





UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 418 : DE CANCÉROLOGIE

Laboratoire : INSERM U1154, CNRS UMR 7196

Structure et instabilité des génomes

Thèse sur travaux

Aspects moléculaires et cellulaires de la biologie

Par

Hind Ghezraoui

Étude du mécanisme moléculaire de formation des translocations chromosomiques dans les cellules humaines

Date de soutenance : 27/03/2015

Composition du jury:

Directrice de thèse Co-directrice de thèse	Dr. Erika Brunet Dr. Carine Giovannangeli	CR1-INSERM. MNHN DR1-CNRS. MNHN
Rapporteurs :	Dr. Pascale Bertrand Dr. Patrick Calsou	CR1-CEA/Fontenay-aux-Roses DR2-INSERM. IPBS Toulouse
Examinateurs :	Dr. Françoise Dantzer Dr. Yvan Canitrot	DR2-CNRS. Irebs Strasbourg CR1-CNRS. LBCMC Toulouse







Remerciements

Je remercie vivement le Docteur Pascale Bertrand et le Docteur Patrick Calsou d'avoir accepté d'être rapporteurs de cette thèse. Merci également au Docteur Françoise Dantzer et au Docteur Yvan Canitrot d'avoir accepté d'être membres du jury et d'évaluer ce travail.

Je remercie La Ligue Contre Le Cancer pour avoir financé cette thèse, je vous exprime ma profonde gratitude pour le soutien que vous m'avez accordé. Je remercie également le Docteur Carine Giovannangeli et le Professeur Jean-François Riou pour m'avoir donné la chance de travailler dans leur unité.

Je tiens à exprimer mes **sincères remerciements** à ma directrice de thèse, le Docteur Erika Brunet pour m'avoir donné l'opportunité de développer ce « super projet », d'avoir cru en moi, et de m'avoir fait confiance. Merci de m'avoir guidée tout au long de la réalisation de cette thèse avec tes conseils et tes directives qui m'ont énormément aidée dans l'orientation de mes objectifs et l'accomplissement de mes travaux. Merci pour ta disponibilité, tu as toujours été là pour répondre à nos questions techniques ou scientifiques. Merci pour ton humanité et ta gentillesse (tes petits mots d'encouragements, les pâtisseries laissées sur le bureau, ou encore les w.end /jours fériés où tu venais récupérer les manips avec nous et bien d'autres choses...). Merci de m'avoir transmis ta rigueur, mais aussi de m'avoir appris à être critique, à exploiter une manip jusqu'au bout pour en tirer le maximum d'informations. Merci de m'avoir tirée vers le haut tout simplement, ce contexte m'a permis de beaucoup apprendre quant au métier de chercheur. Je ne pouvais pas mieux tomber, Je crois que je vais arrêter car je n'en aurai jamais fini ...**MERCI POUR TOUT**

Je remercie également ma co-direcrtice de thèse le Docteur Carine Giovannangeli de m'avoir accepté dans son équipe, pour sa disponibilité et pour m'avoir apporté son soutien et sa confiance. Je remercie vivement le Professeur Maria Jasin avec qui nous avons collaboré.

Je remercie chaleureusement toutes les personnes que j'ai côtoyées pendant ces 4 années et qui ont fait que les journées de travail ont été si agréables : je remercie bien évidemment tous les djeuuuun'sss, les anciens que j'ai eu la chance de connaitre : Céline, Layal, Fanny, Najah, Asta, Jean...et les doctorants/post-doctorants encore présents : Armêl, Karine, Marion, Lorena, Marine, Astrid, et Antoine. Merci pour votre bonne humeur, merci pour tous les moments qu'on a pu partager : les midis en salle café, les gouters qui sont devenus indispensables ...Je remercie également les autres personnes du laboratoire avec qui c'était très agréable de travailler et d'interagir ; merci à mon voisin de paillasse Loïc Perrouault pour sa sympathie, pour toutes nos discussions autour de tes voyages qui me faisaient rêver, et pour toutes les fois où il fallait descendre et remonter les machines PCR. Merci aux deux François (Loll et Strauss) pour votre extrême gentillesse, votre calme et votre bonne humeur quotidienne. Merci à Anne De Cian pour tous tes conseils de manips et pour toutes les discussions scientifiques si enrichissantes qu'on a pu avoir. Merci à Gildas alias « le chimiste qui fait de la vraie science », merci pour ta bonne humeur et pour toutes les fois où tu prenais le temps de m'écouter râler et surtout de m'aider avec les machines PCR. Un grand merci à Patrizia pour sa gentillesse, ses conseils et sa bonne humeur. Merci à Jean-Baptiste Renaud pour ton extrême gentillesse et pour tous les T7 que tu as gentiment accepté de faire pour moi. Merci à Benjamin pour sa « zen attitude » et sa sympathie. Un très grand merci pour ma Charlotte, miss Amaxa, ma jumelle de musique et Dr Techay pour les intimes, merci pour ta gentillesse, ta sympathie, ton humour, tes conseils concernant la culture cellulaire, FACS ou autre. Merci pour toutes les aprèms L2 qui rendaient les manips si agréables, pour toutes les soirées et fous rires partagés. Merci d'avoir été mon infirmière à plusieurs reprises, à refaire mes pansements. Merci pour ta présence et pour toutes les fois où tu m'as écoutée râler, me plaindre, ou me confier ...

Un grand merci à mes collègues de bureau : Céline qui s'est délocalisée pour quelques mois pour rédiger dans notre bureau, merci pour ta sympathie, ma jumelle d'humour avec toutes les répliques de Gad, un jour nous aussi on sera des crèmes brûlées. Merci à Marion, ma jumelle de projet, avec laquelle j'ai partagé l'amour des translocations, et des nucléases. On était les ovnis des journées ROSCOFF : les filles du muséum qui font de la cancéro !!! Hé oui ça existe !!! Merci pour tous les gouters et pauses café et after labo et congrès qu'on a partagé. Merci à Lorena pour ta sympathie, ta gentillesse j'ai été heureuse de te côtoyer, au labo ou ailleurs. Merci au club des « LUC » : Armel, Marine, Charlotte, Karine et Anthony. Merci pour tous les midis rock 'n' roll, les fous rires, pour tous les bons moments partagés au labo ou ailleurs, j'en garderai un très bon souvenir. Merci pour votre soutien et bonne humeur.

Je remercie les autres membres du laboratoire avec qui j'ai pu interagir au quotidien : Loïc Ponger, Emmanuelle, Carole, Tula, Judith, Jean-Baptiste boulet, Patrick, Sacha, Chantal, Danièle, et Laureline. Un grand merci aux filles du secrétariat : Fara, Patricia et Corinne.

Je voudrais terminer en disant un grand merci à ma famille : tout particulièrement à mes parents, de m'avoir laissé faire mes propres choix, merci de m'avoir permis de quitter mon Algérie natale pour réaliser mon rêve, sans vous je n'aurais jamais pu aller aussi loin. Merci de m'avoir fait confiance et d'avoir cru en moi. Merci pour votre amour et soutien indéfectible, cette thèse est aussi la votre. Un grand merci à ma sœur qui a été là dans les moments difficiles, merci pour ton soutien inconditionnel et surtout merci de veiller sur moi. Je remercie également mes frères pour leur soutien et leur confiance. Enfin je remercie mon fiancé de m'avoir soutenu au jour le jour, d'avoir partagé avec moi les moments de stress et de doute. Merci pour ton extrême gentillesse et ton réconfort.

Je dédie cette thèse à mes parents

Abréviations

ADN	Acide DésoxyriboNucléique
AID	Activation Induced Cytosine Deaminase
Ala	Alanine
ALCL	lymphome anaplastique à grandes cellules (Anaplastic Large-Cell Lymphoma)
ALK	Anaplastic Lymphoma Kinase
alt-NHEJ	NHEJ alternatif
AR	Récepteur Androgénique
ARN	Acide Ribonucléique
ARNm	ARN messager
APAF-1	Apoptotic Peptidase Activating Factor 1
Asp	Acide aspartique
ATRIP	ATR Interacting Protein
ATM	Ataxia-Telangectasia Mutated
ATR	A-T and Rad3 Related
BAD	Bcl-2-Associated Death promoter
BARD1	BRCA1-associated RING Domain 1 protein
BAX	Bcl-2–Associated X protein
BCL2	B-Cell Lymphoma 2
BCL-XL	BCL2-Like 1 isoform 1
BLM	Bloom syndrome, RecQ helicase-like
BRCT	BRCA1 Carboxyl-Terminal
BRCA1 et 2	Breast Cancer Protein 1 et 2
Cas9	CRISPR-associated
CCR5	C-C chimiokine de type 5
CDB	Cassure Double-Brin
Cdc	Cell-division-cycle
CDK	Cyclin Dependent Kinase
cDNA	ADN complémentaire
Der	Chromosome dérivatif

Chk 1 et 2	Checkpoint Kinase1 et 2
СНО	Chinese Hamster Ovary
C-NHEJ	NHEJ canonique
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNAs	CRISPR RNAs
CSB	Cassure Simple-Brin
CSR	Class Switch Recombination
C-terminal	Carboxy-terminal
CtIP	CTBP-Interacting Protein
DDR	DNA Damage Response
DBD	DNA Binding Domain
DNA-PK	DNA-Dependent Protein Kinase complex
DNA-PKcs	DNA-dependent Protein Kinase catalytic subunit
EXO1	Exonucléase 1
EWS	Ewing Sarcoma breakpoint region 1
FAT	FRAP, ATM, TRRAP
FATC	FAT C-terminal
FHA	Forkhead-Associated
FLI1	Friend Leukemia-virus Integration 1
gRNA	RNA guide
HDR	Homology Directed Repair
HEAT	Huntingtin Elongation-A-subunit TOR
H2AX	Variant X de l'histone H2A
γΗ2ΑΧ	H2AX phosphorylé
IR	Radiation Ionisante
iPS	cellules souches pluripotentes induites
KI	Knock-In
КО	Knock-Out
Ligl	Ligase I
LigIII	Ligase III
LigIV	Ligase IV
MDC1	Mediator of DNA Damage Checkpoint 1

MRE11	Meitoic Recombination 11 homolog
MRN	Mre11 Rad50 Nbs1
nCas9	Cas9 nickase
NF-κB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NLS	Nuclear Localization Domain
NOXA	Phorbol-12-myristate-13-acetate-induced protein 1
NPM	Nucleophosmin 1
NTase	Nucléotidyltransférase
NTD	N-Terminal Domain
OBD	Oligonucleotide/oligosaccharide-fold Binding Domain
PALB2	Partner And Localizer of BRCA2
PAM	Protospacer Adjacent Motif
PARP 1et 2	Poly ADP-Ribose Polymerases 1,2
PAXX	PAralog of XRCC4 and XLF
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PIKK	Phosphoinositide 3-Kinase related Kinases
PIP Box	PCNA-Interacting Protein
Plk3	Polo-like kinase 3
PNKP	PolyNucléotide Kinase/phosphatase
Pol μ et Pol λ	Polymérase μ et λ
pre-crRNA	précurseur CRISPR RNA
PUMA	p53 Up-Regulated Modulator of Apoptosis
Rad51	Radiation sensitive 51
RFN	RNA-guided Fokl-Cas9 Nucleases
RH	Recombinaison Homologue
RIF1	RAP Interacting Factor 1
ROS	Reactive Oxygen Species
RPA	Protein Replication A
RSS	Recombination Signal Sequences
RVD	Repeat-Variable-Diresidus
SCID	Déficit Immunitaire Combiné Sévère

SOD	SuperOxyde Dismutase
TALEN	Transciption Activator-Like Effector Nuclease
TdT	Terminal.desoxynucleotidyl Transferase)
tracrRNA	trans-activating crRNA
UDR	Ubiquitylation-Dependent Recruitment
wt	Wild type
XLF	XRCC4 Like Factor
XRCC1	X-ray Repair Complementing defective repair in Chinese hamster cells 1
XRCC4	X-ray repair Complementing defective repair in Chinese hamster cells 4
ZEP	protéine à doigts de zinc
ZFN	Zinc Finger Nuclease
5'-dRP/AP	5'-deoxyribose-5-phosphate/apurinique ou apyrimidique
53BP1	TP53BP1 tumor suppressor p53 binding protein 1

Table des matières

1	Intro	duction générale	1
1.1	L Les	Cassures Double-Brin de l'ADN (CDB)	3
	1.1.1	Les Cassures Double-Brin programmées	4
	1.1.2	Les Cassures Double-Brin accidentelles	5
1.2	2 Sigr	nalisation des Cassures Double-Brin de l'ADN	6
	1.2.1	Reconnaissance et détection des CDB	6
	1.2.2	Arrêt du cycle cellulaire après CDB	7
	1.2.3	L'apoptose en réponse aux CDB 1	10
1.3	3 Mé	canisme de réparation des Cassures Double-Brin de l'ADN 1	1
	1.3.1	Le Non Homologous End Joining canonique (C-NHEJ) 1	1
	1.3.2	Le NHEJ alternatif (alt-NHEJ) 2	23
	1.3.3	La Recombinaison Homologue (RH)	30
1.4	1 Que	elle voie choisir pour réparer une CDB ? 3	35
	1.4.1	Influence du cycle cellulaire sur le choix de la réparation et l'initiation de la	
	résect	ion	35
	1.4.2	Régulation de l'initiation de la résection par les protéines 53BP1 et BRCA1	36
1.5	5 Les	conséquences des CDB : cas des translocations chromosomiques	39
	1.5.1	Influence de la position des chromosomes sur la formation des translocations	
	chrom	osomiques 4	ł1
	1.5.2	Causes des CDB qui sont à l'origine de l'apparition des	
	translo	ocations chromosomiques4	14
	1.5.3	Mécanismes de réparation et de formation des translocations chromosomiques. 4	ł6
1.6	5 Ind	uction des translocations chromosomiques avec des nucléases artificielles4	ł7
	1.6.1	Les nucléases à doigts de Zinc (Zinc Finger) 4	18
	1.6.2	Les Transcription Activator-Like Effector (TALEs) 4	19
	1.6.3	Les CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)5	50
	1.6.4	Applications des nucléases artificielles	53
	1.6.5	Cytotoxicité et effets hors-cible des nucléases artificielles	54
2	Proie	et de thèse et résultats	57
2.1	I Arti	icle I : Étude des translocations chromosomiques dans les cellules humaines5	59
	2.1.1	Résultats supplémentaires : Knock down de XRCC4 par des ARN interférents 10)3
2.2	2 Arti	icle II : Établissement de modèles cellulaires porteurs de translocations)9
3	Discu	ission et perspectives 12	:4
4	Conc	lusions 13	19
5	Anne	exe : Article méthodes 14	11
6	Biblio	ographie16	5 4

Table des Figures

Figure 1 : Schéma des différentes sources de Cassures Double-Brin (CDB) de l'ADN et des
réponses cellulaires induites par ces dommages4
Figure 2 : Modèle d'activation de la voie de réponse aux CDB
Figure 3 : Schéma de la mise en place des points de contrôle du cycle cellulaire 10
Figure 4 : Structure cristallographique de l'hétérodimère Ku70/80 humain en présence de
l'ADN
Figure 5: Principaux domaines et clusters de phosphorylation de la DNA-PKcs
Figure 6: Principaux domaines des protéines appartenant au complexe de ligation
XRCC4/Ligase IV/XLF
Figure 7 : Les différentes étapes de réparation des CDB par le C-NHEJ.
Figure 8 : Réparation des CDB RAG-induites lors de la recombinaison V(D)J par le C-NHEJ 22
Figure 9: Utilisation des microhomologies lors du alt-NHEJ 27
Figure 10 : Représentation schématique des différentes étapes du alt-NHEJ 29
Figure 11 : Réparation par Recombinaison Homologue
Figure 12 : Régulation du choix de la voie de réparation
Figure 13 : Formation des translocations chromosomiques suite à une réparation "illégitime"
de deux CDB 41
Figure 14 : Les deux modèles proposés pour la formation des translocations
chromosomiques
Figure 15 : Induction des translocations chromosomiques par des nucléases artificielles 47
Figure 16 : Représentation schématique de l'organisation d'une nucléase à doigt de zinc
(ZFN)
Figure 17 : Représentation schématique de l'organisation d'une TALEN
Figure 18 : Représentation schématique du système CRISPR de type II
Figure 19 : Système CRISPR/Cas9 (nCas9) 52
Figure 20: Les différentes applications du genome editing en utilisant les nucléases
artificielles

Figure 21: Limitation des sites hors-cible des ZFN ou TALEN avec le développement de
variants du domaine catalytique <i>Fokl</i> 55
Figure 22 : Détection de la protéine XRCC4 par Western blot dans les lignées HCTwt et MES
Figure 23 : Taille des délétions et répartition des microhomologies des lignées traitées avec
le siXRCC4 104
Figure 24 : Moyenne du nombre de cellules comptées à 24h et 72h après expression du
couple de nucléases (ZFN ^{EWS} et ZFN ^{p84}) ou (TAL ^{NPM} et TAL ^{ALK}) dans les lignées wt,
hétérozygotes et KO pour le complexe de ligation128
Figure 25 : Analyse des cellules HCT L4 ^{-/-} Ku ^{-/flox} traitées ou non à l'hydroxytamoxifène 131
Figure 26 : Analyse des jonctions de translocations retrouvées chez les patients

1 Introduction générale

Une des plus grandes découvertes en biologie du XXe siècle a été l'élucidation en 1953 par le biologiste américain James Watson et le physicien britannique Franck Crick de la nature chimique du matériel génétique, l'ADN. L'histoire de sa découverte commence avec les travaux de G.Mendel sur les petits pois. En 1865, il publia ses résultats dans un article intitulé « Recherche sur des hybrides végétaux ». De ses résultats il en sortit 3 lois qui sont les fondements de la génétique moderne. En 1871, Friedrich Miescher, l'homme qui posa la première pierre, isole une substance blanche non protéique riche en phosphore à partir de leucocytes qu'il appela «nucléine». En 1929, Phoebus Levene à l'Institut Rockefeller identifie les éléments qui composent une molécule d'ADN (bases, sucre et phosphate). Et ce n'est que des années plus tard qu'Oswald Theodore Avery et ses collègues, en cherchant à purifier le facteur transformant du pneumocoque démontrent et confirment que l'ADN est le support de l'information génétique (Mccarty, 1943). En 1950, Chargaff publie ses travaux sur le contenu en bases azotées de l'ADN chez diverses espèces et montre que le rapport C/G ou A/T est constant et quasiment égal à un chez toutes les espèces étudiées. Cette dernière observation sera déterminante pour l'élaboration du modèle de la structure de l'ADN par Watson et Crick quelques années plus tard. Ces deux chercheurs disposent alors de la composition chimique de l'ADN (désoxyribose, bases azotées, et groupements phosphate), et surtout des clichés de diffraction aux rayons X d'ADN cristallisé, clichés pris principalement par Rosalind Franklin et Maurice Wilkins. Ainsi le 25 avril 1953, un article de la célèbre revue scientifique Nature décrit pour la première fois la structure de la molécule d'ADN.

L'ADN compte parmi les plus grosses molécules chimiques présentes chez l'homme. Cette molécule forme une pelote microscopique d'un filament de 2 mètres, de ce fait il doit être compacté à plusieurs niveaux. L'ADN est enroulé autour d'histones (protéines) pour former les nucléosomes qui sont eux-mêmes plus ou moins compactés. L'ensemble de ces sous-unités constitue la fibre de chromatine. Selon la phase du cycle cellulaire, la fibre chromatinienne est repliée en une structure plus ou moins condensée et forme alors les chromosomes.

L'ADN est le support de l'information génétique, il est présent dans toutes les cellules sous forme de double hélice. Ce patrimoine génétique est transmis à la descendance et donc doit maintenir son intégrité de génération en génération. Cependant, cette double hélice est

régulièrement mise à mal par des facteurs environnementaux ou métaboliques, ce qui peut conduire à des cassures. Ces lésions ne touchent parfois qu'un seul brin et peuvent être facilement réparées. En revanche, les cassures qui touchent les deux brins de l'hélice sont plus difficiles à réparer et peuvent conduire à une instabilité génomique.

1.1 Les Cassures Double-Brin de l'ADN (CDB)

On estime qu'une cellule subit 10⁵ lésions d'ADN par jour (Hoeijmakers, 2009), et qu'environ 10 de ces lésions seraient des cassures Double-Brin (CDB) (Lieber, 2010). Ce nombre est nettement plus bas que les autres types de dommages, mais l'impact que peut avoir une CDB est considérable.

Les CDB sont générées quand les deux brins complémentaires de l'hélice d'ADN se rompent simultanément, ces CDB de l'ADN sont considérées comme des lésions sévères et délétères pour la cellule et sont dites «génotoxiques», puisqu'elles menacent l'intégrité du génome. Un défaut de réparation de ces CDB peut causer des pertes dans l'information génétique (perte de morceaux de chromosome), des réarrangements chromosomiques (translocations, réversion etc) qui participeraient à la cancérogenèse par instabilité génomique.

Les origines des CDB de l'ADN sont multiples, elles peuvent être d'origine physiologiques et sont dites « programmées », ou accidentelles provenant de sources endogènes comme les espèces oxygénées réactives (ROS, Reactive Oxygen Species) produites par le métabolisme cellulaire, ou par des expositions exogènes telles que des agents environnementaux, chimiques ou physiques comme les ultraviolets (UV) et les rayonnements ionisants (Figure 1).



<u>Figure 1</u> : Schéma des différentes sources de Cassures Double-Brin (CDB) de l'ADN et des réponses cellulaires induites par ces dommages En bleu apparaissent les CDB programmées et en rouge les CDB accidentelles.

1.1.1 Les Cassures Double-Brin programmées

Ce sont les CDB Physiologiques. En effet, certains processus cellulaires normaux qui comportent des réarrangements de l'ADN introduisent des CDB à des sites spécifiques par des nucléases cellulaires. Ces CDB programmées surviennent par exemple au cours de la méiose *via* la nucléase Spo11 nécessaire à la ségrégation réductionnelle des chromosomes (Neale and Keeney, 2006). Elles surviennent également au cours des processus de diversification des immunoglobulines des lymphocytes B et T mis en place au cours de la recombinaison V(D)J par les enzymes RAG1 et RAG2, ainsi que lors du CSR (Class Switch Recombination) *via* l'enzyme AID (Activation Induced Cytosine Deaminase) (pour revue : Goodarzi and Jeggo, 2013).

1.1.2 Les Cassures Double-Brin accidentelles

En dehors des processus physiologiques normaux, les CDB peuvent aussi être accidentelles et peuvent être provoquées par des facteurs exogènes ou endogènes.

La majorité des altérations touchant l'ADN génomique sont des lésions spontanées et sont la conséquence de l'activité métabolique normale des cellules en présence d'oxygène. Les espèces oxygénées réactives (ROS, Reactive Oxygen Species) induisent la formation des CDB. En effet, au cours d'un processus respiratoire normal les mitochondries convertissent 0.1 à 1% de l'oxygène (O_2) en superoxide O_2^- . Les superoxydes dismutases des mitochondries SOD2, ou SOD1 dans le cytosol convertissent les O_2^- en ROS qui peuvent oxyder l'ADN et produire des cassures simple-brin et double-brin. Environ 10^{22} radicaux libres sont produits par jour dans un corps humain ce qui représenterait 10^9 ROS pour chaque cellule/heure (Hoeijmakers, 2009; Lieber, 2010).

Un autre stress endogène est l'arrêt de la fourche de réplication lors de la phase S qui, lorsqu'il est prolongé, peut induire des CDB (Saintigny et al., 2001). Au cours de la synthèse d'ADN, la progression de la fourche de réplication peut être compromise par des barrières qui ralentissent ou bloquent les fourches. Ces barrières peuvent êtres dues aux structures secondaires de l'ADN, notamment les régions hybrides ADN-ARN, ou les structures en hairpins, ou encore à la collision entre les machineries de réplication et de transcription (Zeman and Cimprich, 2014).

À ces facteurs endogènes, s'additionne l'effet potentiellement non spécifique des enzymes nucléaires présentes dans les cellules lymphoïdes. C'est le cas des enzymes RAG1 et RAG2 de la recombinaison V(D)J et de AID du CSR, responsables de la diversification du répertoire immunitaire. Ces enzymes induisent des coupures site-spécifiques sur l'ADN, mais parfois elles coupent accidentellement dans des sites hors de leur cible (*ie* en dehors du locus du gène du récepteur de l'antigène) (Mahowald et al., 2008). Cet effet hors-cible serait à l'origine d'environ la moitié des translocations chromosomiques retrouvées dans les lymphomes.

Aux facteurs endogènes provenant du métabolisme cellulaire viennent s'ajouter les facteurs exogènes et environnementaux. Les CDB peuvent être induites de façon directe par des agents physiques tels que les radiations ionisantes naturelles de l'environnement, incluant les rayons X et rayons Gamma. Environ 300 millions de particules ionisantes traversent chaque personne toutes les heures. Ces rayonnements tout au long de leur cheminement dans le corps, peuvent à la rencontre de molécules d'eau (H₂O) former des radicaux libres et endommager l'ADN en créant des CDB (Lieber, 2010).

Certains agents chimiques peuvent induire des CDB en altérant certains processus cellulaires telle que la réplication. Parmi ces agents chimiques, il y a le cisplatine et l'hydroxyurée. Les inhibiteurs des topoisomérases, tels que la camptothécine pour la topoisomérase 1 provoquerait également un stress réplicatif en affectant les fourches de réplication et éventuellement des CDB. Ces molécules sont très utilisées en thérapie antitumorale et certaines peuvent être la cause de leucémies secondaires (Burhans and Weinberger, 2007; McClendon and Osheroff, 2007; Pommier et al., 2010).

1.2 Signalisation des Cassures Double-Brin de l'ADN

1.2.1 Reconnaissance et détection des CDB

Dés lors qu'une CDB est détectée, les cellules mettent en place un ensemble de réponses qui se traduisent par : le remodelage de la chromatine (par phosphorylation, ubiquitination, sumoylation, acétylation et méthylation des histones), l'arrêt du cycle cellulaire, la sénescence ou même l'apoptose (Ciccia and Elledge, 2010). La réponse aux dommages à l'ADN est principalement médiée par des protéines de la famille des protéines kinases PIKKs (phosphoinositide 3-kinase related kinases): ATM, ATR et DNA-PKcs, mais également par des membres de la famille des poly (ADP-ribose) polymerases (PARP) (Harper and Elledge, 2007). La famille PARP compte 16 membres, mais uniquement PARP1 et 2 ont été impliqués directement dans la réponse aux dommages à l'ADN (Schreiber et al., 2006).

Lors de lésions de type CDB, des détecteurs sont rapidement recrutés au site de la cassure. Selon un modèle proposé, les tout premiers détecteurs recrutés au niveau de ces lésions seraient PARP 1 et PARP2 (Poly-ADP Ribose Polymerase 1 et 2), qui via leurs motifs doigt de zinc reconnaissent et lient les CDB, ils catalysent alors l'addition de chaînes Poly-ADP-Ribose (PAR) sur les histones au niveau de la lésion, et sur PARP1 elle-même (Schreiber et al., 2006). Cette PARylation conduit à un remodelage de la chromatine qui induirait le recrutement du complexe MRN (constitué des protéines MRE11-RAD50 et NSB1) (Haince et al., 2008). Ce complexe va ensuite entrainer l'activation d'ATM en induisant son autophosphorylation sur plusieurs sérines et notamment sur la serine S1981, ce qui lui confère son activité kinase (Bakkenist and Kastan, 2003; Lee and Paull, 2005). Contrairement à ATM, ATR est activée principalement lors du blocage des fourches de réplication. En effet, ces lésions sont reconnues et stabilisées par RPA (Replication Protein A) qui va ensuite recruter la protéine ATRIP (ATR Interacting Protein), qui interagit avec ATR et l'active (Cimprich and Cortez, 2008). Les voies ATM ou ATR activées, le variant d'Histone H2AX est phosphorylé au niveau de la sérine S139, devenant ainsi yH2AX (Rogakou et al., 1998). Cette phosphorylation permet le recrutement d'un autre médiateur, MDC1 (Mediator of DNA damage Checkpoint 1) via son domaine BRCT. Le signal yH2AX va alors être amplifié jusqu'a plusieurs kilo-bases (Rogakou et al., 1999) suite à la fixation de MDC1 au domaine FHA (Forkhead-Associated) d'ATM. Cela va créer une boucle rétroactive positive qui permet de maintenir ATM sous sa forme active à proximité des sites des dommages (Lou et al., 2006). MDC1 et yH2AX servent ainsi de plateforme pour le recrutement d'autres protéines telles que 53BP1 (p53 Binding Protein 1) et BRCA1 (Breast Cancer 1) qui sont importantes dans la mise en place du NHEJ (Non Homologous End Joining) et de la Recombinaison Homologue (RH) respectivement (Harper and Elledge, 2007) (Figure 2).

1.2.2 Arrêt du cycle cellulaire après CDB

Afin de déterminer le taux de dommages causés par les CDB, un mécanisme de surveillance se met en place en activant les points de contrôle ou «checkpoints» des phases G1/S et G2/M du cycle cellulaire. Soit il y a réparation des CDB et reprise du cycle normal, soit lorsque les dommages induits sont trop importants et non réparable, il y a entrée en sénescence ou en apoptose.



<u>Figure 2</u>: D'après (Ciccia and Elledge, 2010). Modèle d'activation de la voie de réponse aux CDB. La formation des CDB active PAPP1, ce dernier va induire le recrutement du complexe MRN/ATM au site de la lésion. L'activation de l'activité kinase d'ATM par MRN conduit à la phosphorylation de CHK2 et p53 en plus de plusieurs facteurs de la DDR. L'induction de la cascade dépendante de YH2AX conduit au recrutement de MDC1, RNF8, RNF168, BRCA1 et 53BP1 aux CDB.

Les principaux acteurs du checkpoint sont Chk1 et Chk2 (Checkpoint 1 et 2). Chk1 est préférentiellement activé par ATR en réponse aux blocages des fourches de réplication, Chk2 est activé par ATM en réponse aux CDB (Poehlmann and Roessner, 2010). Les voies ATM/Chk2 et ATR/Chk1 activées, régulent alors l'arrêt du cycle cellulaire en inactivant les phosphatases de la famille Cdc25 par phosphorylation. Cela engendre une accumulation de complexe Cdk-Cycline sous forme inactif et un arrêt du cycle cellulaire. La famille des Cdc25 compte trois isoformes dont les fonctions diffèrent : Cdc25A, Cdc25B et Cdc25C. Cdc25A contrôle la transition G1/S du cycle cellulaire, tandis que Cdc25B et Cdc25C contrôlent l'entrée en mitose (Donzelli and Draetta, 2003).

