained from CLB clustered together, we took the RT mean of these two clones for further analysis. By looking at whole genome scale, we were able to distinguish RT modified regions between each category of cells in different part of the genome. Surprisingly these modifications were not restricted to derivative chromosomes but were spread genome wide. A global view of a 30 Mb region

located on chromosome 6 is presented Figure 2C and show major modifications of RT when comparing Translocated cells to wild-type cells. We next calculated the percentage of changes between each category (Figure 2D). To do so we performed a multiple t-test on each 50 kb windows and calculated the percentage of the genome with significant RT modifications. As suggested by the clustering results, the most important changes were observed between Translocated and wild-type cells, with 22% changes and no bias between Early to Late and Late to Early regions. Translocated-KO cells were only marginally different from either wild type cells or Translocated-KO cells (1,7 to 4,1%). As Translocated-KO clones present an intermediate RT profile it was difficult to observe significant changes compared to wild-type or Translocated clones expressing NPM1-ALK. We next analyzed the size of modified regions (Figure S2). Mean of modified regions was around 650 kb whereas non modified domains had a size around 4 Mb. This size matches to the developmental domains size which get modified in RT during differentiation described in (Hiratani and Gilbert, 2010; Hiratani et al., 2008) (between 200 and 800 kb). We analyzed more precisely the RT variation and we were able to identified few domains modified in both Translocated and Translocated-KO cells (Figure 2E), but most of the domains were modified only in Translocated cells (Figure 2F). Surprisingly we also found regions modified specifically in Translocated-KO cells even if they were rare (0,4%) (Figure 2G). RT data on Double-translocated clones will enable us to further validate these region modifications. As both Translocated and Translocated-KO cells had t(2,5) translocation, we hypothesized that NPM1-ALK fusion protein could compensate some translocation induced changes. This last result is further discussed from RNAseq data.

Ongoing work: Global RepliSeq on Double-Translocated clones.

t(2;5) translocated clones recapitulate patient RNAseq signature

We performed RNAseq analysis on our different clones. Differential gene expression analysis reveals a clear transcriptional signature of Translocated clones (Figure 3A et S3). Gene ontology enrichment analysis, performed with GOrilla (Eden et al., 2009), reveals an enrichment in cell migration and adhesion processes in upregulated signature, and enrichment in cell division processed in down regulated signature. We confirmed the activation of migration in these cells by wound-scratch assay (Figure S4). In Translocated cells we observed single-cell migration invading the scratch, whereas in Wild-type cells we

observed a front-line progression over time. Overall Translocated-KO and Double-Translocated clones have very few up-regulated genes as expected but strikingly we observed a strong down-regulation signature mainly associated with cell adhesion processes. The down-regulation signature was even stronger in Double-Translocated cells (Figure 3A et S3). These results suggested that NPM1-ALK fusion protein have an activating effect whereas t(2,5) translocation itself have potentially a down-regulating effect in these cells. Moreover, nearly all down-regulated genes in Translocated-KO (20/21) were also found in Double-translocated (Figure S5A). Importantly, in the absence of NPM1-ALK expression, 18 genes (for Translocated-KO cells) and 60 genes (for Double-translocated cells) were down-regulated as found in t(2;5) NPM1-ALK positive translocated cells (Figure S5A). This provide circumstantial evidences that inducing a double translocation leads to transcriptional modifications that may have been masked by the presence of wild-type allele in simple-translocated cells. It is worth noting that 20 down-regulated genes were common uniquely between Translocated and Double-Translocated cells. This last observation could indicate that some genes are down-regulated both by NPM1-ALK protein but also to a less extend by the translocation itself. In absence of fusion protein, these genes can be detected only in presence of two translocations. We next compared our results with patient cell signatures using GSEA (Gene Set Enrichment Analysis) software (Mootha et al., 2003; Subramanian et al., 2005) (Figure 3B). This analysis reveals a strong correlation between Translocated clones and patient cells particularly for up-regulated clones (FDR < 0.01 for up-regulated signature and FDR = 0.048 for down-regulated signature). Strikingly for Translocated-KO and Double-Translocated cells, we observed an even stronger enrichment for down-regulated genes compared to patient cells (FDR < 0.01 and FDR = 0.01 respectively), probably reflecting that the translocation formation induces down-regulation of ALCL relevant genes. We further compared down-regulated genes in Translocated-KO and Double-Translocated with ALCL down-regulated genes (Figure S5B). It is worth noting that RPE-1 cells are phenotypically nothing alike T cells (ALCL is a T cell lymphoma) and already activate pathways also stimulated by NPM1-ALK (ERK and AKT) probably masking some signature of ALCL. Howbeit, 10 genes in Translocated-KO and 24 genes in Double-Translocated were also present in ALCL patient down-regulation signature. These results support the relevance and the robustness of our translocation models using CRISPR/Cas9 and support a direct role of the translocation itself in cancer signature, independently of the expression of fusion protein.

Ongoing work: We are now modeling the translocation in primary T cells from healthy donors. These cells are the supposed cells of origin in ALCL, so we hope to observe a signature closer to patient signature with this model. We will also be able to confirm some modifications we observed in RPE-1.

Replication Timing modifications are correlated with genes expression changes

It has been shown that the activation of transcription can modulate RT, notably in mouse ESC and neural progenitors, in chicken DT40 cells, and in human HCT116 cells (Blin et al., 2019; Therizols et al., 2014). Particularly, cell differentiation is correlated with modification in RT concomitant with gene expression program changes (Hiratani et al., 2008). In our models, we analyzed the correlation between modifications in RT and changes in gene expression. We evaluated the modification rate in RT at the promoter site for each gene using protocol from (Ryba et al., 2012). We display the results as a scatterplot assigning " ΔRT " to the horizontal axis, and "logFoldChange" of gene expression to the vertical axis (Figure 3C). Regression curve was determined using a linear model method. We isolate the 2% of the most modified Early to Late population and Late to Early population from the rest of the genes. We next compared the mean of each edge population to the middle one by Welch Two Sample t-test method. We observed a significant change in the mean for both populations in Translocated cells (p-value = $1.2e-12$ and $2.2e-16$ for E to L and L to E populations respectively). Meanwhile, we also observed significant changes in the mean of both populations in Translocated-KO cells but with more significant p-value for E to L population (p-value = $2.1e-6$) than L to E population (p-value = 0.04). As conclusion of this analysis, we observed globally an advance in RT for up-regulated genes and a delay for down-regulated genes. We found this correlation both in Translocated and Translocated-KO cells, suggesting that significant expression changes independent of NPM1-ALK protein expression is also correlated to RT changes. Correlation was stronger for E to L genes in Translocated-KO cells. This observation was in agreement with RNAseq signature that was stronger for down-regulated genes in these cells. These results brought new arguments in favor to a link between Replication Timing and gene expression modifications

Ongoing work: After performing RepliSeq in Double-Translocated cells we will further validate this correlation. We expect a better correlation for down-regulated genes as RNAseq analysis reveal a stronger down-regulation signature compared to simple-translocated cells.

Analysis of RT for ALCL associated genes

We next analyzed the modifications in RT specifically restrained to known ALCL associated genes. To that end, we compared different sets of RNAseq data: data from our Translocated clones, data from ALCL patient's cells (compared to ganglionic wild-type CD4+ cells) but also data from CD4+ overexpressing NPM1-ALK (compared to CD4+ isogenic cells). These last datasets were provided by our collaborator Dr Meggetto (Figure 3D and S5). Importantly, 39 genes were modified in all three cell types, and as expected from previous GSEA analysis we found number of genes modified both in our Translocated clones and in patient cells (259). We also found specific modified genes between the NPM1-ALK overexpression model and our Translocated clones (82). Possibly this genes could also be true ALCL specific genes but missing of ALCL patient signature due to the failure to compare with the proper "physiological" cells of origin. From this analysis, we could separate genes specific of the fusion protein (all genes present in overexpression model, highlight by #) and genes specific of the translocation (genes present only in ALCL patient's cells and in our Translocated clones, highlight by *). We then search for RT modifications in our Translocated clones for the 259 ALCL specific genes (Figure 3E). Interestingly 32% of modulated genes showed significant modification in RT compared to wild-type cells. This percentage is similar for genes specific to NPM1-ALK expression and genes specific to the translocation. This result further supports the link between RT and gene expression. As already reported (Rhind and Gilbert, 2013), we observed that up-regulated genes are mainly changed from Late to Early modified (only 2/214 up-regulated genes were Early to Late). Concerning the down-regulated genes, we observed only 30% of the ALCL genes switching from Early to Late (16/127) when a vast majority changed from Late to Early (36/127). So we couldn't directly associate a later RT with a ALCL specific down-regulation in gene expression. These results indicated that a third of ALCL are associated genes are modified both in expression and RT, but only up-regulated genes exhibit an earlier RT.

Ongoing work: This analysis will be further extended to Translocated-KO and Double-Translocated cells. Interestingly, the relation between down-regulation and later RT in Double

Translocated cells, where the transcriptional signature is mainly a strong down-regulation signature, will be investigated.

C. Conclusion/Discussion

In this paper we analyzed Replication Timing in CRISPR induced NPM1-ALK translocated cells. We separated the role of the fusion protein and the consequence of the translocation itself. To do so, we inactivated NPM1-ALK fusion protein in our translocated clones and also generated Double-Translocated clones. We could highlight local changes in RT at the two breakpoint junctions, mediated by the translocation and independent of the fusion protein expression. These results are consistent with previous studies (Ryba 2012, Dixon 2018).

More interestingly we showed global modifications in RT spread genome wide. These modifications are mainly driven by NPM1-ALK fusion protein expression. In Translocated cells we observed 22% modified regions. This value is in agreement with previous results obtained in leukemia cells (Ryba 2012). In this study Dr. Gilbert's lab pointed out that modifications in RT could be observed without chromosomal rearrangement. Our results indicated that RT modifications are generated mainly by oncogene expression. Moreover, the effect of translocation itself is low and can be difficult to detect. Remarkably the use of Double-Translocated clones helps us to decipher more precisely the contribution of the translocation formation itself.

We then analyzed transcriptional profiles of our clones. We showed a strong up-regulation signature in presence of NPM1-ALK protein similar to patient cell signature. A previous study even showed that STAT3 pathway mainly drives the transcriptomic signature in ALCL (ALK+) (Piva et al., 2010). Thereby even if RPE-1 cells that have constitutional activation of ERK and AKT pathways, the activation of STAT3 in our RPE-1 translocated clones appears to be sufficient to recapitulate most of the major effects of tumor cells, at least at the gene expression level.

More interestingly we showed that translocation in absence of NPM1-ALK expression induced a substantial down-regulation effect, and that this effect is even stronger in Double-

Translocated cells. Overexpression models failed to reproduce these gene expression modulations that could be directly linked to oncogenesis. The down-regulation signature is stronger in absence of NPM1-ALK and more similar to patient ALCL signature suggesting that NPM1-ALK protein and the translocation have an opposite effect on some pathways. This was also observed in RT with some regions modified only in Translocated-KO cells. It is also possible that this opposite effect was specific to RPE-1 cells, where ERK and AKT pathways are already activated. Some compensatory mechanisms can appear, as over-activation of oncogenic pathways can be deleterious for the cells. In absence of NPM1-ALK fusion protein these compensatory mechanisms are no longer necessary.

We also observed a good correlation between Replication Timing changes and Gene expression modifications. The most modified genes in RT are also the most modified genes in RNAseq. This correlation is observed with or without fusion protein, with a better correlation for down-regulated genes in absence of NPM1-ALK. This is consistent with RNAseq data. We are now waiting for RepliSeq in Double-Translocated clones to validate our results. Overall these results provide further new evidences for the association between genes expression and Replication Timing as previous studies have demonstrated it (Blin et al., 2019; Therizols et al., 2014).

Results from this article draft are further discussed in the “Discussion and Perspective” part of the Thesis manuscript.

D. Material and methods

Cell culture

RPE-1 cells were cultured in DMEM/F-12 medium (Gibco 10565018) supplemented with 10% FBS (Gibco 26140079).

CRISPR/Cas9 plasmids

The gRNAs vectors were cloned in MLM3636 plasmid (Addgene plasmid 43860). gRNAs sequences are listed below. gRNAs targeting human NPM1 and ALK genes have been previously described (Ghezraoui et al., 2014; Renouf et al., 2014, Babin et al., 2018). All gRNAs were designed using the CRISPOR designer tool (<http://crispor.tefor.net/crispor.py>). Cas9-2A-GFP was expressed from pCas9-GFP vector (Addgene plasmid 44719). gRNA sequences are list in supplementary material.

RPE-1 transfections and sorting

RPE-1 cells were transfected with Lonza 4D nucleofactor using P3 solution (V4XP-3024) and EA-107 program. 1.10⁶ cells were transfected with 3 µg Cas9-GFP and 3 µg of each guides in nucleocuvette vessels. At 48 h post transfection 10% most GFP+ cells were sorted on Sony SH800 and seeded in 96 wells plate at 10 cells per well.

DNA extraction

DNA was extracted using EznA Tissue DNA kit (D3396-01 Omega Bio-Tek) following manufacturer's instruction.

RNA extraction

Habitually 10.10⁶ cells were used for RNA extraction. Extraction was performed using NucleoSpin[®] RNA kit (Macherey Nagel 740955.50) following manufacturer's instructions.

Translocation detection and sequencing

Breakpoint junctions were detected by PCR as previously described in (Babin et al., 2018). Sanger sequencing was performed from PCR product. Primers sequences are listed in supplementary material.

Western blot analysis

Whole cell extracts were prepared with protein lysis buffer (50 mM Tris-HCl at pH 7.4, 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT prepared from a 1 M DTT stock), with addition of Complete cocktail protease inhibitor tablets (Roche) and Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific). Typically, 30 µg of protein extracts were run on an 8% (w/v) Tris-HCl SDS PAGE gel. Gel was blotted, and probed with Phospho-ALK (Tyr1604), ALK (31F12), Phospho-Stat3 (Tyr705), and Stat3 antibodies from Cell Signaling. Alpha-Tubulin mouse antibody (Sigma-Aldrich) was used as a loading control. Phospho-ALK and Phospho-Stat3 antibodies were used at 1/200 dilution. Other antibodies were used according to manufacturer's recommendations. IRDye 800CW secondary antibodies were used, and blot developed with Odyssey® CLx Imaging System (LI-COR).

Fluorescence in situ hybridization

FISH was performed on metaphases from translocated RPE-1 clones. Human ALK and NPM1 human genes were labeled using respectively ALK Breakapart LPS 019 probe (Cytocell) and homemade NPM1 probe from BAC RP11-546B8. Metaphase spreading, probe labeling, hybridization, washing and fluorescence detection were performed according to standard procedures. Images were taken with the Zeiss Spinning Disk system.

Wound scratch assay

Wound scratch assay was performed using IncuCyte system (Essen BioScience) following manufacturer's recommendations. Typically, 1000 cells/well were seeded in 96 wells plate. Scratch was performed when cells reached confluence and migration was followed during 24 hours. Percentage of relative wound density was determined every 3 hours and compared between samples with t-test.

Replication timing sequencing

Protocol were adapt from (Marchal et al., 2018). Cells were incubated 1h with 50 µM BrdU (ThermoFischer Scientific B23151) at 37°C. Then they were washed twice with PBS and fixed in cold EtOH 70% at 1.106 cells/mL. Cells can be stored at -20°C. Before sorting cells were washed twice with PBS and incubated with 1mg/mL RNaseA (Invitrogen 12091021) in PBS at 37°C 30 min. Propidium Iodide (Sigma P4170) was added at 20 µg/mL and cells were incubated 30 min on ice in the dark. Cells were sorted with Sony SH800 based on S phase fractions.

Usually 200 000 cells were sorted in S1 and S2 phases. Cells were resuspended in lysis buffer (Tris-HCl pH8 50 mM, EDTA pH8 10 mM, SDS 0.8%, proteinase K 0.2 mg/mL) and incubated 2 hours at 55°C. Lysed cells can be stored at -20°C. Phenol-chloroform DNA extraction was performed. DNA was resuspended in TE buffer at 5ng/μL and sonicated on Picobioruptor system (Diagenode B01060010). 3 cycles were done: 30 s ON/ 30 s OFF. Short centrifugation was performed and 3 additional cycles were done. DNA fragments size was assessed on fragment analyzer. DNA was purified using Zymo Clean and Concentrator -5 kit (D4013 Zymo Research) following manufacturer instructions. Indexing for sequencing was performed using NEBNext® Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs) (New England Biolabs E6440S) following manufacturer instructions. DNA was purified using Zymo Clean and Concentrator -5 kit. DNA was resuspended in 1 mL TE buffer and BrdU immunoprecipitation was performed: DNA was heated at 95°C for 10 min and cooled on ice for 2 min. 50 μL adjusting buffer (Sodium Phosphate Buffer pH7 110 mM, NaCl 1540 mM, TritonX-100 0,55%) were added and 20 μL BrdU antibody (Anti-BrdU (B44) 347580 BD Biosciences). DNA was incubated 2h with rotation at RT. 40 μL AG Plus agarose beads were added to the solution and 1h incubation was performed with rotation at RT. DNA was pelleted at 4°C maximum speed 5 min. 1 wash with 750 μL IP buffer (Sodium phosphate buffer pH 7 10 mM NaCl, 140 mM TritonX-100 0,05%) was performed with 5 min incubation with rotation at RT. DNA was pelleted at 4°C maximum speed 5 min. Pellet was resuspended in 200 μL digestion buffer (Tris-HCl pH 8 50 mM, EDTA 10 mM, SDS 0,5%, proteinase K 0,2 mg/mL) and incubated at 37°C overnight. 100 μL digestion buffer were added and incubation was followed at 50°C 1h. Phenol-chloroform DNA extraction was performed. DNA was purified using Zymo Clean and Concentrator -5 kit, elution was done in 16 μL H₂O. PCR cycles number was determined by Q-PCR with 1/10^e of final PCR volume. 1,5 μL library, 2,5 μL NEBNext Q5 HotStart Hifi master mix, 1 μL uniq primers Dual 96, 0,25 μL EvaGreen 20X were mixed. Q-PCR was performed: 98°C 30s, 40x(98°C 10 s, 65°C 75 s), 65°C 5 min. Optimal cycles number was determined for each sample when amplification curve reach ¼ of maximum value. PCR was performed with optimal cycles number mixing 15 μL library, 25 μL NEBNext Q5 HotStart Hifi master mix, and 10 μL uniq primers Dual 96. PCR products were purified with AMPure XP Beads following manufacturer's instruction with ratio 1:9. Second round purification was performed to eliminate primers dimers with the same protocol. Elution was done in 33 μL 10mM Tris-HCl pH 8. Libraries were

sequenced on NovaSeq 6000 Sequencing System (Illumina). Replication Timing values were determined as described in (Marchal et al., 2018).

RNAseq analysis

RNA was sequenced using NovaSeq 6000 Sequencing System (Illumina). Data were aligned on hg19 human genome using Bowtie 2 aligner on Galaxy Europe platform. Statistical analyses were performed using Deseq2 software on Galaxy Europe platform. GSEA analysis was performed using GSEA software (Mootha et al., 2003; Subramanian et al., 2005). In this analysis we used ALCL patient signature from Dr Meggetto lab's generated by comparison of RNAseq data between wild type RLN and ALCL ALK tumors. Gene ontology enrichment analysis was performed with GOrilla online tool (Eden et al., 2009) using ranked gene list.

Data analysis

Percent changes analysis in Replication Timing was performed using multiple t-test on each 50 kb windows. Clustering approach was performed using pvclust R package. Size of modified regions was determined based on (Ryba et al., 2011) protocol. Assignment of replication timing values to promoter was performed based on (Ryba et al., 2011) protocol. We used position of the 5'UTR for each gene based on RefSeq data from UCSC Genome Browser. Correlation between RT and LogFoldChange was analyzed as follow: scatterplot was performed using ggplot2 R package, 4% most modified genes in replication timing were selected (2% Late to Early and 2% Early to Late, n=345 in each group) and Welch Two Sample t-test was performed comparing mean of LtoE/EtoL with mean of the non-modified genes. Regression curve was determined by lm method. Data from overexpression model and ALCL patient were provided by Dr. Meggetto from CRCT (Toulouse). Venn diagrams were obtained with R package "VennDiagram"

E. Figures legends

Figure 1

A) Schematic representation of NPM1-ALK inactivation strategy in t(2;5) Translocated clones. Sequencing of the obtained Der5 junctions are presented for each Translocated-KO clone. **B)** Activation of NPM1-ALK: Western-blot analysis of Phospho-NPM1-ALK, NPM1-ALK, Phospho-STAT3, STAT3, NPM1 from wild-type (blue), Translocated (red), Translocated-KO (green), and Double-Translocated (black) cells. **C)** FISH experiment with a break-apart probe against *ALK* (red and green) and *NPM1* (blue) loci for each Translocated-KO clone. **D)** Schematic representation of double-translocation strategy. *NPM1* locus and surrounding genomic regions are presented. TAD domains (From data in (Darrow et al., 2016)) were represented, as well as Replication Domains in wild-type RPE-1. The 1st breakpoint was used for NPM1-ALK translocation formation and the second breakpoint is used to induce the second translocation (highlight with dark lines). **E)** FISH experiment with probe against *ALK* (red and green) and *NPM1* (blue) loci for each Double-Translocated clone.

Figure 2

A) Reconstitution of RT at the two junctions of derivative chromosomes. Overlay of RT profile of wild-type in blue, Translocated in red and Translocated-KO in green. Positions of *NPM1-ALK* and *ALK-NPM1* genes are represented with arrows. **B)** Hierarchical clustering of different clones based on RT profiles. **C)** Example RT profiles of a 30 Mb region on chromosome 6 (wild-type in blue, Translocated in red and Translocated-KO in green). Major changes are highlighted with arrows. **D)** Percent changes in RT calculated on whole genome. **E)** Examples of regions with modified RT profiles for both Translocated and Translocated-KO cells. **F)** Examples of regions with modified RT profiles only for Translocated cells. **G)** Examples of regions with modified RT profile only for Translocated-KO cells.

Figure 3

A) Volcano plots from differential gene analysis performed between Translocated and wild-type cells (Left), Translocated-KO and wild-type cells (Middle) and Double-Translocated and wild-type cells (Right). Red color was assigned to non-significantly modified genes and blue to modified genes with thresholds “log₂ Fold Change” > 2 and padj < 0.05. **B)** GSEA enrichment

plot for Translocated, Translocated-KO and Double-Translocated cells with associated FDR. ALCL signature from Dr. Meggetto was used as reference gene list. Up-regulated genes in ALCL signature are presented on the left and down-regulated genes on the right. **C)** Scatterplot with Δ RT on horizontal axis and Log 2 Fold Change in vertical axis for Translocated (left) and Translocated-KO (right) compared to wild-type cells. Red color was assigned to genes with no modified RT and blue to the 2% most Early to Late genes and the 2% most Late to Early genes. Boxplot of these 2 populations were indicated in each side of the graph with the associated p-value (Welch Two Sample t-test between each population and the non modified population). **D)** Venn Diagram of modified genes in ALCL patient cells, Translocated RPE-1 clones and Overexpression T cells model (Data from Dr. Meggetto). Genes specific to NPM1-ALK fusion protein are labeled with # and translocation specific genes with *. **E)** Analysis of Replication Timing for ALCL associated genes.

Figure 1

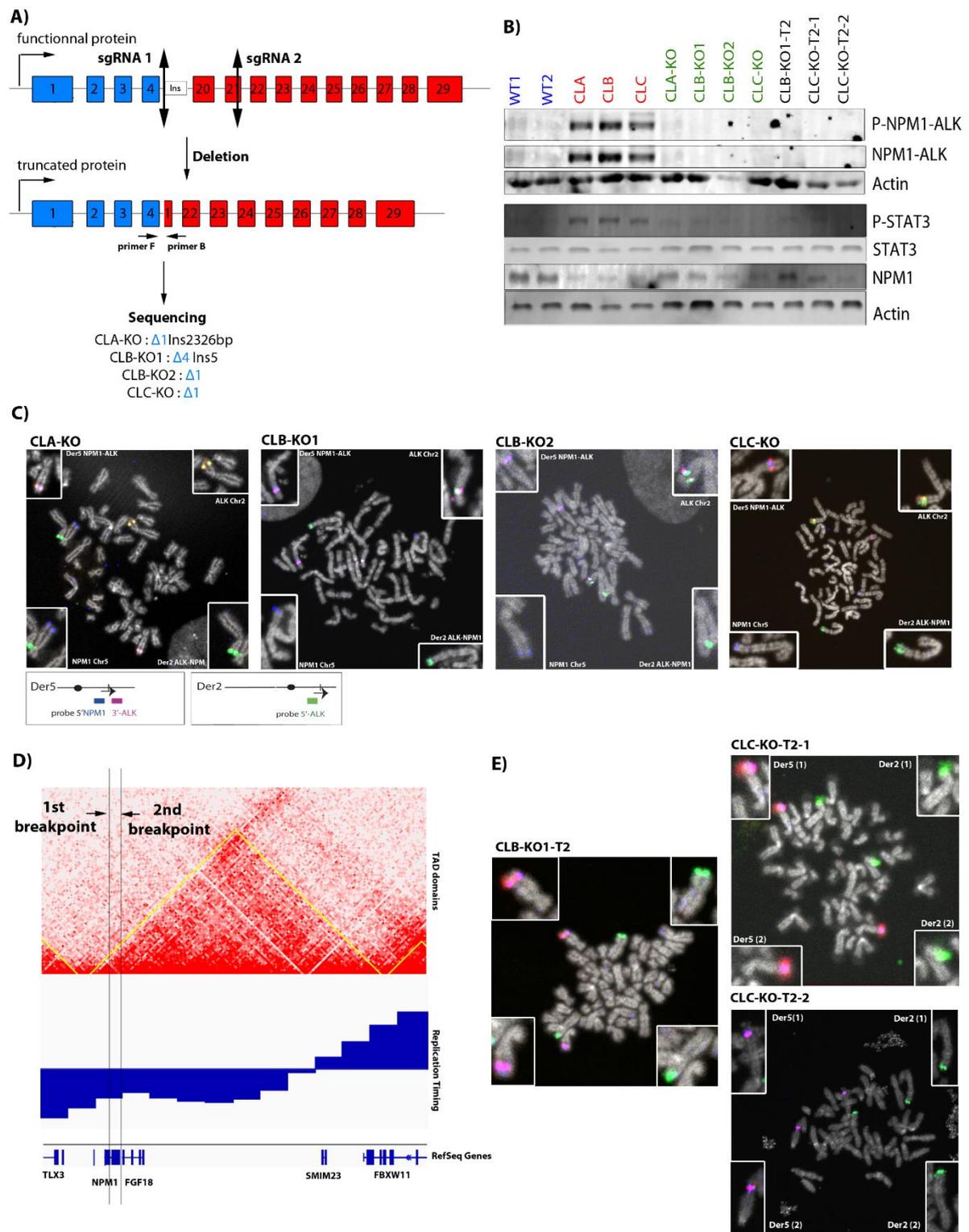


Figure 2

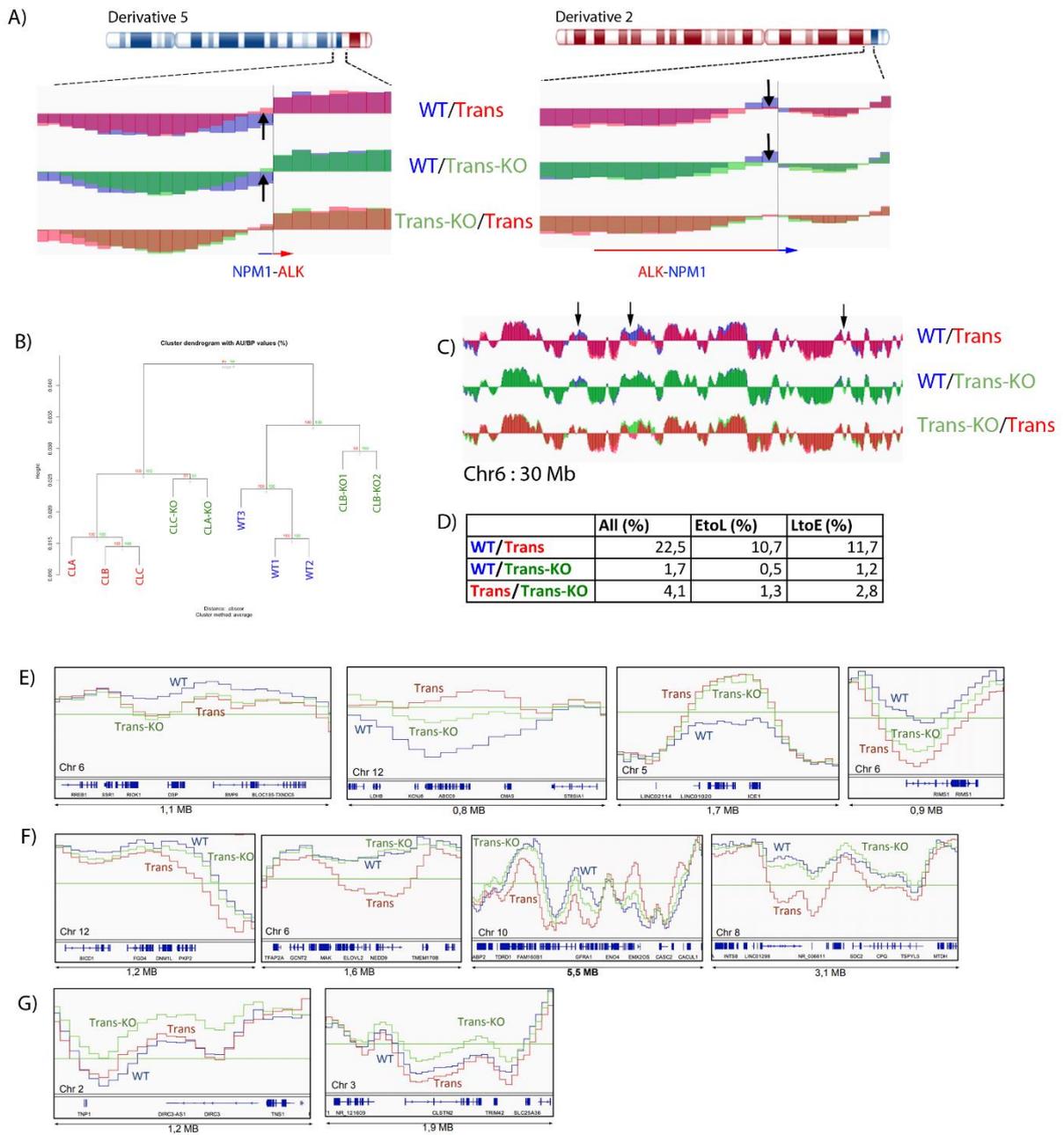
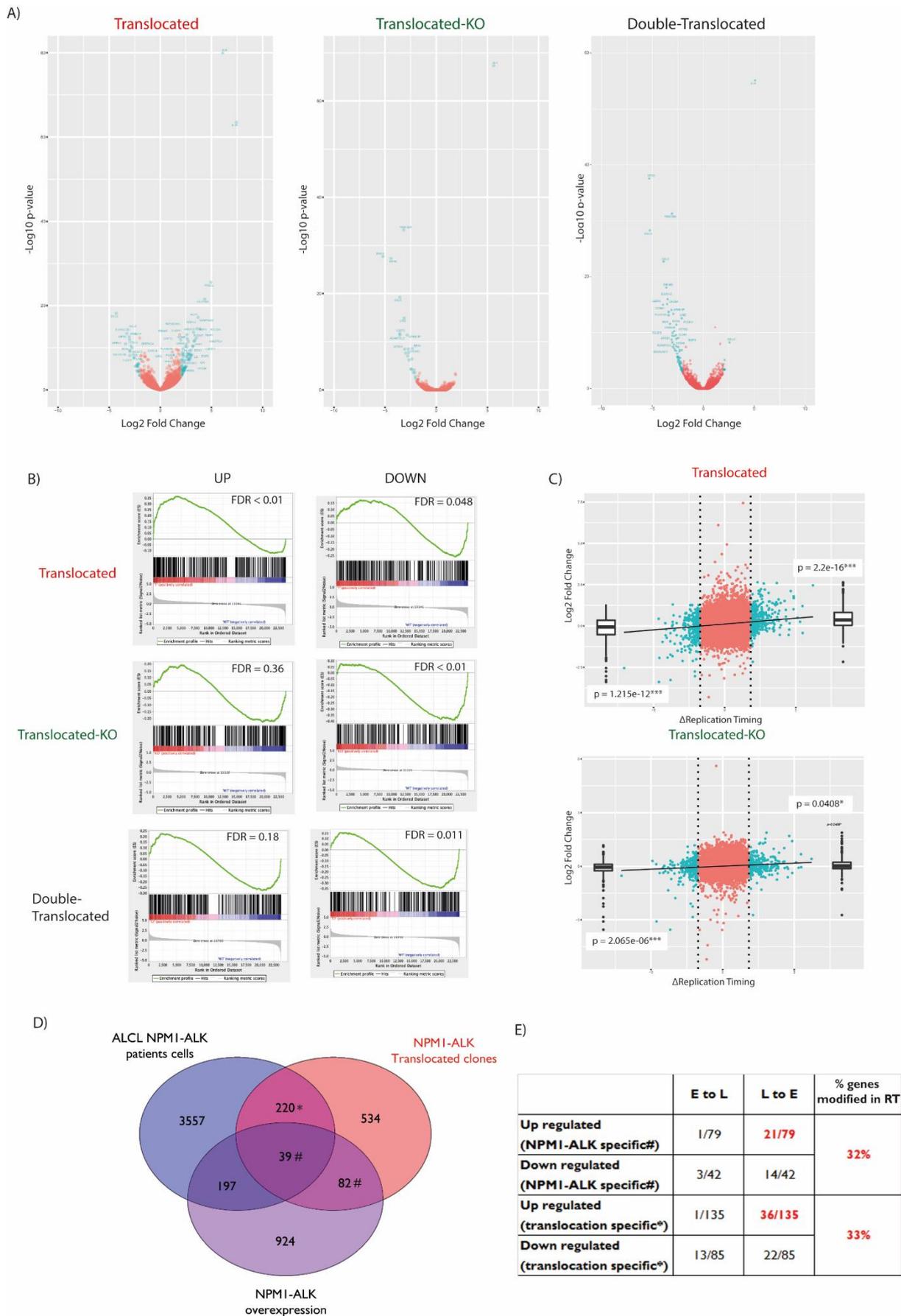


Figure 3



F. Supplementary Data

Supplementary material

gRNA

Target gene	Sequence
ALK	GTCGGTCCATTGCATAGAGG
NPM1	GTGAACCCAGTAGCAGTTCG
NPM1-CLA for KO	CTGTTGCTCATCAGCGTTCG
NPM1-CLB/CLC for KO	GCTTGCTATATCCTCGAATC
ALK(Exon 21) for KO	CGTTGGGCATTCCGGACACC
Intergenic (NPM1/FGF18)	CAAGAGACCACCTGTACGAT
ALK Breakpoint 2	GGTCGGTCCATTGCATAGGA

Primers

Target gene	Forward primer	Reverse primer
ALK-NPM1	GAGACATGCCAGGACAGAT	AGAGCACATGGGAAAAGGAA
NPM1-ALK	GAACTCCTGGCCTTAACGTG	CCACCCTCTAGGGTTGTCAAT
ALK-intron18 (droplet)	CTGTGTGAAAGGTGGGTCAT	TGGGCATGTTAAGTGGAGG
NPM1-ALK KO	GAACTCCTGGCCTTAACGTG	TCGTCCTGTTAGAGCACAC

Supplementary figure legends

Figure S1 (Related to Figure 1)

Junction sequences of the two derivative chromosomes (derivative 2 and derivative 5) for each Translocated-KO and Double-Translocated clone.

Figure S2 (Related to Figure 2)

A) Domain sizes for unmodified domains (None), Early to Late domains (EtoL) and Late to Early domains (LtoE) in Translocated (Left) and Translocated-KO (Right) compared to wild-type cells.

B) Local modifications in RT on Der2 in 5'ALK gene and Schematic representation of Der2 RT profil from Wild-type (blue, no translocation), compared to Translocated cells (red, one translocation) and theoretical RT profile with double translocation

Figure S3 (Related to Figure 3)

Heat-maps of most modified genes in RNAseq. Translocated versus wild-type (left), Translocated-KO versus wild-type (middle) and Double-Translocated versus wild-type (right)

Figure S4 (Related to Figure 3)

A) Typical pictures from wound scratch assay at 0, 12 and 24h from Wild-type and Translocated RPE-1 cells. **B)** Graph of Relative wound density percentage over time in Translocated cells in blue and WT RPE-1 cells in black. One star (*) indicates p-value < 0.05, two stars (**) indicate p-value < 0.01.

Figure S5 (Related to Figure 3)

A) Venn Diagram for down-regulated genes in Translocated, Translocated-KO and Double-Translocated cells. **B)** Venn Diagram for down-regulated genes in ALCL patient cells, Translocated-KO and Double-Translocated cells.