Lors de la transition G1/S, Cdc25A sous sa forme active (non phosphorylée), va activer le complexe Cdk2-CyclineE en le déphosphorylant et induire ainsi son activation. Ceci entraine la progression du cycle cellulaire de la phase G1 à S. Cependant lorsque un dommage à l'ADN se produit, Chk1 ou Chk2 sont activées et phosphoryle Cdc25A sur le résidu S123, qui est alors ubiquitinylée et degaradée par le protéasome. Cette dégradation de Cdc25A aura pour conséquence une accumulation du complexe Cdk2-CyclineE sous sa forme inactive hyperphosphorylée, empêchant ainsi toute progression du cycle cellulaire (pour revue : Poehlmann and Roessner, 2010). La transition G2/M quant à elle est régulée par la voie Chk1/Chk2-Cdc25C/Cdk1. Ce point de contrôle est important car il permet d'éviter la propagation des dommages mal réparés avant l'entrée en mitose. De la même manière, Chk1/Chk2 va inactiver Cdc25C en phosphorylant le résidu Sérine S216, empêchant l'activation du complexe Cdk1-CyclineB et donc toute progression du cycle cellulaire. Cdc25C phosphorylée se lie à la protéine 14-3-3, celle-ci l'exclue du noyau et la séquestre dans le cytoplasme (Reinhardt and Yaffe, 2009) (Figure 3).

Le facteur de transcription p53 est également activé par les kinases ATM, ATR, DNA-PKcs ou Chk1 /Chk2. La protéine p53 régule les transitions G1/S et G2/M du cycle cellulaire. En effet, en réponse aux dommages à l'ADN, p53 est phosphorylée sur plusieurs résidus, ce qui induit la transcription de nombreux gènes cibles tel que p21^{WAF1}. Ce dernier va inhiber le complexe Cdk2-CyclineE, induisant un arrêt du cycle en phase G1 et la formation du complexe mitotique Cdk1-CyclineB induisent l'arrêt du cycle en G2 (pour revue: Poehlmann and Roessner, 2010).

9



Figure 3 : Schéma de la mise en place des points de contrôle du cycle cellulaire. (A): Suite aux dommages à l'ADN les voies ATM/ATR sont activées, celles-ci phosphorylent et activent Chk2/Chk1 respectivement. Chk2/Chk1 phosphoryle à leur tour Cdc25A sur le résidu serine S123 et l'inactive, ce qui accélère sa dégradation par la voie ubiquitine-dépendante. Le complexe Cdk2-CyclineE inactif sous sa forme hyperphosphorylée s'accumule alors dans le noyau, empêchant toute progression du cycle cellulaire de la phase G1 à S. (B): Lors de la transition G2/M, Chk2/Chk1 phosphorylent Cdc25C sur le résidu sérine S216 et l'inactive, maintenant ainsi le complexe Cdk2-CyclineB inactif, et cause un arrêt en phase G2 du cycle cellulaire. Enfin Cdc25C phosphorylée se lie alors à la protéine 14-3-3 qui l'exclue du noyau et la séquestre dans le cytoplasme.

1.2.3 L'apoptose en réponse aux CDB

Si les cellules ne parviennent pas à réparer les CDB de l'ADN, celles-ci arrêtent de proliférer et rentrent en sénescence, un arrêt irréversible du cycle cellulaire. Elles peuvent également enclencher un mécanisme de mort cellulaire programmée, l'apoptose.

Les voies ATR/Chk1 et ATM/Chk2 activées, phosphorylent la protéine p53 qui est l'acteur majeur de l'apoptose. La protéine p53 fortement activée, s'accumule dans le noyau et induit la transcription de protéines pro-apoptotiques, telles que BAX, BID, PUMA et NOXA, et antagonise les protéines anti-apoptotiques, BCL2 et BCL-XL. La translocation de BAX à la mitochondrie entraine la libération du cytochrome C dans le cytosol, ce dernier va se lier à la

protéine APAF-1 (Apoptotic peptidase activating factor 1) et à la pro-caspase 9 formant ainsi l'apoptosome. L'apoptosome induit l'activation des caspases, notamment la caspase 3 qui active des endonucléases qui dégradent l'ADN (Nowsheen and Yang, 2012; Roos and Kaina, 2013).

Enfin le facteur de transcription NF-κB en réponse aux dommages à l'ADN, notamment en présence excessive d'espèces réactives oxygénées, aurait un rôle pro-apoptotique en induisant la transcription du ligand FAS, stimulant ainsi l'apoptose (Ryan et al., 2000).

1.3 Mécanisme de réparation des Cassures Double-Brin de l'ADN

Les CDB de l'ADN sont parmi les lésions les plus dangereuses, du fait de leurs conséquences potentiellement délétères pour la cellule. Des mécanismes spécifiques de réparation sont donc indispensables et sont finement régulés pour détecter et réparer ces lésions afin de maintenir la stabilité génétique. Deux voies majeures sont décrites pour réparer les CDB : la Recombinaison Homologue (RH), et le Non Homologous End Joining canonique (C-NHEJ). Il existe également une troisième voie souvent appelée le NHEJ alternatif (alt-NHEJ), qui est notamment active en cas d'inactivation du C-NHEJ.

1.3.1 Le Non Homologous End Joining canonique (C-NHEJ)

Ce mécanisme de réparation est actif tout au long du cycle cellulaire et n'a pas besoin d'homologie de séquences, il consiste simplement à "recoller" les extrémités séparées par la cassure (Lieber, 2008).

1.3.1.1 Les acteurs du C-NHEJ

L'hétérodimère Ku70/80 : (Figure 4)

L'hétérodimère Ku70/80, l'un des acteurs majeurs du C-NHEJ était initialement décrit comme un autoantigène présent dans le sérum de patients atteints de maladies autoimmunes (Mimori et al., 1981). Ce complexe formé de deux protéines de 70 et 86 kDa (Ku70 et Ku80 respectivement) est très abondant dans les cellules humaines avec environ 4.10^5 molécules/cellule dans les cellules HeLa et possède une très grande affinité pour les extrémités d'ADN avec un K_D de l'ordre du 10^{-9} M (Blier et al., 1993; Mimori et al., 1986). Cet hétérodimère se présente sous forme d'un anneau asymétrique qui entoure la molécule d'ADN, cette interaction ne dépend pas de la séquence (Walker et al., 2001). Ku peut lier une large variété d'extrémités de CDB, telles que les extrémités à bout-francs (blunt), 3' ou 5' sortantes (overhangs), ou encore les extrémités en épingle à cheveux (hairpins), faisant de ce complexe le candidat idéal pour l'initiation de la réparation (Downs and Jackson, 2004; Mimori et al., 1986; Mimoris and Hardin, 1986). Ku protège les extrémités des cassures de la dégradation non spécifique des nucléases. Une étude menée *in vitro* montre que Ku bloquerait l'action d'EXO1 et de MRN (voir partie 1.3.3.1) (Sun et al., 2012). Ku est également impliqué dans le maintien et la stabilité des télomères. Il a été montré qu'un KO de Ku80 dans les cellules humaines, induisait la mort cellulaire par perte massive de télomères (Wang et al., 2009). À l'inverse les cellules de souris sont viables en l'absence de Ku (Taccioli et al., 1993; Zhu et al., 1996).



<u>Figure 4</u> : D'après (Walker et al., 2001) structure cristallographique de l'hétérodimère Ku70/80 humain en présence de l'ADN. Ku70 est indiqué en jaune, Ku80 en rouge et l'ADN en gris.

DNA-PKcs : (Figure 5)

C'est une sérine thréonine kinase de 465 kDa appartenant à la famille des PIKKs (phosphoinositide 3-kinase related kinases) dont l'activité est totalement dépendante de son interaction avec l'ADN. La DNA-PKcs interagit avec l'hétérodimère Ku et forme l'holoenzyme DNA-PK qui entraine l'activation de l'activité kinase de DNA-PKcs (Gottlieb and Jackson, 1993). Ce complexe a plusieurs cibles parmi les acteurs du C-NHEJ (dont Ku-XRCC4/LigaseIV/XLF). Néanmoins il a été montré que la phosphorylation de ces acteurs n'est pas toujours nécessaire pour la réussite du C-NHEJ (Neal and Meek, 2011). La seule cible indispensable décrite à ce jour est la DNA-PKcs elle-même, avec plus de 40 sites potentiellement phosphorylés (Dobbs et al., 2010; Meek et al., 2008). L'autophosphorylation des deux clusters ABCDE et PQR semble réguler l'accès des extrémités aux facteurs du C-NHEJ. En effet, la phosphorylation du cluster ABCDE entraîne un changement de conformation de la DNA-PKcs et va permettre l'ouverture du domaine de liaison à l'ADN, permettant ainsi aux autres acteurs de la voie NHEJ d'accéder aux brins d'ADN et de réparer la lésion. Au contraire, la phosphorylation du cluster PQR empêche l'accès aux protéines de réparation (Neal and Meek, 2011). Il a été montré que l'activité kinase de la DNA-PKcs humaine était environ 50 fois plus élevée par rapport à celle des cellules de souris (Beamish et al., 2000; Finnie et al., 1995; Meek et al., 2001).

La DNA-PKcs a également un rôle important au cours de la recombinaison V(D)J. En effet, les souris déficientes en DNA-PKcs sont incapables de résoudre les structures en hairpins, intermédiaires de la recombinaison V(D)J, produisant alors des souris immunodéficiences (Rooney et al., 2004).

Les facteurs de maturation :

Les extrémités des CDB causées par les radiations ionisantes (IR), par les espèces oxygénées réactives, ou celles générées lors de la recombinaison V(D)J (les jonctions codantes) ne sont pas directement ligables. Le C-NHEJ nécessite d'autres enzymes pour préparer ces extrémités et les rendre compatibles avec l'action de la Ligase. Parmi ces enzymes on trouve :

Artemis : est une enzyme dont le gène a été identifié comme muté chez les patients atteints de déficit immunitaire combiné sévère (SCID) (Moshous et al., 2001). Il a été montré qu'Artémis seule possédait une activité exonucléase simple-brin 5'-3'. Une fois complexée à la DNA-PKcs, elle est phosphorylée et acquiert une activité endonucléase sur les extrémités sortantes 3' ou 5', lui permettant ainsi de résoudre les extrémités en épingle à cheveux générées par le complexe RAG au cours de la recombinaison V(D)J (Ma et al., 2002).



<u>Figure 5</u>: Principaux domaines et clusters de phosphorylation de la DNA-PKcs. La région Nterminale est composée d'une répétition de domaine Heat (Huntingtin Elongation-A-subunit TOR), qui sert d'interface pour les interactions protéine-protéine. La région C-terminale est composée du domaine PI3 Kinase, ce dernier est flanqué du coté N-terminal du domaine FAT (FRAP, ATM, TRRAP) qui lie Ku, et du coté C-terminal par le domaine FATC (FAT C-terminal). La DNA-PKcs possède deux clusters de phosphorylation, qui sont indispensables à l'activation du C-NHEJ. La phosphorylation du cluster ABCDE, entraine un changement de conformation de DNA-PKcs permettant ainsi l'accès aux extrémités aux protéines du C-NHEJ. La phosphorylation du cluster PQR quant à elle maintient la DNA-PKcs aux extrémités des CDB empêchant ainsi tout accès aux protéines de la voie.

PNKP (Polynucléotide kinase/phosphatase) : c'est une enzyme bi-fonctionnelle possédant une activité 3'-phosphatase et 5'-kinase, lui permettant d'enlever un groupe 3'-phosphate et/ou d'ajouter un groupe 5'-phosphate respectivement afin de générer des extrémités compatibles et permettre la ligation (Chappell et al., 2002).

Les polymérases : certaines polymérases sont nécessaires pour remplir les brèches (gaps) avant la ligation. Trois polymérases de la famille Pol X ont été associées au C-NHEJ: la TdT, Polµ, et Polλ. Ces polymérases possèdent un domaine BRCT qui leur permet d'interagir avec

les facteurs du C-NHEJ. Pol μ et Pol λ , sont ubiquitaires et contrairement à Pol μ , le domaine lyase de Pol λ est fonctionnel, lui permettant ainsi d'éliminer les bases endommagées (Ramsden, 2011). La TdT est uniquement exprimée dans les cellules active pour la recombinaison V(D)J, elle catalyse l'addition de nucléotides aux extrémités 3' de l'ADN (Gauss and Lieber, 1996). Il a été montré que la TdT était capable d'ajouter des nucléotides au niveau des jonctions des CDB non RAG-induites (dans des cellules non lymphoïdes), et que la présence des protéines Ku80 et XRCC4 du C-NHEJ était indispensable à son activité (Boubakour-Azzouz et al., 2012).

Une étude très récente réalisée *in vitro* a montré que Ku avait une activité 5'-deoxyribose-5phosphate/apurinique ou apyrimidique (5'-dRP/AP) lyase, et était capable d'exciser les sites abasiques présents aux extrémités 5' sortantes des CDB. Cela suggère que Ku pourrait aussi être impliqué dans la maturation des extrémités des cassures (Roberts et al., 2010).

Le complexe de ligation XRCC4/Ligase IV/XLF: (Figure 6)

L'étape finale du C-NHEJ est la ligation des extrémités des cassures d'ADN par le complexe XRCC4/Ligase IV/XLF. La Ligase IV est formée d'un domaine catalytique N-terminal et de deux domaines BRCT répétés en tandem du coté C-terminal (Shell et al., 1995) (Figure 6A). Cette région de la Ligase IV est impliquée dans l'interaction avec l'hélice- α d'XRCC4, formant ainsi un complexe très stable (Sibanda et al., 2001; Wu et al., 2011). La protéine XRCC4 quant à elle, possède une tête globulaire (head), une hélice- α allongée (stalk) et un domaine C-terminal non structuré contenant la séquence de localisation nucléaire, cette protéine se présente sous forme d'homodimère (Junop et al., 2000). La protéine XRCC4 n'a pas d'activité enzymatique connue, mais celle-ci est nécessaire à la stabilité de la Ligase IV et stimule son activité (Bryans et al., 1999; Grawunder et al., 1997) (Figure 6B).

XLF/Cernunos a été découvert simultanément par deux groupes en 2006. Les mutations du gène *XLF* sont associés à un retard de croissance, à une microcéphalie, et à un déficit immunitaire combiné sévère (SCID) (Ahnesorg et al., 2006; Buck et al., 2006). XLF a une structure similaire à celle d'XRCC4, ces deux protéines interagissent *via* leur tête globulaire (Figure 6C). *In* vitro ces deux protéines forment un filament pour faciliter l'étape de ligation, en particulier en favorisant la réparation d'extrémités non cohésives (Andres et al., 2007).

XLF est également essentiel pour le remplissage de brèches (gaps) par Pol λ et Pol μ , suggérant qu'elle joue un rôle majeur dans l'alignement des deux extrémités d'ADN avant la ligation (Akopiants et al., 2009).



<u>Figure 6</u>: Principaux domaines des protéines appartenant au complexe de ligation XRCC4/Ligase IV/XLF. (A) : La Ligase IV est composée d'un domaine catalytique N-terminal, et de deux domaines BRCT du coté C-terminal, et c'est entre ces deux domaines que se fait l'interaction avec XRCC4. (B), (C): XRCC4 et XLF ont une structure similaire, ils possèdent tous deux une tête globulaire (Head) et une tige allongée qui est une Hélice- α (Stalk), ces deux protéines interagissent via leur tête globulaire.

La protéine PAXX (PAralog of XRCC4 and XLF):

Très récemment, en janvier 2015 une nouvelle protéine, PAXX (PAralog of XRCC4 and XLF, également appelée C9orf142) a été identifiée comme un nouveau facteur du C-NHEJ. En effet, deux équipes ont identifié simultanément cette protéine comme membre de la famille d'XRCC4. L'analyse cristallographique de PAXX montre que celle-ci possède une structure similaire à celle d'XRCC4 et d'XLF. PAXX interagit directement avec l'hétérodimere Ku *via* son domaine C-terminal et est ainsi recrutée aux sites des dommages. En coopérant avec XLF, PAXX stimulerait également la ligation des extrémités non cohésives *in vitro*. D'un point de vue mécanistique, PAXX stabiliserait l'assemblage de Ku et des autres protéines du NHEJ au niveau des cassures, et stimulerait l'étape de ligation (Ochi et al.; Xing et al., 2015).

1.3.1.2 Mécanisme du C-NHEJ

Le C-NHEJ débute par la reconnaissance et la fixation de l'hétérodimère Ku aux extrémités des CDB (Mari et al., 2006). Ce complexe est sous forme d'un anneau asymétrique, dans lequel l'extrémité de l'ADN s'enfile (Walker et al., 2001). Tout en maintenant les extrémités l'une prés de l'autre, l'hétérodimère Ku recrute la sous-unité catalytique de la DNA-PKcs via le domaine C-terminal de Ku80 (Singleton et al., 1999). Cela induit la translocation de Ku à l'intérieur de l'hélice de l'ADN, de façon à ce que la DNA-PKcs se fixe aux extrémités des cassures (Yoo and Dynan, 1999). L'interaction de deux molécules de DNA-PKcs aux extrémités des CDB, induit la formation de la synapse et stimule l'activité kinase de la DNA-PKcs (DeFazio et al., 2002). Le complexe Ku-DNA-PKcs forme l'holoenzyme DNA-PK (Yaneva et al., 1997), qui va avoir plusieurs cible dont la DNA-PKcs elle-même qui va s'autophosphosphoryler sur plusieurs sites. Cela entraine des changements conformationnels permettant de réguler l'accès des extrémités aux protéines (partie 1.3.1.1). Ce complexe recrute également d'autres protéines dont le complexe XRCC4/LigaseIV/XLF. D'ailleurs une étude très récente montre que ce complexe agirait en amont et stimulerait l'autophosphorylation de la DNA-PKcs (Cottarel et al., 2013). Si besoin les extrémités de ces CDB vont subir l'action des nucléases et/ou de polymérases (ie Artemis, PNKP /Polu, Polλ) avant d'être religuées. Enfin le complexe XRCC4/Ligase IV/XLF lie les deux extrémités de la cassure (Davis et al., 2014) (Figure7).

Il est à noter que le modèle présenté ci-dessus est un modèle séquentiel. Il est majoritairement décrit dans la littérature comme étant le modèle dominant du C-NHEJ, et dans lequel les facteurs impliqués sont recrutés de façon progressive et séquentielle. Cependant ce modèle est remis en cause, des études d'imagerie suivant la mobilité de GFP-Ku, GFP-XRCC4, GFP-LigIV ou GFP-XLF s'accordent sur le fait que l'hétérodimère Ku est le premier facteur recruté au site de la CDB, et que celui-ci est capable de recruter les autres protéines de la voie, indépendamment de la DNA-PKcs (Mari et al., 2006; Reynolds et al., 2012; Yano and Chen, 2008). En revanche le recrutement de la DNA-PKcs en seconde position semble être controversé, remettant en cause le modèle séquentiel. En effet, la DNA-PKcs ne semble pas être nécessaire pour le recrutement des autres facteurs du C-NHEJ au site de la CDB (*ie* XRCC4, Ligase IV, et XLF) (Mari et al., 2006; Yano and Chen, 2008).

L'ordre de recrutement peut être flexible et peut dépendre de la nature et de la complexité des extrémités des CDB. En effet, une étude récente montre que seules les extrémités complexes des CDB, et donc cinétiquement plus longues à réparer nécessitent le recrutement de la DNA-PKcs. Les extrémités « simples » sont reconnues par l'hétérodimère Ku, et sont directement ligables par le complexe de ligation XRCC4/LigaseIV/XLF (Reynolds et al., 2012). De plus, il a été montré que la protéine XLF avait un rôle majeur dans l'alignement des deux extrémités d'ADN avant la ligation, et qu'elle est essentielle pour le remplissage de brèches (gaps) par Pol λ et Pol μ (Akopiants et al., 2009). Toutes ces données tendent vers une intervention précoce du complexe de ligation XRCC4/LigaseIV /XLF lors du C-NHEJ. D'ailleurs une étude très récente appuie cette hypothèse. En effet, le complexe de ligation XRCC4/LigaseIV/XLF régulerait l'autophosphorylation de la DNA-PKcs. En plus de la ligation des extrémités des CDB, il contribuerait à la stabilisation de la DNA-PK au cours de la synapse (Cottarel et al., 2013). Ces données sont à l'encontre d'un assemblage séquentiel des facteurs du C-NHEJ, et suggèrent la formation d'un complexe supramoléculaire au niveau des CDB dans lequel les complexes de ligation et le complexe DNA-PK coexistent (Cottarel et al., 2013).

1.3.1.3 Pathologies associées au C-NHEJ

Les patients ayant une déficience dans l'un des facteurs du C-NHEJ peuvent présenter plusieurs symptômes : radiosensibilité, retard de croissance, microcéphalie, ou un déficit immunitaire combiné sévère (SCID). La radiosensibilité combinée à un déficit immunitaire sévère (SCID) est nommé RS-SCID, il représente le phénotype le plus sévère.

À ce jour, on a identifié 4 gènes mutés qui sont responsables de l'apparition d'un ou de plusieurs de ces symptômes. On trouve le gène *LIG4* (codant pour la DNA Ligase IV), *XLF/ Cernunnos* (codant XLF), *PRKDC* (codant DNA-PKcs) et *DCLRE1C/SNM3/Artemis* (codant Artemis) (pour revue (Woodbine et al., 2014) (Tableau 1).



<u>Figure 7</u>: Les différentes étapes de réparation des CDB par le C-NHEJ. (A) : L'hétérodimère Ku70/80 reconnaît et lie les extrémités des CDB, il recrute la DNA-PKcs qui favorise la synapse. (B) : Formation de l'holoenzyme DNA-PK, celle-ci va s'autophosphoryler sur plusieurs sites, induisant la libération des extrémités qui deviennent accessibles aux protéines de la voie. (C) : Si les extrémités ne peuvent pas être liées directement, des nucléases et/ou polymérases sont nécessaires pour maturer les extrémités des CDB et les rendre compatibles. (D) : Enfin le complexe de ligation XRCC4/Ligase IV/XLF lie les deux extrémités de la cassure.

Protéines du C-NHEJ	Mutations identifiées	Syndrome	Symptômes
DNA-PKcs	L3062R A3574V G2013∆ Exon 16∆	Déficit Immunitaire Combiné sévère (SCID)	Lymphopénie
Ligase IV	R278H A3V+T9I +R278H R278H+E427A R814X+R580X H282L+D423X	Syndrome Ligase IV	Radiosensibilité Microcéphalie Retard de croissance Pancytopénie Lymphomes
XLF/Cernunnos	R178X R57G G267∆	Radiosensibilité-Déficit Immunitaire Combiné sévère (RS-SCID)	Microcéphalie Retard de croissance Lymphopénie
Artemis	Exon1-3 Δ+D451K M1T+H35D L70 Δ +G126D Exon14 Δ	Radiosensibilité-Déficit Immunitaire Combiné sévère (RS-SCID)	Lymphopénie

Tableau 1 : Les syndromes associés aux mutations des protéines du C-NHEJ. *A: Délétion

1.3.1.4 Mécanismes physiologiques utilisant le C-NHEJ : exemple de la recombinaison

V(D)J

La capacité du système immunitaire à montrer une réponse efficace contre une large variété d'agents pathogènes repose sur la génération d'une population diverse d'anticorps qui expriment un récepteur unique. Ces récepteurs sont formés de deux chaines lourdes (IgH) et deux chaines légères (IgL), et chaque chaine IgH et IgL comporte une région C-terminale constante et une région N-terminale variable. Les exons codant pour ces régions variables sont assemblés par un processus de réarrangement génétique appelé V(D)J.

La recombinaison V(D)J, responsable de l'assemblage des régions variables nécessite le complexe RAG (RAG1 et RAG2) pour induire des CDB (Oettinger et al.) (Figure 8) ; la TdT pour générer de la diversité aux jonctions (Alt and Baltimore, 1982) et les protéines du C-NHEJ pour lier les segments V, D et J réarrangés (Taccioli et al., 1993). Le complexe RAG génère des CDB entre deux Séquences Signal de Recombinaison (RSS) et deux séquences codantes pour les segments V, D et J. Les RSS sont composés d'un heptamètre et d'un nonamère (AT riche) très conservés et qui sont séparés par une séquence de 12 ou 23 pb. Le complexe RAG lie et clive une paire RSS selon la règle 12/23 (lie une séquence 12 RSS et une 23 RSS) et forme ainsi une synapse (Tonegawa, 1983). Le complexe RAG clive un seul des deux brins entre l'extrémité 5' de l'heptamètre de la RSS et la région codant le récepteur d'antigène. Le groupe 3'OH libre va attaquer une liaison phosphodiester du brin antiparallèle, induisant également le clivage de ce brin. Cela va aboutir à la formation de deux CDB avec des extrémités franches du coté des séquences signales RSS, non codantes (Fraser et al., 1995).

Ces extrémités sont prises en charge par le C-NHEJ. En effet, les extrémités franches non codantes sont directement jointes formant la jonction signal (jonction RSS). Les extrémités en épingle à cheveux codantes sont également liées pour former la jonction codante, cette réaction est plus complexe car elle nécessite une étape de maturation. L'heterodimère Ku reconnaît et lie les extrémités en épingles à cheveux. Il recrute ensuite la DNA-PKcs, à deux ils forment l'holoenzyme DNA-PK, qui par phosphorylation (autophosphorylation et phosphorylation des autres protéines de la voie) va recruter les autres protéines de la voie, notamment Artemis, la TdT et le complexe de ligation XRCC4/Ligase IV/XLF. Artemis complexée à la DNA-PKcs acquiert une activité endocnucléase 3' ou 5' qui va lui permettre de résoudre les extrémités en épingle à cheveux. Une fois ces extrémités ouvertes, de la diversité supplémentaire est introduite au niveau des jonctions codantes par l'action de la TdT, spécifique des lymphocytes, mais aussi par l'action des polymérases μ et λ . Enfin les extrémités sont liées par le complexe de ligation XRCC4/LigaseIV/XLF (Schatz and Swanson, 2011) (Figure 8).



Figure 8: D'après (Alt et al., 2013). Réparation des CDB RAG-induites lors de la recombinaison V(D)J par le C-NHEJ. Le complexe RAG génère des CDB entre deux Séquences Signal de Recombinaison (RSS) et deux séquences codantes pour les segments V, D et J. Ce clivage donne lieu à deux types d'extrémités : les extrémités franches « non codantes » qui seront directement jointes par le C-NHEJ formant ainsi la jonction signal (jonction RSS). Les extrémités en épingle à cheveux « codantes » quant à elles sont d'abord prises en charge par la protéine Artemis qui par son activité endonucléase résout les structures en hairpins. Ces extrémités ouvertes, la TdT augmente la diversité aux jonctions et enfin le complexe de ligation lie les extrémités pour former la jonction codante.

1.3.1.5 Le C-NHEJ est conservatif

Les études de la voie NHEJ se sont beaucoup basées sur l'analyse de la recombinaison V(D)J. Cela explique notamment pourquoi ce mécanisme est souvent considéré comme étant générateur de mutations (error-prone). Comme décrit ci-dessus lors de la génération du répertoire immunitaire, la recombinaison V(D)J fait intervenir des protéines spécialisées autres que les protéines charnières du C-NHEJ (*ie* Ku-XRCC4/LigaseIV/XLF) afin d'introduire de la variabilité au niveau des jonctions. Cette variabilité est l'œuvre de facteurs spécifiques (*ie* Artémis et la TdT) et non du C-NHEJ tout seul (Bétermier et al., 2014).

La mise en évidence de l'existence de la voie alt-NHEJ, notamment dans les cellules murines déficientes pour l'un des composants du C-NHEJ (tels que : Ku, XRCC4) et dans les mécanismes physiologiques comme le V(D)J ou le CSR, a permis de mettre la lumière sur la mutagenèse qui était attribuée au C-NHEJ et qui provient en réalité du alt-NHEJ plus mutagène (voir partie 1.3.2) (Bétermier et al., 2014; Corneo et al., 2007; Yan et al., 2007).

Lorsque les extrémités des CDB sont parfaitement compatibles, la réparation sera extrêmement précise. Mais il arrive souvent que la jonction de réparation soit refermée de façon imprécise, c'est par exemple le cas des CDB causées par les radiations ionisantes (IR) où les extrémités générées sont complexes et nécessitent donc des étapes de maturation afin de les nettoyer et de les rendre propres à l'action de la ligase, pouvant ainsi entrainer la perte de nucléotides. Le C-NHEJ est donc conservatif quand les extrémités sont parfaitement complémentaires, mais lorsque celles-ci ne le sont pas, le C-NHEJ s'adapte à la structure des extrémités des CDB, ce qui influence la qualité de la réparation (Bétermier et al., 2014; Rass et al., 2012).

1.3.2 Le NHEJ alternatif (alt-NHEJ)

Dans la littérature cette voie est retrouvée sous de nombreuses appellations : B-NHEJ, alt-NHEJ ou même MMEJ (microhomology-mediated end-joining) pour son usage très fréquent de microhomologies (petites régions homologues présentes de part et d'autre de la CDB (Figure 9). Ces voies ne sont finalement qu'une seule et même voie indépendante des facteurs du C-NHEJ (*ie* Ku, Ligase IV, et XRCC4).

Les premières évidences de l'existence du alt-NHEJ provenaient des études menées sur des cellules de levures et sur des lignées cellulaires de mammifères déficientes pour les facteurs du C-NHEJ. Ces cellules étaient capables de réparer de façon efficace des substrats plasmidiques linéaires (Boulton and Jackson, 1996; Kabotyanski et al., 1998). Cependant

l'existence de cette voie était remise en cause, notamment du fait que ces études mesuraient la ligature de plasmides épisomiques, donc en dehors d'un contexte chromosomique.

Les premières indications claires sur le fonctionnement du alt-NHEJ dans un contexte chromosomique, proviennent des études menées sur des cellules de levures et de mammifères déficientes pour les facteurs du C-NHEJ, en utilisant des substrats intrachromosomiques dans lesquels une CDB unique est induite par des méganucléases. Ainsi, un mécanisme indépendant de Ku nommé MMEJ (microhomology-mediated endjoining) a été mis en évidence dans les cellules de levure. Ce mécanisme était capable de réparer les CDB générées par la méganucléase HO (Ma et al., 2003). D'autres études menées sur des cellules de hamster (CHO) déficientes pour les protéines Ku80 et XRCC4, ont également permis de mettre en évidence une voie de réparation efficace et indépendante de Ku pour joindre les CDB générées par la méganucléase I-Scel. Les séquences des jonctions de réparation montraient de grandes délétions et une utilisation possible de microhomologies, soulignant le caractère infidèle de cette voie de réparation (Guirouilh-Barbat et al., 2004, 2007). Il a été ensuite montré que cette voie était présente même lorsque le C-NHEJ était fonctionnel. En effet, celle-ci serait notamment impliquée dans la formation des translocations chromosomiques dans les cellules murines (Simsek and Jasin, 2010; Weinstock et al., 2007). Une étude utilisant un mutant de la protéine RAG, montre que cette voie est également active lors de la recombinaison V(D)J (Corneo et al., 2007). D'autres études réalisées sur des lymphocytes B murins déficients pour les protéines Lig IV ou XRCC4, montrent que le niveau du CSR (Class Switch Recombination) atteint 50% de la fréquence observée dans les cellules sauvages, maintenant ainsi de manière robuste le CSR dans un contexte C-NHEJ déficient. Le alt-NHEJ joint parfois le locus Igh à d'autres loci générant ainsi des translocations chromosomiques (Soulas-Sprauel et al., 2007; Yan et al., 2007).