Figure S1 (Related to Figure 1) : junctions sequences

Translocated-KO clones

CLA-KO
Der5
AGCTTGCTATATCCTCG Δ I Ins2326bp GTCCGGAATGCCCAACGACC
Der2
ACGGTCGGTCCATTGCATAG Δ I GCTACTGGGTTACCTCAGC

CLB-KO1
Der5
CGGCTGGAGTGCAGTGATGT Δ 4 Ins5 CCGGAATGCCCAACGACCCA
Der2
ACGGTCGGTCCATTGCATAG G>C CACTGCTACTGGGTTACCT

CLB-KO2
Der5
ATAGCTTGCTATATCCTCGA Δ I GTCCGGAATGCCCAACGACC
Der2
ACGGTCGGTCCATTGCATAG G>C CACTGCTACTGGGTTACCT

CLC-KO
Der5
ATAGCTTGCTATATCCTCGA Δ I GTCCGGAATGCCCAACGACC
Der2
ACGGTCGGTCCATTGCATAG Δ 3 TGCTACTGGGTTACCTCAG

Double translocated clones

CLB-KO1-T2
Der5
CGGCTGGAGTGCAGTGATGT Δ 4 Ins5 CCGGAATGCCCAACGACCCA
Der2
ACGGTCGGTCCATTGCATAG G>C CACTGCTACTGGGTTACCT
Der5 (2)
CCTTATAGTCTGGGCCATC InsA GGAGGGAAGTACACATAG
Der2 (2)
CACGGTCGGTCCATTGCATA Δ A Δ CG GTACAGGTGGTCTCTTGCTA

CLC-KO-T2-1
Der5 (1)
ATAGCTTGCTATATCCTCGA Δ I GTCCGGAATGCCCAACGACC
Der2 (1)
ACGGTCGGTCCATTGCATAG Δ 3 TGCTACTGGGTTACCTCAG
Der5 (2)
CTTATAGTCTGGGCCATCG Δ 2 Δ 2 AGGAGGGAAGTACACATA
Der2 (2)
TGATCACGGTCGGTCCATTG Δ 6 CGTACAGGTGGTCTCTTGCT

CLC-KO-T2-2
Der5 (1)
ATAGCTTGCTATATCCTCGA Δ I GTCCGGAATGCCCAACGACC
Der2 (1)
ACGGTCGGTCCATTGCATAG Δ 3 TGCTACTGGGTTACCTCAG
Der5 (2)
CTTATAGTCTGGGCCATCG Δ 2 GAGGGAAGTACACATAGC
Der2 (2)
TGATCACGGTCGGTCCATTG Δ 6 CGTACAGGTGGTCTCTTGCT

Figure S2 (Related to Figure 2) : Replication domains

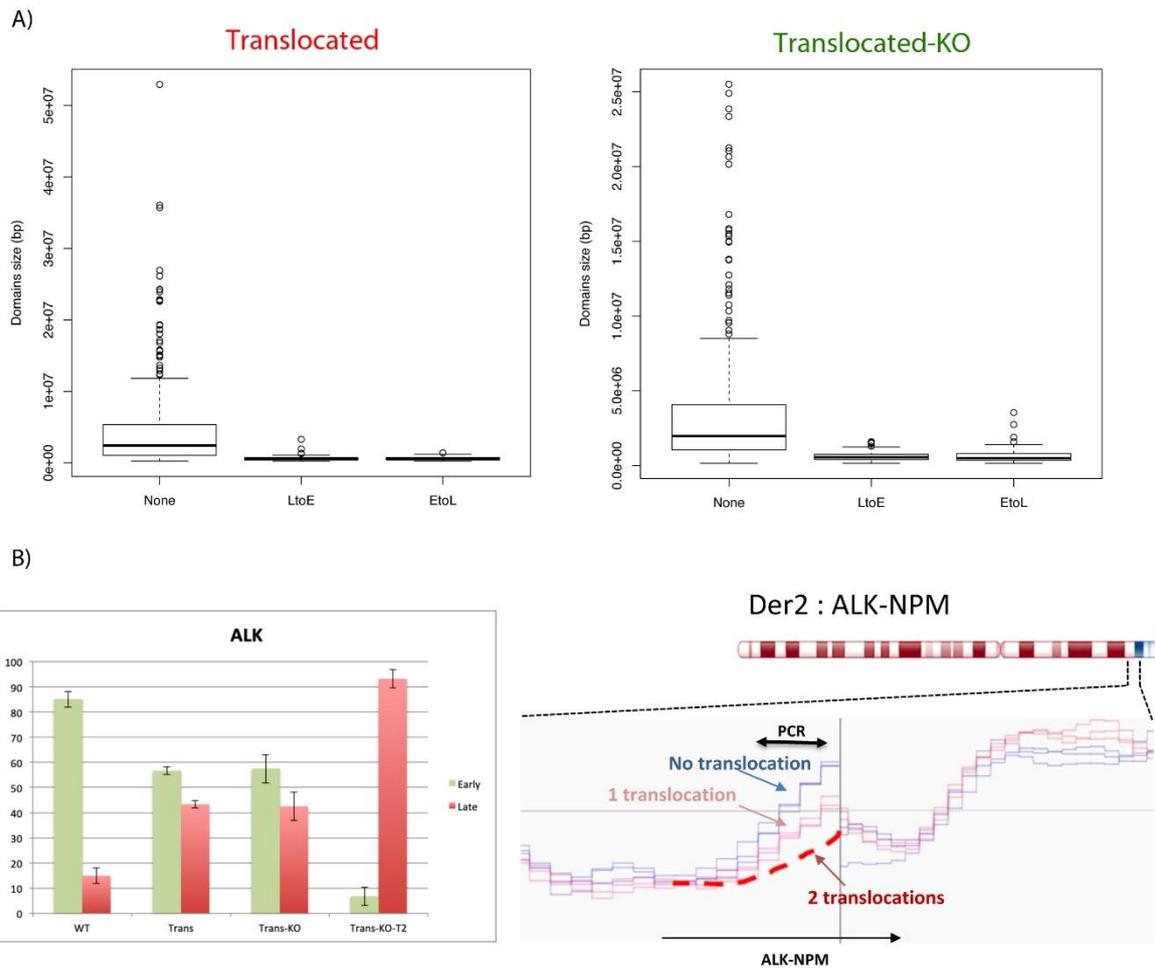


Figure S3 (Related to Figure 3) : Heatmap of RNAseq Datas

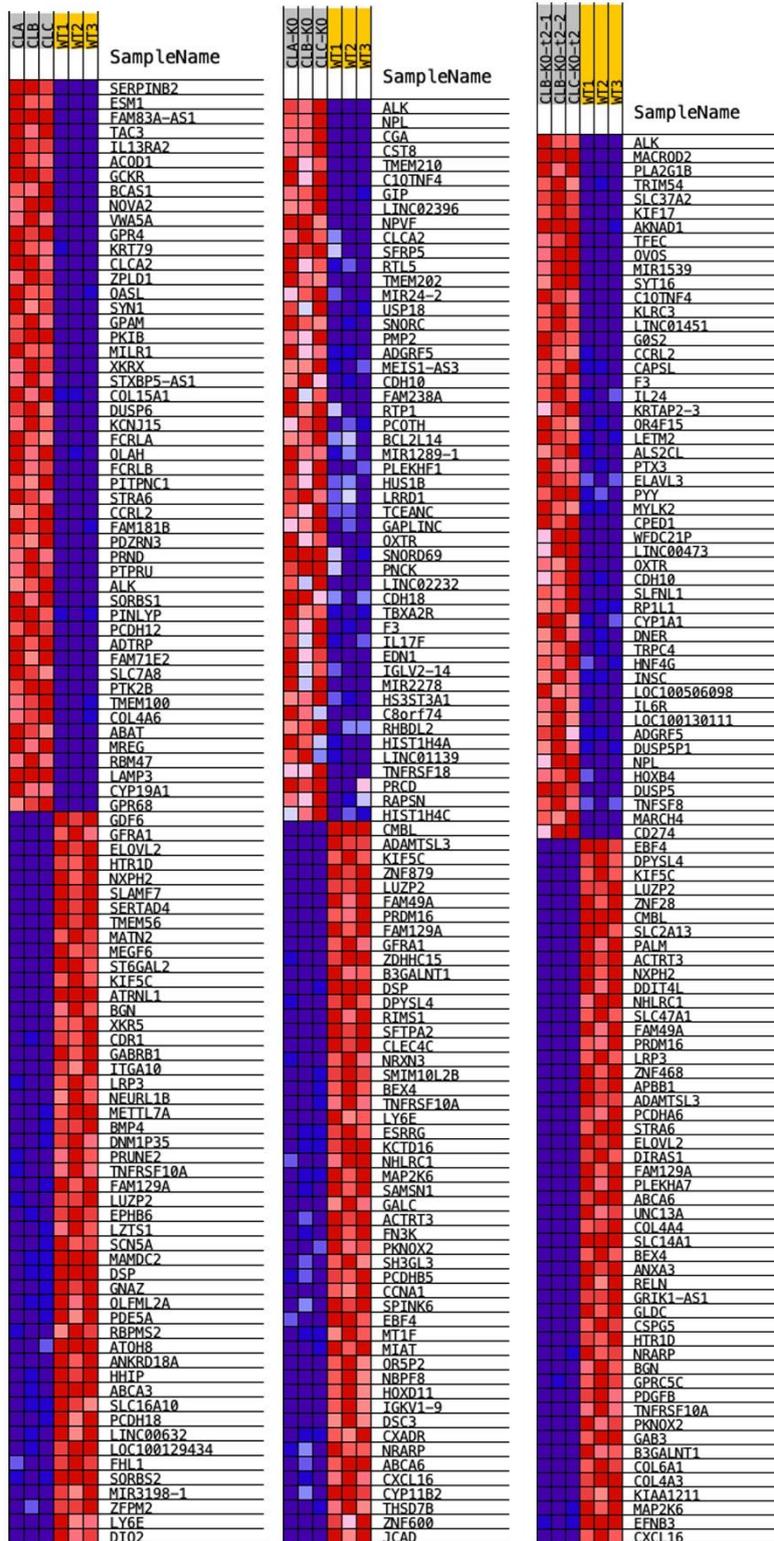
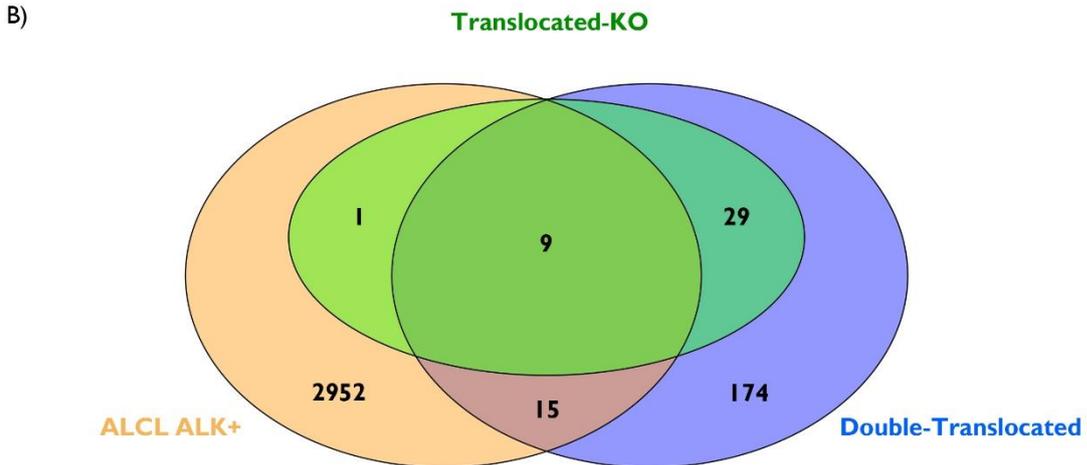
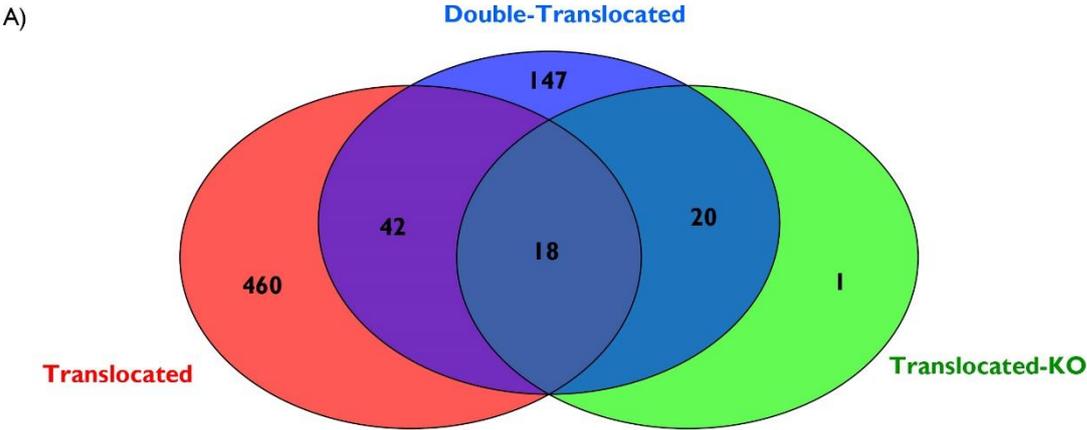


Figure S5 (Related to Figure 3) : Down-regulated signatures comparison



G. Bibliography

- Adolph, S., Hameister, H., and Schildkraut, C.L. (1992). Molecular analysis of the aberrant replication banding pattern on chromosome 15 in murine T-cell lymphomas. *Chromosoma* 101, 388-398.
- Amin, H.M., McDonnell, T.J., Ma, Y., Lin, Q., Fujio, Y., Kunisada, K., Leventaki, V., Das, P., Rassidakis, G.Z., Cutler, C., *et al.* (2004). Selective inhibition of STAT3 induces apoptosis and G(1) cell cycle arrest in ALK-positive anaplastic large cell lymphoma. *Oncogene* 23, 5426-5434.
- Babin, L., Piganeau, M., Renouf, B., Lamribet, K., Thirant, C., Deriano, L., Mercher, T., Giovannangeli, C., and Brunet, E.C. (2018). Chromosomal Translocation Formation Is Sufficient to Produce Fusion Circular RNAs Specific to Patient Tumor Cells. *iScience* 5, 19-29.
- Bai, R.Y., Ouyang, T., Miething, C., Morris, S.W., Peschel, C., and Duyster, J. (2000). Nucleophosmin-anaplastic lymphoma kinase associated with anaplastic large-cell lymphoma activates the phosphatidylinositol 3-kinase/Akt antiapoptotic signaling pathway. *Blood* 96, 4319-4327.
- Blin, M., Le Tallec, B., Nahse, V., Schmidt, M., Brossas, C., Millot, G.A., Prioleau, M.N., and Debatisse, M. (2019). Transcription-dependent regulation of replication dynamics modulates genome stability. *Nature structural & molecular biology* 26, 58-66.
- Breger, K.S., Smith, L., and Thayer, M.J. (2005). Engineering translocations with delayed replication: evidence for cis control of chromosome replication timing. *Human molecular genetics* 14, 2813-2827.
- Chen, Y.H., Keegan, S., Kahli, M., Tonzi, P., Fenyo, D., Huang, T.T., and Smith, D.J. (2019). Transcription shapes DNA replication initiation and termination in human cells. *Nature structural & molecular biology* 26, 67-77.
- Chiarle, R., Simmons, W.J., Cai, H., Dhall, G., Zamo, A., Raz, R., Karras, J.G., Levy, D.E., and Inghirami, G. (2005). Stat3 is required for ALK-mediated lymphomagenesis and provides a possible therapeutic target. *Nature medicine* 11, 623-629.
- Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.C., Sanborn, A.L., Machol, I., Shamim, M., *et al.* (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proceedings of the National Academy of Sciences of the United States of America* 113, E4504-4512.
- Dixon, J.R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V.T., Yardimci, G.G., Chakraborty, A., Bann, D.V., Wang, Y., *et al.* (2018). Integrative detection and analysis of structural variation in cancer genomes. *Nature genetics* 50, 1388-1398.

- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics* *10*, 48.
- Eul, J., Gronemeyer, H., Adolph, S., and Hameister, H. (1988). Suppression of tumorigenicity in T-cell lymphoma hybrids is correlated with changes in myc expression and DNA replication of the myc chromosomal domain. *Chromosoma* *96*, 248-254.
- Fritz, A., Sinha, S., Marella, N., and Berezney, R. (2013). Alterations in replication timing of cancer-related genes in malignant human breast cancer cells. *Journal of cellular biochemistry* *114*, 1074-1083.
- Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C., *et al.* (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Molecular cell* *55*, 829-842.
- Hassan-Zadeh, V., Chilaka, S., Cadoret, J.C., Ma, M.K., Boggetto, N., West, A.G., and Prioleau, M.N. (2012). USF binding sequences from the HS4 insulator element impose early replication timing on a vertebrate replicator. *PLoS biology* *10*, e1001277.
- 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., Yokochi, T., Schwaiger, M., Chang, C.W., Lyou, Y., Townes, T.M., Schubeler, D., and Gilbert, D.M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS biology* *6*, e245.
- Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E. (2018). Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell reports* *24*, 1-10 e14.
- Marchal, C., Sasaki, T., Vera, D., Wilson, K., Sima, J., Rivera-Mulia, J.C., Trevilla-Garcia, C., Nogues, C., Nafie, E., and Gilbert, D.M. (2018). Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nature protocols* *13*, 819-839.
- Marzec, M., Kasprzycka, M., Liu, X., Raghunath, P.N., Wlodarski, P., and Wasik, M.A. (2007). Oncogenic tyrosine kinase NPM/ALK induces activation of the MEK/ERK signaling pathway independently of c-Raf. *Oncogene* *26*, 813-821.
- Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., *et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics* *34*, 267-273.

Morris, S.W., Kirstein, M.N., Valentine, M.B., Dittmer, K.G., Shapiro, D.N., Saltman, D.L., and Look, A.T. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 263, 1281-1284.

Panning, M.M., and Gilbert, D.M. (2005). Spatio-temporal organization of DNA replication in murine embryonic stem, primary, and immortalized cells. *Journal of cellular biochemistry* 95, 74-82.

Piva, R., Agnelli, L., Pellegrino, E., Todoerti, K., Grosso, V., Tamagno, I., Fornari, A., Martinoglio, B., Medico, E., Zamo, A., *et al.* (2010). Gene expression profiling uncovers molecular classifiers for the recognition of anaplastic large-cell lymphoma within peripheral T-cell neoplasms. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28, 1583-1590.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., *et al.* (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature* 515, 402-405.

Pourkarimi, E., Bellush, J.M., and Whitehouse, I. (2016). Spatiotemporal coupling and decoupling of gene transcription with DNA replication origins during embryogenesis in *C. elegans*. *eLife* 5.

Renouf, B., Piganeau, M., Ghezraoui, H., Jasin, M., and Brunet, E. (2014). Creating cancer translocations in human cells using Cas9 DSBs and nCas9 paired nicks. *Methods in enzymology* 546, 251-271.

Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. *Cold Spring Harbor perspectives in biology* 5, a010132.

Ryba, T., Battaglia, D., Chang, B.H., Shirley, J.W., Buckley, Q., Pope, B.D., Devidas, M., Druker, B.J., and Gilbert, D.M. (2012). Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome research* 22, 1833-1844.

Ryba, T., Battaglia, D., Pope, B.D., Hiratani, I., and Gilbert, D.M. (2011). Genome-scale analysis of replication timing: from bench to bioinformatics. *Nature protocols* 6, 870-895.

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.

Seller, C.A., and O'Farrell, P.H. (2018). Rif1 prolongs the embryonic S phase at the *Drosophila* mid-blastula transition. *PLoS biology* 16, e2005687.

Siefert, J.C., Georgescu, C., Wren, J.D., Koren, A., and Sansam, C.L. (2017). DNA replication timing during development anticipates transcriptional programs and parallels enhancer activation. *Genome research* 27, 1406-1416.

Sima, J., Chakraborty, A., Dileep, V., Michalski, M., Klein, K.N., Holcomb, N.P., Turner, J.L., Paulsen, M.T., Rivera-Mulia, J.C., Trevilla-Garcia, C., *et al.* (2019). Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. *Cell* 176, 816-830 e818.

Simon, I., Tenzen, T., Mostoslavsky, R., Fibach, E., Lande, L., Milot, E., Gribnau, J., Grosveld, F., Fraser, P., and Cedar, H. (2001). Developmental regulation of DNA replication timing at the human beta globin locus. *The EMBO journal* 20, 6150-6157.

Slupianek, A., Nieborowska-Skorska, M., Hoser, G., Morrione, A., Majewski, M., Xue, L., Morris, S.W., Wasik, M.A., and Skorski, T. (2001). Role of phosphatidylinositol 3-kinase-Akt pathway in nucleophosmin/anaplastic lymphoma kinase-mediated lymphomagenesis. *Cancer research* 61, 2194-2199.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15545-15550.

Therizols, P., Illingworth, R.S., Courilleau, C., Boyle, S., Wood, A.J., and Bickmore, W.A. (2014). Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* 346, 1238-1242.

Williams, R.R., Azuara, V., Perry, P., Sauer, S., Dvorkina, M., Jorgensen, H., Roix, J., McQueen, P., Misteli, T., Merkschlager, M., *et al.* (2006). Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *Journal of cell science* 119, 132-140.

Yaffe, E., Farkash-Amar, S., Polten, A., Yakhini, Z., Tanay, A., and Simon, I. (2010). Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS genetics* 6, e1001011.

Zamo, A., Chiarle, R., Piva, R., Howes, J., Fan, Y., Chilosì, M., Levy, D.E., and Inghirami, G. (2002). Anaplastic lymphoma kinase (ALK) activates Stat3 and protects hematopoietic cells from cell death. *Oncogene* 21, 1038-1047.

III. Résultats préliminaires : induction de la translocation dans des cellules T primaires

L'ensemble des expériences présentées précédemment a été réalisé dans des cellules modèles : la lignée RPE-1. Les résultats obtenus étant très encourageants, nous avons choisi de réaliser des expériences similaires dans un contexte plus physiologique, en utilisant des cellules hématopoïétiques primaires. L'origine cellulaire de l'ALCL est très discutée, avec deux origines hypothétiques principales (voir introduction générale du manuscrit) : des cellules T matures (majoritairement CD4+) ou des progéniteurs thymiques immatures. Nous avons décidé d'explorer l'hypothèse d'une origine mature dans un premier temps. En effet, il a déjà été montré qu'il était possible d'induire une transformation dans ces cellules par l'utilisation de lentivirus permettant la surexpression de NPM1-ALK (Zhang et al., 2013a). Cependant, cette stratégie présente une efficacité de transformation très limitée. En effet, l'équipe du Dr. Wasik induit une surexpression de la protéine de fusion dans plus de 50% des cellules, mais n'obtient que quelques clones transformés. De même l'équipe du Dr. Meggetto (avec qui nous avons collaboré pendant ma thèse) a obtenu des résultats similaires (données en cours de publication) : à partir de 15 donneurs différents, ils ont obtenus 9 clones exprimant NPM1-ALK et proliférant quand les cellules primaires de départ mourraient, mais un unique clone a été réellement immortalisé. Leurs transductions sont réalisées à partir de 10 millions de cellules par donneur avec une efficacité de transduction supérieure à 50%. Ainsi à partir de 75 millions de cellules exprimant NPM1-ALK initialement, ils ont obtenu un seul clone immortalisé, soit une efficacité de $1,3 \cdot 10^{-8}$. Cette faible efficacité peut s'expliquer par le dosage de NPM1-ALK dans la cellule, car une régulation fine du niveau d'expression est nécessaire, une trop forte expression devenant toxique (Ceccon et al., 2016). Cependant, il est aussi possible que la protéine de fusion seule peine à récapituler la maladie, et que les modifications structurelles dues à la formation de la translocation ainsi que l'éventuelle haplo-insuffisance de NPM1 soient nécessaires à la transformation. Afin de répondre à ces questions nous avons utilisé notre stratégie d'induction de la translocation par CRISPR/Cas9 dans des cellules T de donneurs sains de l'EFS (Etablissement du Sang Français).

Nous avons isolé des PBMC (Cellules mononucléées sanguines périphériques). Les cellules exprimant le marqueur CD4 ont ensuite été triées et activées via CD3/CD28 afin d'induire leur prolifération (**Figure 1**). Nous avons choisi d'induire la translocation dans des cellules CD4+ car il a été montré que 65% des tumeurs exprimaient ce marqueur (Juco et al., 2003). Après cinq jours d'activation les cellules ont été transfectées avec la Cas 9 sous forme protéique et les guides sous forme ARN, dirigés contre *NPM1* et *ALK*, puis mises en culture en présence d'IL2 pour garder l'activation des cellules. Un guide contre *P53* a également été ajouté dans la moitié des expériences afin d'aider à une éventuelle transformation cellulaire *in vitro*. Des transfections contrôles avec uniquement le guide ARN *NPM1*, *ALK* ou *P53* sont également réalisées pour chaque expérience. Ces contrôles permettent d'une part de vérifier qu'une mutation simple des gènes *NPM1*, *ALK* ou *P53* ne permet pas l'immortalisation cellulaire, mais surtout de s'affranchir des éventuels effets off-target des ARN guides. Les transfections sont réalisées en triplicat pour chaque donneur.

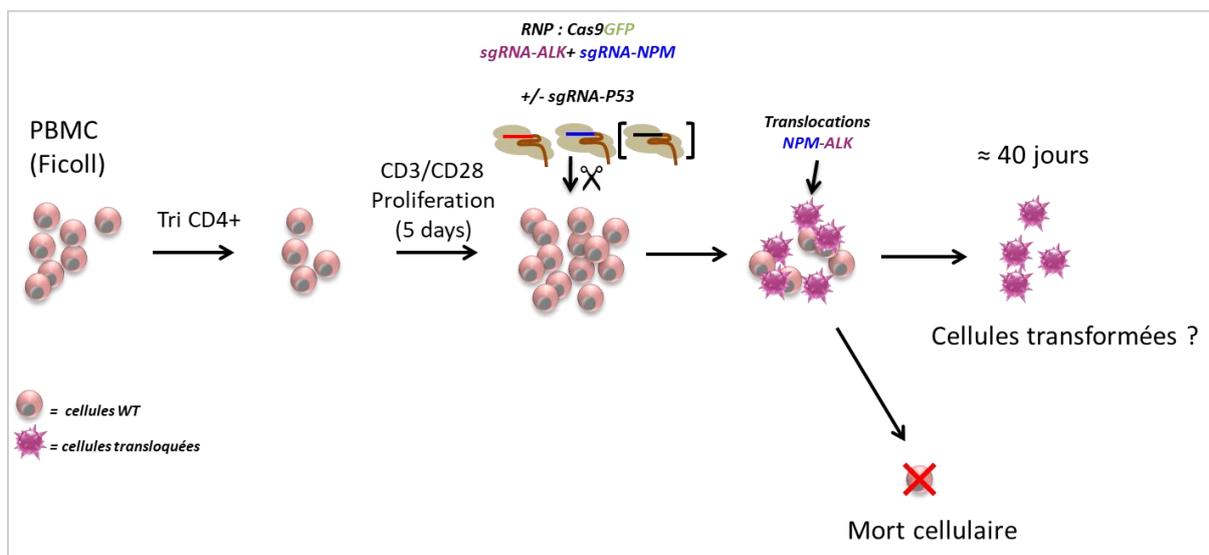


Figure 1. : Schéma des expériences réalisées sur les cellules T CD4+ triées à partir de sang d'individus sains. Les cellules sont activées via CD3/CD28 pendant 5 jours, puis transfectées avec la protéine Cas9 et les guides ARN. Les cellules sont ensuite maintenues en culture environ 40 jours. Passé ce délai les cellules WT meurent et on peut observer la survie de certains clones transloqués.

L'efficacité d'induction de translocation est vérifiée à cinq jours post-transfection pour tous les donneurs (**Figure 2**). Une bonne fréquence de translocation est observée ($f > 1/500$ cellules transloquées, avec une fréquence identique entre les donneurs). Différentes jonctions sont détectables en PCR pour l'ensemble des donneurs 5 jours post-transfection.

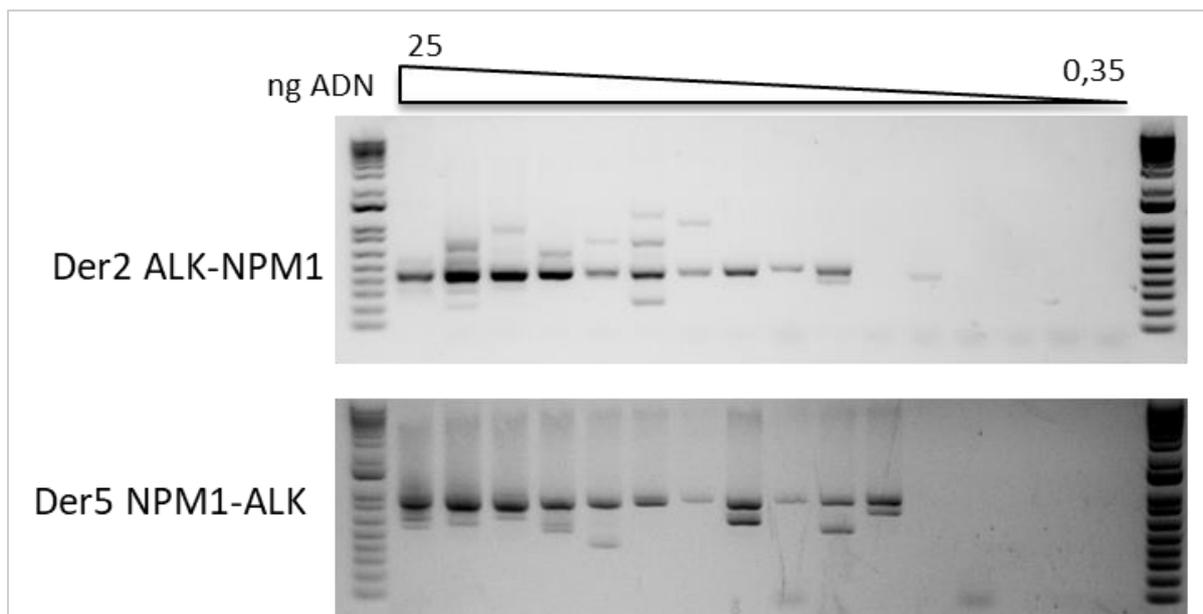


Figure 2. : Détection de la translocation par PCR 5 jours après transfection pour les 2 dérivatifs (Der2 et Der5). Les PCR sont réalisées sur des dilutions d'ADN en duplicats de 25 à 0,35 ng. Les différentes bandes observées correspondent à différentes jonctions porteuse d'insertions/délétions.

Les expériences ont été réalisées à partir de 17 donneurs sains différents. Après un mois de culture, on observe une mortalité cellulaire massive ce qui est communément observé pour des cellules T activées en culture. Cependant quelques « pools » de cellules semblent émerger dans des transfections réalisées à partir de 5 donneurs différents (**Figure 3**). En les caractérisant, nous avons montré que les cellules qui prolifèrent sont toutes porteuses de la translocation (**Figure 4A**). Parmi ces expériences, seules 5 transfections ont donné naissance à des cellules immortalisées (nous les avons gardées en culture plusieurs mois) (**Figure 3**). De façon surprenante, nous avons même obtenu des cellules immortalisées dans toutes les transfections effectuées à partir d'un donneur particulier (Donneur 5), ce qui est a priori fort improbable. Nous avons réalisé une première caractérisation des cellules immortalisées. La translocation a pu être détectée dans l'ensemble des cellules par PCR et FISH (**Figure 5A**), ainsi que l'expression de la protéine de fusion par Western-Blot (**Figure 5B**). De manière surprenante, pour les donneurs D4 et D5 plusieurs jonctions de translocation ont pu être détectées à partir d'une même transfection, indiquant une origine non clonale des cellules transformées (**Figure 4B**). Parmi les clones immortalisés, seul un présente une mutation dans le gène *P53* avec une perte de l'expression la protéine P53 (**Figure 3**). Il semble donc que la perte de ce gène ne soit pas nécessaire à l'immortalisation cellulaire comme c'est le cas chez

les patients où elle est rarement mutée. En effet, le gène P53 est muté dans moins de 10% des tumeurs (Cesarman et al., 1993; Rassidakis et al., 2005).

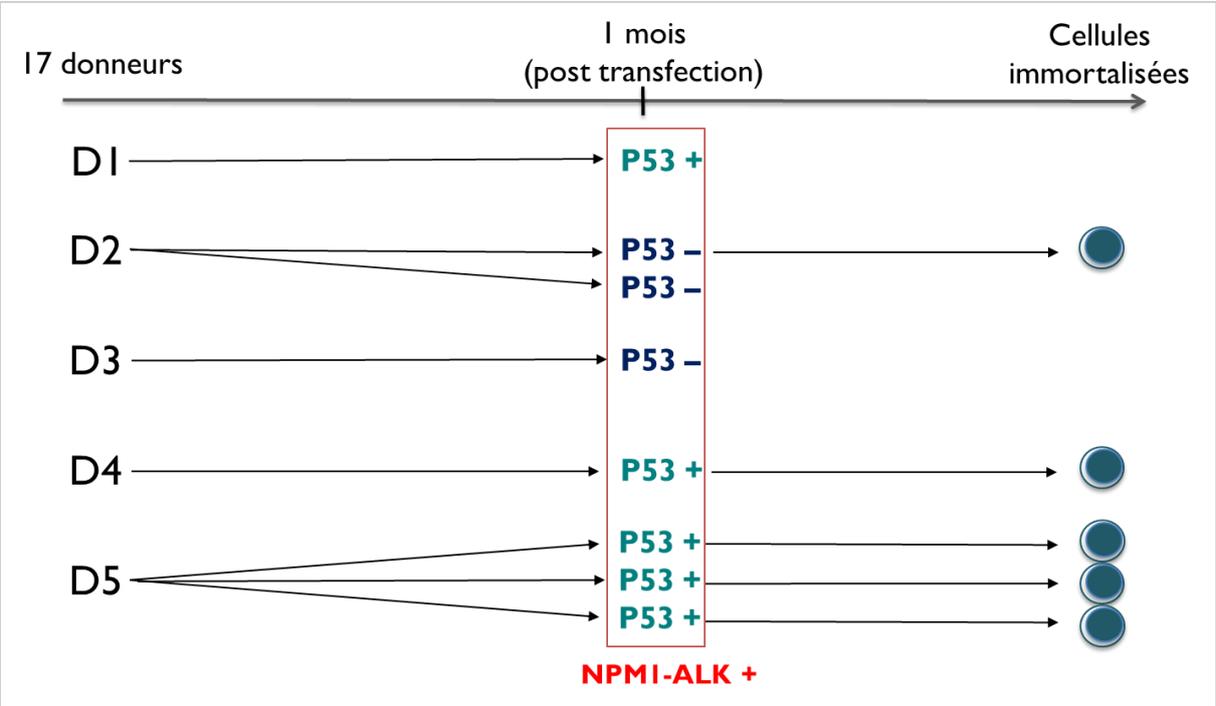


Figure 3. : Schéma des expériences ayant permis l'isolation de cellules immortalisées. Sur 17 donneurs seuls 5 ont permis d'obtenir des cellules proliférant 1 mois post transfection, transloquées NPM1-ALK. Parmi les cellules survivantes certaines étaient mutées P53 (P53 -), d'autres non (P53 +). Seuls 3 donneurs ont permis d'obtenir des cellules immortalisées.

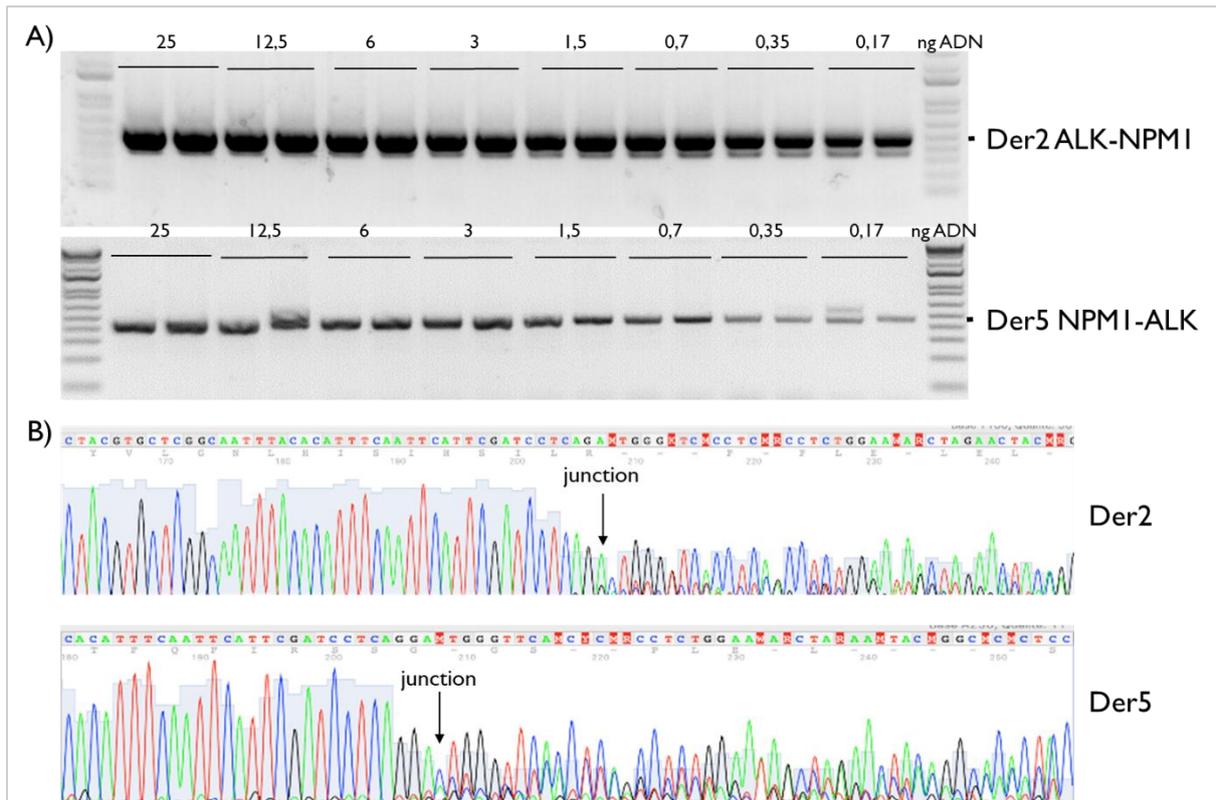


Figure 4. : A) Détection de la translocation par PCR dans les cellules immortalisées pour les 2 dérivatifs (Der2 et Der5) (Exemple du donneur 5, résultats représentatifs des cellules immortalisées pour les donneurs 4 et 5). Les PCR sont réalisées sur des dilutions d'ADN en duplicats de 25 à 0.17 ng. Les différentes bandes observées correspondent à différentes jonctions porteuse d'insertions/délétions. B) Séquençage des jonctions Der2 et Der5. Le point de jonction est représenté par une flèche. On peut observer la présence de plusieurs séquences correspondant aux différentes jonctions avec insertions/délétions/mutations.