1.3.2.1 Les acteurs du alt-NHEJ

Cette voie de réparation n'est pas totalement caractérisée et peu de choses sont connues sur les partenaires du alt-NHEJ.

1.3.2.1.1 Les facteurs impliqués dans la reconnaissance et le maintien des CDB

Les extrémités des CDB doivent être reconnues et maintenues ensemble assez longtemps pour permettre la ligation. Plusieurs travaux s'accordent sur le rôle de la protéine PARP1 dans la reconnaissance des CDB lors du alt-NHEJ.

<u>La protéine PARP1</u> : Est une protéine nucléaire abondante, membre de la famille des poly (ADP-ribose) polymerase. Cette protéine impliquée dans la réponse aux dommages à l'ADN, lie aussi bien les cassures simple-brin (CSB) que les CDB (Amé et al., 2004). Lors de la réparation, PARP1 se lie aux cassures *via* son domaine ZnF présent dans la sa partie Nterminale. Une fois activée elle catalyse l'addition de chaînes Poly-ADP-Ribose (PAR) sur les histones au niveau de la lésion, et sur PARP1 elle-même (Amé et al., 2004). Cette PARylation semble nécessaire pour le recrutement des autres protéines impliquées dans la voie telles que XRCC1 et la ligase IIIα (Leppard et al., 2003; Masson et al., 1998). L'implication de PARP1 dans alt-NHEJ était initialement basée sur des expériences biochimiques et de ligature de plasmide dans des cellules déficientes en Ku (Audebert et al., 2004; Wang et al., 2006). Un *knockdown* de PARP1 dans les cellules CHO déficientes pour Ku, montre une altération de la réparation des CDB induites par I-Scel, suggérant que PARP1 pourrait promouvoir le alt-NHEJ en absence de Ku (Mansour et al., 2010). PARP1 pourrait également servir de plateforme de recrutement d'autres protéines impliquées dans le alt-NHEJ, telles que la Ligase IIIα, XRCC1, et MRN (Audebert et al., 2004; Della-Maria et al., 2011).

1.3.2.1.2 Facteurs impliqués dans la résection des extrémités des CDB

Le alt-NHEJ utilise fréquemment des microhomologies qui sont des petites séquences complémentaires présentes de part et d'autre de la cassure. Ces petites séquences homologues sont situées directement au niveau des CDB (overhangs) et peuvent directement s'apparier (avec peu de perte de nucléotides) pour stabiliser la CDB. La réparation se termine par l'ajout éventuel de nucléotides probablement par une polymérase (Figure 9A). Ces microhomologies peuvent aussi être éloignées de la cassure, nécessitant une dégradation des extrémités afin de les révéler. Une fois révélées ces microhomologies vont s'apparier et former un complexe très stable qui permettra la réparation par simple ligation (Figure 9B). Enfin, ces microhomologies peuvent résulter de l'addition aléatoire de

25
nucléotides par des polymérases (TdT et Polµ) au niveau de la jonction de la cassure, créant ainsi de nouvelles microhomologies (Figure 9C).

Le alt-NHEJ partage avec la RH l'étape de résection des extrémités des cassures, les deux nucléases Mre11 et CtIP ont été associées à cette résection (Boboila et al., 2012a). Le rôle de Mre11 a été démontré aussi bien dans les cellules murines que dans les cellules humaines en utilisant un substrat chromosomique qui permet d'évaluer l'efficacité du NHEJ. En effet, l'inhibition de Mre11 *via* un ARN interférant ou *via* un inhibiteur chimique montre une diminution de l'efficacité de réparation des CDB, aussi bien dans les cellules wt que dans les cellules déficientes pour les facteurs du C-NHEJ (Rass et al., 2009; Xie et al., 2009; Zhuang et al., 2009). La nucléase CtIP est également nécessaire pour la résection en coopérant avec Mre11. En effet, un *knockdown* de CtIP dans les cellules HEK293 montre une diminution significative du alt-NHEJ (Bennardo et al., 2008). Son inhibition dans les cellules B murines a également montré une diminution du CSR mais également de la taille des microhomologies au niveau des jonctions S μ -S α (Lee-Theilen et al., 2011). Dans les cellules murines, le alt-NHEJ est responsable de la formation des translocations chromosomiques (Simsek and Jasin, 2010). La déplétion de CtIP entraine une diminution de la fréquence de translocation et de la taille des délétions dans les cellules wt et Ku-déficientes (Zhang and Jasin, 2011).

1.3.2.1.3 Facteurs impliqués dans la ligation des extrémités des CDB

<u>La Ligase IIIα :</u>

Le gène *Lig3* code plusieurs formes de Ligase III dont les fonctions cellulaires sont différentes (Lakshmipathy and Campbell, 1999; Tomkinson and Sallmyr, 2013). La Ligase III possède une forme nucléaire et une forme mitochondriale, cette dernière est l'unique ligase présente dans cet organelle et elle est indispensable à la survie des cellules (Simsek et al., 2011a). Il a été proposé que la localisation nucléaire de la Ligase III dépendrait de son association avec la protéine XRCC1, qui elle, possède un signal d'adressage vers le noyau (Caldecott, 2003; Parsons et al., 2010). La ligase III possède un domaine BRCT du coté C-terminal, un domaine catalytique contenant un domaine de liaison à l'ADN (DBD), un domaine nucléotidyltransférase (NTase), et domaine oligonucleotide/oligosaccharide-fold binding domain (OBD), et un domaine zinc-finger du coté N-terminal (Tomkinson and Sallmyr, 2013).



<u>Figure 9</u>: Utilisation des microhomologies lors du alt-NHEJ. (A) : Présence de microhomologies directement à la cassure sur la partie simple brin (overhang), ces bases peuvent directement s'apparier (avec peu de perte de nucléotides) pour stabiliser la CDB. Enfin la réparation se termine par l'ajout éventuel de nucléotides (en orange) probablement par une polymérase. (B) : Présence de microhomologies éloignées de la cassure nécessitant une dégradation des extrémités avant réparation. Cette fois-ci pour permettre l'appariement des séquences homologues, il y a nécessité de dégrader le brin opposé. Ainsi découvertes, ces microhomologies peuvent s'apparier et former un complexe très stable qui permettra la réparation par simple ligation. (C) : Lorsque ces microhomologies sont absentes, des polymérases telles que la TdT ou Polµ peuvent ajouter aléatoirement des nucléotides au niveau de la jonction de la cassure, créant alors de nouvelles microhomologies.

La Ligase I :

La ligase I est la ligase impliquée lors de la réplication. Tout comme la ligase III, la Ligase I possède un domaine catalytique conservé (DBD,NTase,ODB), et un motif PIP (PCNAinteracting protein) du coté N-terminal qui semble ne pas avoir d'activité catalytique (Ellenberger and Tomkinson, 2008). Néanmoins ce domaine est indispensable pour l'interaction avec PCNA (proliferating cell nuclear antigen) lors de ligation les fragments d'Okazaki (Levin et al., 2000). <u>XRCC1</u>: C'est la première protéine à avoir été identifiée comme partenaire de la Ligase III, ces deux protéines forment un complexe stable dans le noyau (Caldecott et al., 1994, 1995). Cette protéine n'a pas d'activité enzymatique connue, et possède deux domaines BRCT (BRCT1, 2), un domaine N-terminal (NTD), et un domaine NLS (Nuclear Localization Domain) (Tomkinson and Sallmyr, 2013). Les protéines XRCC1 et Ligase IIIα interagissent via leur domaine C-terminal BRCT formant ainsi un hétérodimère (Cuneo et al., 2011).

Des expériences biochimiques et de ligation de plasmide ont permis de mettre en évidence le rôle de la Ligase III dans cette voie (Audebert et al., 2004; Della-Maria et al., 2011; Wang et al., 2006). Les cellules ES murines déficientes en ligase III (forme nucléaire) ont montré une diminution de la fréquence de translocation entre deux CDB induites par les nucléases à doigts de zinc (ZFN) (Simsek et al., 2011b). La ligase I a également été impliquée dans cette voie. En effet, la déplétion de cette ligase dans les cellules wt (wild-type) n'altère pas la fréquence de formation des translocations chromosomiques après induction de CDB par les ZFN, en revanche son inactivation dans les cellules déficientes en ligases III abolit la formation de translocations chromosomiques (Simsek et al., 2011b). Les cellules doublement déficientes pour XRCC1 et Ligase III n'altèrent pas le alt-NHEJ lors du CSR (Boboila et al., 2012b). Le rôle de la ligase I a également été montré dans les cellules DT40 déficientes pour la Ligase III et IV (Paul et al., 2013). La Ligase I contribue au alt-NHEJ lorsque la Ligase III est absente, cela suggérerait l'existence de deux voies alt-NHEJ, une voie Ligase III-dépendante qui utiliserait les microhomologies et une voie Ligase I-dépendante n'utilisant pas les microhomologies (Simsek et al., 2011b).

Le rôle d'XRCC1 dans le alt-NHEJ semble être plus controversé, bien que des expériences biochimiques ont montré son implication dans le alt-NHEJ (Audebert et al., 2004; Della-Maria et al., 2011), d'autres expériences montrent aucune altération du alt-NHEJ en absence d'XRCC1. Le CSR ne semble pas être altéré dans les cellules doublement déficientes pour XRCC1/ligase IV ou pour XRCC1/Ligase III (Boboila et al., 2012b; Han et al., 2012). Suggérant que la protéine XRCC1 ne serait pas indispensable lors du CSR. Il a également été montré que la déplétion du domaine BRCT d'XRCC1 n'affecte pas la fréquence de formation des translocations dans les cellules murines, ce qui suggère que l'interaction entre la Ligase III et XRCC1 ne semble pas nécessaire lors du alt-NHEJ (Simsek et al., 2011b).

1.3.2.2 Mécanisme du alt-NHEJ

Le modèle actuel suggère que le alt-NHEJ est initié par le détecteur PARP1 qui reconnait les extrémités restées libres et forme la synapse de réparation (Audebert et al., 2004). Les extrémités des cassures sont ensuite dégradées par le complexe MRN-CtIP (Bennardo et al., 2008; Rass et al., 2009) et les microhomologies seront révélées. Celles-ci vont s'hybrider, et les extrémités sortantes non appariées vont être clivées, entrainant des délétions plus ou moins importantes (Guirouilh-Barbat et al., 2004; Rass et al., 2012). Enfin la réparation se termine par l'ajout éventuel de nucléotides probablement par une polymérase. La ligation se fait grâce à la Lig III, et en son absence la Lig I prend le relai, celle-ci contrairement à la Lig III n'utiliserait pas les microhomologies (Simsek et al., 2011b) (Figure 10).



<u>Figure 10</u>: Représentation schématique des différentes étapes du alt-NHEJ. (A): Les CDB sont reconnues par le détecteur PARP1. (B): Les extrémités subissent une résection par le complexe MRN (Mre11)-CtIP. Cette résection permet de révéler des séquences complémentaires qui s'hybrident afin de stabiliser la cassure. (C): L'hybridation des microhomologies distantes de la cassure, conduit au clivage des zones non appariées provoquant ainsi des délétions. (D) Ligation des extrémités par la Lig III ou la Lig I.

1.3.3 La Recombinaison Homologue (RH)

La Recombinaison Homologue, est l'une des voies majeures qui intervient dans la réparation des CDB, celle-ci repose sur l'utilisation de la chromatide non endommagée comme substrat (template) afin de restaurer l'information perdue. Cette voie fonctionne principalement en fin S/G2 du cycle cellulaire. Lorsque la chromatide sœur est disponible, la réparation est donc totalement fidèle à la séquence sauvage (Johnson and Jasin, 2000). La RH est aussi majoritairement utilisée pour le redémarrage des fourches de réplication lors d'un blocage du à un dommage par exemple (Adamo et al., 2010; Petermann and Helleday, 2010).

1.3.3.1 Les acteurs de la RH

<u>Le complexe MRX/MRN</u>: Le complexe MRX/N est formé des protéines: MRE11, RAD50 et Xrs2/Nsb1. Ces protéines sont parmi les premiers acteurs à se localiser au niveau des CDB et sont impliquées dans la résection des extrémités lors de la RH et du alt-NHEJ (Rass et al., 2009; Xie et al., 2009). Il a été montré que Mre11 possédait une activité 5'-3' endonucléase et 3'-5' exonucléase *in vitro*. *In vivo* cette protéine coopère avec CtIP pour initier la résection des extrémités d'ADN (Paull, 2010; Symington, 2014).

<u>Sae2/CtIP</u>: Plusieurs études menées chez la levure ou dans les cellules de mammifères ont démontré que CtIP coopère avec MRX/MRN afin d'initier la résection des CDB (Sartori et al., 2007; Symington and Gautier, 2011; Takeda et al., 2007). Ces protéines sont phosphorylées par les CDKs, et la phosphotylation de Sae2 sur le résidu S267 et de CtIP sur le résidu T847 sont nécessaires pour la résection (Huertas and Jackson, 2009; Huertas et al., 2008).

<u>EXO1</u>: Est une exonucléase 5'-3' qui appartient à la famille RAD2/XPG. Elle agit sur les extrémités 5' sortantes. La résection génère des extrémités 3' sortantes qui serviront pour l'invasion de brin de la chromatide sœur (Symington, 2014).

<u>Sgs1/BLM-DNA2</u>: Sgs1 et son orthologue humain BLM appartiennent à la famille des 3'-5' RecQ hélicases impliquées notamment dans la résolution des jonctions de Hollyday. (Symington, 2014). <u>RPA (Protein Replication A)</u> : C'est une protéine hétérotrimérique composée de RPA1, RPA2 et RPA3 (70 kDa, 32 kDa, et 14 kDa respectivement). RPA possède une très forte affinité pour les extrémités simple-brin (Wold, 1997). Après la résection, RPA lie les extrémités 3' simple-brin sortantes générées par les nucléases précédentes afin de les protéger, et d'empêcher la formation de structures secondaire qui inhiberaient l'assemblage des filaments Rad 51 (Ferretti et al., 2013).

<u>Rad 51</u>: La protéine Rad 51 lie aussi bien l'ADN simple et double-brin, c'est une ATPase qui catalyse la recherche d'homologie et l'échange de brin lors de la Recombinaison Homologue (RH). Rad 51 va déplacer RPA pour former à son tour un filament et permettre l'invasion de brin (Li and Heyer, 2008).

<u>BRCA1</u>: La protéine BRCA1 (Breast Cancer 1) est composée de 1863 acides aminés et joue un rôle important dans le maintien de l'intégrité du génome. Les études biochimiques et structurales ont révélé un domaine «RING» N-terminal possédant une activité E3 ubiquitine ligase et un domaine BRCT C-terminal, qui interagit avec les protéines phosphorylées (Roy et al., 2012). Le domaine E3 ubiquitine ligase est activé suite à son association avec la protéine BARD1 (BRCA1-associated RING domain 1 protein). Ce complexe est impliqué notamment dans l'ubiquitinylation de CtIP (Roy et al., 2012). Le recrutement de BRCA1 aux CDB se fait par son interaction avec le macro-complexe Abraxas-RAP80, qui se lie aux histones ubiquitylées (Wang et al., 2007).

<u>BRCA2</u>: La protéine BRCA2 (Breast Cancer 2) est une très grande protéine de 3418 acides aminés dépourvue d'activité enzymatique dont le rôle est complexe. BRCA2 possède plusieurs domaines : l'extrémité N-terminale lie la protéine PALB2, 8 répétitions BRC qui sont responsables du recrutement de Rad51, et un domaine de liaison à l'ADN (DBD) qui lie la protéine DSS1. L'extrémité C-terminale de BRCA2 contient un site de phosphorylation des CDKs (S3291) reconnu également par Rad 51, et une séquence NLS (Roy et al., 2012).

<u>PALB2</u>: La protéine PALB2 (Partner And Localizer of BRCA2) est une protéine de 1186 acides aminés qui est indispensable pour le recrutement de BRCA2 aux sites des CDB (Xia et al., 2006). En l'absence de PALB2 les protéines BRCA2 et Rad51 ne sont plus recrutées aux cassures (Xia et al., 2006, 2007). La protéine PALB2 lie BRCA1 à BRCA2. En effet, BRCA1 se lie à PALB2 via le « coiled-coil motif » présent sur chacune de ces deux protéines et la recrute aux sites des dommages. PALB2 se lie ensuite au domaine N-terminal de BRCA2 *via* son domaine WD40 et est alors recrutée. Une fois recrutée la protéine BRCA2 à son tour va fixer et recruter Rad51 (Park et al., 2014). Il a été montré que la protéine PALB2 pouvait lier directement Rad51 indépendamment de BRCA2, mais le mécanisme par lequel elle est recrutée reste encore inconnu (Buisson et al., 2010; Dray et al., 2010).

1.3.3.2 Mécanisme de la Recombinaison Homologue (RH)

La recombinaison homologue est initiée par la résection des extrémités des CDB afin de créer des extrémités 3' sortantes qui sont capables d'envahir l'ADN double brin contenant une séquence homologue (Mimitou and Symington, 2009). Des études réalisées chez Saccharomyces cerevisiae suggerent que la résection qui permet l'initiation de la RH se fait en deux étapes. En effet, le complexe MRX (MRE11, RAD50 et XRS2 (NBS1 chez les eucaryotes) ainsi que la protéine Sae2 (CtIP), sont nécessaires pour l'initiation de la résection. Les extrémités 3' générées suite à cette résection ménagée, sont alors substrats de nucléases plus processives, telles que EXO1 ou du couple helicase/nucléase BLM/DNA2 (Mimitou and Symington, 2008; Zhu et al., 2008). Le suppresseur de tumeur BRCA1 semble interagir avec le complexe MRN et CtIP, ce qui suggère sa probable implication dans la résection (Schlegel et al., 2006; Stark et al., 2004). Chez les mammifères, il a été montré que l'exonucléase EXO1 pouvait réséguer l'ADN in vitro, son activité est stimulée par l'hélicase BLM, l'orthologue de Sgs1 (Nimonkar et al., 2008). La protéine hétérotrimérique RPA lie et stabilise ces extrémités 3' simple-brin sortantes générées par la résection, évitant ainsi la formation de structures secondaires qui inhiberaient l'étape suivante (San Filippo et al., 2008). La protéine BRCA1 semble interagir avec la protéine BRCA2 via un une protéine partenaire PALB2. Une mutation de cette protéine alerterait l'interaction entre ces deux protéines, induisant une diminution de la RH. BRCA1 permet donc l'accumulation de PLAB2 qui à son tour va recruter BRCA2 (Sy et al., 2009a, 2009b; Zhang et al., 2009). Le complexe BRCA2-PALB2 recrute alors l'ATPase Rad51, cette protéine catalyse la recherche d'homologie et l'échange de brin, induisant la formation d'un filament nucléoprotéique qui va s'enrouler sur l'extrémité de l'ADN (Sung, 1994; West, 2003). Ce filament nucléoprotéique est stabilisé par Rad54, cette dernière stimule la recherche d'homologie et le remodelage de la chromatine pendant la recherche et l'échange de brin (Heyer et al., 2006; Raoul Tan et al., 2003). Le filament nucléoprotéique Rad51 va envahir la séquence homologue formant ainsi la D-loop, ainsi plusieurs scénarios sont envisageables (San Filippo et al., 2008). En effet, dans un premier modèle appelé DSBR model (Double Strand Break Repair), l'extrémité 3' de la D-loop est rallongée par une ADN polymérase en utilisant le brin complémentaire comme matrice et le brin envahissant peut se lier à l'extrémité 5' de la CDB « capture de la seconde extrémité », ce qui va former des structures appelées jonctions de Holliday. Ces structures sont caractéristiques avec deux brins des deux molécules d'ADN impliqués dans la recombinaison qui se croisent (West, 2003). Chez l'homme la résolution de ces jonctions peut se faire par des endonucléases (Mus81-Eme1, BLM-TOPIIIa, Gen1 ou SXL4-SXL1) avec ou sans crossing-over. Dans le second modèle SDSA (Synthesis Dependent Strand Annealing), le brin envahissant va s'apparier à la séquence complémentaire exposée par la résection et poursuivre la synthèse d'ADN pour réparer la cassure sans générer de jonctions de Holliday (Kass and Jasin, 2010; Klein and Symington, 2009; Mimitou and Symington, 2009) (Figure 11).



<u>Figure 11</u>: Réparation par Recombinaison Homologue. (a): La réparation est initiée par la résection ménagée impliquant le couple MRN-CtIP, ce qui génère des extrémités qui vont être substrats d'autres nucléases plus processives telles que EXO1 et DNA2. Les extrémités simple-brin générées vont être recouvertes par la protéine RPA. Les médiateurs BRCA2-PALB2 permettent le recrutement de l'ATPase Rad51, celle-ci déplace RPA et favorise la formation du nucléofilament. Ce dernier envahit la séquence matrice à la recherche d'homologie et forme la D-loop. La poursuite de la RH se fait en fonction de la capture de la seconde extrémité. (b): SDSA : sans capture de la seconde extrémité, le brin envahissant va s'apparier à la séquence complémentaire exposée par la résection et poursuivre la synthèse d'ADN pour réparer la cassure sans générer de jonctions de Holliday. (c): DSBR : l'extrémité 3' de la D-loop est rallongée par une ADN polymérase en utilisant le brin complémentaire comme matrice et le brin envahissant peut se lier à l'extrémité 5' de la CDB « capture de la seconde extrémité », ce qui va former une double jonction de Holliday. Cette structure peut être résolue par des endonucléases.

1.4 Quelle voie choisir pour réparer une CDB ?

Les CDB sont des lésions très dangereuses pour la cellule, pouvant entraîner des événements mutagènes si elles sont réparées de façon inappropriées ou encore la mort cellulaire si celles-ci ne sont pas réparées. Les cellules utilisent deux voies majeures afin de réparer ces lésions: le Non Homologous End Joining (NHEJ) et la Recombinaison Homologue (RH). Comment la cellule fait elle pour choisir entre ces deux voies ? Quels sont les facteurs qui régissent cette régulation?

Ces deux voies de réparation sont régulées par plusieurs facteurs, notamment la phase du cycle cellulaire dans laquelle se produit la cassure et les protéines recrutées (protéines pivots).

1.4.1 Influence du cycle cellulaire sur le choix de la réparation et l'initiation de la résection

Le choix d'utiliser le NHEJ ou la RH pour réparer des CDB semble être influencé par la phase du cycle cellulaire. En effet la RH est généralement restreinte aux phases S et G2 du cycle cellulaire, lorsque une chromatide sœur est disponible pour servir de matrice de réparation, assurant ainsi une réparation fidèle et efficace. Le NHEJ quant à lui est actif tout au long du cycle cellulaire, et semble être plus fréquent en phase G1 (Symington and Gautier, 2011). En 2004, deux études réalisées chez Saccharomyces cerevisiae ont montré pour la première fois que Cdk1 (Cycline Dependant Kinase 1) était essentielle pour le choix de la voie de réparation en favorisant la résection en phase G2 du cycle cellulaire (Aylon et al., 2004; Ferretti et al., 2013; Ira et al., 2004). Il a été montré que la résection pouvait être inhibée en bloquant les cellules de levures en phase G1 du cycle, ou encore inhibant les CDKs (Cycline Dependant Kinase) (Aylon et al., 2004; Ira et al., 2004). Ce mécanisme semble être conservé dans les cellules humaines, où de la même manière les CDKs limiteraient la résection aux phases S/G2 du cycle cellulaire (Jazayeri et al., 2006). Une autre étude, montre que la stimulation de la résection passerait par la phosphorylation de la protéine CtIP et Sae2 (Saccharomyces cerevisiae) par les CDKs en S/G2. Cela induit l'émergence d'un modèle où les CDKs phosphorylerait CtIP à la transition G1/S pour activer la résection (Huertas and Jackson, 2009; Huertas et al., 2008). Deux études très récentes ont monté que les extrémités complexes joueraient un rôle crucial dans le choix de la voie de réparation lors de la phase G1 du cycle cellulaire. Ces extrémités complexes nécessitant une résection, orienteraient la réparation vers le alt-NHEJ. Les nucléases CtIP, Mre11, et EXO1 seraient indispensables pour la réparation de ces extrémités. L'une de ces deux étude a montré que la phosphorylation de CtIP sur les résidus T847 et S327 passerait par la Plk3 (Polo-like kinase 3) (Averbeck et al., 2014; Barton et al., 2014).

1.4.2 Régulation de l'initiation de la résection par les protéines 53BP1 et BRCA1

Les protéines 53BP1 et BRCA1 ont émergé en tant que régulateurs pivots du choix de la réparation des CDB par le NHEJ ou la RH respectivement. Brièvement la protéine 53BP1 (TP53BP1, tumor suppressor p53 binding protein 1) est une protéine de 1972 acides aminés, qui est dépourvue d'activité enzymatique connue et est organisée en multidomaines. Cette protéine possède un domaine BRCT (BRCA1 carboxy-terminal) du coté C-terminal qui lui permet de se lier à la protéine p53, et un large domaine N-terminal qui s'étend sur plus de la moitié de la séquence de 53BP1 contenant 28 sites S/T-Q, qui sont phosphorylés par ATM et/ou ATR lors de dommages à l'ADN. Ces sites de phosphorylation ne sont pas nécessaires pour le recrutement de 53BP1 aux niveaux des foyers de dommages à l'ADN mais sont nécessaires pour la réparation en elle-même, parce qu'ils permettent le recrutement d'autres facteurs tels que RIF1 (RAP Interacting Factor 1) (Panier and Boulton, 2014). La protéine 53BP1 est recrutée aux sites des dommages en se fixant aux histones modifiés. En effet, 53BP1 reconnaît l'histone H4 mono ou diméthylé sur la lysine 20 (H4K20me2) via son « Tudor motif », et interagit avec l'histone H2AK15 ubiquitinylé (H2AK15ub) via son motif UDR (ubiquitylation-dependent recruitment) (Panier and Boulton, 2014). Par exemple durant le CSR (Class Switch Recombination), AID introduit des CDB séparées de 60 à 200 Kb dans les régions Switch du gène codant pour les chaines lourdes des immunoglobulines (IgH). Ces CDB ont lieu au cours de la phase G1 du cycle cellulaire et sont prises en charge par le NHEJ (C-NHEJ/alt-NHEJ), ce qui est essentiel pour l'assemblage des anticorps de différentes classes (Stavnezer et al., 2008). Une déficience de la protéine 53BP1 entraine un défaut du CSR. En effet, les souris déficientes en 53BP1 montrent une baisse de plus de 90% de cellules B par rapport aux souris sauvages (Difilippantonio et al., 2008; Manis et al., 2004). Deux propriétés majeures permettent à la protéine 53BP1 de stimuler la réparation des CDB distales par le NHEJ: sa capacité à inhiber la résection des extrémités, mais aussi à maintenir à proximité des extrémités très éloignées afin de faciliter la ligation en créant une synapse (Chapman et al., 2012). Plusieurs travaux récents, ont permis de mieux comprendre le rôle de la protéine 53BP1 dans l'inhibition de la résection. En effet, celle-ci afin de promouvoir le NHEJ en inhibant la résection en phase G1, coopère avec une autre protéine, qui est RIF1 (RAP Interacting Factor 1). Une déficience des protéines 53BP1, de RIF1, ou des deux protéines au même temps affecte de façon comparable le NHEJ, confirmant que ces protéines collaborent afin de promouvoir le NHEJ (Chapman et al., 2013; Escribano-Díaz et al., 2013; Di Virgilio et al., 2013).

Il a été montré que la protéine BRCA1 interagissait directement avec CtIP *via* son domaine BRCT. La phosphorylation de CtIP sur le résidu S327 est essentielle à cette interaction, mais aussi indispensable pour la RH. Suggérant une implication de BRCA1 dans la régulation de la résection (Yu and Chen, 2004; Yun and Hiom, 2009). Cependant, d'autres études remettent en cause toute implication de BRCA1 dans la régulation de la résection. En effet, la phosphorylation de CtIP est effectivement nécessaire pour l'interaction BRCA1-CtIP, en revanche elle n'est pas requise pour l'initiation de la résection (Nakamura et al., 2010; Reczek et al., 2013). Des études réalisées *in vivo*, montrent que la protéine BRCA1 joue un rôle crucial dans la compétition qui a lieu entre la RH et le NHEJ. Les anomalies chromosomiques ainsi que la tumorigenèse associée à la perte de BRCA1 sont restaurées en délétant la protéine 53BP1 (Bouwman et al., 2010; Bunting et al., 2010).

Il est important de noter que la perte de BRCA1 conduit au recrutement de 53BP1 aux sites des CDB en phase G2 du cycle cellulaire, ce qui suggère que 53BP1 est alors recrutée au cours de la transition S/G2. De même, le *knockdown* de 53BP1 conduit à l'apparition de foyers BRCA1 en phase G1, ce qui indique que le potentiel de recrutement de BRCA1 reste intacte en phase G1, mais que celui-ci est bloqué d'une manière 53BP1 dépendante (Escribano-Díaz et al., 2013).

En 2013, plusieurs équipes ont montré l'existence d'une relation antagoniste entre les protéines 53BP1 et BRCA1 lors du choix de la voie de réparation des CDB. En effet, lors de la phase G1 du cycle cellulaire, le NHEJ est la voie utilisée par les cellules pour réparer les CDB. Selon le modèle proposé, RIF1 cofacteur de 53BP1 bloque le recrutement de BRCA1 aux sites des CBD, inhibant ainsi la résection et donc l'initiation de la RH. A l'inverse, au cours de la transition S/G2 en présence de la chromatide sœur, les cellules utilisent la RH afin de réparer

les CDB. BRCA1 et CtIP empêchent le recrutement de RIF1 aux extrémités des CDB, celles-ci deviennent accessibles à la résection, permettant ainsi la RH (Chapman et al., 2013; Escribano-Díaz et al., 2013; Di Virgilio et al., 2013) (Figure 12).

L'environnement chromatinien autour du site de la cassure pourrait être un facteur important pour le choix de la voie de réparation des CDB particulièrement la voie alt-NHEJ. La chromatine subit différentes modifications en réponse aux dommages à l'ADN afin de faciliter l'accès aux extrémités des cassures aux protéines de réparation. Dans un modèle proposé, l'échange séquentiel du variant d'histone H2AZ semble réguler et restreindre l'initiation de la résection médiée par CtIP. En effet, il a été observé que l'échange de ce variant par l'ATPase P400 au niveau des CDB est requis pour l'acétylation de l'histone H4, l'ubiquitination, et le recrutement de facteurs tels que BRCA1. Lorsque l'échange du variant d'histone H2AZ n'est pas catalysé, la chromatine reste compacte. Le recrutement de BRCA1 et de Ku70/80 aux extrémités est réduit, et par conséquent une baisse de la RH et du C-NHEJ est observée. Un *knockdown* de CtIP restaure la capacité des cellules à recruter l'hétérodimère Ku70/80 aux CDB. En absence de ce variant les extrémités des cassures subissent une résection non contrôlée *via* CtIP conduisant à une augmentation d'extrémités simple-brin, la baisse de recrutement de BRCA1 et de l'hétérodimère Ku70/80, orienterait la réparation des CDB vers le alt-NHEJ (Price and D'Andrea, 2013; Xu et al., 2012).

Enfin lors de la mitose, il y aurait plusieurs niveaux d'inhibition du NHEJ afin de prévenir l'instabilité génomique. En effet, les kinases mitotiques Cdk1 et plk1 phosphorylent la protéine 53BP1 et la E3 ubiquitine Ligase RN8, empêchant ainsi leur recrutement au niveau des chromosomes mitotiques. La protéine XRCC4 est également phosphorylée par ces mêmes kinases inhibant ainsi l'activité du C-NHEJ (Orthwein et al., 2014; Terasawa et al., 2014).



<u>Figure 12</u>: Régulation du choix de la voie de réparation. En phase G1 du cycle cellulaire, RIF1 qui agit en aval de 53BP1 empêche l'accumulation de BRCA1 aux CDB bloquant ainsi la résection, ce qui favorise le C-NHEJ. Lors des phases S/G2, BRCA1/CtIP empêche l'accumulation de RIF1 aux CDB, les extrémités de ces cassures n'étant plus protégées subiront la résection, permettant l'initiation de la RH.

1.5 Les conséquences des CDB : cas des translocations chromosomiques

Les translocations chromosomiques sont des évènements cellulaires rares qui consistent en la fusion de deux chromosomes hétérologues. Il a été démontré dans les cellules mammifères, que les translocations peuvent se former après l'induction de deux CDB simultanées sur deux chromosomes hétérologues, suivie d'une ligation inappropriée. Cela conduit à deux chromosomes remaniés appelés chromosomes dérivatifs (Richardson and Jasin, 2000; Weinstock et al., 2006) (Figure 13). Ces translocations chromosomiques ont été décrites pour la première fois par Karl Sax. Ce dernier en utilisant comme modèle la plante *Tradescantia reflexa*, a montré une corrélation entre l'exposition à des rayonnements ionisants (IR) et les anomalies chromosomiques. Il y a décrit notamment des fusions entre différents chromosomes, maintenant connues comme des translocations chromosomiques (Sax, 1938). Dans les années 1960, Peter Nowell, Janet Rowley, et David Baltimore ont démontré pour la première fois qu'une anomalie chromosomique spécifique est liée à un type particulier de leucémie humaine, la leucémie myéloïde chronique caractérisée par la translocation t(9 ;22) (Chandra et al., 2011). À l'heure actuelle, plus de 300 gènes impliqués dans ces translocations ont été décrits dans la littérature (Mitelman et al., 2007). Ces translocations ont été identifiées comme signatures de tumeurs malignes. Il en a été recensé dans plus de 50000 tumeurs, qu'il s'agisse de leucémies, lymphomes, sarcomes et plus récemment dans certains carcinomes comme ceux de la prostate et les cancers du poumon (Rowley, 2008).

Les translocations retrouvées dans les cancers conduisent très souvent, soit à la surexpression d'un oncogène existant (proto-oncogène) situé près de la jonction entre les deux chromosomes transloqués, soit à l'expression d'un nouveau gène de fusion. En effet, certaines translocations sont à l'origine de nouveaux gènes de fusion transcrits en un ARNm chimérique, lui-même traduit en une nouvelle protéine hybride qui peut avoir des propriétés oncogéniques (Rowley, 2008).

La formation des translocations nécessite trois étapes fondamentales (i) : l'induction de CDB sur des chromosomes hétérologues, (ii) : le contact physique des extrémités des cassures, et enfin (iii) : la ligation illégitime de ces extrémités. L'implication des translocations dans de nombreux cancers nécessite d'étudier les mécanismes moléculaires à l'origine de la formation des translocations afin de comprendre comment ces étapes se produisent *in vivo*. Des travaux récents combinant des approches d'imageries et des analyses de genome-wide suggèrent que l'organisation spatiale des chromosomes, ainsi que les caractéristiques et les modifications de la chromatine sont les éléments clés dans la formation des translocations (Roukos et al., 2013a).



<u>Figure 13</u> : Formation des translocations chromosomiques suite à une réparation "illégitime" de deux CDB, donnant naissance à un gène de fusion : nouvel oncogène ou surexpression d'un oncogène déjà existant.

1.5.1 Influence de la position des chromosomes sur la formation des translocations chromosomiques

L'ADN interphasique a longtemps été considéré comme une pelote disposée de manière aléatoire dans le noyau. Cependant plusieurs études ont apporté des preuves d'une organisation spatiale, où les chromosomes seraient disposés en territoires chromosomiques (Cremer and Cremer, 2001; Lieberman-Aiden et al., 2009; Meaburn et al., 2007; Parada et al., 2002). Cette organisation spatiale du génome est tissu et cellule spécifique, et peut être influencée par l'activité transcriptionelle des cellules, ou encore par leur prolifération (Parada et al., 2002; Roukos et al., 2013a).

L'organisation spatiale du génome a émergé comme un facteur déterminant dans la formation des translocations chromosomiques (Meaburn et al., 2007). Ce rôle présumé de l'organisation spatiale du génome dans la formation des translocations, provient d'études de corrélations entre l'emplacement des chromosomes et leur fréquence de translocation. En effet, les cellules lymphoblastiques murines sont caractérisées par la présence de diverses

combinaisons de translocations entre les chromosomes 12, 14 et 15 (Liyanage et al., 2000). Les positions de ces chromosomes ont été cartographiées dans le noyau en interphase de cellules lymphoblastiques et cellules "normales" murines. Il a été observé que ces chromosomes formaient des clusters, et sont très proches dans le noyau. Cela suggérerait que la proximité physique pourrait probablement favoriser les réarrangements entre ces différents chromosomes (Parada et al., 2002). Des observations similaires ont été relevées dans d'autres types de translocations, notamment la translocation t(5 ;6) de l'hépatome de souris (Parada et al., 2004), ou encore BCR-ABL t(9 ;22) et PML-RAR t(15 ;17) chez l'homme (Neves et al., 1999). Le rôle de la proximité spatiale dans la détermination des partenaires de translocation est corrélé par les observations faites dans le lymphome de Burkitt. La proximité spatiale de *Myc* avec ses différents partenaires de translocation les plus fréquents (*IGH, IGK* et *IGL*) est corrélée à la fréquence de translocation observée chez les patients (Roix et al., 2003).

Toutes ces observations font état d'un modèle appelé "contact-first model" dans lequel les translocations se produisent préférentiellement entre les chromosomes qui sont spatialement proches dans le noyau. La proximité physique des loci lors de la survenue des CDB serait un facteur déterminant dans le choix de leur partenaire de translocation (Soutoglou and Misteli, 2008) (Figure 14). Cependant, un autre modèle nommé "breakage-first-model" s'oppose au premier. Ce second modèle, suggère que les extrémités cassées des chromosomes pourraient diffuser sur de longues distances dans le noyau à la recherche d'autres extrémités cassées, non réparées, qui serviraient de partenaire pour la translocation (Soutoglou and Misteli, 2008) (Figure 14). Ce modèle a notamment été appuyé par une étude réalisée chez *S.Cerevisiae*, où la réparation intrachromosomique est aussi efficace qu'une réparation interchromosomique. Ces résultats démontrent que les chromosomes de levures ayant subi une CDB, sont libres de chercher un partenaire approprié sur tout le génome (Haber and Leung, 1996).

Bien que ces études cytogénétiques fournissent des éléments sur la contribution de l'organisation spatiale du génome dans la formation des translocations chromosomiques, elles sont menées rétrospectivement et sont corrélatives, sans démonstration directe. Afin de s'affranchir de ces limites plusieurs études récentes ont été mises en place, en utilisant des approches genome-wide mapping, notamment par high-throughput genome-wide

translocation sequencing (HTGTS) ou encore par Translocation Capture Sequencing (TC-Seq). Les études cytogénétiques ont été confirmées par ces approches de genome-wide. En effet, des expériences où une CDB a été induite au niveau du locus *c-myc* ou *Igh* dans des lymphocytes B primaires murins ont montré après séquençage que les translocations se forment principalement entre la CDB et les régions proches sur le même chromosome (Chiarle et al., 2011; Klein et al., 2011). Il a été également montré dans les lymphocytes B activés, que le nombre de CDB induites par AID, ainsi que leur proximité influencerait la fréquence de translocation entre *Igh* et *c-myc* (Hakim et al., 2012; Rocha et al., 2012).

Enfin plus récemment, une étude est allée encore plus loin. Les CDB ont été marquées individuellement sur des chromosomes distincts, afin et de visualiser la formation des translocations. Il est ainsi possible de suivre le sort des ces CDB et de décrire les évènements qui conduisent à la formation des translocations dans chaque cellule vivante. Cette approche a permis de montrer que plus de 80% des chromosomes transloqués étaient très proche les uns des autres. Cela démontre que la formation des translocations chromosomiques se produit majoritairement entre des CDB proches dans l'espace (Roukos et al., 2013b).

Les observations cytogénétiques, et celles reportées par les analyses genome wide suggèrent que la formation des translocations pourrait être expliquée par le "contact-first model". Cependant, il n'est exclu que les CBD distales puissent également former des translocations, mais probablement avec une fréquence plus réduite (Roukos and Misteli, 2014).



Figure 14 : D'après (Misteli and Soutoglou, 2009). Les deux modèles proposés pour la formation des translocations chromosomiques. À gauche le « contact-first model » où les translocations se produisent entre des chromosomes qui sont spatialement proches les uns des autres. À droite le « breakage-first model », une fois les CDB générées sur des chromosomes éloignés, elles vont diffuser dans le noyau à la recherche d'un partenaire de translocation.

1.5.2 Causes des CDB qui sont à l'origine de l'apparition des translocations chromosomiques

A l'heure actuelle, il est difficile de définir avec certitude l'origine exacte des translocations chromosomiques. Comme mentionné dans la première partie de ma thèse, les CDB peuvent être d'origines multiples. Elles peuvent être le résultat de coupures endogènes d'une nucléase ayant un rôle physiologique telle que la topoisomérase 2, ou encore des enzymes nucléaires des cellules lymphoïdes telles que RAG1 et RAG2 de la recombinaison V(D)J et de AID du CSR, respectivement. Elles peuvent être également dues aux facteurs environnementaux tels que les radiations ionisantes (IR), la chimiothérapie anticancéreuse,

etc (voir partie 1.1) (Tsai and Lieber, 2010).

Les enzymes RAG1 et RAG2 de la recombinason V(D)J et de AID du CSR, responsables de la diversification du répertoire immunitaire, peuvent reconnaitre des sites hors de leur cible physiologique (*ie* en dehors du locus du gène du récepteur de l'antigène) (Mahowald et al., 2008). Par exemple RAG peut reconnaître des séquences qui ressemblent aux séquences signal RSS et en induisant des CDB, induire de l'instabilité génomique et conduire à la formation de translocations chromosomiques. Cet effet hors-cible serait à l'origine d'environ la moitié des translocations chromosomiques retrouvées dans les lymphomes (Marculescu et al., 2002; Numata et al., 2010; Robbiani et al., 2008).

La nucléase AID semble être impliquée dans la formation des translocations chromosomiques impliquant les gènes *TMPRSS2*, *ERG*, et *ETV1* dans le cancer de la prostate. Les récepteurs androgéniques (AR) jouent deux rôles majeurs lors de la formation de ces évènements oncogéniques. D'une part la fixation de ces AR au niveau de leur séquence cible « AR-binding DNA séquences », qui sont adjacentes aux séquences impliquées dans la translocation, entrainerait des mouvements chromosomiques qui créent une proximité spatiale des partenaires de la translocation. D'autre part, lors d'un stress génotoxique, ces AR induisent aussi le recrutement de l'enzyme AID au site du binding ce qui sensibiliserait ces séquences aux CDB et conduirait à la formation de translocations chromosomiques (Lin et al., 2009).

Les Îlots CpGs ont été décrits comme étant sujets aux CDB. L'analyse d'une base de données comptant plus de 1700 jonctions de translocations les plus fréquentes retrouvées dans les leucémies et les lymphomes humains, a montré que ces îlots, qui ne représentent que 1% du génome humain, représenteraient 40 à 70% des jonctions de translocations retrouvées dans les pro/pré lymphocytes B, particulièrement dans les régions proches des gènes *bcl-2* et *bcl-*1. Ces ilots CpGs pourraient être notamment une cible des endonucléases RAG et AID (Tsai et al., 2008).

Les sites fragiles communs (SFC) ont également été associés aux translocations chromosomiques. Une récente analyse à grande échelle de 746 lignées cellulaires cancéreuses, a révélé que les délétions homozygotes causant certains cancers co-localiseraient avec les sites fragiles (Bignell et al., 2010).

1.5.3 Mécanismes de réparation et de formation des translocations chromosomiques

Beaucoup d'études ont été menées afin de caractériser le mécanisme impliqué dans la réparation illégitime des CDB conduisant à l'apparition de translocations chromosomiques. Ces études ont été majoritairement réalisées dans les cellules murines, et ont démontré que même en présence du C-NHEJ, le alt-NHEJ est la voie de réparation majeure responsable de l'apparition des translocations chromosomiques dans ces cellules. En effet, il a été montré que les souris déficientes en Ku70/80, XRCC4 ou LigIV avec un fond génétique p53-/développent des tumeurs avec des translocations, dont les points de jonctions sont caractérisés par la présence de microhomologies (Difilippantonio et al., 2002; Zhu et al., 2002). Par ailleurs, les jonctions de réparation retrouvées dans les lymphocytes B murins matures déficients pour Ku70 ou pour Ku70 et LigIV (Boboila et al., 2010), dans les CHO déficients pour Ku80 (Guirouilh-Barbat et al., 2004, 2007), ou encore dans les ES murines KO pour Ku70 ou XRCC4 (Simsek and Jasin, 2010; Weinstock et al., 2007) présentent toutes des délétions et un biais vers l'utilisation des microhomologies, caractéristique d'un mécanisme alt-NHEJ. L'étude du mécanisme de formation des translocations dans les cellules ES murines, en utilisant des nucléases afin d'introduire des CDB, a démontré que les jonctions des translocations abritaient des délétions et un biais vers l'utilisation de microhomologies aussi bien dans les cellules sauvages que dans les cellules déficientes pour les facteurs du C-NHEJ. Il a été également montré que la fréquence de translocation augmentait en absence des facteurs du C-NHEJ, suggérant un rôle protecteur du C-NHEJ contre l'apparition de ces aberrations chromosomiques (Simsek and Jasin, 2010; Weinstock et al., 2007). Ces études dans les ES murines ont également permis de lever le voile sur le mécanisme du alt-NHEJ en confirmant le rôle de CtIP dans l'étape de résection (Zhang and Jasin, 2011), ou encore le rôle de la Ligase III et I dans la ligation des extrémités (Simsek et al., 2011a).

L'étude du mécanisme de formation des translocations dans les cellules humaines est limitée à deux études qui apportent des conclusions contradictoires. En effet, l'inhibition de PARP1 (olaparib, rucaparib, et siRNA) dans les HEK293 ou encore dans les Jurkat, entraine une baisse dans la formation des translocations. Les auteurs suggèrent l'implication de PARP1 et donc un rôle du alt-NHEJ dans la formation des translocations dans les cellules humaines (Wray et al., 2013). En revanche, l'inhibition de PARP1 dans les cellules HCT116 wt

réduit modestement la fréquence de translocation, alors que celle-ci est diminuée de moitié dans cellules HCT LigIV KO (Soni et al., 2014). Cette dernière étude, ne suggère qu'une implication minime de PARP1 dans la formation des translocations lorsque le C-NHEJ est actif.

1.6 Induction des translocations chromosomiques avec des nucléases artificielles

Dans cette partie je ne vais pas faire une revue exhaustive sur les nucléases artificielles, mais je vais plutôt décrire les propriétés qui en font des ciseaux moléculaires capables d'induire des CDB en un site choisi, et que j'ai exploitées pendant ma thèse. Pour étudier les translocations chromosomiques dans les cellules humaines, nous avons utilisé cette technologie des nucléases artificielles. Ainsi, j'ai eu la chance d'avoir entre les mains toutes les nucléases phares du moment, notamment les Zinc Finger (ZFN), TALEN et plus récemment les CRISPR/Cas9. En induisant deux CDB simultanées sur deux chromosomes hétérologues choisis, une réparation inappropriée de ces deux CDB peut conduire à la formation de deux chromosomes dérivatifs transloqués (Figure 15).



<u>Figure 15</u>: Induction des translocations chromosomiques par des nucléases artificielles. Deux nucléases artificielles sont transfectées dans des cellules humaines et coupent spécifiquement et simultanément deux chromosomes A et B, induisant potentiellement la translocation d'intérêt t(A ;B).

1.6.1 Les nucléases à doigts de Zinc (Zinc Finger) (Figure 16)

Les doigts de zinc Cys2-His2, découverts pour la première fois chez le Xénope (Miller et al., 1985) sont des domaines protéigues de liaison à l'ADN retrouvés abondamment chez les eucaryotes. D'ailleurs environ 2% des gènes humains codent pour un domaine à doigt de Zinc (Lander et al., 2001; Tupler et al., 2001; Venter et al., 2001). Chaque doigt de zinc consiste en environ 30 acides aminés qui se replient autour d'un ion de zinc pour former une structure stable et compacte capable de reconnaître environ 3 pb d'ADN via leurs hélice α (Pavletich and Pabo, 1991). Plusieurs doigts de zinc peuvent être liés en tandem afin de reconnaître spécifiquement une large gamme de séquences d'ADN (Choo et al., 1994). Pour avoir une nucléase à doigt de zinc, il faut un domaine de reconnaissance à l'ADN et un domaine de clivage. Les premiers travaux concernant l'ingénierie des nucléases à doigt de zinc (ZFN) ont été réalisés par le Dr Chandrasegaran et ses collaborateurs en mettant en évidence les domaines fonctionnels de la nucléase FokI. Cette enzyme, présente chez la Flavobacterium okeanokoites est une endocnuléase de type IIS composée d'un domaine de reconnaissance à l'ADN et d'un domaine de clivage. Contrairement aux autres enzymes de restriction de type II dont le site de clivage est inclus dans le site de reconnaissance, les endocunélases de type IIS ont des domaines de liaison et de clivage physiquement séparées, leur permettant ainsi de couper l'ADN double-brin à des distances précises de leur le site de reconnaissance (Li et al., 1992). Le domaine catalytique de Fokl n'ayant pas de spécificité de séquence apparente, il pouvait être utilisé avec des domaines de reconnaissance de l'ADN autres que ceux présents naturellement (Kim and Chandrasegaran, 1994; Kim et al., 1996). Un domaine protéique à doigts de zinc (ZFP) peut être ainsi couplé au domaine catalytique de l'enzyme de restriction FokI, créant ainsi une nucléase spécifique d'une séquence d'ADN (Kim et al., 1996; Smith et al., 1999). La dimérisation de Fokl est par ailleurs requise pour le clivage de l'ADN cible (Bitinaite et al., 1998). Ainsi pour former une nucléase à doigt de zinc, il faut deux monomères (2 ZFP+ 1 Fokl) qui se dimérisent par Fokl à une distance de 5-7 pb. Un ZFP peut comporter 3 à 6 motifs à doigts de zinc de part et d'autre du site de clivage et peut donc lier jusqu'à 36 paires de bases permettant ainsi de cibler une séquence unique dans le génome (Pauwels et al., 2014).



<u>Figure 16</u>: D'après (Kim and Kim, 2014). Représentation schématique de l'organisation d'une nucléase à doigt de zinc (ZFN). ZFP-R représente le ZFP de droite et ZFP-L, celui de gauche. Chaque ZFN peut être composée de 3 à 6 motifs à doigts de zinc, qui reconnaissent et lient la séquence cible de 9 à 18 pb respectivement. Ici 4 motifs sont représentés. L'espace (spacer) entre les domaines de liaison varie entre 5-7 pb. Des variantes du domaine *FokI* sont utilisées pour obliger la formation d'un hétérodimère.

1.6.2 Les Transcription Activator-Like Effector (TALEs) (Figure 17)

Tout comme les ZFN, les TALEN possèdent le même domaine de clivage de l'ADN qui est l'endonucléase *Fokl*, par contre leur domaine de reconnaissance et de liaison à l'ADN est différent. Ce domaine de liaison (Transcription activator-like effector (TALEs)) est dérivé de facteurs de transcription des phytopathogènes de la famille de *Xanthomonas*, qui une fois dans le noyau activent la transcription de gènes cibles chez l'hôte (Boch et al., 2009). Les TALEs sont composées d'une vingtaine de répétitions en tandem de 33-35 acides aminés, et chaque répétition reconnaît une seule paire de base, la dernière répétition quant à elle n'est composée que de 20 résidus (Deng et al., 2012; Mak et al., 2012). Au niveau de ces répétitions, la spécificité de liaison aux nucléotides est conférée par les acides aminés en position 12 et 13, ces deux résidus sont appelés repeat-variable-diresidus (RVD) (Boch et al., 2009; Moscou and Bogdanove, 2009). Chaque monomère de la paire de TALEN reconnait une séquence de 14-20 pb, ces deux séquences sont séparées par un spacer de 12-21 pb qui permet la dimérisation de *Fokl*. C'est au sein de ce spacer que la cassure CDB se produit (Miller et al., 2011; Mussolino et al., 2011). Les TALENs peuvent être conçues pour cibler pratiguement n'importe quelle séquence d'ADN donnée, car la seule contrainte de la

séquence ciblée par ces nucléases est de posséder un T en 5' : 5'-T-N(45-55) A-3' (Mak et al., 2012). Plusieurs méthodes d'assemblage ont été développées afin de générer efficacement des TALENs, et rendre leur production accessible aux laboratoires de recherche à des coûts relativement faibles. L'une des méthodes utilisée est basée sur le clonage modulaire "*Golden gate*" qui utilise des enzymes de restriction de type IIS, permettant d'assembler jusqu'à 10 répétitions en une seule réaction de ligation (Cermak et al., 2011; Li et al., 2011b). on peut également citer celle de unit assembly (utilisée pour mes travaux) (Huang et al., 2011).



<u>Figure 17</u>: D'après (Kim and Kim, 2014). Représentation schématique de l'organisation d'une TALEN. Chaque TALE est composée d'une répétition d'un motif de 33-35 acides aminés, au sein de ces répétitions les acides aminés 12 et 13 (RVD) reconnaissent un nucléotide spécifique. L'espace entre deux TALEs varie entre 12-21 pb.

1.6.3 Les CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)

(Figure 18, 19)

S'ajoutant aux ZFNs et aux TALENs, un nouveau système appelé CRISPR/Cas9 a été introduit comme outil pour l'ingénierie des génomes. Le locus CRISPR, pour Clustered Regularly Interspaced Short Palindromic Repeats, est impliqué dans un système immunitaire procaryote utilisé par les bactéries et les archées et qui leur confrère une résistance contre les ADN étrangers (Barrangou et al., 2007; Horvath and Barrangou, 2010). Ce système permet d'intégrer de courts fragments d'ADN étrangers (virus, plasmides, phages, etc...) dans le locus CRISPR afin d'armer ces organismes contre de nouvelles attaques de ces mêmes agents pathogènes donnant lieu à une forme d'immunité adaptative (Barrangou et al., 2007; Horvath and Barrangou, 2010). Le locus CRISPR contient une série de répétitions d'environ 20-50 pb, et chaque répétition est suivie d'un spacer d'ADN de la même taille (Barrangou et al., 2007; Horvath and Barrangou, 2010). Les spacers proviennent de segments d'ADN exogènes (protospacers) et chacun est associé à un PAM (Protospacer Adjacent Motif) dont la reconnaissance est spécifique du système CRISPR, ainsi pour la Cas9 de *Streptococcus poygenes* le PAM est 5'-**NGG**.

Ce locus est transcrit en précurseur CRISPR RNA (pre-crRNA), qui sera directement clivé par l'ARNaseIII conduisant alors à la formation de petits crRNAs qui iront s'hybrider sur la séquence complémentaire de l'ADN étranger en formant un hétéroduplexe ADN-ARN de 20 pb (Deltcheva et al., 2011; Haurwitz et al., 2010; Jinek et al., 2012). Chaque crRNA constitue un complexe de surveillance afin de protéger l'environnement intracellulaire de toute invasion de virus ou plasmides. Ces crRNAs vont former un complexe avec la nucléase Cas9 (CRISPR-associated), et les tracrARN. Ce complexe ira s'hybrider aux répétitions et dégrader l'ADN étranger (Barrangou et al., 2007; Horvath and Barrangou, 2010; Jinek et al., 2012) (Figure 18).

Le système gRNA/Cas a été inspiré de ce système procaryote, et s'ajoute ainsi à la liste des outils disponibles pour l'édition du génome. Ce système est formé d'un ARN guide (gRNA), dont 20 nucléotides sont complémentaires à la séquence cible, et d'une nucléase Cas9. Cette Cas9 clive l'ADN cible sur les deux brins en aval de la séquence PAM (5'-NGG) (Jiang et al., 2013) (Figure 19A). Néanmoins, il semblerait que certains mésappariements sont tolérés dans la reconnaissance ADN/gRNA, ce qui augmenterait le nombre de sites potentiellement hors-cible et donc diminuerait la spécificité (Fu et al., 2013). Pour contrecarrer ces effets hors-cible, la Cas9 peut être mutée sur l'un de ses domaines nucléasiques. En effet, une substitution de l'acide aminé Asp en Ala position 10 (D10A) entraîne l'inactivation d'un des sites catalytiques de la nucléase convertissant la Cas9 en Cas9 nickase, celle-ci induit une cassure simple brin (nick) sur le brin qui lie le gRNA (Jinek et al., 2012). Ces nCas9 peuvent également induire des CDB en utilisant deux gRNAs qui couperont sur des brins opposés (Mali et al., 2013; Ran et al., 2013) (Figure 19B).



Figure 18: D'après (Belhaj et al., 2015). Représentation schématique du système CRISPR de type II. La séquence du CRISPR est composée de petites répétitions palindromiques (losanges) séparées par des « protospacers » (carrés en couleur). La séquence du CRISPR est transcrite en pre-crARN. Ce dernier sera clivé par la RNaseIII. Un complexe crARN, tracrARN et Cas9 est alors formé et ira s'hybrider sur la séquence cible de l'ADN étranger. Cet ADN double-brin est alors clivé par la Cas9.



<u>Figure 19</u>: Système CRISPR/Cas9 (nCas9). (A) : La protéine Cas9 (vert) se lie au gRNA qui reconnaît une séquence de 20 pb, formant un complexe ribonucléoprotéique. Ce complexe va couper sur les deux brins en aval d'une séquence PAM (NGG), indispensable à ce clivage. (B) : Dans le cas de la nCas9, la Cas9 est mutée sur l'un de ses deux sites endonucléasiques. Celle-ci se lie aux deux gRNAs, ces complexes coupent en simple-brin sur les deux brins opposés et créent alors une CDB avec un long 5' overhangs (quelques dizaines de nucléotides).

1.6.4 Applications des nucléases artificielles

La commercialisation des nucléases, ainsi que la facilité de leur conception ont rendu ces outils accessibles à un grand nombre de laboratoires, ouvrant ainsi de larges champs d'applications mais aussi de nombreuses perspectives thérapeutiques. En effet, les nucléases induisent des CDB à un site spécifique du génome. La réparation de cette CDB par le NHEJ peut conduire à des insertions ou délétions d'un nombre variables de nucléotides. Ces mutations (indels) peuvent créer un décalage du cadre de lecture lorsqu'elles sont situées dans les régions codantes du génome. Ces nucléases ont donc été largement utilisées pour le « gene editing » notamment l'inactivation de gène (KO) en induisant des petites mutations. En présence d'un ADN donneur, il est aussi possible d'induire des modifications contrôlées du génome par HDR (Homology Directed Repair) en introduisant de nouvelles séquences dans le gène cible Knock-In (KI). Cela permet d'étudier la fonction de ces gènes modifiés, ou encore de générer des animaux modèles pour l'étude de maladies humaines (pour revue : Gaj et al., 2013; Perez-Pinera et al., 2012) (Figure 20). Elles ont également été utilisées afin d'induire des réarrangements plus complexes en créant deux CDB ciblées dans le génome aussi bien sur le même chromosome, induisant des délétions et ou inversions (Lee et al., 2010, 2012), que sur des chromosomes hétérologues conduisant à des translocations chromosomiques (Brunet et al., 2009; Simsek et al., 2011b). Ces nucléases, ont également été utilisées à visée thérapeutique, afin de corriger directement les mutations pathogènes associés par exemple au déficit immunitaire sévère (SCID) (Urnov et al., 2005), à l'hémophilie (Li et al., 2011a), ou encore à la drépanocytose (Sebastiano et al., 2011). En combinaison avec la technologie des cellules souches pluripotentes induites (iPS), les nucléases semblent prometteuses dans la médecine régénératrice, c'est le cas notamment pour la maladie de parkinson (Soldner et al., 2011). Certaines de ces études à visée thérapeutiques se sont avérées particulièrement puissante et sont en essai clinique. C'est le cas notamment des ZFN qui ont été utilisées afin de bloquer l'entrée du VIH dans les cellules T et des cellules souches hématopoïétiques en inactivant le corécepteur des chimiokines CCR5 (Biosciences, 2014; Holt et al., 2010; Perez et al., 2008).



Figure 20 : Les différentes applications du genome editing en utilisant les nucléases artificielles.

1.6.5 Cytotoxicité et effets hors-cible des nucléases artificielles

La notion « séquence spécifique » attribuée aux nucléases n'est pas absolue et des effets non spécifiques peuvent êtres observés. En effet, cet effet hors-cible peut être dû à une faible spécificité de liaison à l'ADN cible, ce qui entrave la capacité de cette nucléase à distinguer entre la séquence cible et d'autres séquences similaires dans le génome. Comparer l'efficacité de ces nucléases entre elles est difficile, car elles doivent cibler exactement le même site ce qui est difficile à cause de leur code de reconnaissance. D'ailleurs, peu d'études dans ce sens sont répertoriées, bien qu'elles pourraient apporter des informations cruciales sur le choix du type de nucléases à utiliser (forte efficacité et faible toxicité). Parmi ces études comparatives, Mussolino et ses collaborateurs ont fait une comparaison entre des ZFN et des TALEN qui ciblent trois différents loci dans les cellules humaines. Cette étude a montré que ces deux nucléases avaient la même efficacité à modifier le génome, toute fois, dans ce cas, les ZFN sembleraient être plus toxiques que les TALEN (Mussolino et al., 2011). Une autre étude a montré que les TALEN généraient plus de mutations que les ZFN chez le zebrafish (Chen et al., 2013), néanmoins il n y a pas de règles générales aujourd'hui. Le principal challenge est d'avoir des outils très actifs et surtout très spécifiques. De ce fait, les nucléases sont continuellement optimisés. Par exemple, dans le cas de *Fokl* sauvage, chaque monomère peut se dimériser avec un autre ou s'auto-dimériser ce qui peut conduire à des effets hors-cible avec deux séquences identiques (Figure 21A). De ce fait, le domaine *Fokl* des ZFN et TALEN a été muté de manière à ce qu'il ne puisse plus s'apparier en homodimère générant ainsi des hétérodimère "contraints" qui réduit les effets hors-cible et la cytoxicité de ces nucléases (Miller et al., 2007; Szczepek et al., 2007) (Figure 21B). Pour augmenter encore plus leur spécificité, Doyon et ses collègues ont conçu des domaines *Fokl* dit hétérodimères orthogonaux, où chaque monomère ne peut s'apparier qu'avec le monomère correspondant à la nucléase désirée. Ce qui est très utile quand plusieurs nucléases sont utilisées simultanément pour créer des réarrangements, comme dans le cas de mon travail (Doyon et al., 2011) (Figure 21C).