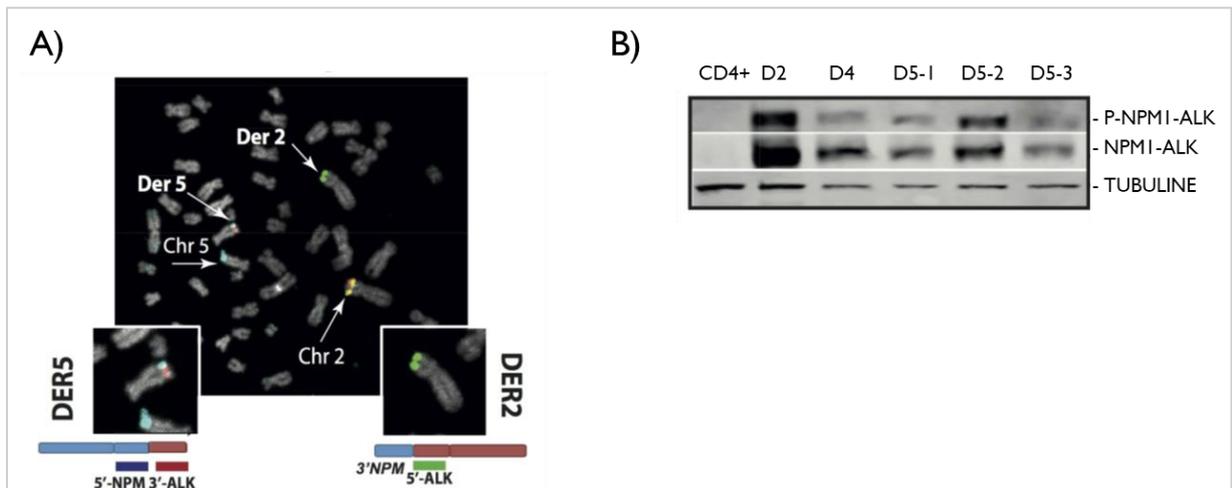


Figure 5. : A) Détection de la translocation par FISH sur étalements de métaphases. B) Détection de la protéine de fusion NPM1-ALK par western-blot

Ces résultats préliminaires ont permis de montrer qu'il était possible d'immortaliser des cellules T CD4+ matures par induction de la translocation t(2,5) à l'aide du système CRISPR/Cas9. L'efficacité d'immortalisation est à priori beaucoup plus importante que dans les modèles de surexpression. En effet, nous avons transfecté $400\,000 \times 3 = 1\,200\,000$ cellules par donneur, soit $17 \times 1\,200\,000 = 20\,400\,000$ cellules au total. L'efficacité d'induction de translocation étant de $1/500$ nous avons obtenu $20\,400\,000/500 = 40\,800$ cellules transloquées. Parmi ces cellules, nous avons obtenu des clones à partir de 3 donneurs différents soit une efficacité de $3/40\,800 = 7.10^{-5}$. Notre système semble donc être 1 000 fois plus efficace que la simple surexpression de la kinase de fusion. Plusieurs hypothèses peuvent expliquer cette différence : le dosage de NPM1-ALK dont la trop forte surexpression est toxique (Ceccon et al., 2016), l'éventuelle haplo-insuffisance de NPM1, ou encore plus globalement les modifications chromosomiques induites par la formation de la translocation et la présence du second chromosome dérivatif.

Nous avons également observé un effet donneur spécifique, avec l'émergence de plusieurs clones pour un même donneur (nous avons obtenu jusqu'à 9 clones différents pour un donneur alors que l'efficacité de formation de la translocation était identique aux autres donneurs). Certains donneurs semblent donc plus « permissifs » à la transformation par la translocation. Ceci pourrait s'expliquer par un contexte génétique particulier (préexistence de mutations par exemple), ou encore par une plus grande abondance d'un type cellulaire plus apte à la transformation.

Travail en cours : Une analyse génétique des gènes fréquemment mutés dans les hémopathies est en cours de réalisation en collaboration avec le Dr Asnafi afin d'identifier d'éventuelles mutations participant à l'oncogenèse chez les donneurs pour lesquels nous obtenons plusieurs clones transformés. Une analyse des réarrangements du TCR sera également réalisée afin de savoir si des réarrangements anormaux sont détectables comme cela est parfois le cas dans les ALCL (Malcolm et al., 2016; O'Connor et al., 1987). Cependant nous avons sélectionné des cellules T matures ayant déjà réalisé leur réarrangement TCR et leur sélection, il est donc peu probable que des anomalies soient présentes. Une caractérisation des marqueurs de surface est également en cours dans ces cellules, avec notamment la détection des marqueurs CD30, CD3, CD4 et CD25. Enfin les cellules immortalisées seront injectées chez la souris par nos

collaborateurs dans le laboratoire du Dr. Meggetto et T. Mercher afin de déterminer le potentiel tumoral des cellules transloquées et éventuellement le phénotype des tumeurs.

Matériel et méthodes

Culture cellulaire

Les cellules PBMC et CD4-NPM1-ALK sont cultivées en RPMI additionné de 20% de serum inactivé (Gibco 26140079 heated 30 min at 55°C) et de Penicilline-Streptomycine (Gibco 15140122).

Isolation des PBMC

Les PBMC sont isolés à partir du sang de donneur sain, en utilisant le système SepMate™-50 PBMC Isolation system (StemCell 15450) selon le protocole du fabricant.

Tri des CD4+

Directement après isolation, un maximum de 10^9 PBMC est trié à l'aide du système MACS® LS Column (Myltenyi Biotec 130-042-401). Les cellules sont rincées deux fois au PBE (PBS additionné de 0.5% BSA et 0,2 mM EDTA). Les cellules sont incubées avec un anticorps primaire antiCD4-PE dilué à 1/10 (Myltenyi Biotec 130-113-214) durant 20 min à 4°C dans le noir. Les cellules sont rincées deux fois au PBE. Les cellules sont incubées avec un anticorps antiPE couplé à des micro-billes dilué à 1/20 (Myltenyi Biotec 130-048-801) durant 20 min à 4°C dans le noir. Les cellules sont passées sur les colonnes, équilibrées avec 3x 2 mL de PBE sur le séparateur QuadroM ACS™ Separator. Les colonnes sont rincées avec 3x 2 mL de PBE. L'élution est réalisée avec 5 mL de PBE en utilisant le piston. Les cellules sont centrifugées à 300 g 8 min et resuspendues à 3.10^6 cells/mL dans du RPMI additionné de 20% de SVF inactivé (Gibco 26140079 chauffé 30 min à 55°C) et de Penicilline-Streptomycine (Gibco 15140122).

Activation CD3/CD28

Une plaque 12 puits est coatée avec un anti-CD3 OKT3 (16-0037-85 eBioscience) dilué à 1/200 dans du PBS. 300 uL sont utilisés par puits. La plaque est incubée 2h à 37°C dans du parafilm. La plaque est lavée deux fois avec du PBS. Les cellules sontensemencées à 3.10^6 cells/mL dans

du RPMI additionné de 20% de SVF inactivé, d'anti-CD28 (16-0289-85 eBioscience) dilué à 1/500, et de Penicilline-Streptomycine (Gibco 15140122).

Transfection RNP

Une semaine après l'activation CD3/CD28 les cellules sont transfectées avec la protéine Cas9 et les gRNA à l'aide du nucleofecteur Lonza 4D selon le protocole du fabricant. Un ratio Cas:gRNA de 1:2 est utilisé. Après transfection les cellules sont cultivées dans du RPMI additionné de 20% de SVF inactivé, de 20U/mL d'IL2 humaine (Proleukin) et de Penicilline-Streptomycine (Gibco 15140122).

Extraction d'ADN

L'ADN est extrait à l'aide du kit « Eza Tissue DNA kit » (D3396-01 Omega Bio-Tek) selon les recommandations du fabricant.

Western blot

Une extraction de protéines totales est réalisée à l'aide de tampon de lyse (50 mM Tris-HCl à pH 7.4, 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 1 mM EDTA, et 1 mM DTT), additionné de « Complete cocktail protease inhibitor tablets » (Roche) et « Halt Protease and Phosphatase Inhibitor Cocktail » (Thermo Fisher Scientific). 30 µg d'extrait protéique sont chargés sur un gel acrylamide 8% (w/v) Tris-HCl SDS PAGE. Les protéines sont transférées sur une membrane. Les protéines d'intérêt sont détectées à l'aide des anticorps Phospho-ALK (Tyr1604), ALK (31F12), de Cell Signaling. Un anticorps Alpha-Tubulin (Sigma-Aldrich) est utilisé comme témoin de charge. Phospho-ALK est utilisé à la dilution 1/200. Les autres anticorps sont dilués selon les recommandations du fabricant. Un anticorps secondaire IRDye 800CW est utilisé, et la révélation est réalisée à l'aide du système « Odyssey® CLx Imaging System » (LI-COR).

FISH (Fluorescence *in situ* hybridization)

L'hybridation est réalisée sur des étalements de métaphases. Les gènes *ALK* and *NPM1* sont marqués respectivement avec une sonde « ALK Breakapart LPS 019 » (Cytocell) et une sonde *NPM1* faite maison à partir du BAC RP11-546B8. Les étalements de métaphases, le marquage, les lavages et la détection de la fluorescence sont réalisés selon les procédures standards. Les images sont réalisées à l'aide du système Zeiss Spinning Disk.

DISCUSSION/PERSPECTIVES

I. Induction de la translocation t(2,5) dans des lignées cellulaires humaines et murines

Au cours de ces travaux, nous sommes parvenus à induire la translocation t(2,5) dans des lignées cellulaires humaines (RPE-1) et murines (BA/F3) à l'aide du système CRISPR/Cas9 dont l'efficacité avait déjà été prouvée au laboratoire (Ghezraoui et al., 2014; Piganeau et al., 2013; Piganeau et al., 2016). Cette stratégie nous a permis d'induire l'expression de la protéine de fusion NPM1-ALK sous contrôle du promoteur endogène de NPM1, conservant ainsi un niveau d'expression physiologique. L'expression de la protéine endogène NPM1 est quant à elle divisée par deux. Nous avons également conservé l'intégralité du contexte chromosomique avec la détection des deux chromosomes dérivatifs par FISH. L'utilisation de la lignée murine BA/F3, bien que très remaniée, nous a permis de confirmer le potentiel oncogénique de la kinase de fusion par levée de la dépendance à l'IL3 comme cela avait été montré précédemment dans des modèles de surexpression (Bai et al., 1998). Dans cette lignée, nous avons pu montrer l'activation des voies de signalisation ERK, AKT et STAT3, associées à l'ALCL chez les patients (Amin et al., 2004; Bai et al., 2000; Chiarle et al., 2005; Marzec et al., 2007; Slupianek et al., 2001; Zamo et al., 2002). La lignée humaine RPE-1 a, quant à elle, été choisie pour son caryotype très peu remanié, en faisant un modèle cellulaire de choix pour l'étude mécanistique des translocations chromosomiques. Les voies ERK et AKT étant déjà activées au niveau basal dans ces cellules, nous avons uniquement pu mettre en évidence l'activation de STAT3 par la protéine de fusion NPM1-ALK. Cependant, nous avons montré que nos modèles cellulaires récapitulaient la signature transcriptionnelle des patients. Nous avons également montré qu'en conditions normales de culture aucun avantage sélectif n'était apporté par la translocation. Ceci peut être expliqué par le type cellulaire (cellules adhérentes de rétine) très éloigné de la cellule d'origine du cancer (lignée hématopoïétique T). Cependant nous avons pu mettre en évidence un avantage prolifératif des cellules transloquées lors d'une réduction de la concentration des facteurs de croissance dans le milieu (culture en 2% sérum), mais également un avantage dans la migration cellulaire par rapport aux cellules sauvages. Ces résultats indiquent que la kinase est capable de prendre le relais lorsque certaines voies de prolifération/survie ne sont plus stimulées par des facteurs extérieurs. Ces différents résultats prouvent ainsi que nos modèles d'étude, bien qu'éloignés des cellules de patients, sont

pertinents pour l'étude de la translocation chromosomique t(2,5) menant à l'expression de la kinase de fusion NPM1-ALK.

II. Etude des ARN circulaires de fusion

Les travaux réalisés par le groupe du Dr Pandolfi (Guarnerio et al., 2016) pendant ma thèse ont montré que des ARN circulaires de fusion pouvaient être formés lors de la présence de translocations dans des cellules tumorales. Dans cette étude, les auteurs ont montré que ces ARN circulaires pouvaient participer à l'oncogenèse en favorisant par exemple la prolifération cellulaire et la résistance aux thérapies médicamenteuses. A l'aide de nos modèles cellulaires nous avons pu également mettre en évidence la formation d'ARN circulaires de fusion (f-circRNA) dans le cas de la translocation *NPM1-ALK*. Des f-circRNA ont pu être détectés à la fois dans les modèles murins et humains. Certains f-circRNA ont été détectés dans tous les clones étudiés (2 f-circRNA chez la souris et 4 chez l'Homme). Les f-circRNA détectés dans les clones RPE-1 ont également pu être identifiés dans des lignées cellulaires de patients ALCL (SUDHL1 et SUPM2), mais aucun f-circRNA n'a été détecté dans les modèles de surexpression. Ces différentes observations montrent d'une part que ces f-circRNA pourraient jouer un rôle dans l'oncogenèse, car ils existent dans les lignées tumorales, et d'autre part qu'il est important d'utiliser des modèles d'études récapitulant au mieux le contexte cellulaire des patients, car dans le cas présent les modèles de surexpression ne permettent pas la formation de f-circRNA en raison de l'absence d'épissage de l'ARNm. En effet, il a été montré que les ARN circulaires étaient dépendants de la machinerie d'épissage (Ashwal-Fluss et al., 2014). De plus, nous avons pu montrer que le niveau d'expression des f-circRNA était corrélé au niveau d'expression de l'ARNm correspondant *NPM1-ALK*. Ce résultat est intéressant car notre système permet de conserver une régulation endogène du transcrit *NPM1-ALK*. Cette stratégie doit également permettre de conserver la régulation des ARN de fusion nouvellement formés. Enfin et de manière intéressante nous avons également pu mettre en évidence un f-circRNA commun entre l'Homme et la Souris (liant les exons 2 de NPM1 et 22 d'ALK). La conservation inter-espèce de cet f-circRNA pourrait indiquer un rôle potentiel dans la cellule et dans l'oncogenèse. La conservation des ARN circulaires avait notamment été montrée par l'étude de Jeck et al (Jeck et al., 2013).

A. Quels sont les mécanismes d'action et le rôle de ces ARN circulaires de fusion ?

Pour d'aller plus loin, il serait intéressant d'étudier le mécanisme d'action de ces f-circRNA et leur rôle dans l'oncogenèse. Afin de répondre à ces questions des modèles de surexpression des f-circRNA ont été générés dans la lignée RPE-1 au laboratoire. Dans ces cellules la protéine de fusion est absente, et chaque f-circRNA est exprimé séparément, ainsi il est possible d'étudier le rôle spécifique d'un f-circRNA particulier. Ces nouveaux modèles pourront permettre d'étudier les cibles éventuelles de ces f-circRNA, en supposant qu'ils puissent agir comme éponges à micro-ARN notamment. En effet ce mécanisme d'action est aujourd'hui le mieux décrit pour les ARN circulaires et constitue une première piste à explorer (Hansen et al., 2013; Memczak et al., 2013). Il sera possible de déterminer *in silico* les micro-ARN cibles potentiels, en utilisant par exemple des logiciels tels que MIRanda (Enright et al., 2003). Les ARNm cibles des micro-ARN sont partiellement connues, il sera alors possible d'analyser l'expression de ces cibles en présence ou en absence de f-circRNA. Afin de confirmer ou infirmer l'effet éponge à micro-ARN des f-circRNA il sera également possible d'étudier la liaison des f-circRNA au complexe RISC. Il est notamment possible d'étudier la liaison à la protéine AGO2, en utilisant une technique d'immunoprécipitation associée à un test luciférase afin de déterminer les micro-ARN cibles, comme cela a été décrit par Li et al. (Li et al., 2018).

Les ARN circulaires possèdent des rôles variés dans l'oncogenèse. Il a par exemple été montré que certains ARN circulaires activaient la prolifération, la migration ou la survie cellulaire (Guarnerio et al., 2016; Meng et al., 2018; Qu et al., 2019; Zhang et al., 2019a; Zheng et al., 2016), d'autres favorisent l'angiogenèse (Zhong et al., 2017), ou encore participent à la résistance aux thérapies (Guarnerio et al., 2016). Dans certains cas, il a également été montré que les ARN circulaires pouvaient agir comme suppresseurs de tumeur (Li et al., 2017a; Wang et al., 2018; Zhang et al., 2019a). A l'aide de nos modèles de surexpression, il sera possible d'étudier l'impact des ARN circulaires sur la prolifération cellulaire ou encore la migration cellulaire à l'aide de tests classiques que nous avons déjà utilisés pour étudier le rôle de la translocation *NPM1-ALK*.

Ces modèles de surexpression présentent cependant certaines limites. Tout d'abord, il est possible que les ARN circulaires aient un effet synergique avec la protéine de fusion, comme cela a été montré par le groupe du Dr. Pandolfi pour la fusion MLL-AF9 (Guarnerio et al., 2016). La simple surexpression d'un seul ARN circulaire ne sera peut-être pas suffisante à l'observation d'un effet. De plus, nos modèles sont réalisés dans des cellules RPE-1, très éloignées de la cellule d'origine du cancer (lignée hématopoïétique). Ces cellules sont immortalisées et indépendantes de NPM1-ALK, ce qui peut rendre difficile l'étude des mécanismes d'oncogenèse et du rôle des ARN circulaires. Un modèle plus proche de la cellule tumorale serait davantage pertinent comme celui des cellules T que nous avons maintenant générés au laboratoire (CF partie « Vers un modèle cellulaire plus physiologique »).

B. Les ARN circulaires de fusion identifiés sont-ils présents chez les patients ?

Lors de notre analyse, nous avons utilisé des lignées ALCL comme contrôles. Nous avons montré que les ARN circulaires de fusion étaient présents dans ces cellules. Il serait également intéressant de comparer nos résultats à des analyses réalisées dans des cellules primaires de patients. Nous pourrions par exemple tenter de détecter dans ces cellules les 4 f-circRNA exprimés dans l'ensemble des clones transloqués humains et dans les lignées tumorales. De plus, nous pourrions envisager un lien entre niveau d'expression des f-circRNA et grade tumoral, comme cela a pu être démontré dans d'autres cancers, tels que les leucémies myéloïdes aiguës ou encore le cancer du poumon (Hirsch et al., 2017; Huang et al., 2019). Ces f-circRNA pourraient alors servir de biomarqueurs d'agressivité.

C. Identification des ARN circulaires et problèmes d'artefacts de PCR

Lors de notre analyse nous avons dû faire face à des problèmes d'artefacts de PCR. En effet lors de l'amplification des ARN circulaires par PCR à l'aide d'amorces divergentes nous avons obtenu un grand nombre de séquences porteuses d'homologies allant de 2 à 8 nucléotides au niveau de la jonction théorique de circularisation. Nous avons dans un premier temps pensé qu'il s'agissait d'une classe d'ARN circulaires particulière. Cependant après avoir réalisé des

analyses par transcription *in vitro* nous nous sommes aperçus que ces produits de PCR étaient formés à partir de l'ARN linéaire seul. Il s'agit donc d'artefacts de PCR produit par un saut de matrice en présence de séquences homologues sur un même ADN ou par recombinaison entre différentes molécules d'ADN. Ce mécanisme est connu depuis longtemps, il avait déjà été montré en 1995 pour l'ADN Polymérase I *Thermus aquaticus* avec des homologies de 15 nucléotides (Odelberg et al., 1995). Plus récemment, il a été montré que ce phénomène pouvait perturber le séquençage de nouvelle génération (Potapov and Ong, 2017; Shafikhani, 2002). Le taux d'erreur est variable d'une enzyme à une autre, allant de 10^{-3} à 1%. Ce phénomène a été montré pour des homologies de minimum 5 nucléotides, dans notre cas nous avons observé un saut de matrice avec seulement 2 nucléotides. Dans le cas de l'utilisation d'amorces divergentes, le phénomène est peut-être amplifié par la faible abondance des ARN circulaires à amplifier. Il est essentiel de garder en tête ce problème d'amplification lors de l'analyse d'ARN circulaires. Notamment, certaines études ont publié des séquences d'ARN circulaires de fusion pour *EWSR1-FLI1* (Guarnerio et al., 2016) ou plus récemment pour *EML4-ALK* (Tan et al., 2018) qui présentent des caractéristiques d'artefacts de PCR et non de f-circRNA biologiques qui résultent d'un épissage inverse.

III. Analyse du timing de réplication et du profil transcriptionnel

Au cours de cette étude nous nous sommes intéressés à l'impact de la translocation t(2,5) sur le timing de réplication. En effet, il a été montré que le timing de réplication était modifié dans les cellules cancéreuses, et notamment au niveau des aberrations chromosomiques (Ryba et al., 2012). De plus, une étude récente a démontré que les translocations chromosomiques avaient un impact local sur l'organisation en TAD de la chromatine (Dixon et al., 2018). Or le timing de réplication est directement lié à cette organisation (Pope et al., 2014; Ryba et al., 2010; Yaffe et al., 2010). Notre modèle d'étude dans la lignée humaine RPE-1 nous a permis d'étudier les modifications du timing de réplication induites suite à la formation de la translocation t(2,5). Nous avons choisi de distinguer le rôle de la protéine de fusion NPM1-ALK d'une part, et celui de la translocation chromosomique elle-même d'autre part. Pour cela nous avons généré des cellules transloquées dans lesquelles nous avons inactivé secondairement

la protéine de fusion. Cependant, dans nos modèles cellulaires l'analyse des résultats est rendue difficile par la subsistance d'un allèle non transloqué. En effet, seuls un chromosome 2 et un chromosome 5 sont transloqués. Dans nos analyses de séquençage, notamment lors de l'analyse du timing de réplication, il est difficile de différencier les données provenant de l'allèle transloqué et celles provenant de l'allèle sauvage. Ainsi l'effet observé est en réalité une moyenne sur les deux chromosomes. Afin de déterminer le timing de réplication réel des chromosomes transloqués nous avons induit une seconde translocation dans les cellules. *NPM1* étant essentiel à la survie cellulaire, nous avons décalé le second point de cassure sur le chromosome 5 dans la région intergénique juste en aval de *NPM1* qui se trouve dans le même domaine de réplication dans les cellules de départ RPE-1. Grâce à cette stratégie nous conservons l'expression de *NPM1* sous contrôle de son promoteur. *ALK* n'étant pas exprimé dans nos cellules nous avons conservé le même point de cassure sur le chromosome 2. Nous avons ainsi pu générer des cellules que nous appelons double-transloquées et qui portent 2 chromosomes der2 et deux chromosomes der5. Nous avons fait le choix de générer la seconde translocation dans les clones transloqués n'exprimant pas la protéine de fusion afin de pouvoir déterminer le rôle de la translocation indépendamment de celui de la kinase de fusion. Nous avons ainsi pu comparer le profil transcriptionnel ainsi que le timing de réplication dans des cellules WT, dans des cellules transloquées exprimant *NPM1-ALK* et dans des cellules transloquées n'exprimant pas la protéine de fusion, porteuses d'une ou de deux translocations.

A. Une signature transcriptionnelle proche de celle des patients ALCL

Nous nous sommes intéressés au profil transcriptionnel des cellules transloquées. Nous avons pu observer une signature similaire à celle des patients ALCL dans nos cellules transloquées exprimant *NPM1-ALK*, aussi bien au niveau des gènes surexprimés que réprimés. Même si dans les cellules transloquées, en absence de la protéine de fusion, nous avons trouvé très peu de gènes surexprimés, une signature similaire à celle des patients a pu être retrouvée pour les gènes réprimés. Cette signature est même davantage conservée dans les cellules double-transloquées. Ces résultats mettent pour la 1^{ère} fois en évidence un rôle de la translocation chromosomique elle-même dans la signature transcriptomique et

potentiellement oncogénique retrouvée chez les patients. Dans le cas de la translocation t(2,5), la protéine de fusion semble jouer principalement un rôle activateur de la transcription (ce qui est normale pour une kinase), alors que la translocation induit majoritairement une répression génique. Ce résultat est d'autant plus intéressant que les études transcriptomiques réalisées dans des cellules de patients se sont principalement concentrées sur les gènes suractivés (Eckerle et al., 2009; Lamant et al., 2007; Piva et al., 2010; Thompson et al., 2005), ou sur les gènes réprimés spécifiquement par ALK, avec notamment des techniques de shRNA. Lamant et al. ont simplement conclu à une répression de facteurs de transcription. Grâce à nos modèles, nous avons montré que les gènes réprimés dans nos cellules transloquées NPM1-ALK KO pouvaient aussi être associés au processus d'adhésion cellulaire, processus directement lié à l'oncogenèse. Nos travaux pourraient mener, à plus long terme, à la découverte de nouvelles cibles thérapeutiques intéressantes.

Ces résultats mettent une nouvelle fois en avant l'importance du modèle utilisé. En effet, nous avons observé un effet très net de la translocation chromosomique indépendamment de la protéine de fusion. Or cette signature transcriptionnelle ne peut être observée dans des modèles de surexpression.

B. Comment expliquer les modifications de timing de réplication à la jonction ?

Nous avons observé un changement local du timing de réplication dans l'ensemble des cellules transloquées indépendamment de l'expression de la kinase de fusion. En effet, les gènes *NPM1* et *ALK* étaient initialement localisés dans des régions au timing différent, la relocalisation des chromosomes transloqués s'accompagne d'un « lissage » du timing de réplication au niveau des deux jonctions, évitant l'apparition d'un changement abrupt de timing dans cette région. Bien que l'analyse intégrale du timing de réplication n'ait pas encore été réalisée dans les cellules double-transloquées, nous avons pu observer un doublement de l'effet sur le dérivatif 2 au niveau de la portion 5' d'ALK. Sur le dérivatif 2, on observe un retard de réplication de la partie 5' d'ALK, alors que sur le dérivatif 5 on observe un avancement de la réplication de la partie 5' de NPM1. Nous attendons les résultats du Repli-seq pour savoir si

les der2 et 5 sont globalement modifiés. Dans notre cas, il ne semble pas y avoir de biais en faveur d'une invasion précoce vers tardif comme cela avait été montré par Yaffe et al. au cours de l'évolution (Yaffe et al., 2010). Cependant, dans notre modèle, ce sont les 2 portions de chromosomes porteuses des centromères des chromosomes dérivatifs qui sont amenées à changer de timing de réplication au niveau de la jonction. Il serait intéressant d'étudier d'autres modèles de translocation afin de voir si cette observation est généralisée ce qui est possible dans notre équipe sur d'autres modèles tels que celui du sarcome d'Ewing. De plus, nous pourrions étudier l'organisation locale des TAD avant et après translocation afin de voir s'il existe une corrélation locale entre modification des contacts et changements de timing de réplication. Dixon et al. ont par exemple montré qu'une réorganisation des contacts en *cis* avait lieu suite à l'apparition de réarrangements chromosomiques (Dixon et al., 2018).

C. La question de la synchronicité de réplication des chromosomes homologues

Dans les cellules double-transloquées, nous allons ensuite étudier les modifications du timing de réplication sur l'ensemble du génome, et en particulier sur les chromosomes dérivatifs. Il a été montré qu'il existait une synchronisation de la réplication entre chromosomes homologues pour les loci dont l'expression est bi-allélique, alors que les loci dont l'expression est mono-allélique peuvent présenter une asynchronie (Boggs and Chinault, 1997; Ensminger and Chess, 2004; Kitsberg et al., 1993). De plus, une étude récente a montré que la synchronisation des loci homologues était plus importante lorsque ces loci étaient répliqués précocement (Duriez et al., 2019). La réplication tardive semble plus stochastique, excepté pour certains loci. Les auteurs ont montré que la localisation nucléaire était un facteur déterminant. Une meilleure synchronisation est observée lorsque les 2 loci sont situés dans la même région nucléaire. Lors de la présence de translocations, il est possible que la localisation nucléaire et la synchronisation des chromosomes transloqués soient perturbées. Dans nos modèles transloqués il serait possible d'étudier ces deux paramètres en présence d'une ou de deux translocation. On pourrait imaginer une meilleure synchronisation lors de la présence de deux translocations car les chromosomes « homologues » sont alors identiques.

D. Comment expliquer les modifications de timing de réplication observées à l'échelle du génome ?

De manière plus surprenante nous avons également observé des modifications significatives du timing de réplication à l'échelle du génome entier. Notamment dans les cellules transloquées exprimant NPM1-ALK, 22% du génome possède un timing de réplication modifié. L'effet semble principalement médié par l'expression de la kinase de fusion. En effet, les cellules transloquées n'exprimant pas NPM1-ALK présentent un profil de réplication intermédiaire plus proche des cellules WT que des cellules transloquées exprimant NPM1-ALK.

Nous avons par la suite tenté d'établir un lien entre profil transcriptionnel et timing de réplication. Pour cela nous avons comparé les modifications d'expression génique observées en RNAseq avec les modifications de timing de réplication calculées au niveau du promoteur de chaque gène. Nous avons ainsi montré une corrélation entre modification d'expression et changement de timing de réplication. Les gènes dont la réplication est avancée sont globalement davantage exprimés et ceux dont la réplication est retardée sont plutôt réprimés. Bien que les modifications observées dans les cellules transloquées n'exprimant pas NPM1-ALK soient moins importantes, aussi bien en timing de réplication qu'en expression génique, la corrélation est retrouvée également dans ces cellules. Ces résultats apportent de nouveaux arguments en faveur d'un lien entre timing de réplication et expression génique. De plus, nous avons montré que les gènes suractivés dans les cellules de patients ALCL présentaient un avancement du timing de réplication dans nos modèles.

Ces différents résultats semblent indiquer que l'activation de la kinase de fusion joue ainsi un rôle activateur de la transcription, qui s'accompagne de modifications de réplication. Il est difficile de savoir si l'avancement du timing de réplication est une cause ou une conséquence de l'activation transcriptionnelle. En effet, notre système d'étude permet uniquement d'observer des corrélations. Cependant on peut faire l'hypothèse que les gènes sont directement activés par la kinase (cible de STAT3, ERK et AKT), même en absence de modification de timing concomitant à l'expression, et que pour ces gènes la variation du timing de réplication est une conséquence de l'activation transcriptionnelle. Pour les autres gènes il est difficile de savoir si l'expression est modifiée avant ou après le timing de réplication.

E. Lien entre timing de réplication, marques épigénétiques et organisation de la chromatine

Afin d'aller plus loin dans les analyses génomiques, il serait bien sûr intéressant d'étudier les modifications épigénétiques associées aux modifications de timing de réplication ainsi qu'aux changements transcriptionnels. En effet, il est connu qu'il existe une corrélation entre timing de réplication et les marques histones, en particuliers entre réplication précoce et marques histones activatrices (Julienne et al., 2013; Thurman et al., 2007; Yaffe et al., 2010). Une analyse en ChIP-seq pour les marques H3K4me1,2,3, H3K9ac, H3K27ac, H3K36me3, H4K20me1 pourrait être réalisée dans nos modèles cellulaires. Il serait également possible de réaliser une analyse en ATAC-seq, qui permettrait de détecter les zones ouvertes de chromatine (Buenrostro et al., 2015). Il faudrait ensuite mettre en perspective ces résultats avec ceux obtenus en RNA-seq et Répli-seq.

Enfin nous pourrions étudier le lien entre les modifications observées et l'organisation de la chromatine. En effet, le timing de réplication est fortement corrélé à l'organisation de la chromatine en compartiments A/B et TAD (Pope et al., 2014; Ryba et al., 2010; Yaffe et al., 2010). Par chance des études de HIC ont déjà été réalisées dans des cellules RPE-1 (Darrow et al., 2016). A partir de ces données nous pourrions rechercher les régions en interactions avec les loci *NPM1* et *ALK* afin de savoir si ces régions correspondent aux régions modifiées en timing de réplication. En effet, on peut émettre l'hypothèse qu'en modifiant l'organisation chromosomique au niveau de ces loci, on perturbe également les régions génomiques présentant un contact avec ces deux régions.

F. Modification du timing de réplication et différenciation cellulaire

Au cours de la différenciation cellulaire des modifications de timing de réplication ont lieu (Ryba et al., 2010). Il a notamment été montré que le timing de réplication est spécifique d'un type cellulaire et relativement bien conservé entre les espèces (Yaffe et al., 2010). Dans nos modèles de translocation, nous avons pu observer des changements globaux de réplication sur l'ensemble du génome. Il est possible que la présence de la translocation génère une

modification de l'identité cellulaire, qui pourrait expliquer les changements de timing observés. Il est notamment connu que les cellules de patients présentent un phénotype intermédiaires entre T matures et immatures (Foss et al., 1996; Hassler et al., 2016; Juco et al., 2003; Malcolm et al., 2016; Moti et al., 2015), avec une inactivation de certains facteurs T par NPM1-ALK (Ambrogio et al., 2009). La kinase pourrait être responsable de ce changement de différenciation.

Cependant, pour pouvoir envisager cette hypothèse il serait nécessaire de refaire les analyses dans des cellules proches de la cellule d'origine du cancer, afin d'être sûr que les modifications observées dans les RPE-1 ne soient pas un biais dû à l'utilisation d'une lignée trop différente de l'ALCL.

IV. Vers un modèle cellulaire plus physiologique

La plupart des résultats obtenus au cours de cette thèse ont été générés dans une lignée cellulaire modèle : la lignée RPE-1. Cette lignée présentait l'avantage d'avoir un caryotype très propre, ce qui n'est pas le cas de la plupart des lignées cellulaires, souvent dérivées de cellules cancéreuses. Ainsi les RPE-1 étaient d'excellents modèles pour étudier les mécanismes moléculaires liés aux translocations chromosomiques. Nous avons même montré que nous pouvions générer une signature transcriptionnelle proche de celle des patients ALCL par simple induction de la translocation t(2,5). Cependant afin d'étudier les mécanismes d'oncogénèse, il est nécessaire de se placer dans un contexte plus physiologique et d'induire la translocation dans la « bonne » cellule. De même, le timing de réplication est lié au type cellulaire (Ryba et al., 2010), il est donc important de confirmer nos résultats dans le « bon » type cellulaire. Pour cela nous avons choisi d'induire la translocation dans des cellules T matures issues de donneurs sains. Nos résultats préliminaires montrent qu'il est possible d'immortaliser les cellules par induction de la translocation t(2,5). Ces résultats confirment ce qui avait été observé lors de la surexpression du gène de fusion dans ces mêmes cellules (Zhang et al., 2013a).

A. Caractérisation cellulaire

La caractérisation des clones transloqués obtenus à partir de cellules T reste à effectuer. Il faudra notamment étudier les marqueurs de surface de ces cellules comme cela a été effectué par Zhang et al dans les modèles de surexpression (Zhang et al., 2013a). Dans cette étude les auteurs ont montré que leurs cellules transformées exprimaient peu le marqueur CD3 mais fortement les marqueurs CD4, CD30 et CD25 comme observé chez les patients ALCL. Ce résultat pourra être comparé à nos modèles.

B. Potentiel tumoral

Lors de nos travaux *in vitro*, nous avons montré que l'efficacité de transformation de la translocation était 1000 fois supérieure à l'efficacité de transformation de la protéine de fusion NPM1-ALK seule ((Zhang et al., 2013a) et données du Dr. Meggetto). Ceci peut s'expliquer par le dosage de NPM1-ALK qui nécessite une régulation fine et précise pour éviter sa toxicité (Ceccon et al., 2016). Il est également possible que l'haplo-insuffisance NPM1 joue un rôle direct dans le processus de transformation. En effet, ce phénomène a déjà été montré dans d'autres cancers touchant les lignées lymphoïdes et myéloïdes (Grisendi et al., 2005; Sportoletti et al., 2008). Enfin, il est possible que le contexte chromosomique soit important et joue un rôle dans l'oncogenèse. Nous avons notamment montré dans la lignée RPE-1 que l'induction de la translocation permettait la formation d'ARN circulaire de fusion. Ces ARN pourraient participer directement au processus tumoral comme cela a été décrit pour d'autres fusions, telles que MLL-AF9 et EML4-ALK (Guarnerio et al., 2016; Tan et al., 2018). Il serait intéressant d'étudier la présence de ces f-circRNA dans nos modèles de CD4+ transloqués. S'ils sont détectés dans nos cellules, il sera possible de les inactiver à l'aide de shRNA afin d'étudier leur rôle potentiel dans la survie cellulaire, la prolifération, ou encore la migration. Nous avons également montré que la translocation chromosomique jouait un rôle, indépendant de la protéine fusion, dans le profil transcriptomique et le timing de réplication. Il est possible que ce rôle soit important pour la transformation cellulaire.

Zhang et al. ont montré qu'il était possible d'induire la formation de tumeurs chez la Souris par injection de cellules CD4 surexprimant NPM1-ALK (Zhang et al., 2013a). Il est important de vérifier le potentiel oncogène de nos cellules transloquées. En effet, il avait été montré que l'utilisation du système CRISPR/Cas9 pouvait mener à des résultats très différents de la surexpression dans le cas de la translocation MLL-ENL. Alors que les modèles de surexpression étaient extrêmement transformants, l'induction de la translocation par CRISPR/Cas9 présentait un phénotype beaucoup moins agressif (Barabe et al., 2007; Reimer et al., 2017). Une injection de nos cellules transloquées chez la souris pourra nous indiquer leur potentiel tumoral. Si des tumeurs sont formées, il faudra analyser leur profil histologique afin de déterminer s'il est semblable aux ALCL.

C. Existence d'un effet donneur spécifique

Nous avons induit la translocation t(2,5) à partir de 17 donneurs différents. Seuls 5 d'entre eux ont permis d'observer une survie cellulaire un mois après transfection, lorsque les cellules sauvages cessent de proliférer et meurent. Par la suite, une réelle immortalisation des cellules a été observée pour 3 donneurs. Parmi ces 3 donneurs, 2 d'entre eux ont permis l'émergence de plusieurs clones transloqués, avec des séquences différentes au niveau des points de jonction. Ces observations semblent donc indiquer qu'il existe un effet donneur spécifique. La translocation pourrait donner un avantage sélectif uniquement dans un fond génétique particulier. Il serait intéressant d'étudier la présence de mutations préexistantes chez les donneurs les plus permissifs. Dans un premier temps, une analyse des gènes fréquemment mutés dans les hémopathies pourrait être un bon point de départ. Il serait également possible de réaliser une analyse d'exome. Si des mutations particulières sont trouvées chez les donneurs permissifs, il serait par exemple possible de les induire par CRISPR/Cas9 lors de l'induction de la translocation chez les donneurs non permissifs. La présence de ces mutations pourrait favoriser l'immortalisation cellulaire par la translocation. Peu de mutations secondaires très récurrentes ont été décrites dans les ALCL, cependant, il a par exemple été montré que les protéines Rb et SHP-1 étaient fréquemment mutées dans ces lymphomes (Honorat et al., 2006; Khoury et al., 2004; Rassidakis et al., 2004). La deuxième hypothèse, avancée par Zhang et al (Zhang et al., 2013a) est que seule une partie des cellules est apte à

être transformée par NPM1-ALK. En effet, parmi les cellules T d'un même donneur il existe une grande variabilité phénotypique. Il est envisageable que seule des cellules d'un phénotype particulier soient capables d'être transformées par la kinase de fusion. Certains donneurs pourraient posséder davantage de cellules de ce type et seraient alors plus permissifs. Il serait intéressant de caractériser les marqueurs de surface du pool de cellules avant transfection, et de les comparer aux marqueurs exprimés par les cellules transformées. Ceci permettrait éventuellement de déterminer les cellules aptes à la transformation.

D. Analyse du timing de réplication

Les modèles transloqués obtenus à partir de cellules T vont nous permettre de confirmer nos résultats obtenus dans la lignée RPE-1. En effet grâce à notre système, nous pouvons comparer les cellules avant et après translocation, ce qui est impossible lors de l'analyse de cellules de patients qui souvent sont issus de tumeur agressives qui ont accumulé des mutations additionnelles loin de la cellule d'origine du lymphome. Le profil transcriptomique ainsi que les analyses de timing de réplication pourront être réalisées dans les cellules T immortalisées, en utilisant les cellules du donneur correspondant, avant transfection, comme contrôle. Ces analyses nous permettront de confirmer l'effet de la kinase de fusion sur le timing de réplication. De plus, nous devrions observer une signature en RNAseq plus proche de celle des patients ALCL étant donné que le type cellulaire sera similaire. Une comparaison de nos résultats avec ceux obtenus dans les modèles de surexpression (obtenu par notre collaborateur le Dr. F. Meggetto au CRCT de Toulouse) nous permettra également d'analyser l'effet de la translocation comparé à celui de la protéine de fusion. En effet, dans ces modèles il nous sera impossible d'inactiver la protéine de fusion, comme nous l'avons fait dans la lignée RPE-1, sans induire la mort cellulaire.

E. Vers l'origine cellulaire de l'ALCL

Enfin, le système développé dans les cellules T pourrait permettre de répondre à la question de l'origine cellulaire du cancer. En effet, les cellules ALCL présentent des caractéristiques à la

fois de T matures et de progéniteurs immatures (Foss et al., 1996; Hassler et al., 2016; Juco et al., 2003; Krenacs et al., 1997; Malcolm et al., 2016; Moti et al., 2015). Nous sommes parvenus à immortaliser des cellules T matures par induction de la translocation t(2,5), ce qui prouve qu'une origine mature est possible. Cependant il serait intéressant d'induire la translocation à différents stades de différenciation afin d'étudier d'une part l'effet sur la différenciation cellulaire, et d'autre part le potentiel transformant. Nous sommes parvenus à induire la translocation dans des progéniteurs hématopoïétiques CD34+ issus de sang de cordon, avec la même efficacité que dans les cellules T matures (données non montrées). Par la suite ces cellules pourront être conservées en culture afin d'étudier l'immortalisation, injectées chez la souris afin d'étudier leur potentiel tumoral, ou encore différenciées *in vitro* et T matures par co-culture OP9/DL-1 (Simons et al., 2018). Cette technique permet d'obtenir des progéniteurs T double-positifs à partir de cellules souches hématopoïétiques CD34+. Il serait intéressant d'induire la différenciation T des cellules CD34+ après induction de la translocation, afin d'étudier un éventuel blocage de différenciation, ou encore des anomalies lors de l'expression de certains marqueurs et notamment du TCR. En effet, il a été montré que seuls 4% des ALCL expriment le TCR (Bonzheim et al., 2004) et certaines tumeurs présentent des réarrangements génomiques anormaux du TCR (Malcolm et al., 2016; O'Connor et al., 1987). L'induction de la translocation dans des cellules T matures ne permet pas ce genre d'analyse, car les réarrangements du TCR ont déjà été réalisés dans ces cellules. Seule une dérégulation éventuelle de l'expression du TCR pourra être observée dans ce cas. C'est pourquoi une analyse dans des cellules CD34+ différenciées en T serait un bon modèle pour répondre à cette question.

Une autre question est liée à l'origine cellulaire : dans quelle cellule la formation de la translocation est-elle facilitée ? En effet, il a été montré que la formation de translocation était facilitée par la proximité nucléaire des loci transloqués (Kuroda et al., 2004; Lukasova et al., 1997; Roix et al., 2003). Or dans les T matures les loci NPM1 et ALK ne sont pas particulièrement proches (Mathas et al., 2009). La différenciation *in vitro* de CD34+ en cellules T pourrait permettre d'analyser la proximité de *NPM1* et *ALK* à différents stades de différenciation et corrélée à la fréquence de translocation que nous savons déterminer simplement dans nos expériences. Si une proximité est observée à un stade particulier il est possible qu'*in vivo* la translocation soit formée dans ce type cellulaire précis. A l'aide du

système CRISPR/Cas9 il serait alors possible d'induire la translocation t(2,5) à ce stade de différenciation afin d'étudier par la suite le potentiel transformant de la translocation.

V. Réflexion personnelle générale

Au cours de ma thèse, j'ai eu la chance de travailler sur un projet de recherche novateur, mais surtout, je me suis formée à une variété de méthodes allant de l'utilisation des CRISPR à l'analyse bio-informatique des données, en passant par de nombreuses techniques de culture cellulaire, et notamment à l'utilisation de cellules primaires (T et CD34+) ou encore d'iPS. J'ai ainsi pu acquérir une solide expertise expérimentale, mais j'ai également eu la chance de participer à différentes collaborations nationales et internationales. J'ai notamment participé aux travaux du laboratoire du Dr Jasin sur l'induction de translocations chromosomiques par CRISPR/Cas9, avec expression conditionnelle du gène de fusion, dans des cellules mésenchymateuses humaines. Ces travaux ont mené à une publication en 2017 dans PNAS : Vanoli F, Tomishima M, Feng W, Lamribet K, Babin L, Brunet E, Jasin M, « CRISPR-Cas9-guided oncogenic chromosomal translocations with conditional fusion protein expression in human mesenchymal cells ». J'ai également participé à une collaboration sur les délétions de l'ADN mitochondrial avec le Dr Sfeir, qui a mené à une publication en 2017 dans Molecular Cell: Phillips AF, Millet AR, Tigano M, Dubois SM, Crimmins H, Babin L, Charpentier M, Piganeau M, Brunet E, Sfeir A, « Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion ». Ces deux articles sont présentés en annexe du manuscrit. J'ai aussi participé à une collaboration avec le laboratoire du Dr Lopez sur les mécanismes favorisant l'apparition de translocations dans les cellules, qui est en cours de publication. Enfin et surtout, mon projet de thèse fait partie d'une collaboration avec les laboratoires des docteurs Mercher (IGR, Villejuif) et Deriano (Institut Pasteur, Paris), collaboration ponctuée tout au long de ma thèse de réunions régulières pour discuter des différents projets dont le mien.

J'ai travaillé avec des technologies récentes comme par exemple le système CRISPR/Cas9, qui sont maintenant incontournables en biologie. Durant mes quatre années de thèse, j'ai bénéficié des avancées dans le domaine de l'édition du génome, avec notamment le développement de nouveaux outils bio-informatiques pour l'analyse des taux de coupure, des off-target ainsi que du développement de nouvelles méthodes de transfections permettant

l'utilisation du CRISPR dans des cellules primaires. L'utilisation de la protéine Cas9 et des guides sous forme ARN a notamment permis de réaliser les expériences dans les cellules T primaires humaines, en limitant la mortalité et augmentant l'efficacité de mutations.

J'ai obtenu des résultats intéressants dans la lignée cellulaire RPE-1 aussi bien pour l'étude des ARN circulaires, du profil d'expression génique, que du timing de réplication. Cependant, un certain nombre de résultats reste à confirmer dans les cellules primaires et notamment dans les cellules T. Je pense que l'origine cellulaire est un point clé de l'étude de l'ALCL et que les expériences à venir dans les cellules T matures et CD34+ pourront participer à la résolution de cette question.

Les résultats obtenus ont montré l'importance de la translocation chromosomique dans l'ALCL, au-delà du simple rôle de la protéine de fusion. Je pense qu'il serait intéressant de poursuivre ces analyses dans le cas d'autres translocations chromosomiques. La translocation EWSR1-FLI1, autre modèle du laboratoire, pourrait par exemple être étudiée.

BIBLIOGRAPHIE

(1997). A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. *Blood* 89, 3909-3918.

Abbondanzo, S.L., Sato, N., Straus, S.E., and Jaffe, E.S. (1990). Acute infectious mononucleosis. CD30 (Ki-1) antigen expression and histologic correlations. *American journal of clinical pathology* 93, 698-702.

Amarante-Mendes, G.P., McGahon, A.J., Nishioka, W.K., Afar, D.E., Witte, O.N., and Green, D.R. (1998). Bcl-2-independent Bcr-Abl-mediated resistance to apoptosis: protection is correlated with up regulation of Bcl-xL. *Oncogene* 16, 1383-1390.

Ambrogio, C., Martinengo, C., Voena, C., Tondat, F., Riera, L., di Celle, P.F., Inghirami, G., and Chiarle, R. (2009). NPM-ALK oncogenic tyrosine kinase controls T-cell identity by transcriptional regulation and epigenetic silencing in lymphoma cells. *Cancer research* 69, 8611-8619.

Amiel, A., Kolodizner, T., Fishman, A., Gaber, E., Klein, Z., Beyth, Y., and Fejgin, M.D. (1998). Replication pattern of the p53 and 21q22 loci in the premalignant and malignant stages of carcinoma of the cervix. *Cancer* 83, 1966-1971.

Amin, H.M., McDonnell, T.J., Ma, Y., Lin, Q., Fujio, Y., Kunisada, K., Leventaki, V., Das, P., Rassidakis, G.Z., Cutler, C., *et al.* (2004). Selective inhibition of STAT3 induces apoptosis and G(1) cell cycle arrest in ALK-positive anaplastic large cell lymphoma. *Oncogene* 23, 5426-5434.

Anastasov, N., Bonzheim, I., Rudelius, M., Klier, M., Dau, T., Angermeier, D., Duyster, J., Pittaluga, S., Fend, F., Raffeld, M., *et al.* (2010). C/EBPbeta expression in ALK-positive anaplastic large cell lymphomas is required for cell proliferation and is induced by the STAT3 signaling pathway. *Haematologica* 95, 760-767.

Anderson, N.D., de Borja, R., Young, M.D., Fuligni, F., Rosic, A., Roberts, N.D., Hajjar, S., Layeghifard, M., Novokmet, A., Kowalski, P.E., *et al.* (2018). Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science* 361.

Aoki, K., Yoshida, T., Matsumoto, N., Ide, H., Sugimura, T., and Terada, M. (1997). Suppression of Ki-ras p21 levels leading to growth inhibition of pancreatic cancer cell lines with Ki-ras mutation but not those without Ki-ras mutation. *Molecular carcinogenesis* 20, 251-258.

Armstrong, F., Duplantier, M.M., Trempat, P., Hieblot, C., Lamant, L., Espinos, E., Racaud-Sultan, C., Allouche, M., Campo, E., Delsol, G., *et al.* (2004). Differential effects of X-ALK fusion proteins on proliferation, transformation, and invasion properties of NIH3T3 cells. *Oncogene* 23, 6071-6082.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evtal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Molecular cell* 56, 55-66.

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., *et al.* (2013). Punctuated evolution of prostate cancer genomes. *Cell* 153, 666-677.

Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473-1475.

Bai, R.Y., Dieter, P., Peschel, C., Morris, S.W., and Duyster, J. (1998). Nucleophosmin-anaplastic lymphoma kinase of large-cell anaplastic lymphoma is a constitutively active tyrosine kinase that utilizes phospholipase C-gamma to mediate its mitogenicity. *Molecular and cellular biology* 18, 6951-6961.

Bai, R.Y., Ouyang, T., Miething, C., Morris, S.W., Peschel, C., and Duyster, J. (2000). Nucleophosmin-anaplastic lymphoma kinase associated with anaplastic large-cell lymphoma activates the phosphatidylinositol 3-kinase/Akt antiapoptotic signaling pathway. *Blood* 96, 4319-4327.

Bakhshi, A., Wright, J.J., Graninger, W., Seto, M., Owens, J., Cossman, J., Jensen, J.P., Goldman, P., and Korsmeyer, S.J. (1987). Mechanism of the t(14;18) chromosomal translocation: structural analysis of both derivative 14 and 18 reciprocal partners. *Proceedings of the National Academy of Sciences of the United States of America* 84, 2396-2400.

Barabe, F., Kennedy, J.A., Hope, K.J., and Dick, J.E. (2007). Modeling the initiation and progression of human acute leukemia in mice. *Science* 316, 600-604.

Barrett, S.P., Wang, P.L., and Salzman, J. (2015). Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *eLife* 4, e07540.

Basecke, J., Griesinger, F., Trumper, L., and Brittinger, G. (2002). Leukemia- and lymphoma-associated genetic aberrations in healthy individuals. *Annals of hematology* 81, 64-75.

Bassing, C.H., Alt, F.W., Hughes, M.M., D'Auteuil, M., Wehrly, T.D., Woodman, B.B., Gartner, F., White, J.M., Davidson, L., and Sleckman, B.P. (2000). Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule. *Nature* 405, 583-586.

Beck, R., Monument, M.J., Watkins, W.S., Smith, R., Boucher, K.M., Schiffman, J.D., Jorde, L.B., Randall, R.L., and Lessnick, S.L. (2012). EWS/FLI-responsive GGAA microsatellites exhibit polymorphic differences between European and African populations. *Cancer genetics* 205, 304-312.

Ben-Neriah, Y., Daley, G.Q., Mes-Masson, A.M., Witte, O.N., and Baltimore, D. (1986). The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. *Science* 233, 212-214.

Benharroch, D., Meguerian-Bedoyan, Z., Lamant, L., Amin, C., Brugieres, L., Terrier-Lacombe, M.J., Haralambieva, E., Pulford, K., Pileri, S., Morris, S.W., *et al.* (1998). ALK-positive lymphoma: a single disease with a broad spectrum of morphology. *Blood* 91, 2076-2084.

Bertwistle, D., Sugimoto, M., and Sherr, C.J. (2004). Physical and functional interactions of the Arf tumor suppressor protein with nucleophosmin/B23. *Molecular and cellular biology* 24, 985-996.

Betermier, M., Bertrand, P., and Lopez, B.S. (2014). Is non-homologous end-joining really an inherently error-prone process? *PLoS genetics* 10, e1004086.

Biernaux, C., Loos, M., Sels, A., Huez, G., and Stryckmans, P. (1995). Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals. *Blood* 86, 3118-3122.

Bilsland, J.G., Wheeldon, A., Mead, A., Znamenskiy, P., Almond, S., Waters, K.A., Thakur, M., Beaumont, V., Bonnert, T.P., Heavens, R., *et al.* (2008). Behavioral and neurochemical alterations in mice deficient in anaplastic lymphoma kinase suggest therapeutic potential for psychiatric indications. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 33, 685-700.

Bischof, D., Pulford, K., Mason, D.Y., and Morris, S.W. (1997). Role of the nucleophosmin (NPM) portion of the non-Hodgkin's lymphoma-associated NPM-anaplastic lymphoma kinase fusion protein in oncogenesis. *Molecular and cellular biology* 17, 2312-2325.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America* 95, 10570-10575.

Blasco, R.B., Karaca, E., Ambrogio, C., Cheong, T.C., Karayol, E., Minero, V.G., Voena, C., and Chiarle, R. (2014). Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR/Cas9 technology. *Cell reports* 9, 1219-1227.

Blin, M., Le Tallec, B., Nahse, V., Schmidt, M., Brossas, C., Millot, G.A., Prioleau, M.N., and Debatisse, M. (2019). Transcription-dependent regulation of replication dynamics modulates genome stability. *Nature structural & molecular biology* 26, 58-66.

Boggs, B.A., and Chinault, A.C. (1997). Analysis of DNA replication by fluorescence in situ hybridization. *Methods* 13, 259-270.

Bohlander, S.K. (2005). ETV6: a versatile player in leukemogenesis. *Seminars in cancer biology* 15, 162-174.

Boi, M., Zucca, E., Inghirami, G., and Bertoni, F. (2015). Advances in understanding the pathogenesis of systemic anaplastic large cell lymphomas. *British journal of haematology* 168, 771-783.

Bolli, N., Nicoletti, I., De Marco, M.F., Bigerna, B., Pucciarini, A., Mannucci, R., Martelli, M.P., Liso, A., Mecucci, C., Fabbiano, F., *et al.* (2007). Born to be exported: COOH-terminal nuclear export signals of different strength ensure cytoplasmic accumulation of nucleophosmin leukemic mutants. *Cancer research* 67, 6230-6237.

Bone, K.M., Wang, P., Wu, F., Wu, C., Li, L., Bacani, J.T., Andrew, S.E., and Lai, R. (2015). NPM-ALK mediates phosphorylation of MSH2 at tyrosine 238, creating a functional deficiency in MSH2 and the loss of mismatch repair. *Blood cancer journal* 5, e311.

Bonzheim, I., Geissinger, E., Roth, S., Zettl, A., Marx, A., Rosenwald, A., Muller-Hermelink, H.K., and Rudiger, T. (2004). Anaplastic large cell lymphomas lack the expression of T-cell receptor molecules or molecules of proximal T-cell receptor signaling. *Blood* 104, 3358-3360.

- Borer, R.A., Lehner, C.F., Eppenberger, H.M., and Nigg, E.A. (1989). Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell* *56*, 379-390.
- Bose, S., Deininger, M., Gora-Tybor, J., Goldman, J.M., and Melo, J.V. (1998). The presence of typical and atypical BCR-ABL fusion genes in leukocytes of normal individuals: biologic significance and implications for the assessment of minimal residual disease. *Blood* *92*, 3362-3367.
- Boue, A., and Gallano, P. (1984). A collaborative study of the segregation of inherited chromosome structural rearrangements in 1356 prenatal diagnoses. *Prenatal diagnosis* *4 Spec No*, 45-67.
- Breger, K.S., Smith, L., and Thayer, M.J. (2005). Engineering translocations with delayed replication: evidence for cis control of chromosome replication timing. *Human molecular genetics* *14*, 2813-2827.
- Bridge, J.A., Kanamori, M., Ma, Z., Pickering, D., Hill, D.A., Lydiatt, W., Lui, M.Y., Colleoni, G.W., Antonescu, C.R., Ladanyi, M., *et al.* (2001). Fusion of the ALK gene to the clathrin heavy chain gene, CLTC, in inflammatory myofibroblastic tumor. *The American journal of pathology* *159*, 411-415.
- Bridger, J.M., Boyle, S., Kill, I.R., and Bickmore, W.A. (2000). Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. *Current biology : CB* *10*, 149-152.
- Brunet, E., Simsek, D., Tomishima, M., DeKolver, R., Choi, V.M., Gregory, P., Urnov, F., Weinstock, D.M., and Jasin, M. (2009). Chromosomal translocations induced at specified loci in human stem cells. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 10620-10625.
- Buechele, C., Breese, E.H., Schneidawind, D., Lin, C.H., Jeong, J., Duque-Afonso, J., Wong, S.H., Smith, K.S., Negrin, R.S., Porteus, M., *et al.* (2015). MLL leukemia induction by genome editing of human CD34+ hematopoietic cells. *Blood* *126*, 1683-1694.
- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology* *109*, 21 29 21-29.
- Callen, E., Jankovic, M., Difilippantonio, S., Daniel, J.A., Chen, H.T., Celeste, A., Pellegrini, M., McBride, K., Wangsa, D., Bredemeyer, A.L., *et al.* (2007). ATM prevents the persistence and propagation of chromosome breaks in lymphocytes. *Cell* *130*, 63-75.
- Camacho, E., Hernandez, L., Hernandez, S., Tort, F., Bellosillo, B., Bea, S., Bosch, F., Montserrat, E., Cardesa, A., Fernandez, P.L., *et al.* (2002). ATM gene inactivation in mantle cell lymphoma mainly occurs by truncating mutations and missense mutations involving the phosphatidylinositol-3 kinase domain and is associated with increasing numbers of chromosomal imbalances. *Blood* *99*, 238-244.
- Canitrot, Y., Lautier, D., Laurent, G., Frechet, M., Ahmed, A., Turhan, A.G., Salles, B., Cazaux, C., and Hoffmann, J.S. (1999). Mutator phenotype of BCR--ABL transfected Ba/F3 cell lines and its association with enhanced expression of DNA polymerase beta. *Oncogene* *18*, 2676-2680.
- Cardoso-Moreira, M., Emerson, J.J., Clark, A.G., and Long, M. (2011). Drosophila duplication hotspots are associated with late-replicating regions of the genome. *PLoS genetics* *7*, e1002340.

Ceccon, M., Merlo, M.E.B., Mologni, L., Poggio, T., Varesio, L.M., Menotti, M., Bombelli, S., Rigolio, R., Manazza, A.D., Di Giacomo, F., *et al.* (2016). Excess of NPM-ALK oncogenic signaling promotes cellular apoptosis and drug dependency. *Oncogene* *35*, 3854-3865.

Cesarman, E., Inghirami, G., Chadburn, A., and Knowles, D.M. (1993). High levels of p53 protein expression do not correlate with p53 gene mutations in anaplastic large cell lymphoma. *The American journal of pathology* *143*, 845-856.

Chen, C.L., Rappailles, A., Duquenne, L., Huvet, M., Guilbaud, G., Farinelli, L., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., Hyrien, O., *et al.* (2010). Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome research* *20*, 447-457.

Chen, C.Y., and Sarnow, P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* *268*, 415-417.

Chen, N., Zhao, G., Yan, X., Lv, Z., Yin, H., Zhang, S., Song, W., Li, X., Li, L., Du, Z., *et al.* (2018a). A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome biology* *19*, 218.

Chen, S., He, H., Wang, Y., Liu, L., Liu, Y., You, H., Dong, Y., and Lyu, J. (2018b). Poor prognosis of nucleophosmin overexpression in solid tumors: a meta-analysis. *BMC cancer* *18*, 838.

Chen, S., Huang, V., Xu, X., Livingstone, J., Soares, F., Jeon, J., Zeng, Y., Hua, J.T., Petricca, J., Guo, H., *et al.* (2019a). Widespread and Functional RNA Circularization in Localized Prostate Cancer. *Cell* *176*, 831-843 e822.

Chen, X., Li, M., Feng, X., and Guang, S. (2015). Targeted Chromosomal Translocations and Essential Gene Knockout Using CRISPR/Cas9 Technology in *Caenorhabditis elegans*. *Genetics* *201*, 1295-1306.

Chen, Y.H., Keegan, S., Kahli, M., Tonzi, P., Fenyo, D., Huang, T.T., and Smith, D.J. (2019b). Transcription shapes DNA replication initiation and termination in human cells. *Nature structural & molecular biology* *26*, 67-77.

Chiarle, R., Gong, J.Z., Guasparri, I., Pesci, A., Cai, J., Liu, J., Simmons, W.J., Dhall, G., Howes, J., Piva, R., *et al.* (2003). NPM-ALK transgenic mice spontaneously develop T-cell lymphomas and plasma cell tumors. *Blood* *101*, 1919-1927.

Chiarle, R., Simmons, W.J., Cai, H., Dhall, G., Zamo, A., Raz, R., Karras, J.G., Levy, D.E., and Inghirami, G. (2005). Stat3 is required for ALK-mediated lymphomagenesis and provides a possible therapeutic target. *Nature medicine* *11*, 623-629.

Chin, L., Tam, A., Pomerantz, J., Wong, M., Holash, J., Bardeesy, N., Shen, Q., O'Hagan, R., Pantginis, J., Zhou, H., *et al.* (1999). Essential role for oncogenic Ras in tumour maintenance. *Nature* *400*, 468-472.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology* *31*, 230-232.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research* 24, 132-141.

Choi, P.S., and Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. *Nature communications* 5, 3728.

Choi, Y.L., Soda, M., Yamashita, Y., Ueno, T., Takashima, J., Nakajima, T., Yatabe, Y., Takeuchi, K., Hamada, T., Haruta, H., *et al.* (2010). EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. *The New England journal of medicine* 363, 1734-1739.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757-761.

Chung, I.H., Lu, P.H., Lin, Y.H., Tsai, M.M., Lin, Y.W., Yeh, C.T., and Lin, K.H. (2017). The long non-coding RNA LINC01013 enhances invasion of human anaplastic large-cell lymphoma. *Scientific reports* 7, 295.

Cilloni, D., and Saglio, G. (2012). Molecular pathways: BCR-ABL. *Clinical cancer research : an official journal of the American Association for Cancer Research* 18, 930-937.

Cocquerelle, C., Mascrez, B., Hetuin, D., and Bailleul, B. (1993). Mis-splicing yields circular RNA molecules. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 7, 155-160.

Colleoni, G.W., Bridge, J.A., Garicochea, B., Liu, J., Filippa, D.A., and Ladanyi, M. (2000). ATIC-ALK: A novel variant ALK gene fusion in anaplastic large cell lymphoma resulting from the recurrent cryptic chromosomal inversion, inv(2)(p23q35). *The American journal of pathology* 156, 781-789.

Colombo, E., Marine, J.C., Danovi, D., Falini, B., and Pelicci, P.G. (2002). Nucleophosmin regulates the stability and transcriptional activity of p53. *Nature cell biology* 4, 529-533.

Colon-Otero, G., Van Wier, S.A., Ahmann, G.J., Braggio, E., Albertie, M.L., Weis, J.A., Ailawadhi, S., Cerhan, J.R., Vishnu, P., Jorgensen, M.S., *et al.* (2017). Prevalence of BCL-2/J(H) Translocation in Healthy African Americans. *Annals of hematology* 96, 51-55.

Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125-1134.

Coussens, M.A., Wendland, R.L., Deriano, L., Lindsay, C.R., Arnal, S.M., and Roth, D.B. (2013). RAG2's acidic hinge restricts repair-pathway choice and promotes genomic stability. *Cell reports* 4, 870-878.

Cremer, T., Cremer, C., Baumann, H., Luedtke, E.K., Sperling, K., Teuber, V., and Zorn, C. (1982). Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Human genetics* 60, 46-56.

- Crockett, D.K., Lin, Z., Elenitoba-Johnson, K.S., and Lim, M.S. (2004). Identification of NPM-ALK interacting proteins by tandem mass spectrometry. *Oncogene* 23, 2617-2629.
- Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of cell biology* 145, 1119-1131.
- Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., *et al.* (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature methods* 10, 361-365.
- Cui, Y.X., Kerby, A., McDuff, F.K., Ye, H., and Turner, S.D. (2009). NPM-ALK inhibits the p53 tumor suppressor pathway in an MDM2 and JNK-dependent manner. *Blood* 113, 5217-5227.
- D'Antoni, S., Mattina, T., Di Mare, P., Federico, C., Motta, S., and Saccone, S. (2004). Altered replication timing of the HIRA/Tuple1 locus in the DiGeorge and Velocardiofacial syndromes. *Gene* 333, 111-119.
- Dahlen, A., Fletcher, C.D., Mertens, F., Fletcher, J.A., Perez-Atayde, A.R., Hicks, M.J., Debiec-Rychter, M., Sciort, R., Wejde, J., Wedin, R., *et al.* (2004). Activation of the GLI oncogene through fusion with the beta-actin gene (ACTB) in a group of distinctive pericytic neoplasms: pericytoma with t(7;12). *The American journal of pathology* 164, 1645-1653.
- Dal Molin, A., Bresolin, S., Gaffo, E., Tretti, C., Boldrin, E., Meyer, L.H., Guglielmelli, P., Vannucchi, A.M., Te Kronnie, G., and Bortoluzzi, S. (2019). CircRNAs Are Here to Stay: A Perspective on the MLL Recombinome. *Frontiers in genetics* 10, 88.
- Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.C., Sanborn, A.L., Machol, I., Shamim, M., *et al.* (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proceedings of the National Academy of Sciences of the United States of America* 113, E4504-4512.
- Day, T.A., Layer, J.V., Cleary, J.P., Guha, S., Stevenson, K.E., Tivey, T., Kim, S., Schinzel, A.C., Izzo, F., Doench, J., *et al.* (2017). PARP3 is a promoter of chromosomal rearrangements and limits G4 DNA. *Nature communications* 8, 15110.
- De, S., and Michor, F. (2011). DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. *Nature biotechnology* 29, 1103-1108.
- Deriano, L., Chaumeil, J., Coussens, M., Multani, A., Chou, Y., Alekseyenko, A.V., Chang, S., Skok, J.A., and Roth, D.B. (2011). The RAG2 C terminus suppresses genomic instability and lymphomagenesis. *Nature* 471, 119-123.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., *et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* 22, 1775-1789.

Dileep, V., Ay, F., Sima, J., Vera, D.L., Noble, W.S., and Gilbert, D.M. (2015). Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. *Genome research* 25, 1104-1113.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376-380.

Dixon, J.R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V.T., Yardimci, G.G., Chakraborty, A., Bann, D.V., Wang, Y., *et al.* (2018). Integrative detection and analysis of structural variation in cancer genomes. *Nature genetics* 50, 1388-1398.

Donato, E.M., Fernandez-Zarzoso, M., Hueso, J.A., and de la Rubia, J. (2018). Brentuximab vedotin in Hodgkin lymphoma and anaplastic large-cell lymphoma: an evidence-based review. *OncoTargets and therapy* 11, 4583-4590.

Dotan, Z.A., Dotan, A., Litmanovitch, T., Ravia, Y., Oniashvili, N., Leibovitch, I., Ramon, J., and Avivi, L. (2000). Modification in the inherent mode of allelic replication in lymphocytes of patients suffering from renal cell carcinoma: a novel genetic alteration associated with malignancy. *Genes, chromosomes & cancer* 27, 270-277.

Druck, T., Podolski, J., Byrski, T., Wyrwicz, L., Zajaczek, S., Kata, G., Borowka, A., Lubinski, J., and Huebner, K. (2001). The DIRC1 gene at chromosome 2q33 spans a familial RCC-associated t(2;3)(q33;q21) chromosome translocation. *Journal of human genetics* 46, 583-589.

Druker, B.J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G.M., Fanning, S., Zimmermann, J., and Lydon, N.B. (1996). Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature medicine* 2, 561-566.

Du, Q., Bert, S.A., Armstrong, N.J., Caldon, C.E., Song, J.Z., Nair, S.S., Gould, C.M., Luu, P.L., Peters, T., Khoury, A., *et al.* (2019). Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nature communications* 10, 416.

Du, W.W., Fang, L., Yang, W., Wu, N., Awan, F.M., Yang, Z., and Yang, B.B. (2017). Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. *Cell death and differentiation* 24, 357-370.

Dumbar, T.S., Gentry, G.A., and Olson, M.O. (1989). Interaction of nucleolar phosphoprotein B23 with nucleic acids. *Biochemistry* 28, 9495-9501.

Dupuis-Sandoval, F., Poirier, M., and Scott, M.S. (2015). The emerging landscape of small nucleolar RNAs in cell biology. *Wiley interdisciplinary reviews RNA* 6, 381-397.

Duriez, B., Chilaka, S., Bercher, J.F., Hercule, E., and Prioleau, M.N. (2019). Replication dynamics of individual loci in single living cells reveal changes in the degree of replication stochasticity through S phase. *Nucleic acids research* 47, 5155-5169.

- Duro, D., Bernard, O., Della Valle, V., Leblanc, T., Berger, R., and Larsen, C.J. (1996). Inactivation of the P16INK4/MTS1 gene by a chromosome translocation t(9;14)(p21-22;q11) in an acute lymphoblastic leukemia of B-cell type. *Cancer research* 56, 848-854.
- Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature methods* 4, 721-726.
- Eckerle, S., Brune, V., Doring, C., Tiacci, E., Bohle, V., Sundstrom, C., Kodet, R., Paulli, M., Falini, B., Klapper, W., *et al.* (2009). Gene expression profiling of isolated tumour cells from anaplastic large cell lymphomas: insights into its cellular origin, pathogenesis and relation to Hodgkin lymphoma. *Leukemia* 23, 2129-2138.
- Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. *Genome biology* 5, R1.
- Ensminger, A.W., and Chess, A. (2004). Coordinated replication timing of monoallelically expressed genes along human autosomes. *Human molecular genetics* 13, 651-658.
- Eshhar, Z., Waks, T., Gross, G., and Schindler, D.G. (1993). Specific activation and targeting of cytotoxic lymphocytes through chimeric single chains consisting of antibody-binding domains and the gamma or zeta subunits of the immunoglobulin and T-cell receptors. *Proceedings of the National Academy of Sciences of the United States of America* 90, 720-724.
- Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., *et al.* (2005). Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *The New England journal of medicine* 352, 254-266.
- Falini, B., Pileri, S., Zinzani, P.L., Carbone, A., Zagonel, V., Wolf-Peeters, C., Verhoef, G., Menestrina, F., Todeschini, G., Paulli, M., *et al.* (1999a). ALK+ lymphoma: clinico-pathological findings and outcome. *Blood* 93, 2697-2706.
- Falini, B., Pulford, K., Pucciarini, A., Carbone, A., De Wolf-Peeters, C., Cordell, J., Fizzotti, M., Santucci, A., Pelicci, P.G., Pileri, S., *et al.* (1999b). Lymphomas expressing ALK fusion protein(s) other than NPM-ALK. *Blood* 94, 3509-3515.
- Fang, J.M., Arlt, M.F., Burgess, A.C., Dagenais, S.L., Beer, D.G., and Glover, T.W. (2001). Translocation breakpoints in FHIT and FRA3B in both homologs of chromosome 3 in an esophageal adenocarcinoma. *Genes, chromosomes & cancer* 30, 292-298.
- Felsher, D.W., and Bishop, J.M. (1999). Reversible tumorigenesis by MYC in hematopoietic lineages. *Molecular cell* 4, 199-207.
- Ferber, M.J., Eilers, P., Schuurig, 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.

Foss, H.D., Anagnostopoulos, I., Araujo, I., Assaf, C., Demel, G., Kummer, J.A., Hummel, M., and Stein, H. (1996). Anaplastic large-cell lymphomas of T-cell and null-cell phenotype express cytotoxic molecules. *Blood* 88, 4005-4011.

Fritz, A., Sinha, S., Marella, N., and Berezney, R. (2013). Alterations in replication timing of cancer-related genes in malignant human breast cancer cells. *Journal of cellular biochemistry* 114, 1074-1083.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature biotechnology* 33, 179-186.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* 31, 822-826.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology* 32, 279-284.

Fujimoto, J., Shiota, M., Iwahara, T., Seki, N., Satoh, H., Mori, S., and Yamamoto, T. (1996). Characterization of the transforming activity of p80, a hyperphosphorylated protein in a Ki-1 lymphoma cell line with chromosomal translocation t(2;5). *Proceedings of the National Academy of Sciences of the United States of America* 93, 4181-4186.

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaepfel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., *et al.* (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature biotechnology* 29, 816-823.

Gambacorti-Passerini, C., Messa, C., and Pogliani, E.M. (2011). Crizotinib in anaplastic large-cell lymphoma. *The New England journal of medicine* 364, 775-776.

Gambacorti Passerini, C., Farina, F., Stasia, A., Redaelli, S., Ceccon, M., Mologni, L., Messa, C., Guerra, L., Giudici, G., Sala, E., *et al.* (2014). Crizotinib in advanced, chemoresistant anaplastic lymphoma kinase-positive lymphoma patients. *Journal of the National Cancer Institute* 106, djt378.

Gascoyne, R.D., Aoun, P., Wu, D., Chhanabhai, M., Skinnider, B.F., Greiner, T.C., Morris, S.W., Connors, J.M., Vose, J.M., Viswanatha, D.S., *et al.* (1999). Prognostic significance of anaplastic lymphoma kinase (ALK) protein expression in adults with anaplastic large cell lymphoma. *Blood* 93, 3913-3921.

Gerber, D.E., and Minna, J.D. (2010). ALK inhibition for non-small cell lung cancer: from discovery to therapy in record time. *Cancer cell* 18, 548-551.

Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C., *et al.* (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Molecular cell* 55, 829-842.

- Glover, T.W., Berger, C., Coyle, J., and Echo, B. (1984). DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Human genetics* 67, 136-142.
- Glover, T.W., and Stein, C.K. (1988). Chromosome breakage and recombination at fragile sites. *American journal of human genetics* 43, 265-273.
- Grinberg-Rashi, H., Cytron, S., Gelman-Kohan, Z., Litmanovitch, T., and Avivi, L. (2010). Replication timing aberrations and aneuploidy in peripheral blood lymphocytes of breast cancer patients. *Neoplasia* 12, 668-674.
- Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K., and Pandolfi, P.P. (2005). Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* 437, 147-153.
- Groschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K., *et al.* (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* 157, 369-381.
- Gruner, H., Cortes-Lopez, M., Cooper, D.A., Bauer, M., and Miura, P. (2016). CircRNA accumulation in the aging mouse brain. *Scientific reports* 6, 38907.
- Grunewald, T.G.P., Cidre-Aranaz, F., Surdez, D., Tomazou, E.M., de Alava, E., Kovar, H., Sorensen, P.H., Delattre, O., and Dirksen, U. (2018). Ewing sarcoma. *Nature reviews Disease primers* 4, 5.
- Gu, T.L., Tothova, Z., Scheijen, B., Griffin, J.D., Gilliland, D.G., and Sternberg, D.W. (2004). NPM-ALK fusion kinase of anaplastic large-cell lymphoma regulates survival and proliferative signaling through modulation of FOXO3a. *Blood* 103, 4622-4629.
- Guarnerio, J., Bezzi, M., Jeong, J.C., Paffenholz, S.V., Berry, K., Naldini, M.M., Lo-Coco, F., Tay, Y., Beck, A.H., and Pandolfi, P.P. (2016). Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations. *Cell* 165, 289-302.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., *et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948-951.
- Guryanova, O.A., and Levine, R.L. (2013). Advances in the development of animal models of myeloid leukemias. *Seminars in hematology* 50, 145-155.
- Haeussler, M., Schonig, K., Eckert, H., Eschstruth, A., Mianne, J., Renaud, J.B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., *et al.* (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome biology* 17, 148.
- Hallberg, B., and Palmer, R.H. (2013). Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nature reviews Cancer* 13, 685-700.

- Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* *495*, 384-388.
- Hassan-Zadeh, V., Chilaka, S., Cadoret, J.C., Ma, M.K., Boggetto, N., West, A.G., and Prioleau, M.N. (2012). USF binding sequences from the HS4 insulator element impose early replication timing on a vertebrate replicator. *PLoS biology* *10*, e1001277.
- Hassler, M.R., Pulverer, W., Lakshminarasimhan, R., Redl, E., Hacker, J., Garland, G.D., Merkel, O., Schiefer, A.I., Simonitsch-Klupp, I., Kenner, L., *et al.* (2016). Insights into the Pathogenesis of Anaplastic Large-Cell Lymphoma through Genome-wide DNA Methylation Profiling. *Cell reports* *17*, 596-608.
- Hauer, J., Borkhardt, A., Sanchez-Garcia, I., and Cobaleda, C. (2014). Genetically engineered mouse models of human B-cell precursor leukemias. *Cell cycle* *13*, 2836-2846.
- Hay, K.A., and Turtle, C.J. (2017). Chimeric Antigen Receptor (CAR) T Cells: Lessons Learned from Targeting of CD19 in B-Cell Malignancies. *Drugs* *77*, 237-245.
- Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nature methods* *11*, 122-123.
- Heisterkamp, N., Jenster, G., Kioussis, D., Pattengale, P.K., and Groffen, J. (1991). Human bcr-abl gene has a lethal effect on embryogenesis. *Transgenic research* *1*, 45-53.
- Hernandez, L., Pinyol, M., Hernandez, S., Bea, S., Pulford, K., Rosenwald, A., Lamant, L., Falini, B., Ott, G., Mason, D.Y., *et al.* (1999). TRK-fused gene (TFG) is a new partner of ALK in anaplastic large cell lymphoma producing two structurally different TFG-ALK translocations. *Blood* *94*, 3265-3268.
- Hesse, J.E., Lieber, M.R., Gellert, M., and Mizuuchi, K. (1987). Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. *Cell* *49*, 775-783.
- 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., Leskovar, A., and Gilbert, D.M. (2004). Differentiation-induced replication-timing changes are restricted to AT-rich/long interspersed nuclear element (LINE)-rich isochores. *Proceedings of the National Academy of Sciences of the United States of America* *101*, 16861-16866.
- Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.W., Lyou, Y., Townes, T.M., Schubeler, D., and Gilbert, D.M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS biology* *6*, e245.
- Hirsch, S., Blatte, T.J., Grasedieck, S., Cocciardi, S., Rouhi, A., Jongen-Lavrencic, M., Paschka, P., Kronke, J., Gaidzik, V.I., Dohner, H., *et al.* (2017). Circular RNAs of the nucleophosmin (NPM1) gene in acute myeloid leukemia. *Haematologica* *102*, 2039-2047.