<u>Figure 21</u>: Limitation des sites hors-cible des ZFN ou TALEN avec le développement de variants du domaine catalytique *FokI*. (a) Lors de l'expression simultanée de deux paires de nucléases dont le domaine *FokI* s'apparie en homodimères (*FokI*-WT) des dimères autres que (Seq1 – Seq2) et (Seq3–Seq4) sont formés. (b) L'utilisation d'un domaine *FokI* formant un hétérodimère (*FokI* hétérodimères) enlève la possibilité de faire des homodimères et limite le type de dimères formés. (c) Dans les nucléases dites orhogonales, les domaines *FokI* des nucléases (*FokI*-hétérodimères) s'apparient en deux hétérodimères différents, et seules les deux nucléases choisies sont formées.

Le système gRNA/Cas9 peut tolérer jusqu'à 5 mismatchs lors de l'appariement ADN/gRNA, en particulier s'ils sont positionnées à moins de 10pb du PAM (Fu et al., 2013). L'utilisation de la Cas9 nickase (nCas9) peut diminuer les effets hors-cible (Mali et al., 2013; Ran et al., 2013). D'autres ont développé des gRNA tronqués (17-18 pb) (Tru-gARN). L'utilisation de ces Tru-gARN semble diminuer les effets hors-cible sans compromettre l'efficacité de ces nucléases (Fu et al., 2014). Enfin, plus récemment, les RNA-guided *Fokl*-Cas9 nucleases (RFNs) ont été développées. Dans ce système la Cas9 a été fusionnée au domaine *Fokl*, et deux gRNA sont nécessaire pour d'induire le clivage augmentant ainsi la longueur de la séquence de reconnaissance. La contrainte majeure de ce système est la nécessité d'avoir deux PAM distants d'environ 55 pb lors de la conception des gRNA (distance entre les 2 sites de fixation des gRNA de 14-17 pb) (Tsai et al., 2014).

2 Projet de thèse et résultats

Mon projet de thèse vise principalement à étudier le mécanisme moléculaire de formation des translocations chromosomiques, leurs conséquences sur la stabilité génomique et s'insert dans un projet plus global qui vise à comprendre leur rôle dans l'oncogenèse. Les translocations sont induites en utilisant des nucléases artificielles de type ZFN, TALEN, et CRISPR/Cas9 pour créer 2 CDB ciblées. Cette approche permet alors l'échange de morceaux de chromosomes transloqués et pas seulement la surexpression du gène de fusion intégré dans le génome.

Les objectifs de ma thèse sont:

(i) Le mécanisme moléculaire induisant l'apparition des translocations a été identifié dans les cellules murines, principalement par l'équipe du Pr Maria Jasin. Ces résultats montrent clairement que la voie impliquée ici est la voie alt-NHEJ. Nous nous sommes alors intéressés plus précisément à ce qui se passe dans les cellules humaines. En effet, celles-ci sont différentes des cellules murines notamment en ce qui concerne le niveau d'expression de certaines protéines du C-NHEJ (*ie* Ku et la DNA-PKcs). Les cellules murines sont également très efficaces pour le ciblage de gène (mécanisme impliquant la RH) ce qui n'est pas le cas des cellules humaines pour lesquelles il est très difficile de faire du ciblage de gène. Comprendre ces mécanismes en identifiant les protéines de réparation impliquées dans l'apparition des translocations dans des cellules humaines permettra d'expliquer pourquoi les modèles tumoraux murins ne correspondent pas toujours à l'apparition de tumeurs chez l'homme. Ainsi, durant ma thèse, nous nous somme principalement intéressés au rôle des protéines du complexe de ligation du C-NHEJ XRCC4/ LigaseIV dans la formation des réarrangements chromosomiques.

(ii) En collaboration avec une autre étudiante en thèse Marion Piganeau, nous nous sommes attachés à développer de nouveaux modèles cellulaires humains porteurs de translocations chromosomiques. Ce projet s'est appuyé sur l'utilisation de ZFN et de TALEN qui ciblent les points de cassure retrouvés chez les patients pour induire les translocations impliquées respectivement dans le sarcome d'Ewing et l'ALCL (leucémie anaplasique à grandes cellules) dans des cellules humaines. Le but est de recréer la translocation au niveau

génomique au lieu de surexprimer le gène de fusion, permettant ainsi d'avoir une approche pas à pas, pour étudier les conséquences cellulaires induites par la formation de ces réarrangements chromosomiques, puis le potentiel tumorigénique des cellules portant la translocation (travail en cours dans l'équipe).

2.1 Article I : Étude des translocations chromosomiques dans les cellules humaines

Dans cette partie, nous nous sommes intéressés à la caractérisation du mécanisme moléculaire impliqué dans la formation des translocations dans les cellules humaines. La réparation des CDB se fait soit par recombinaison homologue (RH) qui repose sur l'utilisation de séquences homologues (en particulier la chromatide sœur) soit le NHEJ (Non Homologous End Joining) qui consiste à réparer la cassure par « simple » ligation. Or les séquences des jonctions des translocations trouvées chez les patients révèlent que moins de 1% de celles-ci présentent de longues homologies de séquences entre les chromosomes transloqués, ce qui favorise l'hypothèse de l'utilisation de la voie NHEJ. Mon travail de thèse consistait à caractériser précisément le rôle du NHEJ dans la formation des translocations chromosomiques. Nous avons utilisé l'approche des nucléases artificielles (ZFN, TALEN, et CRISPR/Cas9) afin d'induire la formation de translocations dans des cellules humaines déficientes pour les protéines du C-NHEJ. Mes résultats montrent que le mécanisme moléculaire de formation de ces événements oncogéniques est distinct de celui reporté pour les cellules murines et utilise une voie de réparation différente. En effet, dans les cellules murines le C-NHEJ protège contre l'apparition de ces réarrangements, alors que celui-ci joue rôle prédominant dans leur apparition dans les cellules humaines. Ceci implique que les cellules humaines utiliseraient des mécanismes spécifiques pour maintenir l'intégrité de leur génome.

Chromosomal Translocations in Human Cells Are Generated by Canonical Nonhomologous End-Joining

Hind Ghezraoui,^{1,2,3} Marion Piganeau,^{1,2,3} Benjamin Renouf,^{1,2,3} Jean-Baptiste Renaud,^{1,2,3} Annahita Sallmyr,⁴ Brian Ruis,⁵ Sehyun Oh,⁵ Alan E. Tomkinson,⁴ Eric A. Hendrickson,⁵ Carine Giovannangeli,^{1,2,3} Maria Jasin,^{6,*} and Erika Brunet^{1,2,3,*}

¹Museum National d'Histoire Naturelle, 43 rue Cuvier, F-75005 Paris, France

²CNRS, UMR7196, 43 rue Cuvier, F-75005 Paris, France

³Inserm, U1154, 43 rue Cuvier, F-75005 Paris, France

⁴Department of Internal Medicine and University of New Mexico Cancer Center, University of New Mexico, Albuquerque, NM 87131, USA ⁵Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota Medical School, Minneapolis, MN 55455, USA ⁶Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

*Correspondence: m-jasin@ski.mskcc.org (M.J.), ebrunet@mnhn.fr (E.B.)

http://dx.doi.org/10.1016/j.molcel.2014.08.002

SUMMARY

Breakpoint junctions of the chromosomal translocations that occur in human cancers display hallmarks of nonhomologous end-joining (NHEJ). In mouse cells, translocations are suppressed by canonical NHEJ (c-NHEJ) components, which include DNA ligase IV (LIG4), and instead arise from alternative NHEJ (alt-NHEJ). Here we used designer nucleases (ZFNs, TALENs, and CRISPR/Cas9) to introduce DSBs on two chromosomes to study translocation joining mechanisms in human cells. Remarkably, translocations were altered in cells deficient for LIG4 or its interacting protein XRCC4. Translocation junctions had significantly longer deletions and more microhomology, indicative of alt-NHEJ. Thus, unlike mouse cells, translocations in human cells are generated by c-NHEJ. Human cancer translocations induced by paired Cas9 nicks also showed a dependence on c-NHEJ, despite having distinct joining characteristics. These results demonstrate an unexpected and striking species-specific difference for common genomic rearrangements associated with tumorigenesis.

INTRODUCTION

Recurrent reciprocal chromosomal translocations are associated with oncogenesis (Mani and Chinnaiyan, 2010; Mitelman et al., 2007). Translocations frequently generate fusion genes with novel properties that drive oncogenesis; to date, more than 300 genes have been associated with oncogenic translocations in both hematological malignancies and solid tumors. In addition to generating fusion genes, translocations can also enhance the expression of proto-oncogenes, the classic example of which results in c-Myc overexpression. Breakpoint junction analysis has demonstrated that oncogenic translocations typically arise by some form of nonhomologous end-joining (NHEJ).

The canonical pathway of NHEJ (c-NHEJ) is required for cellular resistance to ionizing radiation as well as for immune system rearrangements and is active throughout the cell cycle (Deriano and Roth, 2013; Goodarzi and Jeggo, 2013; Pannunzio et al., 2014). Critical components of c-NHEJ include the Ku70/ 80 heterodimer, DNA-PKcs, DNA ligase IV (LIG4), and XRCC4. Loss of c-NHEJ components does not, however, completely abrogate NHEJ (Delacôte et al., 2002; Kabotyanski et al., 1998; Liang and Jasin, 1996), suggesting that there are alternative ways to join ends, referred to as alt-NHEJ. Whether alt-NHEJ is a distinct, regulated pathway(s) or involves the co-opting of non-c-NHEJ proteins with some c-NHEJ components is a subject of debate (Deriano and Roth, 2013; Goodarzi and Jeggo, 2013; Pannunzio et al., 2014). Junctions that form by alt-NHEJ have more microhomology and longer deletions than junctions formed by c-NHEJ (Fattah et al., 2010; Guirouilh-Barbat et al., 2007; Kabotyanski et al., 1998; Oh et al., 2013; Simsek and Jasin, 2010; Smith et al., 2003). Proteins that promote alt-NHEJ include the end resection factor CtIP (Bennardo et al., 2008) and LIG3a (Wang et al., 2005).

Most studies analyzing translocation formation have been performed in mouse cells, in particular in lymphoid cells involving programmed double-stranded breaks (DSBs) and in embryonic stem cells using I-Scel or zinc finger nuclease (ZFN)-generated DSBs (Boboila et al., 2012; Nussenzweig and Nussenzweig, 2010; Weinstock et al., 2007; Simsek et al., 2011). These studies uniformly demonstrated that c-NHEJ suppresses translocation formation at nonhomologous sequences. Thus, in the absence of Ku, LIG4, or XRCC4, translocations are increased in frequency. Since alt-NHEJ proteins CtIP and LIG3 promote translocation formation (Zhang and Jasin, 2011; Simsek et al., 2011) and translocation junction sequences in wild-type and c-NHEJ mutants have similar characteristics, it appears that translocations in mouse cells typically arise by alt-NHEJ.

In contrast to mouse cells, translocation junctions in human tumors do not always show significant lengths of microhomology



Translocations in Human Cells Are c-NHEJ Dependent



Figure 1. Intrachromosomal DSB Repair in c-NHEJ-Deficient Human Cells Is Inefficient and Shows a Shift toward Longer Deletions and Microhomology

(A) Absence of LIG4 in both $L4^{-/-}$ and $X4^{-/-}$ HCT116 cells.

(B) The ZFN^{EWS} cleavage site overlaps an Asel restriction site.

(C) After ZFN^{EWS} expression, Asel-resistant PCR products (*: 724bp) were purified and reamplified (examples shown for *wt*, L4^{+/-}, and L4^{-/-} cells) for sequencing.

(D) Asel-resistant NHEJ junctions at the ZFN^{EWS} site from $L4^{-/-}$ and $X4^{-/-}$ cells demonstrate a shift toward longer deletions and increased microhomology (wt, n = 4; L4, n = 2; X4, n = 3). Junctions from L4 and X4 heterozygous or deficient cell lines are combined here and below, although results from individual cell lines were similar. The median deletion length is indicated, and each value represents the combined deletion from both ends of an individual junction. In the
(Gillert et al., 1999; Langer et al., 2003; Zucman-Rossi et al., 1998; Mattarucchi et al., 2008). Similarly, cancer and model translocations induced by nucleases in several human cell lines also show little or no microhomology at translocation junctions (Brunet et al., 2009; Piganeau et al., 2013). Studies in human cells deficient in c-NHEJ are limited. Ionizing radiation, a potent inducer of translocations in rodent cells, does not significantly induce translocations in a LIG4 mutant human cell line (Soni et al., 2014). In contrast, knockdown of c-NHEJ components did decrease androgen-induced translocations, although junction analysis was not reported (Lin et al., 2009).

To address the role of NHEJ pathways in the joining phase of chromosomal translocation formation, we took advantage of nucleases designed to introduce site-specific DSBs at endogenous loci in human cells (Gaj et al., 2013; Urnov et al., 2010) to induce translocations (Brunet et al., 2009; Piganeau et al., 2013). Using multiple cell lines and different nucleases to provoke DSBs, we found that the translocation frequency was often reduced in human cells in the absence of LIG4 or XRCC4, in stark contrast to results from c-NHEJ-deficient mouse cells. The translocations that were formed in human c-NHEJ mutants had frequent microhomologies and long deletions. Consistent with a requirement for c-NHEJ, loss of alt-NHEJ components did not affect translocation formation unless c-NHEJ was also impaired. We also found that different types of end structures gave rise to different joining characteristics in wild-type cells. Translocations induced by wild-type Cas9 frequently had precisely joined ends, indicating that c-NHEJ can be highly accurate, whereas those induced by Cas9 nickase (nCas9) had more varied junctions; in either case, the absence of LIG4 led to greater inaccuracy in joining. Thus, our studies reveal an unexpected and striking species-specific difference in the generation of these oncogenic rearrangements.

RESULTS

Intrachromosomal DSB Repair Is Altered in c-NHEJ-Deficient Human Cells

To analyze the repair of chromosomal DSBs in c-NHEJ-deficient human cells, we used LIG4 and XRCC4 mutant HCT116 cell lines (Oh et al., 2013; B.R. and E.A.H., unpublished data). LIG4-null $(L4^{-/-})$ cells are deficient in LIG4 but maintain XRCC4 expression; XRCC4 mutant ($X4^{-/-}$) cells are deficient in both XRCC4 and LIG4, because LIG4 is unstable in the absence of XRCC4 (Bryans et al., 1999) (Figure 1A). We expressed ZFN^{p84}, which cleaves the *p84/AAVS1* locus (Brunet et al., 2009), and estimated the insertion/deletion (indel) frequency resulting from NHEJ at this site. In both $X4^{-/-}$ and $L4^{-/-}$ cells, the indel frequency was approximately half that of control cells (Figure S1A available online), suggesting reduced NHEJ. A similar reduction in indel frequency was obtained at a second locus with other nucleases (TALENs, Cas9, and nCas9) (see below, Figure 4C).

Junction characteristics were examined in the c-NHEJ mutants using ZFN^{EWS}, which cleaves the *EWS* locus at an Asel restriction site (Piganeau et al., 2013) (Figures 1B and 1C). Asel-resistant junctions from $X4^{-/-}$ and $L4^{-/-}$ cells had longer deletions compared to control cells (p < 0.0001, Mann-Whitney test; Figures 1D and S1B). Microhomology at the junctions was also altered in the $X4^{-/-}$ and $L4^{-/-}$ cells such that 49% of junctions had \geq 3 bp microhomology—in contrast to only ~20% of junctions in control cells (p = 0.003; Figure 1D). For both mutant and control cells, the microhomology distribution was different from that expected by the random occurrence of microhomology; however, the deviation from random was especially apparent for the mutant cells (p < 0.0001), providing strong evidence that microhomology is critical in the joining process in the absence of LIG4.

To determine if repair of two DSBs by deletion formation has the same dependency on c-NHEJ as the repair of a single DSB, we induced DSBs 3.2 kb apart within the FLI1 gene with ZFNs (Figure 1E). A fragment corresponding to the 3.2-kb deletion product was PCR amplified using primers that flanked the two DSBs. By limiting dilution, the deletion product was estimated to be ~4- to 8-fold less abundant in the mutant cells (Figure 1F). The PCR product showed minimal deletion from the two DNA ends in control cells, whereas it was frequently shorter in $X4^{-/-}$ and $L4^{-/-}$ cells (Figure 1E). Sequencing confirmed that many of the junctions from control cells occurred with minimal processing of the DNA ends, whereas long deletions were common in the mutant cells (median = 96 bp; Figures 1G and S1C). Further, \geq 3 bp microhomology was infrequent at junctions from control cells but common at junctions from $X4^{-/-}$ and $L4^{-/-}$ cells (\leq 20% and \sim 70% of junctions, respectively). Thus, as with repair of a single DSB, intrachromosomal repair of two DSBs in human cells requires the LIG4:XRCC4 complex, similar to rodent cells (Guirouilh-Barbat et al., 2007; Simsek and Jasin, 2010).

Chromosomal Translocations Rely on c-NHEJ in Human Cells

To determine the mechanism by which DSBs give rise to chromosomal translocations in human cells, we coexpressed

See also Figure S1.

microhomology analysis, junctions that contain insertions were not included. Deletion and microhomology distributions here and below were compared by Mann-Whitney analysis. For random microhomology distribution, the probability that a junction will have microhomology by chance assumes an unbiased base composition (Roth et al., 1985). ***p < 0.0001.

⁽E) Two DSBs 3.2 kb apart were introduced by ZFN^{FLI-A} and ZFN^{FLI-B} in a *FLI1* gene intron. To detect the 3.2-kb deletion, an ~890-bp fragment was PCR amplified using primers flanking the two DSBs.

⁽F) Limiting dilution of genomic DNA to estimate the frequency of deletions after expression of ZFN^{FLI-A} and ZFN^{FLI-B}. PCR amplification of the 3.2-kb deletion product was performed with serial dilutions of genomic DNA (100, 50, 25, 12.5, 6.25, 3.125, and 1.56 ng). The number of times a PCR fragment was detected from three independent amplifications is indicated below the gel.

⁽G) Junctions from joining two DSBs 3.2 kb apart at the *FLI1* locus from $L4^{-/-}$ and $X4^{-/-}$ cells demonstrate a shift toward longer deletions and increased microhomology.



Figure 2. c-NHEJ Generates Chromosomal Translocations in HCT116 Human Cells

(A) Induction of chromosomal translocations with sequence-specific nucleases. Derivative chromosomes Der19 and Der22 were detected by PCR using primers that flank the cleavage sites after ZFN^{p84} and ZFN^{EWS} expression.

ZFN^{p84} and ZFN^{EWS} to induce concomitant DSBs on Chr19 and Chr22, respectively (Figure 2A). The translocation frequency was determined by PCR screening of small pools of cells (Brunet et al., 2009; Piganeau et al., 2013), amplifying a >900 bp fragment for each derivative chromosome. Translocation frequencies were 5- to 6-fold lower for $L4^{-/-}$ and $X4^{-/-}$ cells compared to control cells (Figure 2B), indicating a reliance on c-NHEJ for efficient joining. Since ZFN protein levels were similar in all cell lines (Figure 2B), the lower translocation frequency was not due to reduced ZFN expression. To confirm that the reduced frequency was not specific to a particular ZFN pair, ZFN^{EWS} was paired with a TALEN, TAL^{p84}, which cleaves close to the ZFN^{p84} site; TAL^{p84} was also paired with another TALEN, TAL^{LAM}, which cleaves a locus on Chr1. Like ZFNs, TALENs generate 5' overhangs, although the overhangs are more variable (Piganeau et al., 2013). Importantly, $L4^{-/-}$ cells also showed lower translocation frequencies with these nuclease pairs (Figure 2B).

We analyzed ~50 Der19 and Der22 junctions (Figure S2). In control cells, most deletions were short (≤ 2 bp median) such that more than half of junctions maintained all or part of the ZFN 5' overhang (Figures 2C and 2D), presumably by fill-in synthesis. The median deletion length in $X4^{-/-}$ cells was much longer (78 bp, p < 0.001), and none of the deletions were restricted to the ZFN overhang. Instead, most junctions had deletions >30 bp (85%), with many >200 bp. Microhomology was also different in the c-NHEJ mutants. Breakpoint junctions recovered from control cells had microhomology distributions only marginally different from that expected from the chance joining of two random sequences (Figure 2E). In contrast, the microhomology distribution was significantly different from random joining in $X4^{-/-}$ cells (p < 0.0001), with \geq 3 bp microhomology observed in 48% of junctions. Insertions were observed in a fraction of junctions from both control and mutant cells (Figure S2). Many insertions were a few bp, but some were longer and were derived from other chromosomes or from Chr19 or Chr22 sequences close to the breakpoints (see also, Brunet et al., 2009; Piganeau et al., 2013).

Chromosomal Translocations in a Human pre-B Cell Line Depend on c-NHEJ

The dependence on c-NHEJ for translocation formation in human cells was quite surprising, given that c-NHEJ suppresses translocations in mouse cells. For comparison, we tested a second human LIG4 knockout cell line, N114, a derivative of the pre-B cell line NALM6 (Grawunder et al., 1998) (Figures 3A and 3B). As with HCT116 $L4^{-/-}$ cells, indels were reduced at the ZFN^{p84} site in N114 cells compared to NALM6 cells (Figure S3A), consistent with a c-NHEJ defect in these cells. Notably, a substantial decrease in the ZFN-induced translocations was observed in N114 cells compared to parental cells (Figure 3C).

In striking contrast to the HCT116 cells, most translocation breakpoint junctions from the parental pre-B cell line had short insertions (67% of junctions; Figures 3D and S3B), which likely arose from terminal deoxynucleotidyl transferase (TdT) activity in the pre-B cells, whereas only 27% of junctions from the mutant cells had insertions (p < 0.0001, Fisher's exact test). The median deletion length in the translocation junctions from NALM6 cells was 15 bp (Figure 3E), and a substantial fraction were deleted only within the ZFN overhang (25%; Figure 3F). By contrast, in N114 cells, deletions were significantly longer (median 64 bp; p < 0.0001) such that all extended beyond the ZFN overhang. Thus, longer deletions appear to be characteristic of translocation junctions in c-NHEJ-deficient human cells.

Microhomology comparisons were more limited, given the high frequency of insertions in the junctions from the parental cells. In the 33% of junctions from NALM6 cells without insertions, \geq 3 bp microhomology was overrepresented compared to the chance joining of two random sequences (35%, p = 0.001; Figure 3G). In the residual translocation junctions from N114 cells, \geq 3 bp microhomology was even more frequently observed (48% of junctions) such that they showed a strongly biased distribution compared to random (p < 0.0001; Figure 3G).

Patient-Derived LIG4 Mutant Cells Have an Intermediate Translocation Phenotype

Hypomorphic LIG4 mutations have been reported in patients with a radiosensitive syndrome (O'Driscoll et al., 2001). The primary skin fibroblast cell line 411BR was derived from a patient containing homozygous LIG4 mutations in the catalytic domain and N terminus (O'Driscoll et al., 2001) (Figures 3A and 3B). The mutant LIG4 protein has substantially reduced catalytic activity but still interacts with XRCC4 and supports nearly normal levels of V(D)J recombination, albeit with altered joining characteristics. Consistent with the V(D)J recombination assays, indels were not reduced at the ZFN^{p84} site in 411BR cells compared to a nonisogenic control skin fibroblast cell line, HDFa (Figure S3A). In addition, the translocation frequency was only mildly reduced, and this reduction was not statistically significant (Figure 3C). As with HCT116 cells, only a fraction of translocation junctions recovered from HDFa and 411BR cells contained insertions (Figures 3D and S3C), reinforcing a role for TdT in generating the frequent insertions observed in pre-B

(E) Microhomology distribution.

See also Figure S2.

⁽B) Translocation frequency is reduced in $L4^{-/-}$ and $X4^{-/-}$ HCT116 cells. Translocations were quantified as follows: ZFN^{EWS} and ZFN^{P84}, Der19 and Der22 (wt, n = 7; L4, n = 3; X4, n = 4); ZFN^{EWS} and TAL^{P84}, Der22 (n = 4); TAL^{LAM} and TAL^{P84}, Der1 (n = 4). Der19 and Der22 were assessed in the same experiment and the frequencies averaged. Error bars ± SEM. Nuclease expression was detected 48 hr after transfection.

⁽C-E)Translocation junction analysis from $X4^{-/-}$ cells demonstrates a shift toward longer deletions and an increased presence of microhomology. Der19 and Der22 junctions derived from ZFN expression were pooled.

⁽C) Deletion lengths from individual Der19 and Der22 junctions are indicated by the triangle and circle, respectively.

⁽D) Junctions are grouped according to whether the deletions were restricted to the overhang or were short (\leq 30 bp) or long (>30 bp) deletions extending outside of the overhang.





Figure 3. Chromosomal Translocation Formation Is Impaired in Human LIG4 Mutant Cells

(A) Locations of LIG4 mutations in N114 pre-B cells and 411BR fibroblasts.

(B) Western blotting for LIG4 and ZFNs.

(C) LIG4-null N114 cells have a significantly reduced ZFN-induced translocation frequency (n = 3), while hypomorphic 411BR cells have only a mild reduction (n = 4). Der19 and Der22 formation is pooled. Error bars ± SEM.

(D) High insertion frequency in Der19 and Der22 junctions from NALM6.

(E-G)Translocation junction analysis from LIG4 mutant cells.

(E and F) Junctions show longer deletions that always (N114) or mostly (411BR) extend beyond the ZFN overhang.



Figure 4. Cancer Translocation Induced by Paired Nicks and DSBs

(A) Induction of NPM-ALK cancer translocations with sequence specific nucleases. Der5 encodes the NPM-ALK fusion.

(B) Der5 is detected only when DSBs or paired nicks are induced on both chromosomes. Wildtype Cas9 with gRNAs NPM1 and ALK1 gives rise to translocations, as does nCas9 with gRNAs NPM1+NPM2 and ALK1+ALK2. Relative positions of cleavage sites are indicated.

(C) Indel formation at the $A\!L\!K$ locus, as monitored by the T7-endonuclease assay.

See also Figure S4.

cells. The median deletion length for 411BR cells was 21 bp, an increase relative to HDFa cells (4 bp; p < 0.0001), but not as much as for LIG4-null N114 cells (64 bp; p < 0.0001; Figures 3E and 3F). Microhomology distribution in junctions from 411BR cells did not differ from either control cells or a random distribution (Figure 3G). Thus, the hypomorphic 411BR cells presented an intermediate phenotype, suggesting that c-NHEJ is functional for translocation formation but altered such that longer deletions result.

NPM-ALK Translocations Induced by Cas9 DSBs and nCas9 Paired Nicks Involve c-NHEJ

We previously induced the NPM-ALK translocation found in anaplastic large cell lymphoma using TALENs that target the NPM and ALK loci on Chr5 and Chr2, respectively (Piganeau et al., 2013) (Figures 4A, 4B, and S4A). To examine the effect of c-NHEJ components on the formation of an oncogenic translocation, TAL^{NPM} and TAL^{ALK} were expressed in the HCT116 mutants. As with ZFN^{p84}, indel formation was reduced at the TAL^{ALK} site in the absence of LIG4 or XRCC4 (Figures 4C and S4B). NPM-ALK translocations were also reduced in the $L4^{-/-}$ cells compared to $L4^{+/-}$ cells (1.5-fold, p = 0.03, Figure 5A, with similar TALEN expression levels; Figure S5A), although not as dramatically as translocations induced at the EWS and p84 loci (Figure 2B). Correlating with the mild reduction in translocation frequency, the level of the NPM-ALK fusion protein was also somewhat reduced in the mutant cells (Figure 5B). As observed with ZFNs, TALEN-induced translocation junctions were dramatically altered by LIG4 or XRCC4 loss such that deletions and microhomologies were significantly longer (Figures 5C-5E and S5B).

We also tested Cas9-dependent induction of DSBs (Cong et al., 2013; Mali et al., 2013b) for translocation formation. Guide RNAs (gRNAs) were designed to result in DSBs near the TALEN cleavage sites upon Cas9 expression (NPM1 and ALK1, Figure 4B). Indel formation was increased at the Cas^{ALK1} site relative to the TAL^{ALK} site, but importantly, it was also reduced in $L4^{-/-}$ cells (Figures 4C and S4B).

Cas9-mediated translocation formation, which required both Cas^{NPM1} and Cas^{ALK1} (Figure 4B), was more efficient than with the cognate TALEN pair (~3-fold; Figure 5A), consistent with the increased indel formation (Figure 4C). The presence of the NPM-ALK fusion protein confirmed the translocations (Figure 5B). Interestingly, only a slight reduction in translocation frequency was observed in the $L4^{-/-}$ cells (Figure 5A). Unlike ZFNs and TALENs, wild-type Cas9 generates blunt DNA ends or a short overhang in vitro (Jinek et al., 2012). Minimal processing of these DNA ends was observed in control cells such that many junctions apparently arose by direct end ligation (Figures 5C–5E and S5C). Overall, the median deletion length was 1 bp, and the microhomology distribution was almost identical to that expected by chance (p = 0.81).

Translocation junctions were substantially altered in the $L4^{-/-}$ cells. None of the junctions arose by direct joining, and the median deletion length was quite long (283 bp; Figures 5C–5E and S5C). The microhomology distribution was significantly different from random (p < 0.0001), with 38% of the junctions having microhomology \geq 3 bp. Thus, NPM-ALK translocations induced by Cas9-generated DSBs arise by c-NHEJ in control cells with little processing of the DNA ends; in the absence of LIG4, alt-NHEJ gives rise to translocations with a remarkably different junction spectrum.

Patient-derived translocations often exhibit deletions and duplications at the breakpoint junctions (e.g., Zucman-Rossi et al., 1998), suggesting that DSBs with blunt DNA ends may not give rise to such translocations. nCas9, a Cas9 nickase due to a D10A mutation (Jinek et al., 2012), can generate DSBs with overhangs if two gRNAs are used that cleave opposite strands (i.e., paired nicks) (Mali et al., 2013a; Ran et al., 2013). We tested gRNAs that would give rise to offset nicks at the NPM and ALK loci, leading to DSBs with 5' overhangs

⁽G) Microhomology distribution at translocation junctions. Microhomology can only be determined at junctions without insertions (only 33% of junctions from wildtype NALM6 pre-B cells). Microhomology distributions from both pre-B cell lines were different from that expected by chance. By contrast, neither wild-type HDFa nor hypomorphic 411BR fibroblasts had microhomology distributions that differed from chance. See also Figure S3.