Hoareau-Aveilla, C., Quelen, C., Congras, A., Caillet, N., Labourdette, D., Dozier, C., Brousset, P., Lamant, L., and Meggetto, F. (2019). miR-497 suppresses cycle progression through an axis involving CDK6 in ALK-positive cells. *Haematologica* *104*, 347-359.

Hoareau-Aveilla, C., Valentin, T., Daugrois, C., Quelen, C., Mitou, G., Quentin, S., Jia, J., Spicuglia, S., Ferrier, P., Ceccon, M., *et al.* (2015). Reversal of microRNA-150 silencing disadvantages crizotinib-resistant NPM-ALK(+) cell growth. *The Journal of clinical investigation* *125*, 3505-3518.

Hoffman, E.A., McCulley, A., Haarer, B., Arnak, R., and Feng, W. (2015). Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome research* *25*, 402-412.

Honorat, J.F., Ragab, A., Lamant, L., Delsol, G., and Ragab-Thomas, J. (2006). SHP1 tyrosine phosphatase negatively regulates NPM-ALK tyrosine kinase signaling. *Blood* *107*, 4130-4138.

Horton, S.J., Jaques, J., Woolthuis, C., van Dijk, J., Mesuraca, M., Huls, G., Morrone, G., Vellenga, E., and Schuringa, J.J. (2013). MLL-AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny. *Leukemia* *27*, 1116-1126.

Hsu, M.T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* *280*, 339-340.

Huang, P.S., Chung, I.H., Lin, Y.H., Lin, T.K., Chen, W.J., and Lin, K.H. (2018). The Long Non-Coding RNA MIR503HG Enhances Proliferation of Human ALK-Negative Anaplastic Large-Cell Lymphoma. *International journal of molecular sciences* *19*.

Huang, X., Zhang, W., and Shao, Z. (2019). Prognostic and diagnostic significance of circRNAs expression in lung cancer. *Journal of cellular physiology* *234*, 18459-18465.

Hustedt, N., and Durocher, D. (2016). The control of DNA repair by the cell cycle. *Nature cell biology* *19*, 1-9.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C., *et al.* (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell reports* *10*, 170-177.

Iwahara, T., Fujimoto, J., Wen, D., Cupples, R., Bucay, N., Arakawa, T., Mori, S., Ratzkin, B., and Yamamoto, T. (1997). Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene* *14*, 439-449.

Jager, R., Hahne, J., Jacob, A., Egert, A., Schenkel, J., Wernert, N., Schorle, H., and Wellmann, A. (2005). Mice transgenic for NPM-ALK develop non-Hodgkin lymphomas. *Anticancer research* *25*, 3191-3196.

Janz, S., Potter, M., and Rabkin, C.S. (2003). Lymphoma- and leukemia-associated chromosomal translocations in healthy individuals. *Genes, chromosomes & cancer* *36*, 211-223.

- Jeck, W.R., and Sharpless, N.E. (2014). Detecting and characterizing circular RNAs. *Nature biotechnology* 32, 453-461.
- Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *Rna* 19, 141-157.
- Jiang, J., Zhang, L., Zhou, X., Chen, X., Huang, G., Li, F., Wang, R., Wu, N., Yan, Y., Tong, C., *et al.* (2016). Induction of site-specific chromosomal translocations in embryonic stem cells by CRISPR/Cas9. *Scientific reports* 6, 21918.
- Jin, W., Yun, C., Hobbie, A., Martin, M.J., Sorensen, P.H., and Kim, S.J. (2007). Cellular transformation and activation of the phosphoinositide-3-kinase-Akt cascade by the ETV6-NTRK3 chimeric tyrosine kinase requires c-Src. *Cancer research* 67, 3192-3200.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821.
- Johnson, R.D., and Jasin, M. (2000). Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *The EMBO journal* 19, 3398-3407.
- Juco, J., Holden, J.T., Mann, K.P., Kelley, L.G., and Li, S. (2003). Immunophenotypic analysis of anaplastic large cell lymphoma by flow cytometry. *American journal of clinical pathology* 119, 205-212.
- Julienne, H., Zoufir, A., Audit, B., and Arneodo, A. (2013). Human genome replication proceeds through four chromatin states. *PLoS computational biology* 9, e1003233.
- Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E. (2018). Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell reports* 24, 1-10 e14.
- Kaneko, Y., Frizzera, G., Edamura, S., Maseki, N., Sakurai, M., Komada, Y., Sakurai, M., Tanaka, H., Sasaki, M., Suchi, T., *et al.* (1989). A novel translocation, t(2;5)(p23;q35), in childhood phagocytic large T-cell lymphoma mimicking malignant histiocytosis. *Blood* 73, 806-813.
- Karanjawala, Z.E., Murphy, N., Hinton, D.R., Hsieh, C.L., and Lieber, M.R. (2002). Oxygen metabolism causes chromosome breaks and is associated with the neuronal apoptosis observed in DNA double-strand break repair mutants. *Current biology : CB* 12, 397-402.
- Kas, K., Voz, M.L., Roijer, E., Astrom, A.K., Meyen, E., Stenman, G., and Van de Ven, W.J. (1997). Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations. *Nature genetics* 15, 170-174.
- Khoury, J.D., Rassidakis, G.Z., Medeiros, L.J., Amin, H.M., and Lai, R. (2004). Methylation of SHP1 gene and loss of SHP1 protein expression are frequent in systemic anaplastic large cell lymphoma. *Blood* 104, 1580-1581.

Kim, S.H., McQueen, P.G., Lichtman, M.K., Shevach, E.M., Parada, L.A., and Misteli, T. (2004). Spatial genome organization during T-cell differentiation. *Cytogenetic and genome research* 105, 292-301.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* 93, 1156-1160.

Kinney, M.C., Collins, R.D., Greer, J.P., Whitlock, J.A., Sioutos, N., and Kadin, M.E. (1993). A small-cell-predominant variant of primary Ki-1 (CD30)+ T-cell lymphoma. *The American journal of surgical pathology* 17, 859-868.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D.J., Nicholls, R.D., and Cedar, H. (1993). Allele-specific replication timing of imprinted gene regions. *Nature* 364, 459-463.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490-495.

Klucher, K.M., Lopez, D.V., and Daley, G.Q. (1998). Secondary mutation maintains the transformed state in BaF3 cells with inducible BCR/ABL expression. *Blood* 91, 3927-3934.

Kocak, D.D., Josephs, E.A., Bhandarkar, V., Adkar, S.S., Kwon, J.B., and Gersbach, C.A. (2019). Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nature biotechnology* 37, 657-666.

Kodama, T., Tsukaguchi, T., Yoshida, M., Kondoh, O., and Sakamoto, H. (2014). Selective ALK inhibitor alectinib with potent antitumor activity in models of crizotinib resistance. *Cancer letters* 351, 215-221.

Kojis, T.L., Gatti, R.A., and Sparkes, R.S. (1991). The cytogenetics of ataxia telangiectasia. *Cancer genetics and cytogenetics* 56, 143-156.

Kojis, T.L., Schreck, R.R., Gatti, R.A., and Sparkes, R.S. (1989). Tissue specificity of chromosomal rearrangements in ataxia-telangiectasia. *Human genetics* 83, 347-352.

Koren, A., and McCarroll, S.A. (2014). Random replication of the inactive X chromosome. *Genome research* 24, 64-69.

Korenstein-Ilan, A., Amiel, A., Lalezari, S., Lishner, M., and Avivi, L. (2002). Allele-specific replication associated with aneuploidy in blood cells of patients with hematologic malignancies. *Cancer genetics and cytogenetics* 139, 97-103.

Krenacs, L., Wellmann, A., Sorbara, L., Himmelmann, A.W., Bagdi, E., Jaffe, E.S., and Raffeld, M. (1997). Cytotoxic cell antigen expression in anaplastic large cell lymphomas of T- and null-cell type and Hodgkin's disease: evidence for distinct cellular origin. *Blood* 89, 980-989.

Kristensen, L.S., Hansen, T.B., Veno, M.T., and Kjems, J. (2018). Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene* 37, 555-565.

Kuefer, M.U., Look, A.T., Pulford, K., Behm, F.G., Pattengale, P.K., Mason, D.Y., and Morris, S.W. (1997). Retrovirus-mediated gene transfer of NPM-ALK causes lymphoid malignancy in mice. *Blood* 90, 2901-2910.

Kurki, S., Peltonen, K., Latonen, L., Kiviharju, T.M., Ojala, P.M., Meek, D., and Laiho, M. (2004). Nucleolar protein NPM interacts with HDM2 and protects tumor suppressor protein p53 from HDM2-mediated degradation. *Cancer cell* 5, 465-475.

Kuroda, M., Tanabe, H., Yoshida, K., Oikawa, K., Saito, A., Kiyuna, T., Mizusawa, H., and Mukai, K. (2004). Alteration of chromosome positioning during adipocyte differentiation. *Journal of cell science* 117, 5897-5903.

Lagutina, I.V., Valentine, V., Picchione, F., Harwood, F., Valentine, M.B., Villarejo-Balcells, B., Carvajal, J.J., and Grosveld, G.C. (2015). Modeling of the human alveolar rhabdomyosarcoma Pax3-Foxo1 chromosome translocation in mouse myoblasts using CRISPR-Cas9 nuclease. *PLoS genetics* 11, e1004951.

Lamant, L., Dastugue, N., Pulford, K., Delsol, G., and Mariame, B. (1999). A new fusion gene TPM3-ALK in anaplastic large cell lymphoma created by a (1;2)(q25;p23) translocation. *Blood* 93, 3088-3095.

Lamant, L., de Reynies, A., Duplantier, M.M., Rickman, D.S., Sabourdy, F., Giuriato, S., Brugieres, L., Gaulard, P., Espinos, E., and Delsol, G. (2007). Gene-expression profiling of systemic anaplastic large-cell lymphoma reveals differences based on ALK status and two distinct morphologic ALK+ subtypes. *Blood* 109, 2156-2164.

Lamant, L., McCarthy, K., d'Amore, E., Klapper, W., Nakagawa, A., Fraga, M., Maldyk, J., Simonitsch-Klupp, I., Oschlies, I., Delsol, G., *et al.* (2011). Prognostic impact of morphologic and phenotypic features of childhood ALK-positive anaplastic large-cell lymphoma: results of the ALCL99 study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 29, 4669-4676.

Lang, G.I., and Murray, A.W. (2011). Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome biology and evolution* 3, 799-811.

Lannon, C.L., and Sorensen, P.H. (2005). ETV6-NTRK3: a chimeric protein tyrosine kinase with transformation activity in multiple cell lineages. *Seminars in cancer biology* 15, 215-223.

Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. *Rna* 20, 1829-1842.

Laurent, E., Talpaz, M., Kantarjian, H., and Kurzrock, R. (2001). The BCR gene and philadelphia chromosome-positive leukemogenesis. *Cancer research* 61, 2343-2355.

Le Beau, M.M., Bitter, M.A., Larson, R.A., Doane, L.A., Ellis, E.D., Franklin, W.A., Rubin, C.M., Kadin, M.E., and Vardiman, J.W. (1989). The t(2;5)(p23;q35): a recurring chromosomal abnormality in Ki-1-positive anaplastic large cell lymphoma. *Leukemia* 3, 866-870.

Lee, J.H., and Paull, T.T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science* 308, 551-554.

Lee, J.K., Louzada, S., An, Y., Kim, S.Y., Kim, S., Youk, J., Park, S., Koo, S.H., Keam, B., Jeon, Y.K., *et al.* (2017). Complex chromosomal rearrangements by single catastrophic pathogenesis in NUT midline carcinoma. *Annals of oncology : official journal of the European Society for Medical Oncology* 28, 890-897.

Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.H., Lee, K., Jung, I., Kim, D., Kim, S., *et al.* (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature communications* 9, 3048.

Lee, J.J., Park, S., Park, H., Kim, S., Lee, J., Lee, J., Youk, J., Yi, K., An, Y., Park, I.K., *et al.* (2019). Tracing Oncogene Rearrangements in the Mutational History of Lung Adenocarcinoma. *Cell* 177, 1842-1857 e1821.

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., *et al.* (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Molecular cell* 66, 22-37 e29.

Lei, Y., Lu, L., Liu, H.Y., Li, S., Xing, F., and Chen, L.L. (2014). CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Molecular plant* 7, 1494-1496.

Lekomtsev, S., Aligianni, S., Lapao, A., and Burckstummer, T. (2016). Efficient generation and reversion of chromosomal translocations using CRISPR/Cas technology. *BMC genomics* 17, 739.

Leoncini, L., Del Vecchio, M.T., Kraft, R., Megha, T., Barbini, P., Cevenini, G., Poggi, S., Pileri, S., Tosi, P., and Cottier, H. (1990). Hodgkin's disease and CD30-positive anaplastic large cell lymphomas--a continuous spectrum of malignant disorders. A quantitative morphometric and immunohistologic study. *The American journal of pathology* 137, 1047-1057.

Levy, D., Bertoldi, E.R.M., Ruiz, J.L.M., Pereira, J., and Bydlowski, S.P. (2017). Presence of t(14;18) translocation in healthy individuals varies according to ethnic background in the Brazilian population. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas* 50, e6172.

Li, J., Zhang, X., Sejas, D.P., Bagby, G.C., and Pang, Q. (2004). Hypoxia-induced nucleophosmin protects cell death through inhibition of p53. *The Journal of biological chemistry* 279, 41275-41279.

Li, J., Zhen, L., Zhang, Y., Zhao, L., Liu, H., Cai, D., Chen, H., Yu, J., Qi, X., and Li, G. (2017a). Circ-104916 is downregulated in gastric cancer and suppresses migration and invasion of gastric cancer cells. *OncoTargets and therapy* 10, 3521-3529.

Li, S., Zhang, X., Zhou, Z., Huang, Z., Liu, L., and Huang, Z. (2017b). Downregulation of nucleophosmin expression inhibited proliferation and induced apoptosis in salivary gland adenoid cystic carcinoma.

Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology 46, 175-181.

Li, Y., Chen, B., and Huang, S. (2018). Identification of circRNAs for miRNA Targets by Argonaute2 RNA Immunoprecipitation and Luciferase Screening Assays. *Methods in molecular biology* 1724, 209-218.

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L., *et al.* (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nature structural & molecular biology* 22, 256-264.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* 79, 181-211.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, 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.

Limpens, J., de Jong, D., van Krieken, J.H., Price, C.G., Young, B.D., van Ommen, G.J., and Kluin, P.M. (1991). Bcl-2/JH rearrangements in benign lymphoid tissues with follicular hyperplasia. *Oncogene* 6, 2271-2276.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., *et al.* (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 139, 1069-1083.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* 362, 709-715.

Lu, J., Li, H., Hu, M., Sasaki, T., Baccei, A., Gilbert, D.M., Liu, J.S., Collins, J.J., and Lerou, P.H. (2014). The distribution of genomic variations in human iPSCs is related to replication-timing reorganization during reprogramming. *Cell reports* 7, 70-78.

Lukasova, E., Kozubek, S., Kozubek, M., Kjeronska, J., Ryznar, L., Horakova, J., Krahulcova, E., and Horneck, G. (1997). Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. *Human genetics* 100, 525-535.

MacPherson, C.R., and Scherf, A. (2015). Flexible guide-RNA design for CRISPR applications using Protospacer Workbench. *Nature biotechnology* 33, 805-806.

Maddalo, D., Machado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P., Crippa, A., Rekhman, N., de Stanchina, E., *et al.* (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* 516, 423-427.

Malcolm, T.I., Villarese, P., Fairbairn, C.J., Lamant, L., Trinquand, A., Hook, C.E., Burke, G.A., Brugieres, L., Hughes, K., Payet, D., *et al.* (2016). Anaplastic large cell lymphoma arises in thymocytes and requires transient TCR expression for thymic egress. *Nature communications* 7, 10087.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* *31*, 833-838.

Martinelli, P., Bonetti, P., Sironi, C., Pruneri, G., Fumagalli, C., Raviele, P.R., Volorio, S., Pileri, S., Chiarle, R., McDuff, F.K., *et al.* (2011). The lymphoma-associated NPM-ALK oncogene elicits a p16INK4a/pRb-dependent tumor-suppressive pathway. *Blood* *117*, 6617-6626.

Marzec, M., Kasprzycka, M., Liu, X., Raghunath, P.N., Wlodarski, P., and Wasik, M.A. (2007). Oncogenic tyrosine kinase NPM/ALK induces activation of the MEK/ERK signaling pathway independently of c-Raf. *Oncogene* *26*, 813-821.

Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E., and Sfeir, A. (2015). Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature* *518*, 254-257.

Mathas, S., Kreher, S., Meaburn, K.J., Johrens, K., Lamprecht, B., Assaf, C., Sterry, W., Kadin, M.E., Daibata, M., Joos, S., *et al.* (2009). Gene deregulation and spatial genome reorganization near breakpoints prior to formation of translocations in anaplastic large cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 5831-5836.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y., *et al.* (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* *316*, 1160-1166.

Matsuyama, H., Suzuki, H.I., Nishimori, H., Noguchi, M., Yao, T., Komatsu, N., Mano, H., Sugimoto, K., and Miyazono, K. (2011). miR-135b mediates NPM-ALK-driven oncogenicity and renders IL-17-producing immunophenotype to anaplastic large cell lymphoma. *Blood* *118*, 6881-6892.

Matutes, E. (2018). The 2017 WHO update on mature T- and natural killer (NK) cell neoplasms. *International journal of laboratory hematology* *40 Suppl 1*, 97-103.

Maura, F., Bolli, N., Angelopoulos, N., Dawson, K.J., Leongamornlert, D., Martincorena, I., Mitchell, T.J., Fullam, A., Gonzalez, S., Szalat, R., *et al.* (2019). Genomic landscape and chronological reconstruction of driver events in multiple myeloma. *Nature communications* *10*, 3835.

Mayer, R., Brero, A., von Hase, J., Schroeder, T., Cremer, T., and Dietzel, S. (2005). Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC cell biology* *6*, 44.

McDuff, F.K., and Turner, S.D. (2011). Aberrant anaplastic lymphoma kinase activity induces a p53 and Rb-dependent senescence-like arrest in the absence of detectable p53 stabilization. *PloS one* *6*, e17854.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., *et al.* (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* *495*, 333-338.

Meng, J., Chen, S., Han, J.X., Qian, B., Wang, X.R., Zhong, W.L., Qin, Y., Zhang, H., Gao, W.F., Lei, Y.Y., *et al.* (2018). Twist1 Regulates Vimentin through Cul2 Circular RNA to Promote EMT in Hepatocellular Carcinoma. *Cancer research* 78, 4150-4162.

Merkel, O., Hamacher, F., Laimer, D., Sift, E., Trajanoski, Z., Scheideler, M., Egger, G., Hassler, M.R., Thallinger, C., Schmatz, A., *et al.* (2010). Identification of differential and functionally active miRNAs in both anaplastic lymphoma kinase (ALK)+ and ALK- anaplastic large-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* 107, 16228-16233.

Mertens, F., Johansson, B., Fioretos, T., and Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. *Nature reviews Cancer* 15, 371-381.

Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M., *et al.* (2009). New insights to the MLL recombinome of acute leukemias. *Leukemia* 23, 1490-1499.

Miething, C., Grundler, R., Fend, F., Hoepfl, J., Mugler, C., von Schilling, C., Morris, S.W., Peschel, C., and Duyster, J. (2003). The oncogenic fusion protein nucleophosmin-anaplastic lymphoma kinase (NPM-ALK) induces two distinct malignant phenotypes in a murine retroviral transplantation model. *Oncogene* 22, 4642-4647.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., *et al.* (2011). A TALE nuclease architecture for efficient genome editing. *Nature biotechnology* 29, 143-148.

Mistry, A.R., Felix, C.A., Whitmarsh, R.J., Mason, A., Reiter, A., Cassinat, B., Parry, A., Walz, C., Wiemels, J.L., Segal, M.R., *et al.* (2005). DNA topoisomerase II in therapy-related acute promyelocytic leukemia. *The New England journal of medicine* 352, 1529-1538.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic acids research* 42, W401-407.

Moore, N.F., Azarova, A.M., Bhatnagar, N., Ross, K.N., Drake, L.E., Frumm, S., Liu, Q.S., Christie, A.L., Sanda, T., Chesler, L., *et al.* (2014). Molecular rationale for the use of PI3K/AKT/mTOR pathway inhibitors in combination with crizotinib in ALK-mutated neuroblastoma. *Oncotarget* 5, 8737-8749.

Mora, J., Filippa, D.A., Thaler, H.T., Polyak, T., Cranor, M.L., and Wollner, N. (2000). Large cell non-Hodgkin lymphoma of childhood: Analysis of 78 consecutive patients enrolled in 2 consecutive protocols at the Memorial Sloan-Kettering Cancer Center. *Cancer* 88, 186-197.

Morris, S.W., Kirstein, M.N., Valentine, M.B., Dittmer, K.G., Shapiro, D.N., Saltman, D.L., and Look, A.T. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 263, 1281-1284.

Morris, S.W., Naeve, C., Mathew, P., James, P.L., Kirstein, M.N., Cui, X., and Witte, D.P. (1997). ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene* 14, 2175-2188.

Mosse, Y.P., Lim, M.S., Voss, S.D., Wilner, K., Ruffner, K., Laliberte, J., Rolland, D., Balis, F.M., Maris, J.M., Weigel, B.J., *et al.* (2013). Safety and activity of crizotinib for paediatric patients with refractory solid tumours or anaplastic large-cell lymphoma: a Children's Oncology Group phase 1 consortium study. *The Lancet Oncology* *14*, 472-480.

Motegi, A., Fujimoto, J., Kotani, M., Sakuraba, H., and Yamamoto, T. (2004). ALK receptor tyrosine kinase promotes cell growth and neurite outgrowth. *Journal of cell science* *117*, 3319-3329.

Moti, N., Malcolm, T., Hamoudi, R., Mian, S., Garland, G., Hook, C.E., Burke, G.A., Wasik, M.A., Merkel, O., Kenner, L., *et al.* (2015). Anaplastic large cell lymphoma-propagating cells are detectable by side population analysis and possess an expression profile reflective of a primitive origin. *Oncogene* *34*, 1843-1852.

Muller, C.A., and Nieduszynski, C.A. (2017). DNA replication timing influences gene expression level. *The Journal of cell biology* *216*, 1907-1914.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* *102*, 553-563.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of biological chemistry* *274*, 18470-18476.

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research* *39*, 9283-9293.

Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2015). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* *31*, 1120-1123.

Nambiar, M., and Raghavan, S.C. (2011). How does DNA break during chromosomal translocations? *Nucleic acids research* *39*, 5813-5825.

Nishimura, Y., Ohkubo, T., Furuichi, Y., and Umekawa, H. (2002). Tryptophans 286 and 288 in the C-terminal region of protein B23.1 are important for its nucleolar localization. *Bioscience, biotechnology, and biochemistry* *66*, 2239-2242.

Nowell, P.C., and Hungerford, D.A. (1960). Chromosome studies on normal and leukemic human leukocytes. *Journal of the National Cancer Institute* *25*, 85-109.

Nozawa, Y., Van Belzen, N., Van der Made, A.C., Dinjens, W.N., and Bosman, F.T. (1996). Expression of nucleophosmin/B23 in normal and neoplastic colorectal mucosa. *The Journal of pathology* *178*, 48-52.

O'Connor, N.T., Stein, H., Gatter, K.C., Wainscoat, J.S., Crick, J., Al Saati, T., Falini, B., Delsol, G., and Mason, D.Y. (1987). Genotypic analysis of large cell lymphomas which express the Ki-1 antigen. *Histopathology* *11*, 733-740.

Odelberg, S.J., Weiss, R.B., Hata, A., and White, R. (1995). Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic acids research* *23*, 2049-2057.

Oettinger, M.A., Schatz, D.G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* *248*, 1517-1523.

Oliveira, A.M., Perez-Atayde, A.R., Dal Cin, P., Gebhardt, M.C., Chen, C.J., Neff, J.R., Demetri, G.D., Rosenberg, A.E., Bridge, J.A., and Fletcher, J.A. (2005). Aneurysmal bone cyst variant translocations upregulate USP6 transcription by promoter swapping with the ZNF9, COL1A1, TRAP150, and OMD genes. *Oncogene* *24*, 3419-3426.

Osheroff, N. (1989). Biochemical basis for the interactions of type I and type II topoisomerases with DNA. *Pharmacology & therapeutics* *41*, 223-241.

Page, S.L., and Shaffer, L.G. (1997). Nonhomologous Robertsonian translocations form predominantly during female meiosis. *Nature genetics* *15*, 231-232.

Panda, A.C., De, S., Grammatikakis, I., Munk, R., Yang, X., Piao, Y., Dudekula, D.B., Abdelmohsen, K., and Gorospe, M. (2017). High-purity circular RNA isolation method (RPAD) reveals vast collection of intronic circRNAs. *Nucleic acids research* *45*, e116.

Panda, A.C., and Gorospe, M. (2018). Detection and Analysis of Circular RNAs by RT-PCR. *Bio-protocol* *8*.

Panning, M.M., and Gilbert, D.M. (2005). Spatio-temporal organization of DNA replication in murine embryonic stem, primary, and immortalized cells. *Journal of cellular biochemistry* *95*, 74-82.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* *412*, 341-346.

Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* *252*, 809-817.

Pelengaris, S., Littlewood, T., Khan, M., Elia, G., and Evan, G. (1999). Reversible activation of c-Myc in skin: induction of a complex neoplastic phenotype by a single oncogenic lesion. *Molecular cell* *3*, 565-577.

Pendleton, M., Lindsey, R.H., Jr., Felix, C.A., Grimwade, D., and Osheroff, N. (2014). Topoisomerase II and leukemia. *Annals of the New York Academy of Sciences* *1310*, 98-110.

Penrose, L.S., Ellis, J.R., and Delhanty, J.D. (1960). Chromosomal translocations in mongolism and in normal relatives. *Lancet* *2*, 409-410.

- Perriman, R., and Ares, M., Jr. (1998). Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *Rna* 4, 1047-1054.
- Piganeau, M., Ghezraoui, H., De Cian, A., Guittat, L., Tomishima, M., Perrouault, L., Rene, O., Katibah, G.E., Zhang, L., Holmes, M.C., *et al.* (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome research* 23, 1182-1193.
- Piganeau, M., Renouf, B., Ghezraoui, H., and Brunet, E. (2016). TALEN-Induced Translocations in Human Cells. *Methods in molecular biology* 1338, 99-117.
- Pileri, S., Falini, B., Delsol, G., Stein, H., Baglioni, P., Poggi, S., Martelli, M.F., Rivano, M.T., Mason, D.Y., and Stansfeld, A.G. (1990). Lymphohistiocytic T-cell lymphoma (anaplastic large cell lymphoma CD30+/Ki-1 + with a high content of reactive histiocytes). *Histopathology* 16, 383-391.
- Piva, R., Agnelli, L., Pellegrino, E., Todoerti, K., Grosso, V., Tamagno, I., Fornari, A., Martinoglio, B., Medico, E., Zamo, A., *et al.* (2010). Gene expression profiling uncovers molecular classifiers for the recognition of anaplastic large-cell lymphoma within peripheral T-cell neoplasms. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28, 1583-1590.
- Poletto, M., Lirussi, L., Wilson, D.M., 3rd, and Tell, G. (2014). Nucleophosmin modulates stability, activity, and nucleolar accumulation of base excision repair proteins. *Molecular biology of the cell* 25, 1641-1652.
- Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., *et al.* (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature* 515, 402-405.
- Potapov, V., and Ong, J.L. (2017). Examining Sources of Error in PCR by Single-Molecule Sequencing. *PloS one* 12, e0169774.
- Prochazkova, J., and Loizou, J.I. (2016). Programmed DNA breaks in lymphoid cells: repair mechanisms and consequences in human disease. *Immunology* 147, 11-20.
- Prykhodzhiy, S.V., Rajan, V., Gaston, D., and Berman, J.N. (2015). CRISPR multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. *PloS one* 10, e0119372.
- Qu, Y., Liu, D., Jia, H., and Yang, Z. (2019). Circular RNA rno_circ_0004002 regulates cell proliferation, apoptosis, and epithelial-mesenchymal transition through targeting miR-342-5p and Wnt3a in anorectal malformations. *Journal of cellular biochemistry*.
- Querfeld, C., Khan, I., Mahon, B., Nelson, B.P., Rosen, S.T., and Evens, A.M. (2010). Primary cutaneous and systemic anaplastic large cell lymphoma: clinicopathologic aspects and therapeutic options. *Oncology* 24, 574-587.
- Rabbitts, T.H., Appert, A., Chung, G., Collins, E.C., Drynan, L., Forster, A., Lobato, M.N., McCormack, M.P., Pannell, R., Spandidos, A., *et al.* (2001). Mouse models of human chromosomal translocations and approaches to cancer therapy. *Blood cells, molecules & diseases* 27, 249-259.

Rabbitts, T.H., and Boehm, T. (1991). Structural and functional chimerism results from chromosomal translocation in lymphoid tumors. *Advances in immunology* 50, 119-146.

Ramos, C.A., Ballard, B., Zhang, H., Dakhova, O., Gee, A.P., Mei, Z., Bilgi, M., Wu, M.F., Liu, H., Grilley, B., *et al.* (2017). Clinical and immunological responses after CD30-specific chimeric antigen receptor-redirected lymphocytes. *The Journal of clinical investigation* 127, 3462-3471.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., *et al.* (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380-1389.

Rassidakis, G.Z., Lai, R., Herling, M., Cromwell, C., Schmitt-Graeff, A., and Medeiros, L.J. (2004). Retinoblastoma protein is frequently absent or phosphorylated in anaplastic large-cell lymphoma. *The American journal of pathology* 164, 2259-2267.

Rassidakis, G.Z., Thomaidis, A., Wang, S., Jiang, Y., Fourtouna, A., Lai, R., and Medeiros, L.J. (2005). p53 gene mutations are uncommon but p53 is commonly expressed in anaplastic large-cell lymphoma. *Leukemia* 19, 1663-1669.

Redaelli, S., Ceccon, M., Antolini, L., Rigolio, R., Pirola, A., Peronaci, M., Gambacorti-Passerini, C., and Mologni, L. (2016). Synergistic activity of ALK and mTOR inhibitors for the treatment of NPM-ALK positive lymphoma. *Oncotarget* 7, 72886-72897.

Redner, R.L., Rush, E.A., Faas, S., Rudert, W.A., and Corey, S.J. (1996). The t(5;17) variant of acute promyelocytic leukemia expresses a nucleophosmin-retinoic acid receptor fusion. *Blood* 87, 882-886.

Reimer, J., Knoss, S., Labuhn, M., Charpentier, E.M., Gohring, G., Schlegelberger, B., Klusmann, J.H., and Heckl, D. (2017). CRISPR-Cas9-induced t(11;19)/MLL-ENL translocations initiate leukemia in human hematopoietic progenitor cells in vivo. *Haematologica* 102, 1558-1566.

Renouf, B., Piganeau, M., Ghezraoui, H., Jasin, M., and Brunet, E. (2014). Creating cancer translocations in human cells using Cas9 DSBs and nCas9 paired nicks. *Methods in enzymology* 546, 251-271.

Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. *Cold Spring Harbor perspectives in biology* 5, a010132.

Richardson, C., and Jasin, M. (2000). Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 405, 697-700.

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.

Rimokh, R., Magaud, J.P., Berger, F., Samarut, J., Coiffier, B., Germain, D., and Mason, D.Y. (1989). A translocation involving a specific breakpoint (q35) on chromosome 5 is characteristic of anaplastic large cell lymphoma ('Ki-1 lymphoma'). *British journal of haematology* 71, 31-36.

Rimokh, R., Rouault, J.P., Wahbi, K., Gadoux, M., Lafage, M., Archimbaud, E., Charrin, C., Gentilhomme, O., Germain, D., Samarut, J., *et al.* (1991). A chromosome 12 coding region is juxtaposed to the MYC protooncogene locus in a t(8;12)(q24;q22) translocation in a case of B-cell chronic lymphocytic leukemia. *Genes, chromosomes & cancer* 3, 24-36.

Rivera-Mulia, J.C., Buckley, Q., Sasaki, T., Zimmerman, J., Didier, R.A., Nator, K., Loring, J.F., Lian, Z., Weissman, S., Robins, A.J., *et al.* (2015). Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome research* 25, 1091-1103.

Rivera-Mulia, J.C., Desprat, R., Trevilla-Garcia, C., Cornacchia, D., Schwerer, H., Sasaki, T., Sima, J., Fells, T., Studer, L., Lemaitre, J.M., *et al.* (2017). DNA replication timing alterations identify common markers between distinct progeroid diseases. *Proceedings of the National Academy of Sciences of the United States of America* 114, E10972-E10980.

Rivera-Mulia, J.C., and Gilbert, D.M. (2016). Replication timing and transcriptional control: beyond cause and effect-part III. *Current opinion in cell biology* 40, 168-178.

Rizo, A., Horton, S.J., Olthof, S., Dontje, B., Ausema, A., van Os, R., van den Boom, V., Vellenga, E., de Haan, G., and Schuringa, J.J. (2010). BMI1 collaborates with BCR-ABL in leukemic transformation of human CD34+ cells. *Blood* 116, 4621-4630.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A., *et al.* (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* 135, 1028-1038.

Robinson, M.J., and Osheroff, N. (1991). Effects of antineoplastic drugs on the post-strand-passage DNA cleavage/religation equilibrium of topoisomerase II. *Biochemistry* 30, 1807-1813.

Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A., and Misteli, T. (2003). Spatial proximity of translocation-prone gene loci in human lymphomas. *Nature genetics* 34, 287-291.

Roulland, S., Kelly, R.S., Morgado, E., Sungalee, S., Solal-Celigny, P., Colombat, P., Jouve, N., Palli, D., Pala, V., Tumino, R., *et al.* (2014). t(14;18) Translocation: A predictive blood biomarker for follicular lymphoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 32, 1347-1355.

Ryba, T., Battaglia, D., Chang, B.H., Shirley, J.W., Buckley, Q., Pope, B.D., Devidas, M., Druker, B.J., and Gilbert, D.M. (2012). Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome research* 22, 1833-1844.

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.

Sabari, J.K., Santini, F.C., Schram, A.M., Bergagnini, I., Chen, R., Mrad, C., Lai, W.V., Arbour, K.C., and Drilon, A. (2017). The activity, safety, and evolving role of brigatinib in patients with ALK-rearranged non-small cell lung cancers. *OncoTargets and therapy* 10, 1983-1992.

Sabattini, E., Bacci, F., Sagrmoso, C., and Pileri, S.A. (2010). WHO classification of tumours of haematopoietic and lymphoid tissues in 2008: an overview. *Pathologica* *102*, 83-87.

Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., and Brown, P.O. (2013). Cell-type specific features of circular RNA expression. *PLoS genetics* *9*, e1003777.

Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS one* *7*, e30733.

Sasaki, T., Okuda, K., Zheng, W., Butrynski, J., Capelletti, M., Wang, L., Gray, N.S., Wilner, K., Christensen, J.G., Demetri, G., *et al.* (2010). The neuroblastoma-associated F1174L ALK mutation causes resistance to an ALK kinase inhibitor in ALK-translocated cancers. *Cancer research* *70*, 10038-10043.

Schmegner, C., Hameister, H., Vogel, W., and Assum, G. (2007). Isochores and replication time zones: a perfect match. *Cytogenetic and genome research* *116*, 167-172.

Schmitz, N., Trumper, L., Ziepert, M., Nickelsen, M., Ho, A.D., Metzner, B., Peter, N., Loeffler, M., Rosenwald, A., and Pfreundschuh, M. (2010). Treatment and prognosis of mature T-cell and NK-cell lymphoma: an analysis of patients with T-cell lymphoma treated in studies of the German High-Grade Non-Hodgkin Lymphoma Study Group. *Blood* *116*, 3418-3425.

Secker-Walker, L.M., Stewart, E.L., and Todd, A. (1985). Acute lymphoblastic leukaemia with t(4;11) follows neuroblastoma: a late effect of treatment? *Medical and pediatric oncology* *13*, 48-50.

Septyarskiy, V.B., Bazykin, G.A., and Soldatov, R.A. (2015). Polymerase zeta Activity Is Linked to Replication Timing in Humans: Evidence from Mutational Signatures. *Molecular biology and evolution* *32*, 3158-3172.

Shafikhani, S. (2002). Factors affecting PCR-mediated recombination. *Environmental microbiology* *4*, 482-486.

Sharma, S.V., Gajowniczek, P., Way, I.P., Lee, D.Y., Jiang, J., Yuza, Y., Classon, M., Haber, D.A., and Settleman, J. (2006). A common signaling cascade may underlie "addiction" to the Src, BCR-ABL, and EGF receptor oncogenes. *Cancer cell* *10*, 425-435.

Shaw, A.T., Felip, E., Bauer, T.M., Besse, B., Navarro, A., Postel-Vinay, S., Gainor, J.F., Johnson, M., Dietrich, J., James, L.P., *et al.* (2017). Lorlatinib in non-small-cell lung cancer with ALK or ROS1 rearrangement: an international, multicentre, open-label, single-arm first-in-man phase 1 trial. *The Lancet Oncology* *18*, 1590-1599.

Shaw, A.T., Kim, D.W., Nakagawa, K., Seto, T., Crino, L., Ahn, M.J., De Pas, T., Besse, B., Solomon, B.J., Blackhall, F., *et al.* (2013). Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *The New England journal of medicine* *368*, 2385-2394.

Shields, L.B., Gercel-Taylor, C., Yashar, C.M., Wan, T.C., Katsanis, W.A., Spinnato, J.A., and Taylor, D.D. (1997). Induction of immune responses to ovarian tumor antigens by multiparity. *Journal of the Society for Gynecologic Investigation* 4, 298-304.

Shima, E.A., Le Beau, M.M., McKeithan, T.W., Minowada, J., Showe, L.C., Mak, T.W., Minden, M.D., Rowley, J.D., and Diaz, M.O. (1986). Gene encoding the alpha chain of the T-cell receptor is moved immediately downstream of c-myc in a chromosomal 8;14 translocation in a cell line from a human T-cell leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 83, 3439-3443.

Shiota, M., Fujimoto, J., Semba, T., Satoh, H., Yamamoto, T., and Mori, S. (1994). Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* 9, 1567-1574.

Sibon, D., Fournier, M., Briere, J., Lamant, L., Haioun, C., Coiffier, B., Bologna, S., Morel, P., Gabarre, J., Hermine, O., *et al.* (2012). Long-term outcome of adults with systemic anaplastic large-cell lymphoma treated within the Groupe d'Etude des Lymphomes de l'Adulte trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30, 3939-3946.

Sima, J., Chakraborty, A., Dileep, V., Michalski, M., Klein, K.N., Holcomb, N.P., Turner, J.L., Paulsen, M.T., Rivera-Mulia, J.C., Trevilla-Garcia, C., *et al.* (2019). Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. *Cell* 176, 816-830 e818.

Simon, I., Tenzen, T., Mostoslavsky, R., Fibach, E., Lande, L., Milot, E., Gribnau, J., Grosveld, F., Fraser, P., and Cedar, H. (2001). Developmental regulation of DNA replication timing at the human beta globin locus. *The EMBO journal* 20, 6150-6157.

Simon, M.P., Pedoutour, F., Sirvent, N., Grosgeorge, J., Minoletti, F., Coindre, J.M., Terrier-Lacombe, M.J., Mandahl, N., Craver, R.D., Blin, N., *et al.* (1997). Deregulation of the platelet-derived growth factor B-chain gene via fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans and giant-cell fibroblastoma. *Nature genetics* 15, 95-98.

Simons, L., Ma, K., de Chappedelaine, C., Moirangthem, R.D., Elkaim, E., Olivre, J., Susini, S., Appourchaux, K., Reimann, C., Sadek, H., *et al.* (2018). Generation of adult human T-cell progenitors for immunotherapeutic applications. *The Journal of allergy and clinical immunology* 141, 1491-1494 e1494.

Simsek, D., Brunet, E., Wong, S.Y., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., *et al.* (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS genetics* 7, e1002080.

Sinclair, P., Harrison, C.J., Jarosova, M., and Foroni, L. (2005). Analysis of balanced rearrangements of chromosome 6 in acute leukemia: clustered breakpoints in q22-q23 and possible involvement of c-MYB in a new recurrent translocation, t(6;7)(q23;q32 through 36). *Haematologica* 90, 602-611.

Sipos, K., and Olson, M.O. (1991). Nucleolin promotes secondary structure in ribosomal RNA. *Biochemical and biophysical research communications* 177, 673-678.

Slymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84-88.

Slupianek, A., Nieborowska-Skorska, M., Hoser, G., Morrione, A., Majewski, M., Xue, L., Morris, S.W., Wasik, M.A., and Skorski, T. (2001). Role of phosphatidylinositol 3-kinase-Akt pathway in nucleophosmin/anaplastic lymphoma kinase-mediated lymphomagenesis. *Cancer research* 61, 2194-2199.

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., *et al.* (2014). Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell stem cell* 15, 12-13.

Sogo, J.M., Lopes, M., and Foiani, M. (2002). Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* 297, 599-602.

Sontakke, P., Carretta, M., Jaques, J., Brouwers-Vos, A.Z., Lubbers-Aalders, L., Yuan, H., de Bruijn, J.D., Martens, A.C., Vellenga, E., Groen, R.W., *et al.* (2016). Modeling BCR-ABL and MLL-AF9 leukemia in a human bone marrow-like scaffold-based xenograft model. *Leukemia* 30, 2064-2073.

Souttou, B., Carvalho, N.B., Raulais, D., and Vigny, M. (2001). Activation of anaplastic lymphoma kinase receptor tyrosine kinase induces neuronal differentiation through the mitogen-activated protein kinase pathway. *The Journal of biological chemistry* 276, 9526-9531.

Sportoletti, P., Grisendi, S., Majid, S.M., Cheng, K., Clohessy, J.G., Viale, A., Teruya-Feldstein, J., and Pandolfi, P.P. (2008). Npm1 is a haploinsufficient suppressor of myeloid and lymphoid malignancies in the mouse. *Blood* 111, 3859-3862.

Spraggon, L., Martelotto, L.G., Hmeljak, J., Hitchman, T.D., Wang, J., Wang, L., Slotkin, E.K., Fan, P.D., Reis-Filho, J.S., and Ladanyi, M. (2017). Generation of conditional oncogenic chromosomal translocations using CRISPR-Cas9 genomic editing and homology-directed repair. *The Journal of pathology* 242, 102-112.

Stamatoyannopoulos, J.A., Adzhubei, I., Thurman, R.E., Kryukov, G.V., Mirkin, S.M., and Sunyaev, S.R. (2009). Human mutation rate associated with DNA replication timing. *Nature genetics* 41, 393-395.

Stein, H., Gerdes, J., Schwab, U., Lemke, H., Mason, D.Y., Ziegler, A., Schienle, W., and Diehl, V. (1982). Identification of Hodgkin and Sternberg-reed cells as a unique cell type derived from a newly-detected small-cell population. *International journal of cancer* 30, 445-459.

Stein, H., Mason, D.Y., Gerdes, J., O'Connor, N., Wainscoat, J., Pallesen, G., Gatter, K., Falini, B., Delsol, G., Lemke, H., *et al.* (1985). The expression of the Hodgkin's disease associated antigen Ki-1 in reactive and neoplastic lymphoid tissue: evidence that Reed-Sternberg cells and histiocytic malignancies are derived from activated lymphoid cells. *Blood* 66, 848-858.

Steinhilber, J., Bonin, M., Walter, M., Fend, F., Bonzheim, I., and Quintanilla-Martinez, L. (2015). Next-generation sequencing identifies deregulation of microRNAs involved in both innate and adaptive immune response in ALK+ ALCL. *PLoS one* 10, e0117780.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62-67.

Stoica, G.E., Kuo, A., Aigner, A., Sunitha, I., Souttou, B., Malerczyk, C., Caughey, D.J., Wen, D., Karavanov, A., Riegel, A.T., *et al.* (2001). Identification of anaplastic lymphoma kinase as a receptor for the growth factor pleiotrophin. *The Journal of biological chemistry* 276, 16772-16779.

Stoica, G.E., Kuo, A., Powers, C., Bowden, E.T., Sale, E.B., Riegel, A.T., and Wellstein, A. (2002). Midkine binds to anaplastic lymphoma kinase (ALK) and acts as a growth factor for different cell types. *The Journal of biological chemistry* 277, 35990-35998.

Storlazzi, C.T., Von Steyern, F.V., Domanski, H.A., Mandahl, N., and Mertens, F. (2005). Biallelic somatic inactivation of the NF1 gene through chromosomal translocations in a sporadic neurofibroma. *International journal of cancer* 117, 1055-1057.

Stransky, N., Cerami, E., Schalm, S., Kim, J.L., and Lengauer, C. (2014). The landscape of kinase fusions in cancer. *Nature communications* 5, 4846.

Subong, E.N., Shue, M.J., Epstein, J.I., Briggman, J.V., Chan, P.K., and Partin, A.W. (1999). Monoclonal antibody to prostate cancer nuclear matrix protein (PRO:4-216) recognizes nucleophosmin/B23. *The Prostate* 39, 298-304.

Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A., and Mayeda, A. (2006). Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic acids research* 34, e63.

Taccioli, G.E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P.A., and Alt, F.W. (1993). Impairment of V(D)J recombination in double-strand break repair mutants. *Science* 260, 207-210.

Takeda, N., Shibuya, M., and Maru, Y. (1999). The BCR-ABL oncoprotein potentially interacts with the xeroderma pigmentosum group B protein. *Proceedings of the National Academy of Sciences of the United States of America* 96, 203-207.

Takemura, M., Ohoka, F., Perpelescu, M., Ogawa, M., Matsushita, H., Takaba, T., Akiyama, T., Umekawa, H., Furuichi, Y., Cook, P.R., *et al.* (2002). Phosphorylation-dependent migration of retinoblastoma protein into the nucleolus triggered by binding to nucleophosmin/B23. *Experimental cell research* 276, 233-241.

Tan, S., Sun, D., Pu, W., Gou, Q., Guo, C., Gong, Y., Li, J., Wei, Y.Q., Liu, L., Zhao, Y., *et al.* (2018). Circular RNA F-circEA-2a derived from EML4-ALK fusion gene promotes cell migration and invasion in non-small cell lung cancer. *Molecular cancer* 17, 138.

Tanaka, M., Sasaki, H., Kino, I., Sugimura, T., and Terada, M. (1992). Genes preferentially expressed in embryo stomach are predominantly expressed in gastric cancer. *Cancer research* 52, 3372-3377.

Tanaka, M., Yamaguchi, S., Yamazaki, Y., Kinoshita, H., Kuwahara, K., Nakao, K., Jay, P.Y., Noda, T., and Nakamura, T. (2015). Somatic chromosomal translocation between Ewsr1 and Fli1 loci leads to dilated cardiomyopathy in a mouse model. *Scientific reports* 5, 7826.

Tarapore, P., Okuda, M., and Fukasawa, K. (2002). A mammalian in vitro centriole duplication system: evidence for involvement of CDK2/cyclin E and nucleophosmin/B23 in centrosome duplication. *Cell cycle* 1, 75-81.

Therizols, P., Illingworth, R.S., Courilleau, C., Boyle, S., Wood, A.J., and Bickmore, W.A. (2014). Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* 346, 1238-1242.

Thiede, C., Koch, S., Creutzig, E., Steudel, C., Illmer, T., Schaich, M., and Ehninger, G. (2006). Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). *Blood* 107, 4011-4020.

Thomas, N.S., Morris, J.K., Baptista, J., Ng, B.L., Crolla, J.A., and Jacobs, P.A. (2010). De novo apparently balanced translocations in man are predominantly paternal in origin and associated with a significant increase in paternal age. *Journal of medical genetics* 47, 112-115.

Thompson, M.A., Stumph, J., Henrickson, S.E., Rosenwald, A., Wang, Q., Olson, S., Brandt, S.J., Roberts, J., Zhang, X., Shyr, Y., *et al.* (2005). Differential gene expression in anaplastic lymphoma kinase-positive and anaplastic lymphoma kinase-negative anaplastic large cell lymphomas. *Human pathology* 36, 494-504.

Thurman, R.E., Day, N., Noble, W.S., and Stamatoyannopoulos, J.A. (2007). Identification of higher-order functional domains in the human ENCODE regions. *Genome research* 17, 917-927.

Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B., *et al.* (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* 448, 595-599.

Torres-Ruiz, R., Martinez-Lage, M., Martin, M.C., Garcia, A., Bueno, C., Castano, J., Ramirez, J.C., Menendez, P., Cigudosa, J.C., and Rodriguez-Perales, S. (2017). Efficient Recreation of t(11;22) EWSR1-FLI1(+) in Human Stem Cells Using CRISPR/Cas9. *Stem cell reports* 8, 1408-1420.

Torres, R., Martin, M.C., Garcia, A., Cigudosa, J.C., Ramirez, J.C., and Rodriguez-Perales, S. (2014). Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nature communications* 5, 3964.

Tort, F., Campo, E., Pohlman, B., and Hsi, E. (2004). Heterogeneity of genomic breakpoints in MSN-ALK translocations in anaplastic large cell lymphoma. *Human pathology* 35, 1038-1041.

Touriol, C., Greenland, C., Lamant, L., Pulford, K., Bernard, F., Rousset, T., Mason, D.Y., and Delsol, G. (2000). Further demonstration of the diversity of chromosomal changes involving 2p23 in ALK-positive lymphoma: 2 cases expressing ALK kinase fused to CLTCL (clathrin chain polypeptide-like). *Blood* 95, 3204-3207.

Trumper, L., Pfreundschuh, M., Bonin, F.V., and Daus, H. (1998). Detection of the t(2;5)-associated NPM/ALK fusion cDNA in peripheral blood cells of healthy individuals. *British journal of haematology* 103, 1138-1144.

- Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nature methods* *14*, 607-614.
- Turner, S.D., Lamant, L., Kenner, L., and Brugieres, L. (2016). Anaplastic large cell lymphoma in paediatric and young adult patients. *British journal of haematology* *173*, 560-572.
- Turner, S.D., Merz, H., Yeung, D., and Alexander, D.R. (2006). CD2 promoter regulated nucleophosmin-anaplastic lymphoma kinase in transgenic mice causes B lymphoid malignancy. *Anticancer research* *26*, 3275-3279.
- Turner, S.D., Tooze, R., Maclennan, K., and Alexander, D.R. (2003). Vav-promoter regulated oncogenic fusion protein NPM-ALK in transgenic mice causes B-cell lymphomas with hyperactive Jun kinase. *Oncogene* *22*, 7750-7761.
- Tycko, B., Reynolds, T.C., Smith, S.D., and Sklar, J. (1989). Consistent breakage between consensus recombinase heptamers of chromosome 9 DNA in a recurrent chromosomal translocation of human T cell leukemia. *The Journal of experimental medicine* *169*, 369-377.
- Vaites, L.P., Lian, Z., Lee, E.K., Yin, B., DeMicco, A., Bassing, C.H., and Diehl, J.A. (2014). ATM deficiency augments constitutively nuclear cyclin D1-driven genomic instability and lymphomagenesis. *Oncogene* *33*, 129-133.
- van Gent, D.C., McBlane, J.F., Ramsden, D.A., Sadofsky, M.J., Hesse, J.E., and Gellert, M. (1995). Initiation of V(D)J recombination in a cell-free system. *Cell* *81*, 925-934.
- Vanoli, F., and Jasin, M. (2017). Generation of chromosomal translocations that lead to conditional fusion protein expression using CRISPR-Cas9 and homology-directed repair. *Methods* *121-122*, 138-145.
- Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell stem cell* *15*, 27-30.
- Vijayvergia, N., and Mehra, R. (2014). Clinical challenges in targeting anaplastic lymphoma kinase in advanced non-small cell lung cancer. *Cancer chemotherapy and pharmacology* *74*, 437-446.
- Voichek, Y., Bar-Ziv, R., and Barkai, N. (2016). Expression homeostasis during DNA replication. *Science* *351*, 1087-1090.
- Wang, C.M., Wu, Z.Q., Wang, Y., Guo, Y.L., Dai, H.R., Wang, X.H., Li, X., Zhang, Y.J., Zhang, W.Y., Chen, M.X., *et al.* (2017). Autologous T Cells Expressing CD30 Chimeric Antigen Receptors for Relapsed or Refractory Hodgkin Lymphoma: An Open-Label Phase I Trial. *Clinical cancer research : an official journal of the American Association for Cancer Research* *23*, 1156-1166.
- Wang, H.Y., and Jaffe, E.S. (2018). An anaplastic T-cell lymphoma mimicking classical Hodgkin lymphoma. *Blood* *131*, 1262.

- Wang, L., Tong, X., Zhou, Z., Wang, S., Lei, Z., Zhang, T., Liu, Z., Zeng, Y., Li, C., Zhao, J., *et al.* (2018). Circular RNA hsa_circ_0008305 (circPTK2) inhibits TGF-beta-induced epithelial-mesenchymal transition and metastasis by controlling TIF1gamma in non-small cell lung cancer. *Molecular cancer* 17, 140.
- Watanabe, M., Sasaki, M., Itoh, K., Higashihara, M., Umezawa, K., Kadin, M.E., Abraham, L.J., Watanabe, T., and Horie, R. (2005). JunB induced by constitutive CD30-extracellular signal-regulated kinase 1/2 mitogen-activated protein kinase signaling activates the CD30 promoter in anaplastic large cell lymphoma and reed-sternberg cells of Hodgkin lymphoma. *Cancer research* 65, 7628-7634.
- Watanabe, Y., Fujiyama, A., Ichiba, Y., Hattori, M., Yada, T., Sakaki, Y., and Ikemura, T. (2002). Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. *Human molecular genetics* 11, 13-21.
- Waters, C.A., Strande, N.T., Pryor, J.M., Strom, C.N., Mieczkowski, P., Burkhalter, M.D., Oh, S., Qaqish, B.F., Moore, D.T., Hendrickson, E.A., *et al.* (2014). The fidelity of the ligation step determines how ends are resolved during nonhomologous end joining. *Nature communications* 5, 4286.
- Weber, C.C., Pink, C.J., and Hurst, L.D. (2012). Late-replicating domains have higher divergence and diversity in *Drosophila melanogaster*. *Molecular biology and evolution* 29, 873-882.
- Wei, J., Wunderlich, M., Fox, C., Alvarez, S., Cigudosa, J.C., Wilhelm, J.S., Zheng, Y., Cancelas, J.A., Gu, Y., Jansen, M., *et al.* (2008). Microenvironment determines lineage fate in a human model of MLL-AF9 leukemia. *Cancer cell* 13, 483-495.
- Weinstein, I.B., Begemann, M., Zhou, P., Han, E.K., Sgambato, A., Doki, Y., Arber, N., Ciaparrone, M., and Yamamoto, H. (1997). Disorders in cell circuitry associated with multistage carcinogenesis: exploitable targets for cancer prevention and therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* 3, 2696-2702.
- Weinstock, D.M., Brunet, E., and Jasin, M. (2007). Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. *Nature cell biology* 9, 978-981.
- Weiss, J.B., Xue, C., Benice, T., Xue, L., Morris, S.W., and Raber, J. (2012). Anaplastic lymphoma kinase and leukocyte tyrosine kinase: functions and genetic interactions in learning, memory and adult neurogenesis. *Pharmacology, biochemistry, and behavior* 100, 566-574.
- Westholm, J.O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S.E., Graveley, B.R., and Lai, E.C. (2014). Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell reports* 9, 1966-1980.
- Wiemels, J.L., Cazzaniga, G., Daniotti, M., Eden, O.B., Addison, G.M., Maser, G., Saha, V., Biondi, A., and Greaves, M.F. (1999). Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* 354, 1499-1503.
- Wright, R.L., and Vaughan, A.T. (2014). A systematic description of MLL fusion gene formation. *Critical reviews in oncology/hematology* 91, 283-291.

Wu, F., Wang, P., Young, L.C., Lai, R., and Li, L. (2009). Proteome-wide identification of novel binding partners to the oncogenic fusion gene protein, NPM-ALK, using tandem affinity purification and mass spectrometry. *The American journal of pathology* *174*, 361-370.

Wu, M.H., and Yung, B.Y. (2002). UV stimulation of nucleophosmin/B23 expression is an immediate-early gene response induced by damaged DNA. *The Journal of biological chemistry* *277*, 48234-48240.

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., and Zhang, B. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* *30*, 1180-1182.

Xie, S., Shen, B., Zhang, C., Huang, X., and Zhang, Y. (2014). sgRNAs9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PloS one* *9*, e100448.

Yaffe, E., Farkash-Amar, S., Polten, A., Yakhini, Z., Tanay, A., and Simon, I. (2010). Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS genetics* *6*, e1001011.

Yoneda-Kato, N., Look, A.T., Kirstein, M.N., Valentine, M.B., Raimondi, S.C., Cohen, K.J., Carroll, A.J., and Morris, S.W. (1996). The t(3;5)(q25.1;q34) of myelodysplastic syndrome and acute myeloid leukemia produces a novel fusion gene, NPM-MLF1. *Oncogene* *12*, 265-275.

Young, L.C., Bone, K.M., Wang, P., Wu, F., Adam, B.A., Hegazy, S., Gelebart, P., Holovati, J., Li, L., Andrew, S.E., *et al.* (2011). Fusion tyrosine kinase NPM-ALK Deregulates MSH2 and suppresses DNA mismatch repair function novel insights into a potent oncoprotein. *The American journal of pathology* *179*, 411-421.

Yu, C.Y., Li, T.C., Wu, Y.Y., Yeh, C.H., Chiang, W., Chuang, C.Y., and Kuo, H.C. (2017). The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency. *Nature communications* *8*, 1149.

Zamo, A., Chiarle, R., Piva, R., Howes, J., Fan, Y., Chilosi, M., Levy, D.E., and Inghirami, G. (2002). Anaplastic lymphoma kinase (ALK) activates Stat3 and protects hematopoietic cells from cell death. *Oncogene* *21*, 1038-1047.

Zhang, M., Wang, S., Tang, L., Wang, X., Zhang, T., Xia, X., and Fang, X. (2019a). Downregulated circular RNA hsa_circ_0067301 regulates epithelial-mesenchymal transition in endometriosis via the miR-141/Notch signaling pathway. *Biochemical and biophysical research communications* *514*, 71-77.

Zhang, Q., Wang, H.Y., Woetmann, A., Raghunath, P.N., Odum, N., and Wasik, M.A. (2006). STAT3 induces transcription of the DNA methyltransferase 1 gene (DNMT1) in malignant T lymphocytes. *Blood* *108*, 1058-1064.

Zhang, Q., Wei, F., Wang, H.Y., Liu, X., Roy, D., Xiong, Q.B., Jiang, S., Medvec, A., Danet-Desnoyers, G., Watt, C., *et al.* (2013a). The potent oncogene NPM-ALK mediates malignant transformation of normal human CD4(+) T lymphocytes. *The American journal of pathology* *183*, 1971-1980.

Zhang, X., Wang, S., Wang, H., Cao, J., Huang, X., Chen, Z., Xu, P., Sun, G., Xu, J., Lv, J., *et al.* (2019b). Circular RNA circNRIP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway. *Molecular cancer* 18, 20.

Zhang, X.O., Wang, H.B., Zhang, Y., Lu, X., Chen, L.L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. *Cell* 159, 134-147.

Zhang, Y., and Jasin, M. (2011). An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nature structural & molecular biology* 18, 80-84.

Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148, 908-921.

Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.L., Yang, L., and Chen, L.L. (2016). The Biogenesis of Nascent Circular RNAs. *Cell reports* 15, 611-624.

Zhang, Y., Zhang, X.O., Chen, T., Xiang, J.F., Yin, Q.F., Xing, Y.H., Zhu, S., Yang, L., and Chen, L.L. (2013b). Circular intronic long noncoding RNAs. *Molecular cell* 51, 792-806.

Zhao, Y., Alexandrov, P.N., Jaber, V., and Lukiw, W.J. (2016). Deficiency in the Ubiquitin Conjugating Enzyme UBE2A in Alzheimer's Disease (AD) is Linked to Deficits in a Natural Circular miRNA-7 Sponge (circRNA; ciRS-7). *Genes* 7.

Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., Luo, Y., Lyu, D., Li, Y., Shi, G., *et al.* (2016). Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nature communications* 7, 11215.

Zheng, X., Oancea, C., Henschler, R., Moore, M.A., and Ruthardt, M. (2009). Reciprocal t(9;22) ABL/BCR fusion proteins: leukemogenic potential and effects on B cell commitment. *PloS one* 4, e7661.

Zhong, Z., Huang, M., Lv, M., He, Y., Duan, C., Zhang, L., and Chen, J. (2017). Circular RNA MYLK as a competing endogenous RNA promotes bladder cancer progression through modulating VEGFA/VEGFR2 signaling pathway. *Cancer letters* 403, 305-317.

Zhu, L.J., Holmes, B.R., Aronin, N., and Brodsky, M.H. (2014). CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PloS one* 9, e108424.

Zou, H.Y., Friboulet, L., Kodack, D.P., Engstrom, L.D., Li, Q., West, M., Tang, R.W., Wang, H., Tsaparikos, K., Wang, J., *et al.* (2015). PF-06463922, an ALK/ROS1 Inhibitor, Overcomes Resistance to First and Second Generation ALK Inhibitors in Preclinical Models. *Cancer cell* 28, 70-81.

ARTICLES EN COLLABORATION

Article 1 : CRISPR-Cas9-guided oncogenic chromosomal translocations with conditional fusion protein expression in human mesenchymal cells.

Vanoli F, Tomishima M, Feng W, Lamribet K, **Babin L**, Brunet E, Jasin M.

expression, all four clones demonstrated the *EWSR1-WT1* fusion separated by the LoxP site (Fig. 1D and E). Before Cre expression, two of the clones (13 and 63) also gave rise to the out-out PCR product expected for the *EWSR1-WT1* fusion containing the aminopyrimidine-resistance gene (Fig. 1C and D). This amplification product was not obtained from the other two clones (20 and 40) upon multiple attempts. Because these two clones show clear evidence of the desired translocation by out-out PCR after Cre expression (see also below), a possible explanation is that the donor plasmid integrated in such a way as to abrogate efficient PCR across the selectable marker. In this strategy, a splice PCR across the EWSR1-*WT1* fusion was used to identify out-out PCR. This, screening clones after Cre expression by out-out PCR may identify additional clones harboring translocations.

To formally demonstrate the (1:122) reciprocal translocation, we performed FISH with a dual-fusion probe. Whereas the parental line shows discrete green and red signals from *EWSR1* and *WT1*, respectively, overlapping signals indicative of the *EWSR1-WT1* were observed with clone 63 (Fig. 1F).

Expression of the *EWSR1-WT1* Fusion in Translocation Clones. We next performed RT-PCR analysis to detect the *EWSR1-WT1* fusion transcript and to determine whether its expression was constitutive or dependent on Cre expression. Using primers to *EWSR1* exon 7 and *WT1* exon 10, the properly spliced fusion transcript between *EWSR1* exon 7 and *WT1* exon 8 was observed in clones 13 and 63 both before and after Cre expression, whereas it was only detected at significant levels in clones 20 and 40 after Cre expression (Fig. 2A and B). Constitutive expression of the *EWSR1-WT1* fusion transcript in clones 13 and 63 indicates that the pyrimidine-resistance gene does not completely interfere with splicing between *EWSR1* exon 7 and *WT1* exon 8, despite the presence of a polyA site in the sequences inserted in the intron. However, Cre-dependent expression in clones 20 and 40 suggests that the donor integration (e.g., a possible tandem integration) markedly reduces splicing of these exons.

To verify that the transcript analysis is reduced in *EWSR1-WT1* protein expression, Western blots were performed using an antibody to *WT1*, which detected the putative 394-Dx fusion protein in clones 13 and 63 but not in clones 20 and 40. Consistent with the RT-PCR results, the *EWSR1-WT1* protein was expressed in clones 13 and 63 but not constitutively expressed in clones 20 and 40. The *WT1* portion of the *EWSR1-WT1* pre-mRNA is subject to alternative splicing that leads to the retention (or not) of the three amino acids lysine, threonine, and serine, +KTS, between two zinc fingers, which profoundly affects the oncogenic function of the protein (27, 32). Both splice variants are present in the *EWSR1-WT1* fusion transcript, with the +KTS variant somewhat

more abundant than the -KTS variant (Fig. 2D). Thus, we were able to generate cell lines expressing the *EWSR1-WT1* fusion that recapitulates the splice products seen in DSRCT.

Forced Conditional Expression of the *EWSR1-WT1* Fusion After Translocation. These studies provide a proof of principle for the HDR-based approach for generating and selecting translocations of interest. However, because integration of the transcription unit of the selectable marker between the *EWSR1* and *WT1* fusion was not sufficient to block fusion protein expression, we modified the approach to incorporate a gene trap, which has been successfully used to knock out a gene using Cas9 (e.g., ref. 33). In this strategy, a splice acceptor site is present in the donor plasmid in such a way as to lead to expression of a selectable marker, in this case hygromycin resistance, from the *EWSR1* promoter upon HDR with the loxP (Fig. 3A). Importantly, the splice acceptor associated with the selectable marker would be predicted to prevent expression of the *EWSR1-WT1* fusion transcript after translocation. As with the prior approach (Fig. 1B), the selectable marker is flanked by LoxP sites and so can be deleted by Cre expression, leading to *EWSR1-WT1* expression.

We tested this modified approach in genetically unmodified, multipotent human mesenchymal (nonimmortalized) cells derived from human embryonic stem cells (34), which have been previously used to generate random *EWSR1-FLI1* fusions in translocated cells (Fig. 1F). These cells maintain a diploid chromosome complement in culture (Fig. 1F). Stable expression of multiple mesenchymal markers (CD44, CD73, and CD105) is also observed (Fig. S1C) (34). Because the +KTS and -KTS isoforms are not induced in these cells, neither *EWSR1-WT1* expression is sufficient to transform the cells into cancerous growth with the Cas9 and gRNA expression vectors for HDR on 1000 cells per well in three 96-well plates and treated with the DNA-PKcs inhibitor for 48 h, 5 d after transfection. Cells were exposed to hygromycin for positive selection. Hygromycin-resistant cells grew in 25 wells and were analyzed by PCR for HDR with the donor plasmid at the *EWSR1* and *WT1* loci (Fig. 3B and C and Fig. S2A). Cells in all 25 wells had undergone HDR with the donor plasmid at the *EWSR1* gene (5' out-in PCR). Ten of the wells were also positive for HDR with the donor plasmid at the *WT1* gene (3' in-out PCR), three of which were positive for fusion of the *EWSR1* and *WT1* genes (out-out PCR). Without the DNA-PKcs inhibitor, fewer wells had hygromycin-resistant cells, and none of these were positive for the out-out PCR, suggesting that the inhibitor may promote the recovery of clones with a translocation.

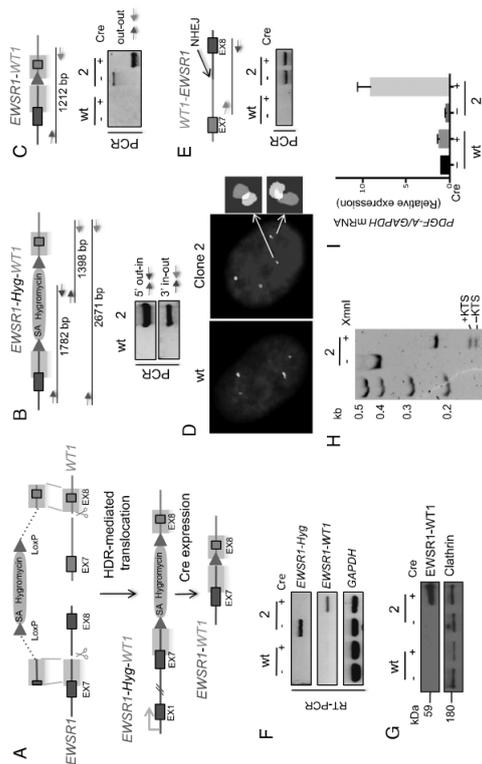


Fig. 3. Gene trap strategy to generate the *EWSR1-WT1* translocation in nonimmortalized human mesenchymal stem cells with conditional fusion protein expression. (A) HDR targeting strategy using a gene trap approach. A splice acceptor (SA) upstream of a hygromycin selectable marker in the donor plasmid enables selection of the *EWSR1-WT1* fusion. (B) The resulting *EWSR1-Hyg-WT1* fusion. (C) RT-PCR analysis of the translocation shows correct HDR with the donor plasmid at the *EWSR1* and *WT1* loci. See Fig. S2A for additional detailed analysis. (D) PCR across the breakpoint junction before and after Cre recombinase expression. A shift in the PCR product size shows loss of the hygromycin selectable marker following Cre expression. (E) Interphase FISH analysis showing the (1:122) reciprocal translocation, *EWSR1-WT1* and *WT1-EWSR1*, and unrearranged chromosomes 11 and 22. The FISH scheme is the same as that in Fig. 1F. (See Fig. S2B and C for sequence analysis of breakpoint junctions.) (F) NHE-mediated reciprocal *WT1-EWSR1* fusion in clone 2, which is unchanged after Cre expression. (G) RT-PCR and (H) Western analysis demonstrating conditional *EWSR1-WT1* expression. Before Cre expression, only the *EWSR1-Hyg* transcript is observed, whereas post-Cre, only the *EWSR1-WT1* transcript is observed, together with the fusion protein. (I) Poly-ubiquitinated EWSR1-WT1 target PDGF-A in clone 2 cells, as demonstrated by quantitative RT-PCR.

Cells in two wells grew poorly and in the third (clone 2) grew slowly but could be expanded for further analysis. Upon Cre expression, the direct fusion of *EWSR1* and *WT1* was confirmed by a shift in the PCR fragment size (Fig. 3C), and sequencing demonstrated retention of the LoxP site at the fusion, as in Fig. 1E.

Cells were tested by FISH before Cre expression to verify the parental cells gave two signals each for chromosomes 11 and 22, clone 2 cells showed only one signal each for initial chromosomes 11 and 22 but additional signals for two derivative chromosomes, indicating a reciprocal translocation.

Whereas the *EWSR1-WT1* fusion forms by HDR, the *WT1-EWSR1* fusion would be expected to form by NHEJ because the donor plasmid does not have homology to these other two chromosomes ends. To determine the nature of the fusion event, PCR with primers flanking the expected breakpoint junction, the *EWSR1* gene, and the *WT1* gene was performed. The *EWSR1* and *WT1* genes were sequenced (Fig. 3E). Sequencing confirmed the *WT1* and *EWSR1* fusion (Fig. S2B). The *WT1-EWSR1* junction apparently arose by NHEJ with an insertion, as found in other oncogenic translocations (e.g., see ref. 11). The *WT1-EWSR1* junction was sequenced from a second clone (clone 1), which also showed an NHEJ event in this case a simple deletion. We also examined the unrearranged chromosomes in clone 2 to determine whether

the gRNA target sites were also cleaved on them; both showed NHEJ junctions at the gRNA sites within the respective *EWSR1* and *WT1* introns which would have abrogated gRNA binding and subsequent rounds of cleavage by Cas9 (Fig. S2C).

Before Cre expression, RT-PCR showed expression of the *EWSR1-Hyg* transcript, whereas the *EWSR1-WT1* fusion was not detectable (Fig. 3F), indicating the effectiveness of the splice acceptor strategy for impeding fusion gene expression. After Cre exposure and deletion of the selectable marker, expression of the *EWSR1-WT1* fusion was apparent both at the mRNA (Fig. 3F) and protein (Fig. 3G) levels, demonstrating that the approach leads to conditional expression. The *EWSR1-WT1* fusion protein product also retained both KTS splice variants (Fig. 3F).

In DSRCT, PDGF-A, a potent fibroblast growth factor, has been reported to be a major target of EWSR1-WT1 (9). Out-of-frame RT-PCR (RT-PCR) confirmed that the PDGF-A transcript (detected in Cre control cells) (Fig. 3I). These results indicate that our cell system recapitulates properties of the *EWSR1-WT1* fusion product seen in DSRCT. However, *EWSR1-WT1* was not sufficient to transform the cells to overcome growth arrest. Whereas both control and clone 2 cells proliferated slowly upon subsequent passages, clone 2 ceased proliferating after Cre expression, suggesting that *EWSR1-WT1* expression in non-immortalized cells is not sufficient for cellular transformation.

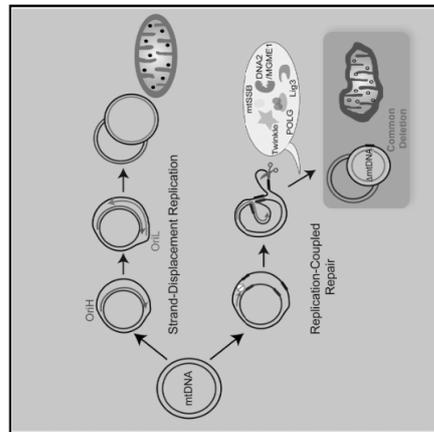
Article 2: Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion.

Phillips AF, Millet AR, Tigano M, Dubois SM, Crimmins H, **Babin L**, Charpentier M, Piganeau M, Brunet E, Sfeir A.

Molecular Cell

Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion

Graphical Abstract



Highlights

- Development of mito-SMARD, a single-molecule approach to study mtDNA replication
- In vivo analysis validates the strand displacement mode of mtDNA replication
- Twinkle helicase mutations promote fork stalling near the common deletion
- Repair of mito-TALEN-induced breaks triggers the common deletion

Phillips et al., 2017, *Molecular Cell* 65, 1–12
 February 2, 2017 © 2016 Elsevier Inc.
<http://dx.doi.org/10.1016/j.molcel.2016.12.014>

CellPress

Article

Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion

Aaron F. Phillips,^{1,4} Armé R. Millet,^{2,3,4} Marco Tigano,¹ Sonia M. Dubois,² Hannah Crimmins,¹ Loelia Babin,^{2,3} Aurine Charpenet,² Marion Piganeau,² Erika Brunet,² and Agnès Steir,^{1,4} *1*INSERM U1154, Centre for Mitochondrial Research, Institut Pasteur, Paris, France; *2*INSERM U1154, CNRS UMR 7168, 75005 Paris, France; *3*Structure et Instabilité des Gélocomes, Madsém National d'Histologie Normale, INSERM U1154, CNRS UMR 7168, 75005 Paris, France; *4*Genome Dynamics in the Immune System Laboratory, INSERM, UMR 1163, Institut Imagine, 75015 Paris, France

*Co-first author

†Lead Contact

Correspondence: erika.brunet@inserm.fr (E.B.), agnel.steir@med.nyu.edu (A.S.)
<http://dx.doi.org/10.1016/j.molcel.2016.12.014>

SUMMARY

Mutations in mtDNA lead to muscular and neurological diseases and are linked to aging. The most frequent aberrancy is the “common deletion,” that involves a 4,977-bp region flanked by 13-bp repeats. To investigate the basis of this deletion, we developed a single-molecule mtDNA combing method. The analysis of replicating mtDNA molecules provided *in vivo* evidence in support of the asymmetric mode of replication. Furthermore, we observed frequent fork stalling at the junction of the common deletion, suggesting that impaired replication triggers the formation of this toxic lesion. In parallel experiments, we employed mito-TALENs to induce breaks in distinct loci of the mitochondrial genome and found that breaks adjacent to the 5' repeat trigger the common deletion. Interestingly, this process was mediated by the mitochondrial replisome independent of canonical DSB repair. Altogether, our data underscore a unique replication-dependent repair pathway that leads to the mitochondrial common deletion.

INTRODUCTION

Human mtDNA consists of a 16.6-kb double-stranded circular genome and encodes 37 genes that are essential for oxidative phosphorylation (Anderson et al., 1981). In addition to point mutations, mtDNA abnormalities comprise several large-scale deletions. The most common aberrancy is a 4,977-bp deletion spanning nucleotides 8,483–13,459 of the mitochondrial genome. The common deletion (also known as mtDNA⁴⁹⁷⁷) was first detected in neuromuscular diseases—Kearns-Sayre syndrome (KSS) and progressive external ophthalmoplegia (PEO) (Shoffner et al., 1989)—and later found to accumulate in aging tissues (Corral-Debrinski et al., 1992; Piko et al., 1988; Yang et al.,

1994). The loss of 12 essential mitochondrial genes within the deleted region cripples mitochondrial biogenesis. As a result, the common deletion is heteroplasmic. In nature, manifesting clinically when mtDNA⁴⁹⁷⁷ molecules exceed a critical threshold (Wallace and Chalkia, 2013). The molecular mechanism underlying the formation of the common deletion remains unknown. The deleted region is flanked by two 13-bp repeats, with only one repeat being retained in mtDNA⁴⁹⁷⁷ (Schon et al., 1988). This characteristic feature, shared by the majority of mtDNA deletions, implicates replication slippage or aberrant double-strand break (DSB) repair during the process of common deletion formation (Kishan et al., 2008; Shrivastava and Moraes, 2005; Figure S1A, available online).