Figure 5. Translocations Induced by Paired Nicks and DSBs Rely on c-NHEJ

(A–E) Der5 induced by expression of the indicated nucleases in wild-type and mutant HCT116 cells. Nucleases: TAL, TAL^{ALK}+TAL^{NPM} (n = 4); Cas9, Cas9+gRNAs (ALK1+NPM1) (n = 4); nCas9, nCas9+gRNAs (ALK1+ALK2 and NPM1+NPM2) (n = 3).

(A) Translocation frequency. Error bars \pm SEM.

(B) The NPM-ALK fusion protein was detected 48 hr after expression of the indicated nucleases. NT, not transfected; SUPM2, ALCL cell line expressing NPM-ALK, 1/40th the extract loaded; I, protein marker ladder.

(C–E) Translocation junction analysis. Junctions from L4^{-/-} cells exhibited longer deletions and more microhomology.

(C) Deletion lengths from individual Der5 junctions. For wild-type Cas9, a large fraction of junctions from control $L4^{+/-}$ cells were direct ligations of the two ends, but these were absent from $L4^{-/-}$ cells. For nCas9, the deletions were quite long, reflecting the long overhangs; however, in $L4^{-/-}$ cells, unlike $L4^{+/-}$ cells, most of the deletions extended well past the overhangs (see last two columns with asterisks for deletion lengths beyond the overhang).



Figure 6. LIG3 Deficiency Does Not Affect Translocations in Human Cells

(A–C) Neither translocation frequency nor junction characteristics are altered by LIG3 loss. Translocations are induced by expression of the indicated nucleases: ZFN, ZFN^{P84}+ZFN^{EWS}, generating Der22 (n = 2); Cas9, Cas9+gRNAs (ALK1+NPM1), generating Der5 (n = 2); nCas9, nCas9+gRNAs (ALK1+ALK2 and NPM1+NPM2), generating Der5 (n = 2).

(A) Translocation frequency. Error bars ± SEM.

(B) Deletion lengths from individual Der22 (ZFNs) and Der5 (Cas9 and nCas9) junctions.

(C) Microhomology distribution. Microhomology in junctions from control and LIG3-deficient cells was similar.

See also Figure S6.

Translocation junctions from paired nicks differed substantially from those from wild-type Cas9-induced DSBs. The median deletion length was 85 bp with the paired nicks in the control cells compared with 1 bp with wild-type Cas9 (Figures 5C, 5D, and S5D). Many deletions involved only the \sim 40 base overhangs at each end (34%). Nevertheless, the median deletion length with the paired nicks was substantially longer in the $L4^{-/-}$ cells (243 bp, p = 0.0004) such that the majority of deletions extended beyond the overhangs into the double-stranded region (88%). Microhomology distribution for $L4^{+/-}$ cells was different from that expected by chance (Figure 5E), suggesting that the presence of long overhangs may promote more

of 41 and 37 bp, respectively (NPM1+NPM2 and ALK1+ALK2, Figure 4B). Indel formation from paired nicks directed to the *ALK* locus was efficient in control cells, consistent with DSB formation, and was reduced in $L4^{-/-}$ cells (Figures 4C and S4B).

Translocations were not recovered when a single nick was introduced on each chromosome but were recovered when paired nicks were introduced (Figure 4B), indicating that DSBs are required to drive translocation formation. Again, the NPM-ALK fusion protein was detected (Figure 5B). The translocation frequency was not as high as with Cas9-induced DSBs, possibly due to the requirement for four cleavage events by nCas9, rather than two by wild-type Cas9 (Figure 5A). In contrast to wild-type Cas9, a clear decrease in nCas9-induced translocations was observed in the $L4^{-/-}$ cells (Figure 5A).

Alt-NHEJ Does Not Affect Translocations in Human Cells Unless c-NHEJ Is Also Deficient

LIG3 is the major DNA ligase involved in translocation formation in mouse cells (Simsek et al., 2011). To determine if LIG3 affects translocation formation in human cells, we tested LIG3-null HCT116 cells (Oh et al., 2014) (Figure S6A). LIG3 deficiency did not affect indel efficiency at ZFN-, Cas9-, or nCas9-induced DSBs (Figure S6A). Similarly, LIG3 deficiency did not significantly affect the recovery of translocations (Figure 6A). Translocation junctions were also not significantly altered for either deletions or microhomology (Figures 6B, 6C, and S6B–S6D). Thus, LIG3

Deletions

p=0.54

Der5

p=0.40

L3+/-

nCas9

L3^{-/}

L3-/

Cas9

microhomology-mediated events even in the presence of LIG4. However, microhomology was much longer in the $L4^{-/-}$ cells (p < 0.04).

⁽D) Junctions are grouped according to whether the deletions extend beyond the overhang (\leq 30 bp or >30 bp) or are restricted to the overhang (or, in the case of wild-type Cas9, are 0 bp).

⁽E) Microhomology distribution. Microhomology in junctions from control $L4^{+/-}$ cells is distributed almost identically to that expected by chance with wild-type Cas9 but is greater with nCas9 which leaves long overhangs. See also Figure S5.



has either no role or a minor role in translocation formation in human cells.

In mouse cells, translocations are thought to arise by CtIPmediated resection of DNA ends, followed by annealing at microhomologies present in the resected DNA, since CtIP knockdown leads to smaller deletions and microhomologies at junctions (Zhang and Jasin, 2011). CtIP knockdowns had minimal effect on indel efficiency in wild-type and $X4^{-/-}$ HCT116 cells (Figure S7A). CtIP depletion also had no discernible effect in wild-type cells on ZFN^{p84}-ZFN^{EWS}-induced translocation frequency or junction characteristics (Figures 7 and S7B). However, in X4^{-/-} cells, translocations were significantly reduced (Figure 7A, p = 0.008). Residual translocations showed smaller deletions (Figures 7B, 7C, and S7C; p = 0.03), indicating that CtIP plays a role in resecting DNA ends in the absence of c-NHEJ components. Interestingly, microhomology was not significantly altered (Figure 7D). Thus, in human cells, CtIP appears to participate in translocations only in the absence of c-NHEJ.

DISCUSSION

We demonstrate here that c-NHEJ is the predominant mechanism for joining DNA ends during translocation formation in human cells. In cells lacking the LIG4/XRCC4 complex, translocation junctions were altered with longer deletions and microhomologies, signatures of alt-NHEJ. Our results suggest that

Figure 7. CtIP Affects Translocations in XRCC4-Deficient but Not Wild-Type Cells

(A) Translocation frequency is reduced with CtIP knockdown in $X4^{-/-}$ but not wild-type human cells (n = 3). Error bars \pm SEM. FLAG-ZFN levels at the time of transfection (t0) and 48 hr later (t2) is shown in the inset.

(B) Deletion lengths from individual Der22 junctions are reduced in $X4^{-/-}$ but not wild-type cells with CtIP knockdown.

(C) Fewer deletions extend \leq 30 bp beyond the overhang in *X*4^{-/-} cells with CtIP knockdown.

(D) Microhomology in junctions from control and $X4^{-/-}$ cells is not affected by CtIP.

See also Figure S7.

patient-derived translocations also occur primarily through c-NHEJ. Translocation frequency was often reduced in c-NHEJ mutants, although the efficiency with which joining was redirected to alt-NHEJ varied. In contrast, loss of alt-NHEJ components did not alter either translocation frequency or outcome unless c-NHEJ was also disrupted.

Different LIG4 Mutants Show Subtly Different Joining Characteristics

In each c-NHEJ mutant cell line tested, translocation junction characteristics differed from wild-type cells but with

subtle differences. HCT116 c-NHEJ mutants had junctions with longer deletions and more microhomology regardless of the nuclease employed (ZFNs, TAL, Cas9, and nCas9). Cas9-generated translocations from LIG4-null cells had the longest median deletion length of any of the translocations analyzed, implying that c-NHEJ is particularly critical for maintaining sequence information at these ends. Translocation frequency was lower in the HCT116 c-NHEJ mutants, but the reduction was greater at the EWS and p84 loci than at NPM and ALK, even when comparing the same type of nuclease (TALENs). These results suggest locus effects in the ability to switch from c-NHEJ to alt-NHEJ; alternatively, the oncogenic NPM-ALK fusion protein formed by the translocation could provide a slight growth advantage (Zhang et al., 2013). Interestingly, Cas9-induced NPM-ALK translocations were not significantly reduced in the absence of LIG4, suggesting that alt-NHEJ can efficiently act on these ends to generate translocations.

In patient 411BR cells, the translocation frequency was not affected, consistent with a report that low levels of LIG4 are sufficient for DSB repair in vitro (Windhofer et al., 2007), but deletion lengths were intermediate between control cells and other LIG4 mutants. The remaining LIG4 in these cells has residual catalytic activity (O'Driscoll et al., 2001), and it may also promote DNA-PKcs bridging activity (Cottarel et al., 2013) to protect ends from extensive degradation. The pre-B cell line NALM6 showed a large number of junctions with small insertions, likely due to TdT activity (Smith et al., 2003). TdT is required for diversification of antigen receptors, and TdT-generated insertions at these and other DSBs are dependent on c-NHEJ components (Boubakour-Azzouz et al., 2012; Smith et al., 2003). Consistent with this dependence, N114 cells had fewer insertions. NALM6 junctions without insertions showed longer microhomology than is typical of wild-type cells, raising the possibility that some of the junctions that did not engage TdT did not join by c-NHEJ; alternatively, TdT may have created microhomology in some of these junctions (Lieber, 2010).

Translocations in Human and Mouse Cells Form by Different Mechanisms

The requirement for c-NHEJ components in chromosomal translocation formation in human cells is in direct opposition to findings from mouse studies. In mice and mouse cell lines, c-NHEJ suppresses translocations while alt-NHEJ promotes their formation. Mice with mutations in c-NHEJ components develop RAG recombinase-mediated pro-B lymphomas on a p53-deficient background with complex Myc translocations involving microhomology (Zhu et al., 2002). Similarly, c-NHEJ mutant B cells undergoing class switching have increased Myc translocations, implying that alt-NHEJ is a robust mechanism for generating translocations (Boboila et al., 2010). In addition to immune-system-generated DSBs, nuclease-induced translocations in mouse embryonic stem cells form by alt-NHEJ such that the translocation junctions show reduced microhomology and deletions in the absence of LIG3 or CtIP (Simsek et al., 2011; Zhang and Jasin, 2011). Consistent with a c-NHEJ-independent mechanism, junctions from wild-type, LIG4/XRCC4, and Ku70 mutant murine cells show similar characteristics (Simsek et al., 2011; Simsek and Jasin, 2010; Weinstock et al., 2007). Thus, in mouse cells, DSB repair leading to translocations differs from intrachromosomal repair of a single DSB (Simsek and Jasin, 2010; Weinstock et al., 2006), whereas in human cells both types of repair are similar and mediated by c-NHEJ.

Although NHEJ components are conserved, human cells have substantially higher DNA-PK activity than rodent cells (Finnie et al., 1995; Beamish et al., 2000; Meek et al., 2001). Another striking difference is that loss of the Ku protein is incompatible with human cell survival (Li et al., 2002) due to an essential role in telomere maintenance (Wang et al., 2009), whereas rodent cells and mice are viable in the absence of Ku (Nussenzweig et al., 1996; Taccioli et al., 1994; Zhu et al., 1996). The enhanced DNA-PK activity and differential Ku-requirements imply that basic DSB metabolism may be different between humans and rodents, a hypothesis supported by our studies.

Nuclease Induction of Oncogenic Translocations: DSBs and Paired Nicks

Developing methodologies to induce oncogenic translocations has ramifications for understanding the role of translocations in tumor formation and their etiology. Here, we extended our previous studies in which cancer-relevant translocations were induced by ZFNs and TALENs (Piganeau et al., 2013) to Cas9 and nCas9 to generate the NPM-ALK translocation associated with lymphoma. Cas9 has also been used in HEK293 cells to induce a translocation found in lung tumors (Choi and Meyerson, 2014).

Cas9 has several advantages as a nuclease, in particular that it is easy to engineer by changing the gRNA sequence. Because mismatches to the gRNA can be tolerated, however, wild-type Cas9 has been associated with off-target mutagenesis (Fu et al., 2013). Paired nicks generated by nCas9 have the potential to substantially increase cleavage specificity, given that two gRNAs are required per DSB (Mali et al., 2013a; Ran et al., 2013), and we have determined that nCas9-generated paired nicks efficiently induce translocations.

Somewhat unexpectedly, we found that the DNA end structure markedly affected translocation junction characteristics. Cas9induced DSBs are blunt or have a small overhang (Jinek et al., 2012), and many of the resultant translocation junctions from control cells appear to be simple c-NHEJ-dependent joining of the DNA ends. As translocation formation destroys the gRNA site, these events must arise from a single joining event, rather than iterative cycles of breakage and rejoining. The precise conservation of DNA end sequences in many translocations seems surprising, given that translocation formation results in a chromosome aberration. However, it is consistent with c-NHEJ protection of DNA ends and suggests that c-NHEJ is not an inherently error-prone process (Bétermier et al., 2014).

The precise joining of Cas9-induced DSBs means that the translocation junctions often do not exhibit the small deletions typical of patient-derived translocations (Gillert et al., 1999; Langer et al., 2003; Reichel et al., 1998; Weinstock et al., 2006; Zhang et al., 2002; Zucman-Rossi et al., 1998), suggesting that cancer translocations may also arise from some other type of DNA end structure. nCas9-induced nicks provide flexibility in DNA end structures. The paired nicks we used to induce translocations were \sim 40 bp apart and thus led to relatively long 5' overhangs. Overhang sequences are retained in some junctions, suggestive of fill-in synthesis, as observed at ZFN and TAL overhangs (this report; Brunet et al., 2009; Piganeau et al., 2013). But deletions are significantly longer than those generated from wild-type Cas9, often encompassing one or both overhangs. Offset nicks with large overhangs that are filled in, however, such as those generated here by nicking, have been proposed to lead to duplications of sequences found in some leukemic translocation junctions (Gillert et al., 1999; Reichel et al., 1998), indicating that the paired nicks may provide a good model to interrogate the factors involved in these translocations.

Insertions ranging from ~20 bp to several hundred bp were found in ~6% of translocation junctions (Figures S2 and S5). These insertions were frequently derived from sequences close to one of the four DNA ends, as reported for mouse cells (Simsek and Jasin, 2010). However, unlike mouse cells, a significant fraction was also derived from other chromosomes, as reported for other human translocations (Piganeau et al., 2013; Sobreira et al., 2011; Zucman-Rossi et al., 1998). A plausible mechanism for their derivation is replication primed by one of the DNA ends using microhomology (Simsek and Jasin, 2010; Weinstock et al., 2006). Microhomology-primed duplication of sequences has similarities to what has been termed microhomology-mediated break-induced replication, which has been postulated as a mechanism to give rise to copy number variants in the germline (Carvalho et al., 2013; Hastings et al., 2009).

CONCLUSIONS

Understanding mechanisms of genome rearrangements that contribute to the accumulation of cancer-driving mutations and secondary mutations resulting from cancer therapies is a critical area of research. The results presented here provide strong evidence that c-NHEJ is the primary pathway for chromosomal translocation formation in human cells. It is noteworthy in this regard that patients with LIG4 mutations (LIG4 syndrome) are not extremely cancer prone, unlike patients with mutations in genes in many other DNA repair pathways (Woodbine et al., 2014). While loss of c-NHEJ components may result in decreased DSB repair with a concomitant increase in apoptosis, our results here suggest that chromosomal translocations that would promote oncogenesis are also less likely to form, providing an alternative (or additional) explanation for the lack of a strong tumor predisposition. In conclusion, while the high level of c-NHEJ activity in human cells compared to rodents may provide a risk for oncogenic translocations, this is offset by the more accurate and efficient DSB repair from c-NHEJ that may restrain cancer incidence while permitting a longer lifespan.

EXPERIMENTAL PROCEDURES

Additional experimental procedures are available online in the Supplemental Information.

Nucleases

All nucleases target intronic sequences. ZFN^{EWS}, ZFN^{FL11}, ZFN^{P84/AAVS1}, TAL^{ALK}, and TAL^{NPM} have been described (Brunet et al., 2009; Piganeau et al., 2013). ZFN^{FL1-A} and ZFN^{FL1-B} (Sangamo BioSciences) have an obligate heterodimer architecture. All ZFNs have FLAG epitope tags. TAL^{ALK} and TAL^{P84} have HA epitope tags; TAL^{LAM}, and TAL^{NPM} have FLAG epitope tags. pCas9_GFP expressing wild-type Cas9 (Addgene plasmid 44719) and pCas9-D10A_GFP expressing nCas9 (Addgene plasmid 44720) express fusions with GFP. The gRNA expression vector was derived from Addgene plasmid 43860 MLM3636. Cas9/gRNA target sequences are underlined with PAM sequences in bold:

NPM: CCTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAG AACTACAGG ALK:

CCTCAGGTAACCCTAATCTGATCACGGTCGGTCCATTGCATAGAGG**AGG**

Cell Lines and Repair Assays

LIG4 mutant ($L4^{+/-}, L4^{-/-}$) cells have been described (Oh et al., 2013); a manuscript detailing the derivation of XRCC4 mutant ($X4^{+/-}, X4^{-/-}$) cells is in preparation (B.R. and E.A.H., unpublished data). HCT116 LIG3^{flox/-} cells containing a conditional LIG3 allele and a deletion allele (Oh et al., 2014) were further engineered to express human LIG1 with a mitochondrial leader sequence (MtLIG1). The human pre-B cell line NALM6 derivative that is LIG4 defective, N114P2, has been described. (Grawunder et al., 1998), as have 411BR primary skin fibroblasts (O'Driscoll et al., 2001). For CtIP knockdown experiments, cells were transfected using LipofectamineTM RNAiMax (Invitrogen) with 40 nmol siRNA following the manufacturer's protocol, and the level of knockdown was evaluated by western blotting.

Cell lines were transfected by Amaxa technology (Lonza); genomic DNA was isolated 48 hr post transfection. T7 endonuclease I assays have been described (Piganeau et al., 2013). For intrachromosomal repair involving a single DSB (Asel assay) or two DSBs, cells were nucleofected with expression

vectors for ZFN^{EWS} or ZFN^{FLI-A} and ZFN^{FLI-B}, respectively. (Note: gRNAs have also been termed sgRNAs.) For translocations, cells were nucleofected with expression vectors for the relevant nuclease combinations. Translocation frequency was calculated from a 96-well screen using small pools of cells and nested PCR to amplify translocation junctions (Brunet et al., 2009; Piganeau et al., 2013). Frequencies were normalized to the number of viable cells 24 hr after transfection. Cells were also counted 48 hr after transfection; overall, mutant cells were reduced in number ~15%. Statistical analyses used a t test for frequency comparisons and Mann-Whitney for deletion and microhomology distributions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi. org/10.1016/j.molcel.2014.08.002.

AUTHOR CONTRIBUTIONS

M.J. and E.B. conceived the study; A.T., E.A.H., C.G., M.J., and E.B. designed experiments; H.G., M.P., B.R., and J.B.R. performed experiments; A.S., B.R., and S.O. developed key reagents; and M.J. and E.B. wrote the paper with A.E.T., E.A.H., and C.G.

ACKNOWLEDGMENTS

We thank Michael Lieber (USC) and Penny Jeggo (Sussex) for cell lines; Jean-Paul Concordet, Marine Charpentier, and Anne de Cian (MNHN) for discussions and technical assistance; Sangamo BioSciences, especially Fyodor Urnov and Lei Zhang, for providing ZFNs; and members of the Jasin laboratory, especially Francesca Cavallo and Gemma Regan-Mochrie. This work was supported by La Ligue Nationale contre le Cancer (H.G), Le Canceropole IDF (M.P.), an ANR grant ANR-12-JSV6-0005 (B.R.), and NIH grants to A.E.T. (GM047251, ES012512, and NCI P30CA118100), E.A.H. (GM088351 and CA15446), and M.J. (GM054668).

Received: June 10, 2014 Revised: July 14, 2014 Accepted: July 29, 2014 Published: September 4, 2014

REFERENCES

Beamish, H.J., Jessberger, R., Riballo, E., Priestley, A., Blunt, T., Kysela, B., and Jeggo, P.A. (2000). The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity. Nucleic Acids Res. 28, 1506–1513.

Bennardo, N., Cheng, A., Huang, N., and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet. *4*, e1000110.

Bétermier, M., Bertrand, P., and Lopez, B.S. (2014). Is non-homologous endjoining really an inherently error-prone process? PLoS Genet. 10, e1004086.

Boboila, C., Jankovic, M., Yan, C.T., Wang, J.H., Wesemann, D.R., Zhang, T., Fazeli, A., Feldman, L., Nussenzweig, A., Nussenzweig, M., and Alt, F.W. (2010). Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. Proc. Natl. Acad. Sci. USA *107*, 3034–3039.

Boboila, C., Alt, F.W., and Schwer, B. (2012). Classical and alternative endjoining pathways for repair of lymphocyte-specific and general DNA doublestrand breaks. Adv. Immunol. *116*, 1–49.

Boubakour-Azzouz, I., Bertrand, P., Claes, A., Lopez, B.S., and Rougeon, F. (2012). Terminal deoxynucleotidyl transferase requires KU80 and XRCC4 to promote N-addition at non-V(D)J chromosomal breaks in non-lymphoid cells. Nucleic Acids Res. *40*, 8381–8391.

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.

Bryans, M., Valenzano, M.C., and Stamato, T.D. (1999). Absence of DNA ligase IV protein in XR-1 cells: evidence for stabilization by XRCC4. Mutat. Res. *433*, 53–58.

Carvalho, C.M., Pehlivan, D., Ramocki, M.B., Fang, P., Alleva, B., Franco, L.M., Belmont, J.W., Hastings, P.J., and Lupski, J.R. (2013). Replicative mechanisms for CNV formation are error prone. Nat. Genet. *45*, 1319–1326.

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

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

Cottarel, J., Frit, P., Bombarde, O., Salles, B., Négrel, A., Bernard, S., Jeggo, P.A., Lieber, M.R., Modesti, M., and Calsou, P. (2013). A noncatalytic function of the ligation complex during nonhomologous end joining. J. Cell Biol. 200, 173–186.

Delacôte, F., Han, M., Stamato, T.D., Jasin, M., and Lopez, B.S. (2002). An xrcc4 defect or Wortmannin stimulates homologous recombination specifically induced by double-strand breaks in mammalian cells. Nucleic Acids Res. *30*, 3454–3463.

Deriano, L., and Roth, D.B. (2013). Modernizing the nonhomologous endjoining repertoire: alternative and classical NHEJ share the stage. Annu. Rev. Genet. 47, 433–455.

Fattah, F., Lee, E.H., Weisensel, N., Wang, Y., Lichter, N., and Hendrickson, E.A. (2010). Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. PLoS Genet. 6, e1000855.

Finnie, N.J., Gottlieb, T.M., Blunt, T., Jeggo, P.A., and Jackson, S.P. (1995). DNA-dependent protein kinase activity is absent in xrs-6 cells: implications for site-specific recombination and DNA double-strand break repair. Proc. Natl. Acad. Sci. USA *92*, 320–324.

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. Nat. Biotechnol. *31*, 822–826.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. *31*, 397–405.

Gillert, E., Leis, T., Repp, R., Reichel, M., Hösch, A., Breitenlohner, I., Angermüller, S., Borkhardt, A., Harbott, J., Lampert, F., et al. (1999). A DNA damage repair mechanism is involved in the origin of chromosomal translocations t(4;11) in primary leukemic cells. Oncogene *18*, 4663–4671.

Goodarzi, A.A., and Jeggo, P.A. (2013). The repair and signaling responses to DNA double-strand breaks. Adv. Genet. 82, 1–45.

Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., and Lieber, M.R. (1998). DNA ligase IV is essential for V(D)J recombination and DNA doublestrand break repair in human precursor lymphocytes. Mol. Cell 2, 477–484.

Guirouilh-Barbat, J., Rass, E., Plo, I., Bertrand, P., and Lopez, B.S. (2007). Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. Proc. Natl. Acad. Sci. USA *104*, 20902–20907.

Hastings, P.J., Ira, G., and Lupski, J.R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet. *5*, e1000327.

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.

Kabotyanski, E.B., Gomelsky, L., Han, J.O., Stamato, T.D., and Roth, D.B. (1998). Double-strand break repair in Ku86- and XRCC4-deficient cells. Nucleic Acids Res. *26*, 5333–5342.

Langer, T., Metzler, M., Reinhardt, D., Viehmann, S., Borkhardt, A., Reichel, M., Stanulla, M., Schrappe, M., Creutzig, U., Ritter, J., et al. (2003). Analysis

of t(9;11) chromosomal breakpoint sequences in childhood acute leukemia: almost identical MLL breakpoints in therapy-related AML after treatment without etoposides. Genes Chromosomes Cancer *36*, 393–401.

Li, G., Nelsen, C., and Hendrickson, E.A. (2002). Ku86 is essential in human somatic cells. Proc. Natl. Acad. Sci. USA 99, 832–837.

Liang, F., and Jasin, M. (1996). Ku80-deficient cells exhibit excess degradation of extrachromosomal DNA. J. Biol. Chem. 271, 14405–14411.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu. Rev. Biochem. 79, 181–211.

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

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

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. Science 339, 823–826.

Mani, R.S., and Chinnaiyan, A.M. (2010). Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. Nat. Rev. Genet. *11*, 819–829.

Mattarucchi, E., Guerini, V., Rambaldi, A., Campiotti, L., Venco, A., Pasquali, F., Lo Curto, F., and Porta, G. (2008). Microhomologies and interspersed repeat elements at genomic breakpoints in chronic myeloid leukemia. Genes Chromosomes Cancer *47*, 625–632.

Meek, K., Kienker, L., Dallas, C., Wang, W., Dark, M.J., Venta, P.J., Huie, M.L., Hirschhorn, R., and Bell, T. (2001). SCID in Jack Russell terriers: a new animal model of DNA-PKcs deficiency. J. Immunol. *167*, 2142–2150.

Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. Nat. Rev. Cancer 7, 233–245.

Nussenzweig, A., and Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. Cell *141*, 27–38.

Nussenzweig, A., Chen, C., da Costa Soares, V., Sanchez, M., Sokol, K., Nussenzweig, M.C., and Li, G.C. (1996). Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. Nature *382*, 551–555.

O'Driscoll, M., Cerosaletti, K.M., Girard, P.M., Dai, Y., Stumm, M., Kysela, B., Hirsch, B., Gennery, A., Palmer, S.E., Seidel, J., et al. (2001). DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. Mol. Cell 8, 1175–1185.

Oh, S., Wang, Y., Zimbric, J., and Hendrickson, E.A. (2013). Human LIGIV is synthetically lethal with the loss of Rad54B-dependent recombination and is required for certain chromosome fusion events induced by telomere dysfunction. Nucleic Acids Res. *41*, 1734–1749.

Oh, S., Harvey, A., Zimbric, J., Wang, Y., Nguyen, T., Jackson, P.J., and Hendrickson, E.A. (2014). DNA ligase III and DNA ligase IV carry out genetically distinct forms of end joining in human somatic cells. DNA Repair (Amst). *21*, 97–110.

Pannunzio, N.R., Li, S., Watanabe, G., and Lieber, M.R. (2014). Non-homologous end joining often uses microhomology: implications for alternative end joining. DNA Repair (Amst.) *17*, 74–80.

Piganeau, M., Ghezraoui, H., De Cian, A., Guittat, L., Tomishima, M., Perrouault, L., René, 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.

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., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

Reichel, M., Gillert, E., Nilson, I., Siegler, G., Greil, J., Fey, G.H., and Marschalek, R. (1998). Fine structure of translocation breakpoints in leukemic

blasts with chromosomal translocation t(4;11): the DNA damage-repair model of translocation. Oncogene *17*, 3035–3044.

Roth, D.B., Porter, T.N., and Wilson, J.H. (1985). Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol. *5*, 2599–2607.

Simsek, D., and Jasin, M. (2010). Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. Nat. Struct. Mol. Biol. *17*, 410–416.

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 Genet. 7, e1002080.

Smith, J., Riballo, E., Kysela, B., Baldeyron, C., Manolis, K., Masson, C., Lieber, M.R., Papadopoulo, D., and Jeggo, P. (2003). Impact of DNA ligase IV on the fidelity of end joining in human cells. Nucleic Acids Res. *31*, 2157–2167.

Sobreira, N.L., Gnanakkan, V., Walsh, M., Marosy, B., Wohler, E., Thomas, G., Hoover-Fong, J.E., Hamosh, A., Wheelan, S.J., and Valle, D. (2011). Characterization of complex chromosomal rearrangements by targeted capture and next-generation sequencing. Genome Res. *21*, 1720–1727.

Soni, A., Siemann, M., Grabos, M., Murmann, T., Pantelias, G.E., and Iliakis, G. (2014). Requirement for Parp-1 and DNA ligases 1 or 3 but not of Xrcc1 in chromosomal translocation formation by backup end joining. Nucleic Acids Res. *42*, 6380–6392.

Taccioli, G.E., Gottlieb, T.M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., and Jeggo, P.A. (1994). Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. Science *265*, 1442–1445.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat. Rev. Genet. *11*, 636–646.

Wang, H., Rosidi, B., Perrault, R., Wang, M., Zhang, L., Windhofer, F., and Iliakis, G. (2005). DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. Cancer Res. *65*, 4020–4030.

Wang, Y., Ghosh, G., and Hendrickson, E.A. (2009). Ku86 represses lethal telomere deletion events in human somatic cells. Proc. Natl. Acad. Sci. USA *106*, 12430–12435.

Weinstock, D.M., Elliott, B., and Jasin, M. (2006). A model of oncogenic rearrangements: differences between chromosomal translocation mechanisms and simple double-strand break repair. Blood *107*, 777–780.

Weinstock, D.M., Brunet, E., and Jasin, M. (2007). Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. Nat. Cell Biol. *9*, 978–981.

Windhofer, F., Wu, W., and Iliakis, G. (2007). Low levels of DNA ligases III and IV sufficient for effective NHEJ. J. Cell. Physiol. *213*, 475–483.

Woodbine, L., Gennery, A.R., and Jeggo, P.A. (2014). The clinical impact of deficiency in DNA non-homologous end-joining. DNA Repair (Amst.) *16*, 84–96.

Zhang, Y., and Jasin, M. (2011). An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat. Struct. Mol. Biol. *18*, 80–84.

Zhang, Y., Strissel, P., Strick, R., Chen, J., Nucifora, G., Le Beau, M.M., Larson, R.A., and Rowley, J.D. (2002). Genomic DNA breakpoints in AML1/ RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t(8;21) leukemia. Proc. Natl. Acad. Sci. USA *99*, 3070– 3075.

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. (2013). The potent oncogene NPM-ALK mediates malignant transformation of normal human CD4(+) T lymphocytes. Am. J. Pathol. *183*, 1971–1980.

Zhu, C., Bogue, M.A., Lim, D.S., Hasty, P., and Roth, D.B. (1996). Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates. Cell *86*, 379–389.

Zhu, C., Mills, K.D., Ferguson, D.O., Lee, C., Manis, J., Fleming, J., Gao, Y., Morton, C.C., and Alt, F.W. (2002). Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. Cell *109*, 811–821.

Zucman-Rossi, J., Legoix, P., Victor, J.M., Lopez, B., and Thomas, G. (1998). Chromosome translocation based on illegitimate recombination in human tumors. Proc. Natl. Acad. Sci. USA *95*, 11786–11791.

Supplemental Information

Supplemental Figure Summary

Figure S1, related to Figure 1. Intrachromosomal repair of a single or two DSBs in LIG4-XRCC4 deficient HCT116 human cells

A. Indel formation: T7 assay at the p84 locus after a ZFN^{p84} DSB.

B. Asel-resistant junction sequences after a ZFN^{EWS} DSB on Chr22.

C. Junction sequences at a 3.2 kb deletion on Chr11 after ZFN^{FLI-A} and ZFN^{FLI-B} DSBs.

Figure S2, related to Figure 2. Translocation junction sequences from ZFN^{EWS} and ZFN^{P84} DSBs derived from wild-type and XRCC4 deficient HCT116 cells.

Figure S3, related to Figure 3. Translocations induced by ZFN^{EWS} and ZFN^{P84} DSBs in pre-B cells and patient-derived cells with LIG4 mutations.

A. Indel formation: T7 assay at the p84 locus after a ZFN^{p84} DSB.

B. Translocation junction sequences from pre-B cells.

C. Translocation junction sequences from patient-derived cells.

Figure S4, related to Figure 4. DSBs induced by TALENs, wild-type Cas9, and paired nCas.

A. Relative positions of nCas9 cleavage sites.

B. Indel formation at the ALK locus after a TAL^{ALK}, Cas9(ALK1) or nCas9(ALK1+ALK2) DSB.

Figure S5, related to Figure 5. NPM-ALK cancer translocation junctions induced by TALENs, wild-type Cas9, and paired nCas.

- A. Nuclease expression is quantified after 48 h by Western blotting for the HA epitope for TALENs or GFP expression for Cas9 and nCas9.
- B. Translocation junction sequences from TALENs.
- C. Translocation junction sequences from wild-type Cas9.
- D. Translocation junction sequences from paired nCas9.

Figure S6, related to Figure 6. LIG3 is not required for translocation formation in human cells.

A. LIG3 null cells: Western blot, PCR genotyping, and ZFN expression, T7 assay at p84 and ALK loci.

B. Translocation junction sequences from ZFNs.

C. Translocation junction sequences from Cas9

D. Translocation junction sequences from paired nCas9.

Figure S7, related to Figure 7. Translocation junction sequences from ZFN^{EWS} and ZFN^{p84} expression after CtIP depletion.

- A. Western blot of CtIP knock-down and T7 assay at p84 locus.
- B. Translocation junction sequences from ZFNs in wild-type cells.
- C. Translocation junction sequences from ZFNs in XRCC4-deficient cells.

Supplemental Figures

Figure S1, related to Figure 1. Intrachromosomal repair of single or two DSBs in LIG4/XRCC4-deficient human HCT116 cells.

A. <u>Indel formation</u>: Indel formation is reduced in both L4^{-/-} and X4^{-/-} cells at the ZFN^{p84} cleavage site, as monitored by the T7-endonuclease assay {n = 7 for wt, n = 10 for het (+/-) and mutant (-/-)}.



B. <u>Asel-resistant junction sequences after a ZFN^{EWS} DSB on Chr22.</u> Duplicate junction sequences were counted twice only if they arose in independent experiments, although results were similar if all sequences were counted. The ZFN recognition sequences at each DNA end are underlined; bps in italics represent the overhangs. Microhomologies (underlined), insertions (green) and lengths of deletions from each end are indicated.

HCT116 WT

Blue : AseI site		
DNA ends: AATAGCTGCCTCCCCACTTTACATTAAT		GAC TGATAG G GAGGCCAAA AACGATGTT
29 AATAGCTGCCTCCCCACTTTACATTA-1	т	TGAC TGATAGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCACTTTACATTA-1		<i>TGA</i> C TGATAGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCACTTTACATTAA-1		-2CTGATAGGGAGGCCAAAAACGATGTT
AATAGCTGCCTCCCCACTTTACATTA-2		-1ACTGATAGGGAGGCCAAAAACGATGTT
AATAGCTGCCTCCCCACTTTACA5		GAC TGATAGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCACTTTACA5		-2C TGATGGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCACTTTACATT3	G	-4GATAGGGAGGCCAAAAAACGATGTT
AATAGCTGCCTCCCCACTTTACATT3	TT	-5 ATAGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCAC9		ATGAC TGATAGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCACTT9	AT	GAC TGATAGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCACTTT <u>AC</u> 6		-3. TGATGGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCACT10 45		-3.TGATAGGGAGGCCAAAAACGATGTT

AATAGCTGCCTCCCACTTTA5		-8GGGAGGCCAAAAACGATGTT
4 AATAGCTGCCTCCCCACT10 52		-4GATAGGGAGGCCAAAAACGATGTT
AATAGCTGCCTCCCCA10	A	-13GCCAAAAACGATGTT
5 AATAGCTGCCTCCCCAC11		-13GCCAAAAACGATGTT
6 AATAGCT <u>GCC</u> 18		-16
AATAGCTGCCTCC <u>CCA</u> 12		-15 AA AACGATGTT
19 AATAGCT GCCTCC<u>CCA</u>12		-17 AA AACGATGTT
41 AATAGCT GCCTC<u>CC</u>14		-16
46		-16
25		
<u>CTG</u> 29 43		-5ATAGGGAGGCCAAAAACGATGTT
AATAGCTGC		-16
AATA <u>G</u> 24 39		-23Atgtt
GATAGCTGGC	+11	-20
40		
10		TOAC I GATAGOGAGOCCARAAACGATGTT
AATAGCTGCCTCCCCA12 37	TG	-44
1 <u>TG</u> 29		-30
AATAGCTGCC <u>TCCCCA</u> 12 15		-59
GC31		-42
AA <u>TAGC</u> 22		-72
<u>CT</u> 30		-79
AATAGCTGCCTCC15		-133
AATAGCTGCCTCCCCA <u>CT</u> 10		-190
<u>A</u> 27		-208
2 / AATAGCTGC19 51		-235
		-13 <u>C</u>AAA AACGATGTT
50 		-12
44		-11
	TCm116 V/1/	

нст	116	X4+/	-

AATAGCTGCCCCCCCCCCCTTTACATTAAT	<i>GACTGATAGGGAGGCCAAAAACGATGTT</i>
36 AATAGCTGCCTCCCCACTTTACATTA-2 27	GAC TGATAGGGAGGCCAAA AACGATGTT
AATAGCTGCCTCCCCACTTTACATTAA-1	-2CTGATAGGGAGGCCAAAAACGATGTT

32 AATAGCTGCCTCCCCACTTTACATT3	Gi
11	
AATAGCTGCCTCCCCACTTTACATTA-2	-2
AATAGCTGCCTCCCCACTTTACAT4	Gž
3	
AATAGCTGCCTCCCCACTTT8 39	G2
AATAGCTGCCTCCCACTTT8	TAT GA
25 AATAGCTGCCTCCCACT =10	_/
17	
AATAGCTGCCTCCCCACTTTACA5	-1
24	
AATAGCTG	ATGA
21 AATAGCTGCCTCCCC _13	_1
33	
AATAG23	-9
8	
AATAGCTGCCTCCCCACTTTACA5	-1
38	
AATAGCTGCCTCC	-1
29	
22	
AATACC _22	_1
23	
	TTTTTGAT -2
15	
<u>CA</u>	-7
7	
	CA -7
40	
<u><u>c</u>100</u>	-7
2 A M A C C T C C C C C C C C T T T A C A 5	
10	-2
AATAGCTGCCTCC	-3
19	
AATAGCTGCCTCCCCAC11	-2
31	
<u>CT</u> 250	-3
34	
GA284	-2
CA	
<u></u>	
	HCT116 I.4+/-

GAC TGATAGGGAGGCCAAA AACGATGTT
-2C TGATAGGGAGGCCAAA AACGATGTT
GAC TGATAGGGAGGCCAAA AACGATGTT
GAC TGATAGGGAGGCCAAA AACGATGTT
GAC TGATAGGGAGGCCAAA AACGATGTT
-4GATAGGGAGGCCAAAAACGATGTT
-13GCCAAAAACGATGTT
ATGAC TGATAGGGAGGCCAAA AACGATGTT
-18 A AACGATGTT
-9GGAGGCCAAAAACGATGTT
-18 A AACGATGTT
-18 A AACGATGTT
-11AGGCCAAAAACGATGTT
-15 CAAA AACGATGTT
-23Atgtt
-75
-77
-73
-226
-225
-229
-35
-22
-335

<u>HCT116 L4+/-</u>

AATAGCTGCCCCCCCCCCCCTTTACATTAAT	GAC TGATAGGGAGGCCAAAAACGATGTT
5	
AATAGCTGCCTCCCCACTTTACATTAA-1	-2CTGATAGGGAGGCCAAAAACGATGTT
19	
AATAGCTGCCTCCCCACTTTACATTA-2	-2CTGATAGGGAGGCCAAAAACGATGTT
13	
AATAGCTGCCTCCCCACTTT8	GAC TGATAGGGAGGCCAAA AACGATGTT
24	
AATAGCTGCCTCCCCACTTT8	-4GATAGGGAGGCCAAAAACGATGTT
3	
AATAGCTGCCTCCCCACTTT8	-5ATAGGGAGGCCAAAAACGATGTT

18 AATAGCTGCCTCCCCACTTTACATTA-2		-12GGCCAAAAACGATGTT
34		
AATAGCTGCCTCCCCACTTTACATT3		-12GGCCAAAAACGATGTT
AATAGCT G 20		ATGAC TGATAGGGAGGCCAAA AACGATGTT
9 AATAGCTGCCTCCCCA		-18 B AACGATGTT
1		
AATAGCT <u>GCC</u> 18 32		-16 AAA AACGATGTT
AATAGCT GCC -18		-16C AAA AACGATGTT
AATAGCT GCC -18		-25GTT
12 AA26	CATC	-18 A AACGATGTT
36		
A25	т	-20ACGATGTT
<u>AA</u> 26		-21CGATGTT
		-18 A AACGATGTT
27 AATAGC22	CAGCTA	-30
17		
2 <u>1</u>		-18 A AACGATGTT
2 <u>2</u> 27		-39
<u>CAA</u> 26		-18 A AACGATGTT
33 AATA24		-44
16		
<u>AA</u> 26		-20ACGATGTT
<u>TTTACT</u> 61		-31
28 TGG33		-144
30		
AATAGCT GCC<u>TCC</u>15 35		-215
		-22GATGTT
<u>TG</u> 411		-24TGTT
20	+22	10 3,3,5,0,5,0,0,00
32bn : 11bn from Homo sonions chr/	TJJ altornato acco	mbly CHW1 1 1 and 22 hp from FWS
inverted duplication downstream DS	r, aiteindte dsse SR	and 22 bp from Ews
AAAATAGCTGCCTATTGCAGGCCACTATGATTT		

HCT116 X4-/-

AATAGCTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GAC TGATAGGGAGGCCAAA AACGATGTT
34 AATAGCT GCCTC -16 17	-18 R AACGATGTT
<u>AAT</u> 25 26	-25GTT
7 <u>AAT</u> 25	-32
AATAG <u>CT</u> 21 20	-37

<u>AATA</u>		-46
	+11	-36
11bp: TGCATACAAAA		
<u>CAA</u> 26		-49
<u>AG</u> 36		-60
<u>CA</u> 65		-39
<u>CATA</u> 62		-46
		-20 <u>A</u> CGATGTT
<u>AA</u> 175 8		-20ACGATGTT
<u>CC</u> 30		-178
	+27	-18 A AACGATGTT
27pb : Duplication from chr22 419bp up	stream DSB	
CTGAGCTCCATAAATCAACACTACATC From Homo	sapiens chromoso	ome 22, GRCh37.p13
5	А	-173
33		
ATGA		-25
AATAGCTGCCTCCCCA12		-226
36		
403		-55
35		126
TAAAAA 20		-130
ATC180		-363
32		
ATGTGT		-214

<u>HCT116 L4-/-</u>

AATAGCTGCCCCCCCCCCTTTACATTAAT		GAC TGATAGGAGGCCAAAAACGATGTT
1		
AATAGCTGCCTCCCCACTTTACA5		-16 AAA AACGATGTT
4		
AATAG <u>CTG</u> 20		-5ATAGGGAGGCCAAAAACGATGTT
13 AATAGCT <u>G</u> 20		-10GAGGCCAAAAACGATGTT
2 AATAGCT <u>GCC</u> 18		-16
42 AATAGCT GC1 9		-16
18		
AAT25 3	TC	-19AACGATGTT
A <u>AT</u> 25		-25GTT
AATAG <u>CTGC</u>		-43
7 <u>AAAA</u> 42		-21CGATGTT
34 <u>CA</u> 42		-45
22 CA74		-41

37		
<u>AC</u> 66 39		-38
<u>TGG</u> 33		-278
	cc	-32
36		-10 AAA AACGATGTT
AATAGCTGCCTCCCCACTTT8		-357
35 		-24TGTT
$\frac{44}{2}$		-109
4.3	ATT	-278
40 <u>TCT</u> 393		-162

C. Junction sequences at a 3.2 kb deletion on Chr11 after ZFN^{FLI-A} and ZFN^{FLI-B} DSBs.

Chr11

HCT116	WT

GGAACC CTAAGCCCTTTCCTTC ATTT(GC	GT)CCC CGATGAAAAGCAGGTTA GGTTGAGGGA
15	
GGAACCCTAAGCCCTTTCCTTCATTTG.	. TCCCCCGATGAAAAGCAGGTTAGGTTGAGGGA
2	
GGAACCCTAAGCCCTTTCCTTCATTT.	CCCCGATGAAAAGCAGGTTAGGTTGAGGGA
5	
CCAACCCTTAACCCCTTTCCTTTCATTT 1	ССС ССА ТСА А А А ССАССТВА ССТРСАСССА
, , , , , , , , , , , , , , , , , , , ,	100000 mC3 3 3 3 CO3 COmma Commo A COC
GGAACCETAAGCCETTTEETTEATTIGE	ICCCGATGAAAAGCAGGTTAGGTIGAGGGA
4	
GGAACCCTAAGCCCTTTCCTTCATTTGC	20CGATGAAAAGCAGGTTAGGTTGAGGGA
1	
GGAACCCTAAGCCCTTTCCTTCAT-2	CCCCGATGAAAAGCAGGTTAGGTTGAGGGA
59	
GGAACCCTAAGCCCTTTCCTTCATT-1.	. – 1CCCGATGAAAAGCAGGTTAGGTTGAGGGA
28	
GGAACCCTAAGCCCTTTCCTTCATTT	3.CGATGAAAAGCAGGTTAGGTTGAGGGA
14	
GGAACCCTAAGCCCTTTCCTTCATTTGC	5ATGAAAAGCAGGTTAGGTTGAGGGA
31	
GGAACCCTAAGCCCTTTCCTTC4	-1CCCGATGAAAAGCAGGTTAGGTTGAGGGA
34	
GGAACCCTAAGCCCTTTCCTTCAT-2	5ATGAAAAGCAGGTTAGGTTGAGGGA
17	
GGAACCCTAAGCCCTTTCC _7	-2000ATCA AAACCACCTTACCTTCACCCA
22	2000410444004001140011040004
CCAACC CTAACCCCTTTCCTTTC ATT 1	
GUAACCEIAAGCCEIIICEIICATT=1.	··- 12····· CAGGTTAGGTTGAGGGA
CC222CCC 10	
GGAACCC	3. CGATGAAAAGCAGGTTAGGTTGAGGGA

HCT116 X4+/-

GGAACCCTAAGCCCTTTCCTTCATTT (GC

GT)CCC**CGATGAAAAGCAGGTTA**GGTTGAGGGA

5	
GGAACCCTAAGCCCTTTCCTTCATTTG.	. TCCCCGATGAAAAGCAGGTTAGGTTGAGGGA
16	
GGAACCCTAAGCCCTTTCCTTCATTT	. TCCCCGATGAAAAGCAGGTTAGGTTGAGGGA
4	
GGAACCCTAAGCCCTTTCCTTCATTT	CCCCGATGAAAAGCAGGTTAGGTTGAGGGA
8	
GGAACCCTAAGCCCTTTCCTTCATT-1.	CCCCGATGAAAAGCAGGTTAGGTTGAGGGA
15	
GGAACCCTAAGCCCTTTCCTTCAT-2	CCCCGATGAAAAGCAGGTTAGGTTGAGGGA
21	
GGAACCCTAAGCCCTTTCCTTCATT-1.	. – 1CCCGATGAAAAGCAGGTTAGGTTGAGGGA
18	
GGAACC CTAAGC C CTTTCCTTC A3	CCCCGATGAAAAGCAGGTTAGGTTGAGGGA
19	
GGAACCCTAAGCCCTTTCCTTCATTTC.	
32	
GGAACCCTAAGCCCTTTCCTTC4	- 1 CCCGATGAAAAGCAGGTTAGGTTGAGGGA
41	
GGAACCCTAAGCCCTTTCCTTCATTTG.	
20	
CCAACCCTAACCCCTTTTCC _7	-2CCGATGAAAAGCAGGTTAGGTTGAGGGA
1	200000000000000000000000000000000000
	СССССА ТСА А А А ССАССТТА ССТТСАСССА
1	
22	4GAIGAAAAGCAGGIIAGGIIGAGGAA
CCAACC CTAA 15	<i>тессесатала за сел сетта</i> сетте лесел
12	. TECECONTEMANOCAGOTIAGOTIGAGOGA
CCAAC 20	
27	4GAIGAAAAGCAGGIIAGGIIGAGGAA
CCAACCCTTAACCCCTTTTCCTTTCATTT 1	25 ACCCA
27	=25AdddA
27 CCAACC CTTAACC 14	20
2	=39
22 CCN 22	40
20	=49
29 20	40
25	=47
23	44
<u>CAGCAG</u>	=44
28	27
<u>AA</u>	=>/

HCT116 L4+/-

GGAACC CTAAGC C CTTTCCTTC ATTT(GC	GT)CCC CGATGAAAAGCAGGTTA GGTTGAGGGA
3	
GGAACCCTAAGCCCTTTCCTTCATTTGC	4GATGAAAAGCAGGTTAGGTTGAGGGA
4 —	
GGAACCCTAAGCCCTTTCCTTCATTT	CCCCGATGAAAAGCAGGTTAGGTTGAGGGA
1	
GGAACCCTAAGCCCTTTCCTTCAT-2	CCCCGATGAAAAGCAGGTTAGGTTGAGGGA
2 -	
GGAACCCTAAGCCCTTTCCTTCATTTGC	4GATGAAAAGCAGGTTAGGTTGAGGGA
7 –	
GGAACCCTAAGCCCTTTCCTTC4	. –1CCCGATGAAAAGCAGGTTAGGTTGAGGGA
8	
GGAACCCTAAGCCCTTTCC7	3.CGATGAAAAGCAGGTTAGGTTGAGGGA
11	
CC47	3.CGATGAAAAGCAGGTTAGGTTGAGGGA
11	
<u>AACC</u> 47	49

HCT116 X4-/-

GGAACC CTAAGC C CTTTCCTTC ATTT(GC		GT)CCC CGATGAAAAGCAGGTTA GGTTGAGGGA
1 GGAACC CTAAG<u>CCC</u>12		4GATGAAAAGCAGGTTAGGTTGAGGGA
5 GGAACC CTAA<u>G</u>15		5 ATGAAAAGCAGGTTA GGTTGAGGGA
46 GGAACC C -19		4GATGAAAAGCAGGTTAGGTTGAGGGA
19 GGAACC CTAAGCC		-14 AGGTTAGGTTGAGGGA
37		22
24		=55
<u>GCAGG</u> 24 33		17 TTA GGTTGAGGGA
<u>AGCAGG</u>		17 TTA GGTTGAGGGA
<u>GGAA</u> 22		31
GG <u>AACC</u>		49
<u>CAGCAG</u> 35		44
AGGTTAG		21GTTGAGGGA
<u>TGAG</u> 69		27GGA
39 <u>TTC</u> 41		· 65 · · · · · · · · · · · · · · · · · ·
46 GGAACC C -19		102
7 GCAAAGC		77
9 GGAACC CTAAGC C CTTTCC -7		169
13 		-198
10		201
8	G	201
GGAACC CTAAGC C CTTTC<u>CT</u>6 31		207
GGAACC CTAAGCC -13 12		219
<u>CAGC</u> 168		133

<u>HCT116 L4-/-</u>

GGAACCCTTAAGCCCTTTCCTTCATTT(GC	GT)CCC CGATGAAAAGCAGGTTA GGTTGAGGGA
16 GGAACC CTAAGC C CTTT<u>CC</u>7	3.CGATGAAAAGCAGGTTAGGTTGAGGGA
10 GGAACC CTA<u>AG</u>15	27GGA
23 GGA23 17	23TGAGGGA
<u>GGAA</u> 22	31

GCAG25		44
CAGCAG		44
<u>TGATT</u> 42		55
		12
	+13	42
25 153		24 <u>GA</u> GGGA
<u>CC</u> 118		69
	CTT	191
<u>TGATT</u> 42		55

<u>Figure S2, related to Figure 2.</u> Translocation junction sequences from ZFN^{EWS} and ZFN^{P84} DSBs derived from wild-type and c-NHEJ mutant HCT116 cells.

Each junction was independently derived from small pool PCR. Deletions ≤ 800 bp could be identified. The ZFN recognition sequences at each DNA end are underlined; bps in italics represent the overhangs. The chromosome 19 end is in black and the chromosome 22 end is in red. Deletions from each end are indicated, as are microhomologies (underlined) and insertions (green). Insertions were observed in a fraction of breakpoint junctions from both control and mutant cells (wild-type, 18%; X4^{+/-}, 14%; X4^{-/-}, 24%).

Der19

HCT116 wt

DNA ends:		
TCTGTCACCAATCCTGTCCCTAGTGGCC		GTCATT<u>AATGTAAAGTGGGGAGGC</u>AGCT
2		
TCTGTCACCAATCCTGTCCCTAGTGGCC		. TCATTAATGTAAAGTGGGGAGGCAGCT
24		
TCTGTCACCAATCCTGTCCCTAGTGGC.		. TCATTAATGTAAAGTGGGGAGGCAGCT
26		
TCTGTCACCAATCCTGTCCCTAGTGGC.		CATTAATGTAAAGTGGGGAGGCAGCT
7,9,12,18,28		
TCTGTCACCAATCCTGTCCCTAGTGG.		CATT AATGTAAAGTGGGGAGGC AGCT
8,20		
TCTGTCACCAATCCTGTCCCTAGTG-1.		_TCATTAATGTAAAGTGGGGAGGCAGCT
22 mcmcmcaacaaaamacmcmccccmacmcccc		
TOTGTCACCAATCCTGTCCCTAGTGGC.		ZTTAATGTAAAGTGGGGAGGCAGCT
J, ZI, JI TCTCTCACCAATCCTCTCTCCTACT_2		Саттаатстаастссссссссссс
32		
TCTGTCACCAATCCTGTCCCTAGT-2		. TCATTAATGTAAAGTGGGGAGGCAGCT
19		
TCTGTCACCAATCCTGTCCCTAGTGGC.		3.TAATGTAAAGTGGGGAGGCAGCT
27		
TCTGTCACCAATCCTGTCCCTAGTGG		3.TAATGTAAAGTGGGGAGGCAGCT
10		
TCTGTCACCAATCCTGTCCCTAGTG-1.	с	3.TAATGTAAAGTGGGGAGGCAGCT
17		
TCTGTCACCAATCCTGTCCCTAG3	GGC	3.TAATGTAAAGTGGGGAGGCAGCT
	mm a a m	
CIGICACCAATCCTGTCCCTAGIGGC.	TTAAT	0TGTAAAGTGGGGAGGCAGCT
0 		26
1		20
GCC		35
16		
GG32		34
23		
	CT	35
5		
<u>GCT</u>		26
14		
		196
25	~	27
	G	2/

HCT116 X4+/-

TCTGTCACCA <u>ATCCTGTCCCTA</u> GTGG <i>CC</i>		<i>GT</i> CATT <u>AATGTAAAGTGGGGAGGC</u> AGCT
53		
TCTGTCACCAATCCTGTCCCTAGTGGC.		GTCATT AATGTAAAGTGGGGAGGC AGCT
36.42.51.56		
TCTGTCACCAATCCTGTCCCTAGTGGC.		. TCATT AATGTAAAGTGGGGGGGGGC AGCT
50,52,54		
TCTGTCACCAATCCTGTCCCTAGTGGC.		CATTAATGTAAAGTGGGGAGGCAGCT
35,38,41,46,49		
TCTGTCACCAATCCTGTCCCTAGTGG		CATTAATGTAAAGTGGGGAGGCAGCT
34,39,47		
TCTGTCACCAATCCTGTCCCTAGTG-1.		. TCATTAATGTAAAGTGGGGAGGCAGCT
45		-
TCTGTCACCAATCCTGTCCCTAGTGGC.		2TTAATGTAAAGTGGGGAGGCAGCT
31		
TCTGTCACCAATCCTGTCCCTAGT-2		CATTAATGTAAAGTGGGGAGGCAGCT
48		
TCTGTCACCAATCCTGTCCCTAGTG-1.	с	2TTAATGTAAAGTGGGGAGGCAGCT
62		
TCTGTCACCAATCCTGTCCCTAGTGGC.		3.TAATGTAAAGTGGGGAGGCAGCT
64		
TCTGTCACCAATCCTGTCCCTAGTGGC.		5ATGTAAAGTGGGGAGGCAGCT
30		
GGC		35
37		
AGG		21CAGCT
33		
TCTGTCACCAATCCTGTC8	т	234
40		
TCTGTCACCAATCCTGTCCCTAGTGGC.	+72	GTCATTAATGTAAAGTGGGGAGGCAGCT
72bp directly from upstream of DS	B on chr22	

<u>CTGCTAGCCCTGCTGTCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCA</u>

HCT116 X4-/-

TCTGTCACCA <u>ATCCTGTCCCTA</u> GTGG <i>CC</i>	<i>GT</i> CATT <u>AATGTAAAGTGGGGAGGC</u> AGCT
83	
TCTGTCACCAATCCTGTCCCTAGTG-1	9AAAGTGGGGAGGCAGCT
74	
TCTGTCACCAATCCTGTCCC6 ACTGTGGG	9AAAGTGGGGAGGCAGCT
8bp directly from downstream of DSB on chr19	
70	
TCTGTCACCAATCCTGTCCCTAGTGG	16 GGAGGC AGCT
78	
TCTGTCACCAATCCTGT9	9AAAGTGGGGAGGCAGCT
80	
TCTGTCACCAATCCTGTCCCTAGTGG	21 C AGCT
84	
TCTGTCA21	2.TTATGTAAAGTGGGGAGGCAGCT
65	
TCTGTCACCA AT -14	11AGTGGGGAGGCAGCT
93	
TCTGTCACCAATCCTGTCCCTAG3	24
69	
TCTGTCACCAATCCT11	26
64	
TCTGTCACCA16	24CT
76	
	13 GGC AGCT
92	

TCTGTCACCAATCCTGTCC7		40
66		
		13TGGGGAGGCAGCT
88		
TCTGTCACCAATCCTGTCCCT5	TTTT	43
81		
TCTGTCA <u>CCA</u>		37
75		
T <u>CT</u> 23		39
86		
<u>GCT</u> 28		26
87		
42	G	23
90		
<u>48.</u> .		17
71		
<u>T</u> 33		46
	110	
TCTGTCACCAATCCTGTCCCTAG3	+10	144
+16DD :GCCAAAAACGATGGCC		
122	2.20	24
70	AAT	=34
124		27
68	9	=37
	T	-158
85	-	150000000000000000000000000000000000000
CT=256		-26
96		
GCAGG		
89		
TCTGTCACCAATC13		299
94		
TCTGTCACCAATCCTGTCCCTAGT-2		311
91		
		10

Der22

HCT116 WT

ATCGT <u>TTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>	GTGGCCCCCCCTGTGGGGTGGAGGGGACAGAT
2,4,7,15,24,36,45	
ATCGT TTTTGGCCTCCCTATCA GTCA <i>TT</i>	GGCCCCACTGTGGGGTGGAGGGGACAGAT
6,18,34,46	_
ATCGT TTTTGGCCTCCCTATCA GTCAT.	GGCCCCACTGTGGGGGTGGAGGGGGACAGAT
19	
ATCGT TTTTGGCCTCCCTATCA GTCAT.	1GCCCCACTGTGGGGTGGAGGGGGACAGAT
10,17,40	
ATCGT TTTTGGCCTCCCTATCA GT-2	GGCCCCACTGTGGGGTGGAGGGGACAGAT
38	
ATCGTTTTTGGCCTCCCTATCA4	IGCCCCCCTGTGGGGTGGAGGGGACAGAT
3	
ATCGTTTTTGGCCTCCCTATCAGTCAT.	5CACTGTGGGGGGGGGGGGGGACAGAT
20 MCCCCCCCCC 13	A CCDCTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
42	4CCACIGIGGGGIGGAGGGGACAGAI
25.27	12
ATCCTTTT _17	-12 GGGTGGAGGGGAGAGAT