Over the years, the mode of mtDNA replication has been a matter of controversy (Figure S1B). The strand-displacement model was put forth in the 1970s based on the visualization of replicating mtDNA intermediates by electron microscopy (EM) (Robberson and Clayton, 1972). According to this model, mtDNA synthesis employs two distinct replication origins marked by unidirectional fork progression. Heavy-strand synthesis initiates from the origin of heavy-strand replication (OH), located in the displacement loop, and advances for ~11 kb before the synthesis of the light chain initiates from the origin of light-strand replication (OL) in the opposite direction. Later, Holt et al. (2000) proposed the strand-coupled model based on 2D gel electrophoresis analyses of mtDNA replication intermediates. According to this model, heavy- and light-strand replication is coupled and proceeds bi-directionally from a broad initiation zone downstream of OH (Holt et al., 2000). The strand-coupled model was later amended to include extensive RNA priming of the lagging strand, which is then replaced by DNA in subsequent maturation steps (Yang et al., 2002; Yasukawa et al., 2006).

mtDNA replication is accomplished with the help of a unique set of enzymes, distinct from the nuclear replisome. The minimal mitochondrial replisome was reconstituted *in vitro* and consists of polymerase gamma (POLG), the hexameric Twinkle helicase, and the mitochondrial single-stranded DNA binding protein (mtSSB) (Korhonen et al., 2004). Additional factors were found to assist mtDNA synthesis, including the transcription and

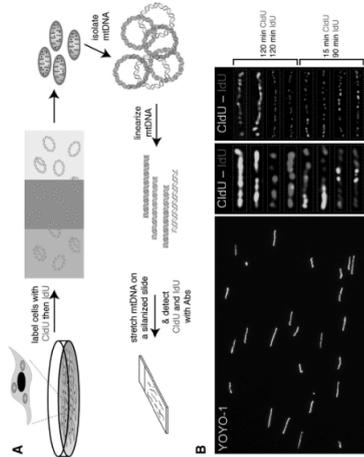


Figure 1. A Single-Molecule DNA Combining Approach to Visualize mtDNA Replication
(A) Schematic of the experimental workflow. Cells were labeled with IdU, mtDNA was stretched on a glass slide, and IdU was combined with CldU. mtDNA was purified and digested with the AflI restriction enzyme, which cleaves upstream of OriH. Linearized DNA was then stretched on a silanized glass slide and stained with antibodies that distinguish CldU (green) from IdU (red).
(B) Left: stretched mtDNA molecules imaged with a Nikon TE2000-SP microscope. Right: super-resolution microscopy (Nikon Eclipse Ti-V) using a 100 \times 1.45 NA objective. Shown are mtDNA molecules replicating in the presence of CldU (green molecules) and IdU (red molecules) as well as those replicating during the switch between CldU and IdU (dualy labeled molecules). Right: mtDNA molecules detected by mito-SMARD and analyzed by POLRMT. Scale bars: 100 nm (left); 50 nm (right). Scale bars are 100 nm (left) and 50 nm (right).

Here we define the underlying molecular mechanism of mtDNA^{del7} formation in mammalian cells and distinguish between the contribution of replication slippage and DSB repair. We developed a single-molecule assay based on DNA combing technology to examine key features (mode, origin, and directionality) of mtDNA replication and investigate stalling of the replisome in the vicinity of the common deletion. In parallel, using mitochondrially targeted transcription activator-like effector nucleases (mito-TALENs) to induce breaks in specific loci of the mitochondrial genome, we investigated the underlying genetic basis of mitochondrial DSB repair. Our results identify a unique replication-dependent repair pathway that triggers the formation of the mitochondrial common deletion. Understanding the basis of this widespread aberration is an essential first step to prevent its accumulation in mitochondrial diseases.

RESULTS

A Single-Molecule Approach to Study mtDNA Replication

We aimed to investigate the dynamics of mtDNA replication at the single-molecule level and with high resolution. To do so, we adapted the single-molecule analysis of replicating DNA (SMARD) technique (Norie and Schlickraut, 2001), a DNA combing method that has been widely applied to study the replication of specific loci within the nuclear genome. This method, which we named mito-SMARD (Figure 1A), is based on labeling cells sequentially with two halogenated thymidine derivatives (5-iodo-2'-deoxyuridine [IdU] and 5-Chloro-2'-deoxyuridine [CldU]) that are incorporated into replicating DNA molecules. After a short recovery period, during which cells are incubated with media not supplemented with any thymidine derivatives, circular mtDNA was isolated from lysed cell extracts using a miniprep kit (Culspise-Tripathy et al., 2013). The DNA was linearized with a restriction enzyme that cleaves upstream of OriH and resolved by gel electrophoresis. The isolated 16.6-kb mtDNA fragment

packaging protein TFAM (Fallenberg et al., 2007), the mitochondrial RNA polymerase POLRMT (Wanrooij et al., 2009), and the mitochondrial exonuclease MGME1 (Kornblum et al., 2013). Mutations in POLG and Twinkle have been implicated in mitochondrial disease and linked to large-scale mtDNA deletions (Szelbrink et al., 2001; Tynjismaa et al., 2005; Van Goethem et al., 2001). 2D gel analysis of mtDNA in cells expressing disease-associated Twinkle mutations highlights the presence of replication intermediates (Goffart et al., 2009; Wanrooij et al., 2007), indicative of replication fork pausing along the major arc of the mitochondrial genome. These observations provide circumstantial evidence in support of replication slippage as the basis of the common deletion (Figure S1A).

Originally, it was assumed that mitochondria lack the basic machinery to fix DSBs. However, recent reports have described DSB repair activities in this organelle (Blicman et al., 2009; Kraysberg et al., 2010; Simsek et al., 2010), and several nuclear DSB repair proteins appear to localize to mitochondria (Omitrieva et al., 2011; Sage et al., 2010; Simsek et al., 2011). Furthermore, introducing DSBs with restriction endonucleases triggers the accumulation of deletions across the mitochondrial genome (Shvastava and Moraes, 2005), thus implicating aberrant DSB repair during the process of deletion formation. The 13-bp repeats flanking the junction of the common deletion are predicted to promote microhomology-mediated end joining (MMEJ) of broken DNA (Samuels, 2004; Shoffner et al., 1989; Shvastava and Moraes, 2005; Figure S1A). MMEJ is an error-prone pathway of DSB repair that is associated with both deletions and the usage of microhomology and is dependent on LIG3, MRE11/CtIP, and PARP1 (Steir and Symington, 2015). Although mitochondrial protein extracts are known to stimulate MMEJ *ex vivo* (Tadi et al., 2016), *in vivo* evidence in support of a role for this erroneous pathway in common deletion formation is lacking.

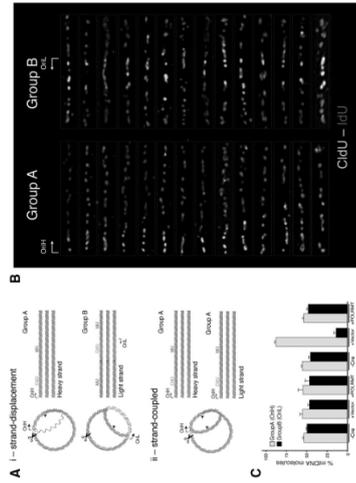


Figure 2. Mito-SMARD Highlights the Mode of mtDNA Replication
(A) Schematic of two proposed models for mtDNA replication based on the strand-coupled or strand displacement models. Taking into consideration dual-labeled molecules, the strand-coupled model is expected to yield one replication pattern in which green is followed by red. In contrast, according to the strand displacement model, two independent groups of molecules will be visualized: one with green followed by red (Group A) and one with green followed by red (Group B).
(B) Mito-SMARD analysis of mtDNA in HT1080 cells yields molecules that are categorized in two distinct groups: group A molecules show CldU (green) followed by IdU (red) labeling, whereas group B molecules show an IdU-CldU labeling pattern.
(C) Visualization of the frequency of group A versus group B molecules in *Poimv*^{+/+} and *Poimv*^{-/-} cells with the indicated treatment. Cells were analyzed 48 hr after induction of a 4-hydroxytamoxifen-inducible Cre-ER⁺. Average frequencies were calculated from three independent experiments ($n \geq 100$ molecules in each experiment) with SD.

that mito-SMARD analysis would allow us to distinguish between the two opposing models for mtDNA replication (Figure 2A). Should the mtDNA be replicated asymmetrically—according to the strand-displacement model—mito-SMARD would yield molecules with two distinct labeling patterns (groups A and B) (Figure 2A). Heavy-strand synthesis from OriH will appear as molecules labeled with CldU in green at the 5' end, followed by IdU in red (group A), whereas replication of light-strand through OriL will result in a set of molecules in which CldU is flanked with IdU on both sides (group B). In contrast, the strand-coupled model, which involves firing of a single origin, is expected to yield one labeling pattern. Specifically, synchronous replication of mtDNA molecules from OriH would result in a labeling pattern consisting of green (CldU) followed by red (IdU) (group A) (Figure 2A). Inspection of dual-labeled mtDNA molecules in human and mouse cells revealed two distinct labeling patterns representing group A and group B (Figures 2B and S2B). The visualization of these two configurations with similar frequency (Figures 2B, 2C, and S2B) underscores efficient firing of both OriH and OriL and is, therefore, consistent with the strand-displacement model of mtDNA replication (Clayton, 1982; Robberson and Clayton, 1972).

The initiation of light-strand synthesis has been proposed to be mediated by POLRMT, an RNA polymerase that acts as an OriL-specific primase upon recognizing a stem-like loop within the lagging-strand origin (Fusie et al., 2010; Wanrooij et al., 2008). To validate the mito-SMARD results and confirm that group B molecules are products of OriL firing, we tested the effect of depleting POLRMT on the efficiency of lagging-strand synthesis *in vivo*. To do so, we treated *Poimv*^{+/+} and *Poimv*^{-/-} mouse embryonic fibroblasts (MEFs) (Kuhl et al., 2014) with Cre recombinase (Figure S3A). POLRMT-deficient and -proficient cells were pulse-labeled with CldU and IdU 2 days after Cre treatment and analyzed by mito-SMARD. We chose to investigate an early

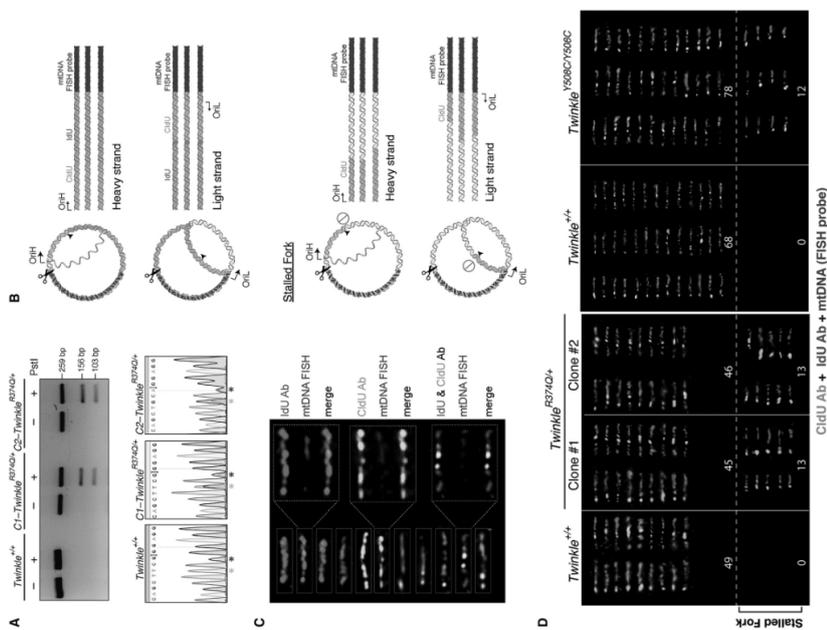


Figure 3. Evidence for Replication Fork Stalling in Cells Expressing Mutant Twinkle
 (A) Schematic PCR to detect heterocytous *Twinkle*^{R374Q/R508C} associated by CRISPR/Cas9 targeting of PPE-1. The PCR product was treated with a restriction enzyme that specifically cleaves the targeted allele. Two independently derived clonal cell lines were generated. Targeting was confirmed by Sanger sequencing of cDNA from the indicated clones. The asterisks mark mutated residues, including the R374Q mutation and a silent mutation in the protospacer adjacent motif (PAM) sequence.
 (B) Schematic for Replication Fork Stalling in Cells Expressing Mutant Twinkle. The PCR product was treated with a restriction enzyme that specifically cleaves the targeted allele. Two independently derived clonal cell lines were generated. Targeting was confirmed by Sanger sequencing of cDNA from the indicated clones. The asterisks mark mutated residues, including the R374Q mutation and a silent mutation in the protospacer adjacent motif (PAM) sequence.
 (C) FISH images showing mtDNA (green) and Cyt b (red) in *Twinkle*^{+/+}, *Twinkle*^{R374Q}, and *Twinkle*^{R508C} cells. Scale bars, 10 μm. The asterisks mark mtDNA molecules that have initiated replication during the IdU pulse to fully complete synthesis (rate of mtDNA synthesis, ~190 nt/min) (data not shown; Clayton, 1982; Kohonen et al., 2004).
 (D) FISH images showing mtDNA (green) and Cyt b (red) in *Twinkle*^{+/+}, *Twinkle*^{R374Q}, and *Twinkle*^{R508C} cells. Scale bars, 10 μm. The asterisks mark mtDNA molecules that have initiated replication during the IdU pulse to fully complete synthesis (rate of mtDNA synthesis, ~190 nt/min) (data not shown; Clayton, 1982; Kohonen et al., 2004).

(legend continued on next page)

time point to avoid the confounding effect of POLRMT inhibition on mtDNA transcription and mtDNA content (Figures S3A–S3E). Depletion of POLRMT impaired firing of OHL, as depicted by a significant reduction in the frequency of group B molecules (Figure 2C). As a control, stable expression of exogenous POLRMT in cells deficient for this RNA polymerase rescued OHL firing (Figures 2C and S3F). The demonstration that POLRMT-mediated priming initiates mtDNA synthesis from OHL provides further evidence in support of the strand-displacement model of mtDNA replication (Figure S1B, top; Clayton, 1982; Robberson and Clayton, 1972) and argues against both strand-coupled mtDNA replication (Figure S1B, center; Holt et al., 2000) and discontinuous lagging-strand synthesis via the RNA incorporation throughout the lagging strand (RTOLIS) model (Figure S1B, bottom; Yang et al., 2002; Yasukawa et al., 2009).

Using CRISPR/Cas9 to Establish Cells with Disease-Associated Twinkle Mutations

Having developed mito-SMARD to explore the link between common deletion and replication infidelity, we next tested whether disease-associated Twinkle mutations, previously linked to the formation of the mitochondrial common deletion (Spelbrink et al., 2001; Tynimäa et al., 2009), induce replication fork stalling in the vicinity of the 4,977-bp region. Pausing of the fork is expected to foster the mis-annealing of the flanking repeats, a key step during the genesis of the common deletion. In the first instance, we utilized the previously established Fip-1-T-Rex-inducible cell lines that express wild-type and mutant variants of Twinkle (Giffart et al., 2009; Figure S4A; data not shown). Induction of Twinkle in this heterocytous setting was achieved by treating cells with doxycycline. Upon examining the mtDNA replication profile, we observed that the expression of either wild-type or mutant Twinkle led to similar impairment of mtDNA replication. Specifically, analysis of the ratio of group A to group B molecules revealed a significant bias toward mtDNA replication through OHL in cells treated with doxycycline (Figure S4B; data not shown). These results underscore the limitation of the Fip-1 system when examining the function of mitochondrial replication factors. It is possible that overexpression of a core component of the replisome sequesters the remaining subunits of the complex, leading to dysregulated mitochondrial replication. Alternatively, tetracycline was recently shown to impair mitochondrial function (Moullan et al., 2015) and could indirectly affect mtDNA replication.

To avoid artifacts emanating from Twinkle overexpression and/or doxycycline treatment and to address the function of Twinkle mutations *in vivo*, we employed the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated

protein 9 (Cas9) gene-editing technology and derived independent *Twinkle*^{R374Q} and *Twinkle*^{R508C} sub-clones from normal human somatic RPE-1 cells. R374Q is a dominant mutation in Twinkle that has been previously identified in patients with mitochondrial disorders and lies in the linker region that separates the helicase domain from the N-terminal primase-like domain (Martin-Negrier et al., 2011). Y508C is a recessive disease-associated Twinkle mutation (Nikali et al., 2005) within the helicase domain. Genotyping PCR and sequence analysis revealed mono-allelic editing for the dominant R374Q mutations and bi-allelic editing for the Y508C mutant (Figures 3A, S4C, and S4D). The mtDNA content in Twinkle-mutant cells, as determined by qPCR, was similar to that of wild-type cells, confirming that the presence of the disease-linked mutations did not completely abolish mtDNA replication (Figure S4E).

Evidence for Replication Fork Stalling in Cells Expressing Mutant Twinkle

We next investigated whether disease-associated mutations in Twinkle lead to replication fork stalling. We anticipated that transient pausing of the replisome during the 10-min IdU pulse or shortly after the switch to IdU would lead to a class of molecules labeled with green because of IdU incorporation but little or no red denoting IdU incorporation (Figure 3B). To distinguish fork stalling events from partially labeled fragments emanating from broken mtDNA molecules, we further amended mito-SMARD and included a library of fluorescence *in situ* hybridization (FISH) probes to mark the end of the stretched mtDNA molecules (spanning nucleotides 10,000 and 16,400 of the mitochondrial genome) (Figures 3C, S4F, and S4G).

Upon analyzing molecules obtained from independent clonally derived cells comprising *Twinkle*^{R374Q} and *Twinkle*^{R508C} alleles, we observed foci that contained OHL but no IdU labeling (Figures 3D and S4H). Importantly, this pattern, which is indicative of fork stalling, was not identified in cells carrying wild-type Twinkle. Although the number of stalled molecules was small, they were solely found in cells with mutant Twinkle and observed in independent experiments (Figures 3D and S4H), providing qualitative evidence in support of fork stalling in the vicinity of the common deletion. In conclusion, based on the analysis of mtDNA replication *in vivo*, we conclude that the onset of the common deletion in human cells is potentially coupled to replication fork stalling adjacent to the 4,977-bp deleted region.

Breaks at the 5' End of the Deletion Breakpoint Trigger mtDNA^{Δ977}

So far, our analysis of mtDNA replication using mito-SMARD provided *in vivo* evidence in support of the replication model of

(B) Schematic for mito-SMARD to assess replication fork stalling. FISH probes that label the control region of the mitochondrial genome were used to mark the end of the linearized mtDNA molecules stretched on glass slides. Top: schematic of molecules representative of fully replicated mtDNA. Bottom: schematic of partially labeled mtDNA molecules indicative of replication fork stalling.
 (C) Representative mtDNA molecules labeled with OHL (green), IdU (red), and Cyt b (blue) (green-red-blue) and stained with mito-FISH probes (blue). IdU incorporation into mtDNA molecules is observed only in mtDNA molecules that have initiated replication during the IdU pulse (90 min) and fully complete synthesis (rate of mtDNA synthesis, ~190 nt/min) (data not shown; Clayton, 1982; Kohonen et al., 2004).
 (D) Representative mtDNA molecules labeled with OHL (green), IdU (red), and Cyt b (blue) (green-red-blue) and stained with mito-FISH probes (blue). IdU incorporation into mtDNA molecules is observed only in mtDNA molecules that have initiated replication during the IdU pulse (90 min) and fully complete synthesis (rate of mtDNA synthesis, ~190 nt/min) (data not shown; Clayton, 1982; Kohonen et al., 2004).

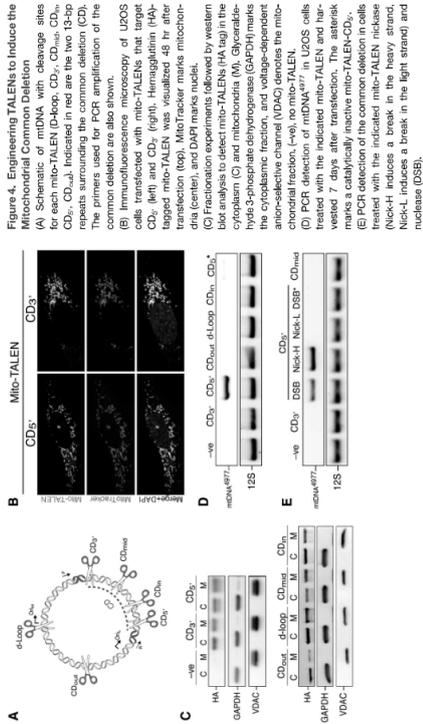


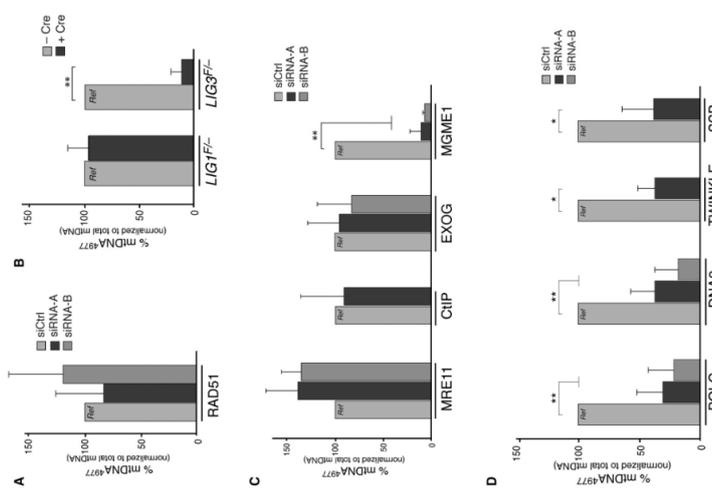
Figure 4. Engineering TALENs to Induce the Common Deletion
(A) Schematic of mtDNA with cleavage sites for each mito-TALEN (D-loop, CD₃, CD₄, CD₅, CD₆, CD₇, CD₈, CD₉, CD₁₀, CD₁₁, CD₁₂, CD₁₃, CD₁₄, CD₁₅, CD₁₆, CD₁₇, CD₁₈, CD₁₉, CD₂₀, CD₂₁, CD₂₂, CD₂₃, CD₂₄, CD₂₅, CD₂₆, CD₂₇, CD₂₈, CD₂₉, CD₃₀, CD₃₁, CD₃₂, CD₃₃, CD₃₄, CD₃₅, CD₃₆, CD₃₇, CD₃₈, CD₃₉, CD₄₀, CD₄₁, CD₄₂, CD₄₃, CD₄₄, CD₄₅, CD₄₆, CD₄₇, CD₄₈, CD₄₉, CD₅₀, CD₅₁, CD₅₂, CD₅₃, CD₅₄, CD₅₅, CD₅₆, CD₅₇, CD₅₈, CD₅₉, CD₆₀, CD₆₁, CD₆₂, CD₆₃, CD₆₄, CD₆₅, CD₆₆, CD₆₇, CD₆₈, CD₆₉, CD₇₀, CD₇₁, CD₇₂, CD₇₃, CD₇₄, CD₇₅, CD₇₆, CD₇₇, CD₇₈, CD₇₉, CD₈₀, CD₈₁, CD₈₂, CD₈₃, CD₈₄, CD₈₅, CD₈₆, CD₈₇, CD₈₈, CD₈₉, CD₉₀, CD₉₁, CD₉₂, CD₉₃, CD₉₄, CD₉₅, CD₉₆, CD₉₇, CD₉₈, CD₉₉, CD₁₀₀. Indicated in red are the two 13-bp repeats surrounding the common deletion (CD). The primers used for PCR amplification of the common deletion are also shown.
(B) Immunofluorescence microscopy of U2OS cells expressing TALEN1 (left) and TALEN2 (right). Heritagelation (HA)-tagged mito-TALEN was visualized 48 hr after transfection (top). MitoTracker marks mitochondria (center), and DAPI marks nuclei.
(C) Fractionation experiments followed by western blot analysis to detect mito-TALENs (HA tag) in the cytoplasm (C) and mitochondria (M). Glyoxal-treated cells were fractionated into soluble and insoluble fractions, and soluble-dependent and insoluble-dependent (VDAC) antibodies were used to detect the mitochondrial fraction (VDAC) (center), and DAPI marks nuclei.
(D) PCR detection of mtDNA⁴⁹⁷⁷ in U2OS cells treated with the indicated mito-TALEN and harvested 7 days after transfection. The asterisk marks a catalytically inactive mito-TALEN-CD₃.
(E) PCR detection of mtDNA⁴⁹⁷⁷ in U2OS cells treated with the indicated mito-TALEN. Nick-I induces a break in the heavy strand, Nick-L induces a break in the light strand and nuclease (DSB).

common deletion formation. In an independent set of experiments, we aimed to investigate the role of DNA break repair by MMEJ during common deletion formation. To do so, we engineered mito-TALENs capable of inducing DSBs in distinct loci within the mitochondrial genome, including D-loop, common deletion (CD₃, CD₄, CD₅, CD₆, CD₇, CD₈, CD₉, CD₁₀, CD₁₁, CD₁₂, CD₁₃, CD₁₄, CD₁₅, CD₁₆, CD₁₇, CD₁₈, CD₁₉, CD₂₀, CD₂₁, CD₂₂, CD₂₃, CD₂₄, CD₂₅, CD₂₆, CD₂₇, CD₂₈, CD₂₉, CD₃₀, CD₃₁, CD₃₂, CD₃₃, CD₃₄, CD₃₅, CD₃₆, CD₃₇, CD₃₈, CD₃₉, CD₄₀, CD₄₁, CD₄₂, CD₄₃, CD₄₄, CD₄₅, CD₄₆, CD₄₇, CD₄₈, CD₄₉, CD₅₀, CD₅₁, CD₅₂, CD₅₃, CD₅₄, CD₅₅, CD₅₆, CD₅₇, CD₅₈, CD₅₉, CD₆₀, CD₆₁, CD₆₂, CD₆₃, CD₆₄, CD₆₅, CD₆₆, CD₆₇, CD₆₈, CD₆₉, CD₇₀, CD₇₁, CD₇₂, CD₇₃, CD₇₄, CD₇₅, CD₇₆, CD₇₇, CD₇₈, CD₇₉, CD₈₀, CD₈₁, CD₈₂, CD₈₃, CD₈₄, CD₈₅, CD₈₆, CD₈₇, CD₈₈, CD₈₉, CD₉₀, CD₉₁, CD₉₂, CD₉₃, CD₉₄, CD₉₅, CD₉₆, CD₉₇, CD₉₈, CD₉₉, CD₁₀₀. We observed that, generation of a DSB internal to the 5' repeat (CD₃) was sufficient to trigger the common deletion. In contrast, breaks incurred within the D-loop, CD₄, CD₅, CD₆, CD₇, CD₈, and CD₉ did not trigger the accumulation of mtDNA⁴⁹⁷⁷ (Figures 4D and 4E). Furthermore, no deletion was formed when a DSB was induced within or external to the 5' repeat (CD₃, CD₄, CD₅, CD₆, CD₇, CD₈, CD₉, CD₁₀, CD₁₁, CD₁₂, CD₁₃, CD₁₄, CD₁₅, CD₁₆, CD₁₇, CD₁₈, CD₁₉, CD₂₀, CD₂₁, CD₂₂, CD₂₃, CD₂₄, CD₂₅, CD₂₆, CD₂₇, CD₂₈, CD₂₉, CD₃₀, CD₃₁, CD₃₂, CD₃₃, CD₃₄, CD₃₅, CD₃₆, CD₃₇, CD₃₈, CD₃₉, CD₄₀, CD₄₁, CD₄₂, CD₄₃, CD₄₄, CD₄₅, CD₄₆, CD₄₇, CD₄₈, CD₄₉, CD₅₀, CD₅₁, CD₅₂, CD₅₃, CD₅₄, CD₅₅, CD₅₆, CD₅₇, CD₅₈, CD₅₉, CD₆₀, CD₆₁, CD₆₂, CD₆₃, CD₆₄, CD₆₅, CD₆₆, CD₆₇, CD₆₈, CD₆₉, CD₇₀, CD₇₁, CD₇₂, CD₇₃, CD₇₄, CD₇₅, CD₇₆, CD₇₇, CD₇₈, CD₇₉, CD₈₀, CD₈₁, CD₈₂, CD₈₃, CD₈₄, CD₈₅, CD₈₆, CD₈₇, CD₈₈, CD₈₉, CD₉₀, CD₉₁, CD₉₂, CD₉₃, CD₉₄, CD₉₅, CD₉₆, CD₉₇, CD₉₈, CD₉₉, CD₁₀₀). Analysis of the junction by next generation (Next-Gen) sequencing confirmed the basis of one of the two 13-bp repeats and the intervening sequences (Figures S5I and S5J). We next assessed whether single-stranded breaks (strand-specific) are sufficient to result in the common deletion. To do so, we used a combination of catalytically inactive and wild-type FokI domains (Wu et al., 2014) and designed mito-TALEN pairs capable of inducing nicks in the heavy or light strands. Our results demonstrate that introducing a single-strand break on the heavy strand of CD₃ led to the accumulation of mtDNA⁴⁹⁷⁷ molecules. However, nicks at CD₄ and

A Critical Role for LIG3 during the Accumulation of mtDNA⁴⁹⁷⁷

The induction of the common deletion in response to mito-TALEN-CD₃ provided us with a robust tool to decipher the genetic requirement for mtDNA⁴⁹⁷⁷ formation. Candidate genes included nuclear DSB repair factors that localize to the mitochondrial matrix as well as mitochondrial-specific factors implicated in mtDNA metabolism. RAD51-mediated recombination has been shown to stimulate mtDNA rearrangements in yeast (Stein et al., 2015), and human RAD51 is found in the mitochondrial matrix (Sage et al., 2010). We observed that the levels of mtDNA⁴⁹⁷⁷ accruing in response to mito-TALEN-CD₃ were not reduced upon depletion of RAD51 (Figures 5A and 5B). This observation potentially ruled out inappropriate recombination as the underlying basis of the common deletion and prompted us to ask whether the deletion is contingent on end-joining activity. Of the three DNA ligases involved in replication and repair in mammalian cells, LIG3 is the only enzyme that localizes to the mitochondria and is crucial for mtDNA integrity (Sage et al., 2011; Spies et al., 2011). To investigate the role of LIG3 during common deletion formation, we introduced a plasmid expressing mito-TALEN-CD₃ into LIG3^{+/+} HCT116 cells. Following depletion of LIG3 with Cre recombinase, we noted a marked decrease in the frequency of mtDNA⁴⁹⁷⁷. As a control, we

Figure 5. Replication-Mediated Repair of mtDNA⁴⁹⁷⁷ by Common Deletion
(A) qPCR analysis of mtDNA⁴⁹⁷⁷ in cells treated with mito-TALEN-CD₃. The amount of mtDNA⁴⁹⁷⁷ was determined 7 days after mito-TALEN transfection and normalized to total mtDNA (12S). The frequency of mtDNA⁴⁹⁷⁷ in cells treated with siRAD51 is presented relative to siControl.
(B) Relative percentage of mtDNA⁴⁹⁷⁷ accumulation in cells after Cre induction and mito-TALEN expression.
(C) qPCR analysis of mtDNA⁴⁹⁷⁷ induced by mito-TALEN-CD₃ in cells depleted for the indicated nucleases and analyzed 7 days after transfection. The frequency of mtDNA⁴⁹⁷⁷ upon inhibition of replication in cells treated with the indicated siRNA and analyzed 7 days after transfection. The frequency of mtDNA⁴⁹⁷⁷ upon inhibition of replication was normalized to mtDNA content (12S) and is presented relative to siControl. Values represent the mean of three independent experiments with SD.

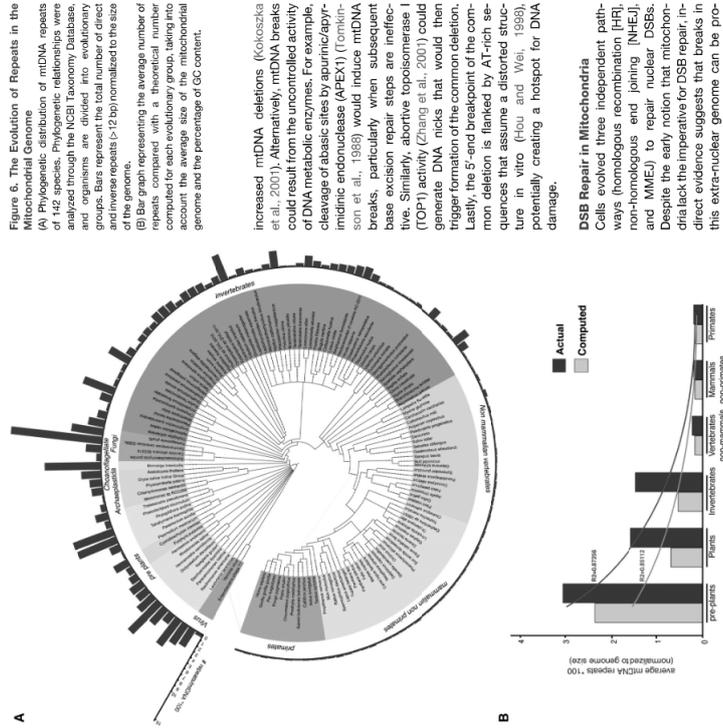


a number of mitochondrion-residing nucleases, including MRE11 and CIP (Otriviera et al., 2011; Tadi et al., 2016)—key components of the MMEJ pathway. In addition, we inhibited EXOG, which is implicated in mitochondrial single-strand break repair (Cymerman et al., 2009), and MGME1, a mitochondrion-specific nuclease that acts together with POLG to process DNA flap intermediates during replication (Köblum et al., 2013). To identify the rescuing nuclease responsible for common deletion formation, we used a small interfering RNA (siRNA)-based approach and depleted MRE11, CIP, and MGME1 from HCT116 cells (Figures 5C, 5E, and 5E). Concurrently, we introduced plasmids encoding mito-TALEN-CD₃ and monitored the frequency of mtDNA⁴⁹⁷⁷. We observed no differences in the levels of mtDNA⁴⁹⁷⁷ upon inhibition of repair-related nucleases—EXOG, CIP, and MRE11, instead, upon depletion of the replicative nuclease MGME1, the levels of common deletion were significantly reduced (Figure 5C). To account for the variation in mtDNA content because of these genetic perturbations, the common deletion was normalized to total mtDNA (Figure S5E).

Given the role of MGME1 during mtDNA replication, we wondered whether other components of the mitochondrial replisome are also implicated in the formation of mtDNA⁴⁹⁷⁷. To that end, we treated HCT116 cells with siRNA to deplete POLG, DNA2, Twinkle, and mtSSB independently. We then expressed mito-TALEN-CD₃ and monitored the frequency of the common deletion by qPCR. Our results reveal an attenuation in mtDNA⁴⁹⁷⁷ upon inhibition of various members of the

showed that the frequency of mtDNA⁴⁹⁷⁷ formation was not affected by the loss of LIG1 (Figures 5B and 5B). It is important to note that the levels of mtDNA⁴⁹⁷⁷ were normalized to the total mtDNA content in each setting (Figure S5B). In summary, these results invoke a LIG3-mediated end joining step during the process of common deletion formation.

Induction of the Common Deletion Is Dependent on mtDNA Replication
 Prior to joining of broken DNA ends, end resection is an essential step to describe the intervening sequence, resulting in mtDNA⁴⁹⁷⁷. To decipher this key processing step, we targeted



stimulate common deletion formation. The mis-annealing of homologous repeats flanking the deleted region represents a crucial step during common deletion formation. The exposed heavy strand is typically covered with mtSSB, but the factor(s) that promote annealing of the complementary junctions remains unknown. Twinkle possesses strand-annealing activity that could potentially promote the annealing of flanking repeats (Sen et al., 2012), although we cannot exclude the involvement of additional factors.

According to the replication-dependent repair mode of the common deletion, nicks anywhere within the ssDNA loop are expected to induce the common deletion. On the contrary, our data indicate that TALEN-induced nicks adjacent to the 3' end of the common deletion are refractory to common deletion formation (Figures 4E and 5SL). This contradiction is inherent to the source of the damage in our experimental setup. TALENs recognize double-stranded DNA and will, therefore, induce breaks prior to the unwinding of the double helix during replication. As a result, breaks at the 3' end of the deleted region are expected to obstruct the progression of the fork and subsequent unwinding of the DNA and, therefore, hinder formation of the common deletion.

A Method to Address a Fundamental Question Related to the Mode of mtDNA Replication

Replication of mtDNA has been classically studied using 2D gels and EM analysis (Clayton, 1982; Robberson and Clayton, 1972), with both methods leading to contradictory results as to what is the exact mode of mtDNA replication. Here we developed a method based on DNA combing to examine mtDNA replication at the single-molecule level and with high resolution. DNA combing has long been used to address key questions related to nucleic acid replication. By adapting this powerful technology to the specifics of DNA synthesis in mitochondria, we observed efficient firing of two replication origins (Figure 2), thus confirming the strand displacement model of mtDNA replication (Clayton, 1982; Robberson and Clayton, 1972; Figure S1B, top). In addition, we provided genetic evidence in support of a role for POLRMT in the initiation of OriL in vivo (Figure 2C). These data are consistent with previous experiments demonstrating that POLRMT primes lagging strand synthesis in vitro by recognizing a specific loop in the ssDNA around OriL (Fujita et al., 2010; Wang et al., 2009) and argues against strand-coupled mtDNA replication (Figure S1B, center; Foll et al., 2009), as well as its discontinuous lagging-strand synthesis via the RITOLS model (Figure S1B, bottom; Tang et al., 2002; Yoshikawa et al., 2009). Whether a protein factor (such as mtSSB) or RNA coats the lagging strand prior to OriL firing remains to be settled.

It is worth noting that the analysis of mtDNA replication was carried out in cells in which DNA synthesis served to maintain steady-state levels of mtDNA molecules following normal turnover. In the future, it will be informative to investigate whether the mode of mtDNA replication is altered in cells undergoing mtDNA repopulation following a sudden drop in mtDNA content or in response to increased energy demand.

The Endogenous Source of DNA Damage

Our study raises an obvious question related to the physiological source of damage that induces breaks in mtDNA. The

mitochondrial replicase (Figures 3D and S6A–S6C). Similarly, treatment of cells with ethidium bromide (EB), known to block mtDNA replication, led to a significant decrease in the frequency of TALEN-induced common deletion (Figures S3D–S3F). In contrast, deletion of PIF1, an ATP-dependent helicase that unwinds secondary structures and ensures progression of the replication fork (Paeschke et al., 2011), led to a 2-fold increase in the common deletion in response to TALEN cleavage of the mtDNA (Figures S5G and S5H). Altogether, our results highlight a replication-dependent form of DNA repair acting in the mitochondria and capable of stimulating the common deletion.

DISCUSSION

The mitochondrial common deletion was among the first identified mitochondrial mutations causative of human diseases and is currently one of 200 documented pathogenic mtDNA deletions (http://www.mitomap.org). Deletions in mtDNA are predominantly located on the major arc of the genome and are typically flanked by direct repeats (Samuels et al., 2004), potentially implicating a common underlying mechanism that triggers their formation. The experiments described here define the mechanism that leads to mitochondrial common deletion in vivo. By deciphering mtDNA replication and repair, we highlight a replication-dependent mode of DNA repair that is unique to mitochondria and leads to accumulation of the toxic common deletion (Figure S1A, bottom).

The Mechanism of Common Deletion Formation

It has been proposed that the mtDNA common deletion is triggered by replication slippage, MMEJ-mediated repair of DSBs or the combination of defective replication coupled with breaks in the DNA (Figure S1A). According to the DSB repair hypothesis, any break internal to the two 13-bp repeats is expected to undergo resection and re-joining, ultimately yielding mtDNA^{del} molecules. On the contrary, our data demonstrate that only breaks adjacent and internal to the 5'-end breakpoint trigger formation of the common deletion (Figures 4D, 4E and S3D–S3H). Furthermore, we provide genetic evidence indicating that formation of the common deletion in response to TALEN-induced breaks is independent of the core MMEJ factors MRE11 and CUP (Figure S5C). Altogether, these results argue against a role for canonical MMEJ during common deletion formation. Instead, our data demonstrate that mtDNA break repair is executed by components of the mitochondrial replicase and coupled to replication (Figures 3D and S6D). According to the replication-dependent repair model (Figure S1A, bottom), frequent stalling of the replication fork facilitates the mispairing of the two 13-bp repeats, thereby exposing a long single-stranded loop on the heavy strand. Damage incurred to the looped DNA triggers loss of the intervening sequences, resulting in an mtDNA^{del} genome. In effect, impaired mtDNA replication combined with aberrant break repair lead to the common deletion.

G4-DNA-forming sequences, typically unwound by PIF1 (Duan et al., 2013), are prevalent in regions proximal to the deletion breakpoints (Bardi et al., 2014) and could, therefore, contribute to stalling of the replicase. Similarly, Twinkle mutations impede fork progression to foster mispairing of the 13-bp repeats and

STAR METHODS
KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-MGME1 polyclonal antibody	Sigma	Cat# HPA040913; RRID: AB_10962835
anti-CtRP monoclonal antibody (clone 14-1)	Active motif	Cat# 611877
anti-human UG3 monoclonal antibody (clone 7)	Becton Dickinson	Cat# ab1517
anti-UG1 monoclonal antibody (clone 10H5)	Abcam	Cat# ab1517; RRID: AB_2138117
anti-HA monoclonal antibody (clone 12CA5)	Roche	Cat# 11-867-423-001
anti-PFI1 polyclonal antibody	Sigma	Cat# S68430117
anti-VDAC1 polyclonal antibody	Abcam	Cat# ab15895
anti-GAPDH monoclonal antibody (clone 6C5)	Abcam	Cat# ab8245; RRID: AB_2107448
anti-RAD51 monoclonal antibody (clone 14BH)	Novus Biologicals	Cat# NB100-148; RRID: AB_10002131
anti-POLG polyclonal antibody	Novus	Cat# NB100-142
anti-POLG polyclonal antibody	Novus	Cat# NB1P-52300; RRID: AB_11027699
anti-DNA2 polyclonal antibody	ProteinTech	Cat# ab194939
anti-MSSB polyclonal antibody	ProteinTech	Cat# 12212-1-AP; RRID: AB_2195520
anti-CMYC monoclonal antibody (clone 9E10)	Millipore	Cat# 05-419; RRID: AB_309725
anti-FLAG monoclonal antibody (clone M2)	Sigma Aldrich	Cat# F884; RRID: AB_262044
anti-γ-Tubulin monoclonal antibody (GTU-88)	Abcam	Cat# ab11316
anti-α-Tubulin monoclonal antibody (clone DM1A)	Sigma	Cat# T9028
mouse anti-BrdU monoclonal antibody	BD Biosciences	Cat# 547580
rat anti-BrdU monoclonal antibody	Abd Serotec	Cat# GB1000CX
Alexa Fluor 488-conjugated goat anti-rat	Molecular Probes	Cat# A-11004
Alexa Fluor 568-conjugated goat anti-mouse	Molecular Probes	Cat# A-11004
Biotinylated anti-streptavidin	Vector Labs	Cat# BA-3500
Chemicals, Peptides, and Recombinant Proteins		
Protease Inhibitor Cocktail	Roche	Cat# 0483159001
T7 Endonuclease I	New England Biolabs	Cat# M0202S
Proteinase K	New England Biolabs	Cat# P6107S
12-O-tetradecanoylphorbol-13-acetate	Sigma	Cat# T7904-SMG
5-Fluoro-2'-deoxyuridine	Sigma	Cat# F06881
5-Iodo-2'-deoxyuridine	Sigma	Cat# I7125
Alexa Fluor 405-conjugated streptavidin	Molecular Probes	Cat# S32351
(3-Aminopropyl)triethoxysilane	Sigma	Cat# A40140
UltraPure low melting point agarose gel	Thermo Fisher	Cat# 16520650
Critical Commercial Assays		
Qiagen Spin Miniprep Kit	QIAGEN	Cat# 27106
RNeasy Mini Kit	QIAGEN	Cat# 74106
Script gDNA Clear cDNA Synthesis Kit	Bio-Rad	Cat# 1725034
High Capacity cDNA Reverse Transcription Kit	Thermo Fisher	Cat# 4368814
SYBR Green PCR Master Mix	Thermo Fisher	Cat# 4309155
Cell Fractionation Kit	Abcam	Cat# Ab109719
BrightView Ultra-Fast SYBR qPCR MasterMix	Agilent	Cat# 600882
Lab-TekChamber Slide system	Sigma	Cat# C17182
Mito-Tracker Red CMXRos	Invitrogen	Cat# M7512
Hoechst 33342	Thermo Fisher	Cat# H3399

(Continued on next page)

Renaud, J.B., Bek, C., Charpentier, M., De Cian, A., Cochennec, J., et al. (2015) The mitochondrial DNA replication machinery is essential for embryonic development. *Development* 142, 1005–1014.

Tan, S.K., Schaefer, R., Dinkel, S., Bhat, B.K., Choudhury, B., and Rajabian, S. (2016) Mitochondrial DNA replication is the critical regulator of double-strand break repair during mitochondrial DNA lesions. *Mol. Biol. Cell* 27, 223–235.

Tonkinson, A.F., Bork, R.T., and Lim, S. (1988) Mitochondrial endonuclease activities specific for apurinic/aprimidinic sites in DNA from mouse cells. *J. Biol. Chem.* 263, 12532–12537.

Tymoczko, H., Migard, K.P., Wainio, L., Lappalainen, L., Vilkko, E., Jalkanen, A., Spelbrink, J.N., Pietari, A., and Suomalainen, A. (2005) Mutant mitochondrial helicase Twinkle causes multiple mtDNA deletions and a late-onset mitochondrial disease in mice. *Proc. Natl. Acad. Sci. USA* 102, 17897–17899.

Vin, G., Goethem, G., Dermaut, B., Lofgren, A., Marin, J.L., and Van Broeckhoven, C. (2001) Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. *Nat. Genet.* 29, 211–212.

Wallace, D.C., and Chalkia, D. (2013) Mitochondrial DNA genetics and the hereditary deafness in evolution and disease. *Cold Spring Harb. Perspect. Biol.* 5, a021220.

Wainio, S., Goffart, S., Poljanski, J.L., Yasukawa, T., and Spelbrink, J.N. (2007) Expression of catalytic mutants of the mtDNA helicase Twinkle and polynucleotide kinase in yeast reveals distinct replication stalling phenotypes. *Nucleic Acids Res.* 35, 3238–3251.

Wainio, S., Fiala, J.M., Fraga, G., Shi, Y., Guschke, C.M., and Spelbrink, J.N. (2008) Human mitochondrial RNA polymerase II is a lagging-strand DNA synthesis *in vitro*. *Proc. Natl. Acad. Sci. USA* 105, 11122–11127.

Wu, Y., Gao, T., Wang, X., Hu, Y., Hu, X., Hu, X., Peng, J., Li, Z., Xue, J., Feng, M., et al. (2014) TALE nuclease mediates high-efficient targeted transgene integration at the human multi-copy ribosomal DNA locus. *Biochem. Biophys. Res. Commun.* 446, 261–266.

Yang, J.H., Lee, H.C., Lin, K.J., and Wei, Y.H. (1994) A specific 497-bp deletion of mitochondrial DNA in human aging skin. *Arch. Dermatol.* 130, 285–289.

Yang, M.Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringoli, E., Jacobs, H.T., and Holt, L.J. (2002) Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* 111, 495–506.

Yasukawa, T., Reyes, A., Quatt, T.J., Yang, M.Y., Bowmaker, M., Jacobs, H.T., and Holt, L.J. (2006) Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *EMBO J.* 25, 3338–3371.

Yu, M., and Misker, W. (2001) T7 single-strand DNA binding protein but not T7 helicase is required for DNA double-strand break repair. *J. Bacteriol.* 183, 1862–1868.

Zhang, H., Barcadi, J.M., Lee, B., Kolthagen, G., Zimonic, D.B., Popescu, N.C., and Palmieri, M. (2006) Human mitochondrial topoisomerase I. *Proc. Natl. Acad. Sci. USA* 103, 10628–10633.

Renaud, J.B., Bek, C., Charpentier, M., De Cian, A., Cochennec, J., et al. (2016) Intra-mitochondrial repair of large-scale deletion of human mitochondrial DNA by TALEN and CRISPR-Cas9 nucleases. *Cell Rep.* 14, 2263–2272.

Roberson, D.L., and Clayton, D.A. (1972) Replication of mitochondrial DNA in mouse L cells and their thymidine kinase-derivatives: displacement replication on a covalently-closed circular template. *Proc. Natl. Acad. Sci. USA* 69, 3810–3814.

Sage, J.M., Gildmeister, O.S., and Knight, K.L. (2010) Discovery of a novel function for human Rad51: maintenance of the mitochondrial genome. *J. Biol. Chem.* 285, 18954–18960.

Samuels, D.C. (2004) Mitochondrial DNA repeats constrain the life span of mammals. *Trends Genet.* 20, 250–259.

Samuels, D.C., Schen, E.A., and Chimney, P.F. (2004) Two direct repeats cause most human mtDNA deletions. *Trends Genet.* 20, 380–386.

Schon, E.A., Hazzuto, R., Moraes, C.T., Nakase, H., Zavanin, M., and DiMauro, S. (1989) A direct repeat is a template for large-scale deletion of human mitochondrial DNA. *Science* 246, 1504–1506.

Song, D., Beck, S., D. T. G. O., and Pohl, S.S. (2012) Human mitochondrial DNA helicase TWINKLE is both an unwinding and annealing helicase. *J. Biol. Chem.* 287, 14545–14556.

Stein, A., and Strimling, L.S. (2015) Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends Biochem. Sci.* 40, 701–714.

Shoffner, J.M., Lott, M.T., Voljace, A.S., Souleian, S.A., Costigan, D.A., and Wallace, D.C. (1989) Spontaneous Kearns-Sayre/chronic external ophthalmoplegia plus syndrome associated with a mitochondrial DNA deletion: a self-replication model and metabolic therapy. *Proc. Natl. Acad. Sci. USA* 86, 7952–7956.

Shrobenko, L., Venediktova, N., Bochkareva, A., Wilson, G.L., and Alexeyev, M.F. (2009) Oxidative stress induces degradation of mitochondrial DNA. *Nucleic Acids Res.* 37, 2332–2346.

Shutt, J.E., and Gray, M.W. (2008) Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet.* 24, 59–65.

Smilek, D., Janda, A., S. K. P., J. Brunel, E., H. G. O., A. K. V. N. P., and S. K. P. (2016) Mitochondrial DNA replication is a lagging-strand DNA ligase III in mitochondria but not in Xist-dependent repair. *Nature* 531, 245–248.

Spelbrink, J.N., Li, F.Y., Tranti, V., Nikali, K., Yuan, O.P., Tarr, M., Wainio, S., Garada, N., Comi, G., Morandi, L., et al. (2001) Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat. Genet.* 28, 223–231.

Shrivastava, S., and Moraes, C.T. (2005) Double-strand breaks of mouse muscle mtDNA promote large deletions similar to multiple mtDNA deletions in humans. *Hum. Mol. Genet.* 14, 889–902.

Stein, A., Kolita, L., and Ska, E.A. (2015) Members of the RAD52 epistasis group contribute to mitochondrial homologous recombination and double-strand break repair in *Saccharomyces cerevisiae*. *PLoS Genet.* 11, e1005064.

Srinivasan, G., Galk, P., Hildebrand, C., W. J. M. M., K. V. G., K. V. G., P. E., M. G., and Wallace, D.C. (2008) MITOCHIP: assessment of gfp-

Continued	REAGENT or RESOURCE	SOURCE	IDENTIFIER
	FastStart Taq DNA Polymerase, dNTPack	Roche	Cat#04738420001
	Lipidexamine RNAMAX	Thermo Fisher	Cat#13778150
	Cal extraction kit	EZNA	Cat#D2900-02
Experimental Models: Cell Lines			
	U2OS	ATCC	HTB-95
	HCT116	ATCC	CCL-247
	HT1080	ATCC	CCL-121
	BJ-H1ERT	ATCC	CRL-4001
	C2C12 myoblasts	ATCC	CRL-1772
	RPE-1-H1ERT WT	ATCC	CRL-4000
	MEF	This paper	N/A
	Twinkle ^{CRISPR} and RPE-1-H1ERT	This paper	N/A
	Twinkle ^{Y508C} and RPE-1-H1ERT	This paper	N/A
	POLRMT ^{+/+} and POLRMT ^{+/+} mouse embryonic fibroblasts	Laboratory of N.G. Larsson	Kuhl et al. (2014)
	HCT116 Lip1 F/	E. Hendrickson	Oh et al. (2014)
	HCT116 Lip3 F/	E. Hendrickson	Oh et al. (2014)
Recombinant DNA			
	pVAX-mitoTALENs, see Table S3	This paper	N/A
	Plasmid: pJM1-Flag-POLRMT-puro	This paper	N/A
	Plasmid: pBabe-Cre-ER ²	Laboratory of G. David	N/A
	Plasmid: pSpCas9(Bbs)Δ2A-GFP	This paper	N/A
Sequence-Based Reagents			
	siRNA sequences, see Table S1	This paper	N/A
	Custom mito-FISH probe library, see Table S2	This paper	MYCarray custom synthesis
	Mito-TALEN target sequences, see Table S3	This paper	N/A
	Primers for common deletion detection and quantification, see Table S4	This paper	N/A
	Primers for muscle cell differentiation qPCR, see Table S5	This paper	Ali et al. (2008)
	Primers for qPCR, see Table S5	PrimerBank	https://pga.mgh.harvard.edu/primerbank/
	Primers for POLRMT genotyping: 3F: 5'-CCACTAACCATGGCTCTGTC-3' and Int3: 5'-GGCACATCTGATACAGCTGG-3'	This paper	N/A
	Primers for semiquantitative PCR of POLRMT: 2F: 5'-GGCAACATGCTCTGAGGGAGTGG-3' and 4R: 5'-GCCAGGATCCTCAGCATGAGG-3'	This paper	N/A
	Twinkle R374Q CRISPR guides: exon1 gRNA #1: 5'-CCATCGTATCTCCGGGAGCT-3'; twinkle exon1 gRNA #2: 5'-GTGAGGAGAGTGTGAAATGTGG-3'	This paper	N/A
	Twinkle R374Q genotyping primers: Fwd: 5'-GGACGAAAGTGAACCCCAACC-3' and Rev: 5'-CCTTGGATGCTCTCAGATAGC-3'	This paper	N/A
	Twinkle Y508C CRISPR guides: twinkle exon3 gRNA #1: 5'-GAAACCTGCAATCATGAT-3'; twinkle exon3 gRNA #2: 5'-ATTGATCTATACAGCTCT-3'	This paper	N/A
	Twinkle Y508C genotyping primers: Fwd: 5'-ACTCACACTGAGTCCCTGTC-3' and Rev: 5'-ATCGCTTGAGGCTAGTATGGG-3'	This paper	N/A

(Continued on next page)

Continued	REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms			
	NCBI Taxonomy database		https://www.ncbi.nlm.nih.gov/taxonomy
	NCBI Organellar database		https://www.ncbi.nlm.nih.gov/genome/organelle/
	UGENE bioinformatics suite		http://www.ugene.net
	Random DNA sequence generator		http://www.faculty.ucr.edu/~mmaduro/random.htm
	Custom python pipeline	Renaud et al., 2016	N/A
Other			
	Nan-Gen Sequencing data	This paper	NCBI Bioproject: PRJNA357534
	Unprocessed image data	This paper	http://dx.doi.org/10.17622/nctmmpipr.1

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by, the lead contact, Dr. Agnet Steir, agnel.steir@med.nyu.edu.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Source of cell lines used in the study is reported in the reagent and resource table.

METHOD DETAILS

Cell culture
 HT1080, BJ-H1ERT, C2C12 myoblasts, RPE-1, and MEF cell lines were grown in DMEM supplemented with 10% FBS (fetal bovine serum), 2 mM L-glutamine (Sigma), 100 U ml⁻¹ streptomycin (Sigma), 0.1 μg ml⁻¹ penicillin (Sigma), 0.1 mM non-essential amino acids (Invitrogen) and 200 μM uridine (Sigma) as needed. C2C12 myoblasts were differentiated into myotubes by replacing the media with DMEM supplemented with 2% horse serum, 2 mM L-glutamine (Sigma), 100 U ml⁻¹ penicillin (Sigma), 0.1 μg ml⁻¹ streptomycin (Sigma), and 0.1 mM non-essential amino acids (Invitrogen) at 90% confluency and cultured for 5–7 days. *Polrmt*^{+/+} and *Polrmt*^{-/-} MEFs were a gift from N.G. Larsson (Kuhl et al., 2014). U2OS and HCT116 cells were cultured in McCoy's 5A medium containing 10% FCS (fetal calf serum), *LG3*^{+/+} and *LG3*^{-/-} and *LG1*^{+/+} HCT116 cells were a gift from E. Hendrickson (Oh et al., 2014). Deletion of *LG3* and *LG1* was induced by treating cells with 10nM 4-hydroxy-tamoxifen for 24 hr. To delete POLRMT, Cre-ERT² was stably expressed in *Polrmt*^{+/+} and *Polrmt*^{-/-} MEFs, and cells were treated with 10 μM 4-hydroxy-tamoxifen (Sigma) for 24 hr. Cells were pulsed with CldU and IdU for mito-SMARD, 48 hr post Cre induction. The primers used for *Polrmt* genotyping are: 3F: 5'-CCACTAACCATGGCTCTGTC-3'; 3R: 5'-GGCACATCTGATACAGCTGG-3'; and Int3: 5'-GGCACATCTGATACAGCTGG-3'. U2OS and HCT116 cells were transfected with 3.5 μg of mito-TALEN using Amaxa Nucleofector technology (Lonza). Typically, 1–2 × 10⁶ cells were transfected with the K1 V and program D-032 Amaxa 2b device or SE. Solution and CM-104 program (Amxaxi-4D). For siRNA transfection, cells were plated at a density of 3 × 10⁵ cells per well in 6-well and transfected 24 hr later using Lipofectamine™ RNAiMax (Invitrogen) following the manufacturer's protocol. The siRNA sequences are listed in Table S1. Cells were then incubated for 24 hr before transfection with Mito-TALEN. mtDNA depletion was achieved by treating cells with 50 ng/ml EB1 in presence of uridine for a total of 5 days, followed by 2 days of chasing in regular media supplemented with 200 μM uridine.

Mito-SMARD

Cells were initially pulsed with CldU for 2 hr, followed by IdU for 2 hr, yielding mostly all green or all red molecules. Pulse-labeling was experimentally optimized in order to capture a greater number of dual-labeled molecules. Human cells by sequentially labeling with 15 min of CldU (30 μM) and 90 min of IdU (30 μM), followed by 2 hr chase in standard media. Mouse cells were pulsed with 30 min CldU, followed by 90 min with IdU. After repeating the procedure twice, cells were collected by trypsinization and mtDNA was purified using the Oligoprep Spin Miniprep Kit (Qiagen) (Ousajpe-Tripathy et al., 2013). mtDNA was eluted from the column with 55 μL of H₂O, mixed with an equal volume of 2% agarose and loaded into a plug mold (Bio-Rad). Plugs were incubated with 10 U of AluI (New England Biolabs) overnight at 37 °C. The plugs were run on a 0.7% UltraPure low melting point agarose gel (Invitrogen) at 30 V for 2 hr. The linearized mtDNA band was excised from the gel and melted in 100 μL melting buffer (1×TAE, 100mM NaCl, and 0.1% beta-mercaptoethanol) at 68 °C for 20 min before being moved to 45 °C and incubated for 2 hr with the addition of 3 U Gelsae. After melting, an additional 10 μL of beta-mercaptoethanol and 0.2 μL of YoYo-1 (Invitrogen) were added to the tube and incubated overnight. The mtDNA was combed by capillary action across microscope cover glasses coated with 3-aminopropyltriethoxysilane (Sigma). The stretched DNA was denatured in an alkali-denaturing buffer (0.1 N NaOH in 70% ethanol and 0.1% beta-mercaptoethanol) for

9 min and fixed by adding 0.5% glutaraldehyde for 5 min. mtDNA FISH probes were hybridized overnight in buffer containing 2x SSC, 40% formamide, and 10% dextran sulfate. The probe was labeled with Alexa Fluor 405-conjugated streptavidin (Molecular Probes). The signal was amplified by two rounds of labeling with biotinylated anti-streptavidin antibody (Vector) and additional Alexa Fluor 405-conjugated streptavidin. The thymidine analogs were detected with a mouse anti-BrdU monoclonal antibody recognizing IdU (Becton Dickinson) and a rat anti-BrdU monoclonal antibody recognizing CldU (Add Serotec). Secondary antibodies used in these studies were Alexa Fluor 488-conjugated goat anti-rat (Molecular Probes) and Alexa Fluor 568-conjugated goat anti-mouse (Molecular Probes). Slides were imaged using an Eclipse T12 with a 100x Apo TIRF oil objective (Nikon) or on a DeltaVision OMX (GE Healthcare Life Sciences).

Generation of mito-FISH probes

A library of mtDNA probes for FISH analysis was generated using MYtag technology. A library of probes labeling nucleotides 10000-16400 of the mtDNA genome were custom synthesized by MYcroarray. The library was prepared according to manufacturer's recommendations (Murgha et al., 2014). Briefly, the oligonucleotide library was transcribed *in vitro* to generate RNA. The RNA fragments were then reverse transcribed using a common biotinylated primer. Degradation of the RNA molecules yielded single stranded oligonucleotides. The sequences for the probes are found in Table S2.

CRISPR/Cas9 targeting to generate Twinkle mutants

Twinkle heterozygote R374Q mutation was introduced into human, retina pigmented epithelium cells (RPE-1) using a CRISPR/Cas9 double nickase strategy. Briefly the plasmid pSpCas9(BB)-2A-GFP (Addgene #48138), harboring a Bsal cloning site for gRNA expression and the Cas9-2A-GFP nickase was reengineered with an additional Bsal gRNA cloning cassette giving rise to pSpCas9(BB)2A-GFP. Two gRNAs targeting Twinkle Exon 1 (Twinkle exon1 gRNA #1: 5'-CCATGATCTTCCGGCAGCT-3', Twinkle exon1 gRNA #2: 5'-GCTAGGAGAACTGTCAATGTGG-3') were cloned sequentially as dsDNA oligos into Bsal and Bsal sites and sequence confirmed. The construct was co-transfected with a repair template consisting of an 180bp ssDNA oligo (IDT) introducing the desired point mutation, a silent point mutation introducing a PstI restriction site for genotyping purposes and silent mutations in the PAM sites, 48 hr post-transfection, GFP-positive cells were sorted as single cells into 96-well plates. Clonal cell lines were confirmed by genotyping and cDNA. The genotyping PCR primers consist of Fwd: 5'-GCAGAAACTGAAAGCCCAAGC-3' and Rev: 5'-CCTTGATGTCCTCCAGATAGC-3'. A similar strategy was used to generate the homozygote Y508C mutation in exon 3. Two gRNAs (Twinkle exon3 gRNA #1: 5'-GACAACTCAGATT CATGAT -3', Twinkle exon3 gRNA #2: 5'-ATTGATCTATTACAGTCTC-3') were cloned sequentially into pSpCas9(BB)2A-GFP and co-transfected with a repair template consisting of an 180bp ssDNA oligo (IDT) introducing the desired point mutation, a silent point mutation introducing a ClaI (NEB) restriction site for genotyping purposes and silent mutations in the PAM sites. Single clones were screened by genotypic PCR followed by ClaI digestion. The genotyping PCR primers consisted of Fwd: 5'-ACTCACACTCAGATCCCTGCC-3' and Rev: 5'-ATCGCTGAGCGCTAGTAGTGG-3'.

Mito-TALEN

For each TALEN subunit, the fragment containing the 16 RVD segment was obtained from single unit vectors: A (NI), T (NG), G (NN) and C (HD), kindly provided by B. Zhang. Mito-TALENs were assembled into pVax vector and included the 5'UTR and mitochondrial localization signal (MLS) derived from SOD2. 3'UTR of SOD2 was also introduced at the C-term end (Bischoff et al., 2013). The target sequences for each TALE domain is described in Table S3. To decrease mitoTALEN off-target effects, we used two obligate heterodimeric domains of FokI: FokI-ELD and FokI-KKR (Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architecture (Doye et al., 2011)). The T2A 'self-cleaving' peptide sequence and GFP reporter were fused to FokI-ELD heterodimeric domain. Similarly, the E2A and RFP reporter coding sequences (kindly provided by T. Auer) were fused to KKR domain. Mito-TALEN nickases were generated using a catalytically inactive FokI domain (D450A) on one of the monomers (Kim et al., 2012).

Detection of the common deletion by nested PCR or qPCR

Cells were trypsinized and were re-suspended in 200 μ L of PBS buffer containing Proteinase K (0.2mg/ml), SDS (0.2%), EDTA (5mM) for 3 hr at 56 C. After isopropanol precipitation, DNA was amplified by PCR. For direct detection, 100 ng of DNA was amplified by nested PCR (PCR1 - 15 cycles; PCR2 - 30 cycles). For quantification of the common deletion formation, 30 ng of DNA is amplified by Q-PCR (Biorad III Ultra-Fast SYBR qPCR Master Mix (Agilent)). A list of primers is provided in Table S4.

RT-qPCR

Cells were collected by trypsinization and RNA was purified using the RNeasy kit (QIAGEN) according to the manufacturer's protocol. cDNA was generated using 1 μ g of RNA and the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Q-PCR was run using SYBR Green PCR Master Mix (Applied Biosystems). Primers listed in Table S5.

Western blotting

Whole-cell extracts were prepared in protein lysis buffer (50 mM Tris-HCl pH 7.4, 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT), with addition of Protease Inhibitor Cocktail (Complete, Roche). Typically 30 μ g of protein extracts

64 Molecular Cell 65, 1–12.e1–e6, February 2, 2017

from cells were resolved on an 8% or 10% SDS-page gel, blotted, and probed with the following antibody: anti-MGME1 polyclonal antibody (Sigma, HPA040913), anti-CTIP monoclonal antibody (Active motif, 61141), anti-human LG3 monoclonal antibody (Becton Dickinson, 611877), anti-LIG1 monoclonal antibody (Abcam, ab815), anti-HA monoclonal antibody (Roche, 11-867-423-001), anti-Pif polyclonal antibody (Sigma, SAB301117), anti-VDAC1 polyclonal antibody (Abcam, ab15895), anti-GAPDH monoclonal antibody (Abcam, ab8245), anti-RAO2 monoclonal antibody (Novus Biologicals, NB100-148), anti-WRE11 polyclonal antibody (Novus, NB100-142), anti-POLG polyclonal antibody (Novus, NBPI-52300), anti-DNA2 polyclonal antibody (Abcam, ab194939), anti-miSSB polyclonal antibody (Novus, NBP1-52300), anti-c-MYC antibody clone 9E10 (Millipore, 05-419), anti-FLAG antibody clone M2 (Sigma Aldrich, F1084), anti- γ -TUBULIN antibody clone GTU-88 (Abcam, ab11316) and anti- α -TUBULIN antibody (Protein, 19029). Cellular fractionation was performed using an Abcam fractionation kit (Ab109719) according to the manufacturer protocol.

Fluorescence in situ hybridization

UZOS cells were cultured in Lab Tek chambers (Sigma). Mitochondria were stained with 62.5 μ M Mito-Tracker Red CMXRos (Invitrogen) for 25 min at 37 C. Cells were then washed with PBS and fixed for 15 min at room temperature in 4% paraformaldehyde. After two washes in PBS, cells were permeabilized with 1% Triton for 5 min. Pre-blocking was performed for 1 hr in 10% FBS (fetal bovine serum). After an additional PBS wash, the slides are incubated with primary HA antibody (Roche) for 1 hr, washed, and incubated with Alexa Fluor 488 (Jackson Immuno Research) for an additional hour. After 3 more washes, nuclei were stained with Hoechst 33342 (Thermo Fisher). Images were captured with the Zeiss Axio Observer Z1 system.

Analysis of repeats content in mtDNA sequences

142 organisms with sequenced mtDNA were chosen to cover the whole eukaryotic reign. Two viruses (*Enterobacteria phage 77* and *Vaccinia virus*) were also included in the analysis due to homology of DNA metabolism enzymes with mitochondrial replication/repair machinery. Phylogenetic relationships were inferred from NCBI Taxonomy database and plotted as circular tree with iTOL interactive tool v3.1 (Letunic and Bork, 2016). The tree was further annotated with repeat content computed as follows. mtDNA sequences were retrieved at NCBI Origami database and repeats content and GC% were analyzed with UGENE bioinformatics suite (Chernichikov et al., 2012). Unique direct and inverse repeats > 12bp with 100% identity were computed for each sequence, normalized to the size of the respective mtDNA (genome in case of viruses) and multiplied by 100. Additional control sequences with defined size and GC% were randomly generated (<http://www.faculty.ucr.edu/~rmmaduro/random.htm>) and analyzed in the same way.

T7 endonuclease assay to determine the efficiency of mito-TALEN cutting

UZOS cells were transfected with 2 μ g mito-TALEN target sequence (PCR amplified from UZOS cells (Table S4)) and cloned using TOPO-TA cloning (Invitrogen) and 3.5 μ g of the corresponding nuclear TALEN expression plasmids. DNA was extracted after 36 hr and indels formation was monitored using a T7 endonuclease as previously described (Glaschin et al., 2010) using the primers T3: GCATTAACCCCTCCTCAAAAGG and T7: TAATACGACTCAGTATAGGG.

Next-Gen sequencing

Cells treated with mito-TALENs were trypsinized and resuspended in lysis buffer (10 mM Tris-MOPS pH 7.4, 1 mM EGTA/Tris pH 7.4, 0.2 M sucrose) for 10 min on ice. After douncing, lysates were centrifuged at 700 g for 10 min at 4 C. Supernatants were incubated with 1,000 Units of DNase I at 37 C for 30-60 min and spin at 1000 g for 30 min at 4 C. Pellets were washed with 1 mL of lysis buffer, centrifuged at 10,000 g for an additional 30 min at 4 C, and resuspended in PBS buffer containing 0.2% SDS, 5 mM EDTA and 0.2 mg/ml Proteinase K. After incubation at 56 C for 3 hr, mtDNA was precipitated and PCR amplification was performed using primers from Table S4 to detect the common deletion. PCR fragments were purified using Gel extraction kit (EZNA). Libraries were prepared with NEBNext DNA Library Prep Master Mix (IDT) and NEBNext Multiplex Oligos Index Primers (set 1 and set 2, IDT) following manufacturer's instructions. A custom python pipeline was used to characterize the indels (Penaud et al., 2016).

QUANTIFICATION AND STATISTICAL ANALYSIS

For Next-Gen sequencing analysis, a custom python pipeline was used to characterize the indels (Penaud et al., 2016). Quantification of the semi-quantitative PCR shown in Figure S3B. Band intensity was measured using Image Lab and the intensity is normalized to the untreated sample.

Quantification of common deletion. The frequency of common deletion in all genetic perturbations was calculated by RT-PCR. Values were normalized to mtDNA content (12S) and presented relative to sControl. Values are mean of three independent experiments with SDs.

Quantification of group A and group B mtDNA molecules: Average frequencies were calculated from three independent experiments (n \geq 100 molecules in each experiment) with SDs.

Molecular Cell 65, 1–12.e1–e6, February 2, 2017

Please cite this article as: Phillips et al., Single-Molecule Analysis of mtDNA Replication Uncovers the Bias of the Common Deletion, *Molecular Cell* (2016), <http://dx.doi.org/10.1016/j.molcel.2016.12.014>

CellPress

DATA AND SOFTWARE AVAILABILITY

Original imaging data have been deposited to [Mendeley Data](http://dx.doi.org/10.17632/htwmpj9r-1) and are available at <http://dx.doi.org/10.17632/htwmpj9r-1>. The accession number for the Next-Gen sequencing data reported in this paper is NCBI BioProject: PRJNA357534 (<https://www.ncbi.nlm.nih.gov/bioproject/357534>).

ADDITIONAL RESOURCES

Random DNA sequence generator: <http://www.faculty.ucr.edu/~rmmaduro/random.htm>
NCBI Organelle database: <https://www.ncbi.nlm.nih.gov/genome/organelle/>
NCBI Taxonomy database: <https://www.ncbi.nlm.nih.gov/taxonomy>
Python pipeline for the analysis of indels: Renaud et al., 2016

Titre : Etude des translocations chromosomiques en utilisant les méthodes d'édition du génome : des mécanismes moléculaires à l'oncogenèse

Mots clés : Translocations chromosomiques – Cancer – CRISPR/Cas9

Résumé : Les translocations chromosomiques sont associées à un grand nombre de cancers. Les translocations chromosomiques sont impliquées dans la tumorigenèse par différents mécanismes : elles conduisent soit à une dérégulation d'un oncogène, soit à la formation d'un nouvel oncogène de fusion. Cependant, le lien direct entre l'apparition d'une translocation chromosomique et la formation d'une tumeur n'est pas totalement établi. Par exemple, plusieurs translocations associées au cancer ont été détectées dans le sang d'individus sains voire dans le sang de cordon des bébés avec une prévalence bien supérieure à celle de la maladie. Ceci suggère que la seule formation de la translocation ne suffit pas toujours à induire l'oncogenèse. La plupart des travaux de recherche antérieurs reposaient sur la surexpression de la protéine de fusion, oncogène supposé. Ces approches présentent de nombreuses limites, la translocation chromosomique est alors absente de même que le contexte chromosomique natif du gène de fusion (promoteur endogène, statut de la chromatine, etc.) ou les éventuels effets d'haplo-insuffisance qui ne sont pas récapitulés. La molécule d'ADN étant organisée de manière non aléatoire dans le noyau, les réarrangements chromosomiques sont également susceptibles d'affecter le statut épigénétique, la réplication et la transcription du chromosome dérivatif entier, en plus des segments

d'ADN nouvellement juxtaposés. Or la technologie CRISPR/Cas9, permet de reproduire la translocation chromosomique in situ, après avoir induit deux cassures double-brin simultanées. Ce travail de thèse a porté spécifiquement sur la translocation t(2,5) (p23, q35) qui induit l'expression de la protéine de fusion NPM1-ALK fréquemment rencontrée dans le lymphome anaplasique à grandes cellules (ALCL). Nous avons reproduit la t(2,5) à la fois dans des lignées cellulaires mais aussi dans des cellules T primaires à la fin de ma thèse. Nous avons pu montrer des modifications significatives du timing de réplication des cellules qui portent la translocation en comparaison des cellules isogéniques de départ (par la méthode du Répli-seq) pouvant avoir un impact sur l'homéostasie des cellules tumorales. En parallèle, nous avons mis en évidence la formation d'ARN circulaires de fusion spécifiques, exprimés à partir du gène de fusion, spécifiques des lignées tumorales. Ces ARN circulaires pourraient donner naissance à de nouveaux biomarqueurs diagnostic/pronostic dans le futur. Ces travaux permettront de mieux comprendre les conséquences des translocations chromosomiques oncogéniques dans les cellules humaines et pourraient mener vers de nouvelles orientations thérapeutiques à l'avenir.

Title : Cancer translocations induction using genome editing: from molecular mechanisms to oncogenesis

Keywords : Chromosomal translocations – Cancer – CRISPR/Cas9

Abstract: Chromosomal translocations are associated with a wide range of cancers. These chromosomal rearrangements are implicated in tumorigenesis by different mechanisms: either they lead to oncogene upregulation or tumor suppressor downregulation. However, the direct link between the appearance of one chromosomal translocation and tumor formation is not always clear. For example, several cancer translocations have been found in PBMCs or in cord blood cells from healthy individuals, suggesting that translocation formation alone is not always sufficient to drive oncogenesis. Most of previous research works on cancer translocation relied on studies using overexpression of the fusion protein. These approaches do not reproduce the chromosome arm translocation nor the chromosomal context of the fusion gene (endogenous promoter, chromatin status etc...) or do not recapitulate a potential haplo-insufficiency of the translocated cells. Because the DNA molecule is organized non-randomly in the nucleus, chromosomal rearrangements are also likely to impact the epigenetic, replication and transcriptional status of the whole rearranged chromosome in addition to the newly juxtaposed gene segments.

Using CRISPR/Cas9 technology, we can recapitulate chromosomal translocation in situ, after inducing 2 concurrent double-strand breaks. In this work, we focus on t(2,5)(p23,q35) leading to NPM1-ALK fusion protein frequently found in Anaplastic Large Cell Lymphoma (ALCL). We could recapitulate t(2;5) in cell lines but more importantly in human primary T cells from healthy donors. We showed significant modifications on Replication Timing in model cell lines compare to isogenic non-translocated cells (using Repli-seq analysis). Importantly, these changes might have a direct impact on tumor cell homeostasis. In parallel, we also highlighted the formation of specific fusion circular RNAs expressed from the fusion gene also found in tumor cells. These circular RNAs could give rise to new diagnostic/prognostic biomarkers in the future. This work will lead to a better understanding of the consequences of cancer translocation in human cells and could give new directions for therapeutics in future.