35		
47	с	22GACAGAT
ATCGT TTTTGGC<u>CT</u> 11		125
ATCGT TTT<u>T</u>17		196
<u>GCC</u> 239		148
ATCGTTTTTGG15	+119	<i>GT</i> GGCC CCACTGTGGGGT GGAGGGGACAGAT
25bp from chr19 inverted duplic	ation from 29bp dow	nstream of DSB
94bp from chr19 directly upstre	am of DSB	
GTCCTAACAGGAGGTGGGGGGTTAGACCCAAT. TGTCCCTA	ATCAGGAGACTAGGAAGGA	GGAGGCCTAAGGATGGGGCTTTTCTGTCACCAATCC
33		
ATCGTTTTTGGCCTCCCTATCAGTC-1.	+134	4CCACTGTGGGGGTGGAGGGGGACAGAT
134bp from Homo sapiens chris,	alternate assembly	HUKEI
AGACTCAAACCCCTCATCGCTTTGGACACGC	TRATACCIGITCITIGCGI TTTTTACCTTTTTCTTTGCGI	Gergererininaniicigenangaenangeene
1	<u>-</u>	
ATCGTTTTTGGCCTCCCTATCAG3	+233	23ACAGAT
233bp from Human DNA sequence f	rom clone RP11-335L	15 on chromosome 9, complete
sequence		
TCAGACTCTTCAACATAGCAAGTCCTTCCCA	GACAATTCCACTCCAATCA	TAAGGACCACTCCTTCCTCTAAGGATTCAAAGCACT
TTGTACTTATAGGGCCTCTAGGGGTACTGCC	CTCACTCTATTGTTAATGG	TATATATTTCTATCTCCCTCACTAGGCTTCCAACTA
TTTAAGAAAAGGAtGCATATTTTTACTCCTC	TETGTGTCTTCAGCACCCT	TCCACAGTGCCTGCAACA
32	1001	63
221bp:	+221	63
-133bp from Homo sapiens chr7,	alternate assembly	HuRef:
TCAGGCCAGAGTGTGGGGAAAGGCTCAAGGA TTTCTCTATGATGACACAGTTGGGAAAGCTG	ATTAGAATTCTCACCTGAT CTCTCCAGGGTCCCCCG	TGTGCCCCCAGCTGTGGACCGTGGCCTCGGCTTCCT
-88bp from Homo sapiens chr10,	alternate assembly	HuRef
GCCACGGTGGGCCCTGTGACCCCTCTCTCTG	CCTGCGCAGAACCTGGAGT	ACTGTATCATGGTCATTGGGGTCCCCAACGTGGGCA
AGTCCT		

HCT116 X4+/-

ATCG <u>TTTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>		<i>GT</i> GGC <u>CCCACTGTGGGGT</u> GGAGGGGACAGAT
34,43,49,58,61 ATCGTTTTTGGCCTCCCTATCAGTCATT		GGCCCCACTGTGGGGTGGGAGGGGACAGAT
50,52,60,62		· <u>-</u>
ATCGTTTTTGGCCTCCCTATCAGTCAT.		GGCCCCCCCCGCGGGGGGGGGGGGGGGGGGGGGGGG
41,48		-
ATCGTTTTTGGCCTCCCTATCAGTCAT.		1GCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
51 ATCGT TTTTGGCCTCCCTATCA GT-2	A	1GCCCCCCCCCGCGGGGGGGGGGGGGGGGGGGGGGGGG
35		
ATCGT TTTTGGCCTCCCTATCA GTCA <u>T</u> .		11GGGGTGGAGGGGACAGAT
		CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
45		GGCCCCCCIGIGGGGIGGAGGGGACAGAI
ATCGT TTT -18		GGCCCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGG
36		-
ATCGT TTTTGGCCTCCTAT 6		16GGAGGGGACAGAT
54		
ATCGT TTTTGGCCTCCCTAT CAGTCA <i>TT</i> 44	GAAGCAACCGA	41

ATCGT TTTTG 16	A	35
ATCGTTTTTGGCCTCCCTATC5 46	ACATTTGCCCC	62
	AG	29
AGTA34		36
<u>CAG</u> 40		31
ATCGT TTTT<u>GG</u> 15		103
+22bp : 17bp from Homo sapiens	+22 chromosome 8, clone	115 CTD-2517M22, complete sequence
A-CTGCGGAGCCCTGATGG-CGGG		
<u>GT</u> 266		GGCCCCCCTGTGGGGGTGGAGGGGGACAGAT
311		324
ATCGTTTTTGGCCTCCCTATCAGTCATT	+276	GTGGCCCCACTGTGGGGGTGGAGGGGGACAGAT

HCT116 X4-/-

ATCG <u>TTTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>		<i>GT</i> GGC <u>CCCACTGTGGGGT</u> GGAGGGGACAGAT
94		
ATCGT TTTTGGCCTCCC -9	A	3.CCCACTGTGGGGTGGAGGGGGACAGAT
70		
A25		2CCCCACTGTGGGGTGGAGGGGGACAGAT
76,89		
<u>CA</u> 25		7CTGTGGGGGTGGAGGGGACAGAT
80 25		26
91		=20
ATCGTTTTTGGCCTCCCTAT6		55
95		
<u>GT</u> 46		16GGAGGGGACAGAT
79		
<u>CA</u> 28		40
93	7.Cmm	27
66	ACTI	=57
CAG40		41
92		
		49
87		
<u>GG</u> 55		62
84		25
85		=35
CAG40		87
68		
GCCCTG65		67
88		
<u>CA</u> 41		97

82		
77		37
	т	68
73		
71	т	8 TGTGGGGTG GAGGGGACAGAT
ATCGTTTTTGGCCTCCCTATCAG3		232
83		
TCC		129
74		
<u>GCC</u> 277		64
65		
<u>GAG</u>		337
/5		220
81		=330
ATC23	+56	276
56bp: Inverted duplication from 229	3bp upstrea	am of DSB on chr19
CTGGACTTCGGCTTTTGTCCCCCCAAGTTTTGGAC	CCCTAAGGGAA	AGAATGAGAAAC
96		
220	+37	43
37bp: AGATTCATTTATACAATGAATACATCATT	GTCTTTGT	
Part of the sequence may come from:		
-18bp from Homo sapiens chr7, alter	nate assemb	oly CHM1_1.1
ATTCATTTATACAATGAA		

-16bp from Homo sapiens v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4 (ERBB4), RefSeqGene on chromosome 2 -ACATCATTGTTCTTTG

Figure S3, related to Figure 3. Translocations induced by ZFN^{EWS} and ZFN^{P84} DSBs in pre-B cells and patient-derived cells.

A. Indel formation: T7 assay on p84 locus after a ZFN^{p84} DSB. Indel formation at the ZFN^{p84} cleavage site, as monitored by the T7-endonuclease assay, in wild-type and LIG4 mutant pre-B cells and fibroblasts.



B. Translocation junction sequences from pre-B cells.

The ZFN recognition sequences at each DNA end are underlined; bps in italics represent the overhangs. The chromosome 19 end is in black and the chromosome 22 end is in red. Microhomologies (underlined), insertions (green) and lengths of deletions from each end are indicated.

Der19

NALM6		
DNA ends: TCTGTCACCA <u>ATCCTGTCCCTA</u> GTGGCC		<i>GT</i> CATT <u>NATGTNAAGTGGGGAGGC</u> AGCT
11 TCTGTCACCA ATCCTGTCCCTA GTGGC. 5 bp from chr19 duplication TGGC	TGGCG G	GTCATT AATGTAAAGTGGGGAGGC AGCT
TCTGTCACCAATCCTGTCCCTAGTGG	TT	GTCATT AATGTAAAGTGGGGAGGC AGCT
23 TCTGTCACCAATCCTGTCCCTAGTGGCC 22 39	с	2TTAATGTAAAGTGGGGAGGCAGCT
TCTGTCACCAATCCTGTCCCTAG3	G	GTCATT AATGTAAAGTGGGGAGGC AGCT
TCTGTCACCAATCCTGTCCCTAG3 76	GG	GTCATT AATGTAAAGTGGGGAGGC AGCT
TCTGTCACCAATCCTGTCCCTAG3 3	AG	GTCATT AATGTAAAGTGGGGAGGC AGCT
TCTGTCACCAATCCTGTCCCTAG3 10	AAG	GTCATT AATGTAAAGTGGGGAGGC AGCT
TCTGTCACCAATCCTGTCCCTAG3 8	GCG	GTCATT AATGTAAAGTGGGGAGGC AGCT
TCTGTCACCAATCCTGTCCCTAG3 59	GCGC	. TCATTAATGTAAAGTGGGGAGGCAGCT

TCTGTCACCAATCCTGTCCCTAG3	GGA	. TCATTAATGTAAAGTGGGGAGGCAGCT
33		
TCTGTCACCAATCCTGTCCCTAG3	GG	CATTAATGTAAAGTGGGGAGGCAGCT
20		
TCTGTCACCAATCCTGTCCCTAGTGGCC	TCG	3.TAATGTAAAGTGGGGAGGCAGCT
83		
TCTGTCACCAATCCTGTCCCTAGTGGC.	GCC	3.TAATGTAAAGTGGGGAGGCAGCT
26		
TCTGTCACCAATCCTGTCCCTAGTGGC.	GC	3.TAATGTAAAGTGGGGAGGCAGCT
35		
TCTGTCACCAATCCTGTCCCTAGT-2	TC	2TTAATGTAAAGTGGGGAGGCAGCT
42		
TCTGTCACCAATCCTGTCCCT5	GAGAG	GTCATTAATGTAAAGTGGGGAGGCAGCT
68		
TCTGTCACCAATCCTGTCCCTAGTGGC.	GCCCTGG	6 TGTAAAGTGGGGAGGC AGCT
72		
TCTGTCACCAATCCTGTCCC6	CGGGA	GTCATT AATGTAAAGTGGGGAGGC AGCT
2		
TCTGTCACCAATCCTGTCCC6	CCCCA	3.TAATGTAAAGTGGGGAGGCAGCT
24		
TCTGTCACCAATCCTGTCCCTA4		5ATGTAAAGTGGGGAGGCAGCT
12		
TCTGTCACCAATCCTG10	CCTGGGAC	. TCATTAATGTAAAGTGGGGAGGCAGCT
60 IO	20000	
TCTGTCACCAATCCTG10	AGGGGC	. TCATTAATGTAAAGTGGGGAGGCAGCT
01 mcmcmchcchhamccm 11	CTTC 3 3 C C C	
101G1CACCAATCCT	CIGAAGCC	. TCATTAATGTAAAGTGGGGGGGGGGGCAGCT
44 TCTCTCACCA ATC _13	TAGCC	СЛСАНТАВЛСТВАВСТСССВССВССТ
12	INGGC	GICATIAAIGIAAAGIGGGGGGGGGGCAGC
тотелерассар атест 11	TTCC	
18.32	lice	S.IAAIGIAAAGIGGGGAGGCAGCI
тстстсассаатсст _11		
31		
TCTGTCACCA ATCC	CGTCC	
21		
TCTGTCACCAATCCTG10	+52	9AAAGTGGGGAGGCAGCT
52bp inverted sequence located 1	1bp downstream of	DSB from chr19
GGCCGGTTAATGTGGCTCTGGTTCTGGGTACT	TTTATCTGTCCCCTCCAC	CCCA
 16		
TCTGTCAC	TTTAC	3.TAATGTAAAGTGGGGAGGCAGCT
54		
тстд22	CGA	3.TAATGTAAAGTGGGGAGGCAGCT
62		
TCTGTCACCAATCCTG10		20 GC AGCT
=		

G

ССТ

тс

..-3.TAATGTAAAGTGGGGAGGCAGCT

..-21.....**C**AGCT

..-13.....<u>TGGGG</u>AGGCAGCT

..-33.....

...-48.....

..-19.....**GGC**AGCT

..-5...ATGTAAAGTGGGGAGGCAGCT

67

34

.....-27..

TCTGTCACCA**A**....-15... 43

.<u>...</u>-35...

<u>53</u>....-32...

<u>TG</u>....-33..

9______118...

36		
TCTG22	cc	233
66		
		12GTGGGGGGGGGGGCAGCT
6		—
	+51+8	23GCT
59bp: 51bp from chr22 17bp upstrea	um of DSB	
CTGCTGTCTTTGGGGGAAGTTGTATGCAGTGAGTA	AATTCAACATCGTT	TTTCGTTTGTG
75		
205	с	3.TAATGTAAAGTGGGGAGGCAGCT
41		
	AAT	283
89		
413		11AGTGGGGAGGCAGCT
4		
256		119
79		
TCTGTCACCAATCCTGTCCC6	+229	7GTAAAGTGGGGAGGCAGCT

N	1	14	1
14	-	43	Ξ.

TCTGTCACCAATCCTGTCCCTAGTGGCC		<u>GTCATTAATGTAAAGTGGGGAGGCAGCT</u>
83		
TCTGTC	TG	8TAAAGTGGGGAGGCAGCT
59		
TCTGTCAC		11AGTGGGGAGGCAGCT
64		
TCTGTCACCA ATC<u>CT</u> 11		26
58		
TCT <u>G</u> 22		21CAGCT
/1 22		24
95		24 <u>CT</u>
-29		-19 GGC ACCT
66		
тт25		29
61		
	GTGGG	23GCT
84		
<u>G</u> 30		33
-29		-35
65		
GG=30		-34
88		
GG31		34
73		
		39
76		
<u>CTA</u>		40
87		
		48
77		
	AA	64
80		
AAG		5/
55		
=123		CATTAATGTAAAGTGGGGAGGCAGCT
81		

	6+27+4	57
37bp: 27 bp sequence from chr22 70	bp upstream of	DSB
CAGTCC ATCAGAAAAAATGTTTTAGACTGCTAG	TTTC	
75		
		12GTGGGGAGGCAGCT
74		—
TC24	+76	126
76bp inverted sequence from chr22	20bp upstream of	of DSB
AACGATGTTGAATTTACTCACTGCATACAACTTC	CCCAAAGACAGCAG	GCTAGCAGTCTAAAACATTTTTTCTGATCA
60		—
A60		106
92		
GCAGG243		34
57		
TCTGTCACCAATCCT11		288
78		
AGA		
<u></u>		

Der22

NALM6			
ATCG <u>TTTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>		<i>GT</i> GGCC <u>CCACTGTGGGGT</u> GGAGGGGACAGAT	
39 ATCGT TTTTGGCCTCCCTATCA GTCA <i>TT</i> 22,41		GGCCCCCCTGTGGGGTGGAGGGGACAGAT	
ATCGT TTTTGGCCTCCCTATCA GTCA <i>TT</i> 44		1GCCCCACTGTGGGGTGGAGGGGACAGAT	
ATCGT TTTTGGCCTCCCTATCA G3 1	GGA	. TGGCCCCACTGTGGGGTGGAGGGGACAGAT	
ATCGTTTTTGGCCTCCCTATC5 46	GGGGG	GTGGCCCCACTGTGGGGTGGAGGGGACAGAT	
ATCGT TTTTGGCCTCCCTA 7 42	AA	GTGGCCCCACTGTGGGGTGGAGGGGACAGAT	
ATCGT TTTTGGCCTCCCTAT 6 12	TAA	5CACTGTGGGGGTGGAGGGGGACAGAT	
ATCGT TTTTGGCCTCC -10 34	AT	3.CCCACTGTGGGGTGGAGGGGACAGAT	
ATCGT TTTTGGCCTCC 10 3 ATCGT TTTTGGCCT<u>CCC</u>9	G	5 CACTGTGGGGT GGAGGGGACAGAT	
ATCGT TTTTGGCCTCCCTATCA 4	TGGCCA	11 ggggt ggagggacagat	
ATCGT TTTTGGCC -13	G	5CACTGTGGGGTGGAGGGGGACAGAT	
ATCGT TTTTGGC -14 38	т	4CCACTGTGGGGTGGAGGGGGACAGAT	
ATCG22 27	CTATGTCCCTA	<i>GT</i> GGCCC CACTGTGGGGT GGAGGGGACAGAT	
26	GTCGGCGAA	. <i>T</i> GGCCC CACTGTGGGGT GGAGGGGACAGAT	
ATC23 20,40	A	5CACTGTGGGGTGGAGGGGACAGAT	
ATCGTTTTTGGCCTCCCTATCAG3 29		27	
ATCGT <u>T</u> 20 33		11 GGGGT GGAGGGGACAGAT	
45	ccc	9 GTGGGGT GGAGGGGACAGAT	
37	GATGGA	16GGAGGGGACAGAT	

	CA	11GGGGTGGAGGGGGACAGAT
8		
ATCGTT		64
25		
<u>AGT</u> 49		35
24		
ATCGT TTTT<u>GGCC</u>13		148
30		
ATCGTTTTTGGCCTCCCTATCAG3		162
6		
ATCGTTTTTGGCCTCCCTATCAG3		242
48		
257	CCTGG	133
47		

31

ATCGTTTTTGGCCTCCCTATCA...4. +47 ...-3.CCCACTGTGGGGTGGAGGGGACAGAT 47bp: 44bp from Homo sapiens caspase 8, apoptosis-related cysteine peptidase (CASP8), RefSegGene (LRG_34) on chromosome 2 CTGTGCCTGGCCATAATTCATTATCAGTCATTATATAACAC-GTC

13

ATCGTTTT......-18.. +539 ...-2CCCCACTGTGGGGGGGGGGGGGGGGGACAGAT 539bp from ZFN^{MSE} coding vector GCGTTTAAACTTAAGCTATATCCACTGTGCGGTGGGGGGATTCGCCACGGGCTACAAAGACCATGACGGTGATTATAAAAATCA

N114

atcgt <u>ttttggcctccctatca</u> gtca <i>tt</i>		<i>GT</i> GGCC <u>CCACTGTGGGGT</u> GGAGGGGACAGAT
79		
ATCGT TTTTGGCCTCCCTATC 5		5CACTGTGGGGGTGGAGGGGGACAGAT
ATCGTTTTTGGCCTCCCT8.		
65		
ATCGT TTTTGG -15	TGAACAT	6ACTGTGGGGGTGGAGGGGACAGAT
68		
ATCGTTTT <u>TGG</u>		13GGTGGAGGGGACAGAT
ATC23	TGT	5CACTGTGGGGTGGAGGGGACAGAT
72		
	TAGTG	5CACTGTGGGGGTGGAGGGGACAGAT
74		4
87		4 <u>CCACTGTGGGGT</u> GGAGGGGACAGAT
ATCGT TTTTGG 15		23ACAGAT
76		
41		9 <u>GTG</u> GGGTGGAGGGGACAGAT
71 52		10 TCCCCC 2C2C2C2C2C
73		10
AA33		33
92		
<u>AGTA</u> 34		36
ა გულ		-56
77		

GAA51		43
144		4CCACTGTGGGGTGGAGGGGGACAGAT
75		
<u></u>		5 <u>C</u> ACTGTGGGGGTGGAGGGGGACAGAT
90 20	CC 7 7	217
86	CCAA	=21/
	TGCC	102
70		
<u>AGTGAG</u>		279
88		
ATCGT TTTTGG 15	A	361
91		
ATCGTT	+90	24CAGAT
90bp from TALEN expression vector		
CGGCTATGACTGGGCACAACAGACAATCGGCTGC	CTCTGATGCCGCCG	IGTTCCGGCTGTCAGCGCAGGGGGCGCCCGGTTCTTTTG
m(3.)		

TCAA 67

ATCGTTTTTGGCCTCCCTATCAG.-3.. +205 ..-35.....

205bp from Homo sapiens chr13, alternate assembly CHM1 1.1

 $\label{eq:charge} Charge carrier of the state of the st$

משתש

C. Translocation junction sequences from patient-derived cells.

Der19

	nbra	
TCTGTCACCA <u>ATCCTGTCCCTA</u> GTGG <i>CC</i>		<i>GT</i> CATT <u>AATGTAAAGTGGGGAGGC</u> AGCT
38,44		
TCTGTCACCAATCCTGTCCCTAGTGGC.		. TCATTAATGTAAAGTGGGGAGGCAGCT
10		
TCTGTCACCAATCCTGTCCCTAGTG-1.		. TCATT AATGTAAAGTGGGGAGGC AGCT
11,40 mcmcmcacca amccmccccm a cmcccc		
6		IAITAATGTAAAGTGGGGGGGGGGGGCAGCT
TCTGTCACCAATCCTGTCCCTAGT=2		TCATTAATGTAAAGTGGGGAGGCAGCT
22.39.45.50		
TCTGTCACCAATCCTGTCCCTAGT-2		CATTAATGTAAAGTGGGGAGGCAGCT
15		
TCTGTCACCAATCCTGTCCCTAGTGGC.		3TAATGTAAAGTGGGGAGGCAGCT
27		
TCTGTCACCAATCCTGTCCCTAGT-2		1ATT AATGTAAAGTGGGGAGGC AGCT
9		
TCTGTCACCAATCCTGTCCCTAGTG-1.	CCGTC	2TT AATGTAAAGTGGGGAGGC AGCT
1		3.TAATGTAAAGTGGGGAGGCAGCT
TCTGTCACCAATCCTGTCCCTA4		CATT AATGTAAAGTGGGGAGGC AGCT
16		
TCTGTCACCAATCCTGTCCCTAGTGGC.		4AATGTAAAGTGGGGAGGCAGCT
2		
TCTGTCACCAATCCTGTCCCTAGT-2		3.TAATGTAAAGTGGGGAGGCAGCT
47		
TCTGTCACCAATCCTGTCCCTAGT-2		3.TAATGTAAAGTGGGGAGGCAGCT
4 mcmcmcaccaaaamcmccccmaa 4		
TUTGTUACUAATCUTGTCUUTA4		ZTTAATGTAAAGTGGGGAGGCAGCT

38 TCTCTCACCA ATCCTC 10		12 GGG3 GG 3 CCT
66		15
TCTGTCACCA ATCC -12	С	12 TGGGGAGGC AGCT
TCTGTCAC18		8 TAAAGTGGGGAGGC AGCT
TCTGTCACCAATCCTGTCCCTAGTGGC.	A	28
33 TCTGTCACCA16		14
81 TCTGTCACCA AT -14		16 GGAGGC AGCT
88 TCTGTCACCAATC13		27
7		2TT AATGTAAAGTGGGGAGGC AGCT
28 TCTGTCACC <u>A</u> 16		19 GGC AGCT
TC24		11AGTGGGGAGGCAGC
<u></u>		8 <u>T</u> AAAGTGGGGAGGCAGC
86 T25		10AAGTGGGGGGGGGGGCAGCT
23 TCTGTCACC17	с	20 GC AGCT
6 <u>T</u> 25		14
67 TCTGTCACCA ATCCTG -10		29
78 T25		17 GAGGC AGCT
51 		22
46 TCTGT21		167
61 -72		-176
<u></u>		=1/0

Der22

HDFa		
ATCGT <u>TTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>	<i>GT</i> GGCC <u>CCACTGTGGGGT</u> GGAGGGGACAGAT	
13		
ATCGTTTTTGGCCTCCCTATCAGTCAT.	GTGGCCCCCCCTGTGGGGTGGAGGGGACAGAT	
83,7		
ATCGT TTTTGGCCTCCCTATCA GTCATT	GGCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGG	
5		
ATCGTTTTTGGCCTCCCTATCAGTCAT.	GGCCCCACTGTGGGGGTGGAGGGGGACAGAT	
81,1,4		
ATCGTTTTTGGCCTCCCTATCAGT-2	GGCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGG	
	2	
10	3.CCCACTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	
ATCCTTTTCCCCTATCAC -3	-1666 66367666667 6636666636363	
82		
ATCGTTTTTGGCCTCCCTATCAGTC-1.	3.CCCACTGTGGGGGTGGAGGGGGACAGAT	
86		
ATCGTTTTTGGCCTCCCTATCAG3	1GCCCCACTGTGGGGTGGAGGGGGACAGAT	
3		
ATCGTTTTTGGCCTCCCTATCAGTC-1.	4CCACTGTGGGGTGGAGGGGGACAGAT	
63		

17 TCTCTCACCAATCCTCTCCCC 6	
5	CATTAATGTAAAGTGGGGAGGCAGCT
TCTGTCACCAATCCTGTCCCTA4	3.TAATGTAAAGTGGGGAGGCAGCT
14,20	
TCTGTCACCAATCCTGTCCCTA4	5ATGTAAAGTGGGGAGGCAGCT
33	
TCTGTCACCAATCCTGTC8	1ATT AATGTAAAGTGGGGAGGC AGCT
12	
TCTGTCACCAATCCTGTCCCTAGTGG	16 GGAGGC AGCT
18	
TCTGTCACCA ATC -13	5ATGTAAAGTGGGGAGGCAGCT
41	
	3.TAATGTAAAGTGGGGAGGCAGCT
	411BR

*GT*CATT<u>AATGTAAAGTGGGGAGGC</u>AGCT

TCTGTCACCAATCCTGTCCCTAGTGGCC

56		
TCTGTCACCAATCCTGTCCCTAGTGGC.		. TCATT AATGTAAAGTGGGGAGGC AGCT
22,63		
TCTGTCACCAATCCTGTCCCTAGT-2		CATT AATGTAAAGTGGGGAGGC AGCT
65		
TCTGTCACCAATCCTGTCCCTAGTGGC.		2TT AATGTAAAGTGGGGAGGC AGCT
10		
TCTGTCACCAATCCTGTCCCTAGTG-1.	ATA	3.TAATGTAAAGTGGGGAGGCAGCT
84		
TCTGTCACCAATCCTGTCCCTAGT-2		2TT AATGTAAAGTGGGGAGGC AGCT
91		
TCTGTCACCAATCCTGTCCC6		. – 1ATTAATGTAAAGTGGGGAGGCAGCT
89		
TCTGTCACCAATCCTGTCCCTA4		5ATGTAAAGTGGGGAGGCAGCT
27,70		
TCTGTCACCAATCCTGTCCCTAGT-2		9AAAGTGGGGAGGCAGCT
12		
TCTGTCACCAATCCTGTCCCTA4		10AAGTGGGGAGGCAGCT
75		
TCTGTCACCAATCCTGTCCCTAGTG-1.	A	13GGGGAGGCAGCT
42,59		
TCTGTCACCAATCCTGTCCCTAGTGG		15 GGGAGGC AGCT
1,24,47,53,76		
TCTGTCACCAATCCTGTCCCTAGTGG		16 GGAGGC AGCT
13		
TCTGTCACC		GTCATAATGTAAAGTGGGGAGGCAGCT
17		
TCTGTCACCAATCCTGTCCCTA4		12 TGGGGAGGC AGCT
13		
TCTGTCACC		GTCAT AATGTAAAGTGGGGAGGC AGCT
39		
TCTGTCACCAATCCTGT9		8TAAAGTGGGGAGGCAGCT
54		
TCTGTCACCAATCCTGT9	G	11GTGGGGAGGCAGCT
82		
TCTGTCACCAATCCTGTC10		10AAGTGGGGAGGCAGCT
11,73		_
TCTGTCACC <u>AAT</u> 14		7GTAAAGTGGGGAGGCAGCT
50		
TCT <u>GTCA</u>		2TT AATGTAAAGTGGGGAGGC AGCT
32		
TCTGTCACCA ATCCT<u>GT</u>9		13GGGGAGGCAGCT
92		
TCTGTCACCAATCC		- 11 AGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

ATCGTTTTTGGCCTCCCTATCAG3		3.CCCACTGTGGGGTGGAGGGGGACAGAT
62		
ATCGT TTTTGGCCTCCCTATC -5		5CACTGTGGGGTGGAGGGGGACAGAT
69,77		
ATCGT TTTTGGCCTCCCTATCA G <u>T</u> -2		9GTGGGGTGGAGGGGGACAGAT
70	_	-
ATCGT TTTTGGCCTCCCTAT -6	т	3.CCCACTGTGGGGGTGGAGGGGGACAGAT
9		_
ATCGT TTTTGGCCTCCCT<u>A</u>/		· · - / · · · · · CTGTGGGGTGGAGGGGGACAGAT
95		10
ATCGTTTTTGGCCTCCCT8		10TGGGGTGGAGGGGGACAGAT
Amccmmmmmaaccmccc 0		2 СССХСТСТСССССССССССССССССССССССССССССС
61		=J.CCCACIGIGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
ATCGTTTTTGGCCTCCCTATCAG. = 3.		
92		
<u>GC</u> 42		171

..-14.....GTGGAGGGGACAGAT

ATCGTTTTTGGCCTCCCTATCAGTCATT GTGGCCCCACTGTGGGGGTGGAGGGGGACAGAT 46 ATCGTTTTTGGCCTCCCTATCAGTC-1. 8 ATCGTTTTTGGCCTCCCTATCAG.-3.. 49 ATCGT**TTTTGGCCTCCCTA**....-7... 5,34 ATCGT**TTTTGGCCTCCCT**....-8.. 10,26,71 ATCGT**TTTTGGCCTCCCTATC...**-5... ..-4..CCACTGTGGGGGTGGAGGGGGACAGAT 29 ATCGTTTTTGGCCTCCCTA....-7.. CC 4 ATCGTTTTTGGCCTCC.....-10.. A 32 ATCGT**TTTTGGCCTCCCTA**....-7.. G ..-3.CCCACTGTGGGGTGGAGGGGGACAGAT 54 ATCGTTTTTGGCCTCCC.....9.. AG ..-3.CCCACTGTGGGGGTGGAGGGGGACAGAT 60 ..-3.CCCACTGTGGGGGTGGAGGGGGACAGAT ATCGTTTTTGGCCTCCC.....9.. CCT 20 ATCGT**TTTTGGCCTCCCTA**....-7.. GTGC ..-7....CTGTGGGGTGGAGGGGGACAGAT 48 ATCGT**TTTTGGCCTCC**....-10.. ..-6....ACTGTGGGGTGGAGGGGGACAGAT 57 ATCGTTTTTGGCCTCCCTATCAGT-2.. ..-16.....GGAGGGGACAGAT 58 ATCGT**TTTTGGCCT**....-12. ..-8.....**TGTGGGGT**GGAGGGGACAGAT 41 ATCG.....-21.. A 70 ATCGT**TTTTGGC<u>CT</u>.....-12**. ..-9.....GTGGGGTGGAGGGGACAGAT 1 ATCGTTTTTGGCCTCCCTAT....-6.. А ..-16.....GGAGGGGACAGAT 17,52,62 ..-11.....GGGGTGGAGGGGACAGAT ATCGTTTTTGGCCT.....-12.. 3 ATCGT**TTTTGGCCTCCC.....9.** ..-15.....**T**GGAGGGGACAGAT 61

ATCGTTTTTGGCCTCC.....-10..

ATCGT TTTTGG 15		10 TGGGGT GGAGGGGACAGAT
6 / ATCGT TTTTGGCCTCCC9		16GGAGGGGACAGAT
51,53		10
27		19GGGGACAGAT
ATCGTTTTTGGCCTCCC9		21GGGGACAGAT
ATCGT TTTTGGCC<u>T</u>12		18GGAGGGGACAGAT
25 ATCGT TTTTGGCCT12		20AGGGGACAGAT
47		
68		21GGGGACAGAT
ATCGT TTTTGGCCT12	A	20AGGGGACAGAT
ATCGTTTTTGGCCT12		25AGAT
69 ATCGT TTTTGGCCTCCC9	c	31
65		
36		8 <u>TG</u> TGGGGTGGAGGGGGACAGAT
ATCGT21		31
ATC <u>G</u> 21		47
55 A27		48
50		
12		11GGGGTGGAGGGGGACAGAT
$\frac{G}{10}$		41
ATCGT21		68
37		45
11	_	
45	A	44
ATCGT TTTTG<u>G</u>15		358

15

411BR

Figure S4, related to Figure 4. DSBs induced by TALENs, wild-type Cas9, and paired nCas.

A. <u>Relative positions of nCas9 cleavage sites.</u> nCas9 gives rise to DSBs with 5' overhangs that are 41 bp (NPM1+NPM2) and 37 bp (ALK1+ALK2) apart. PAM sequences are underlined; bps in italics represent overhangs.

nCas9 : Sequence after nicking both strands:

Wild-type Cas9 gives rise to DSBs with blunt ends or short overhangs. The TAL^{ALK} and TAL^{NPM} cleavage sites are between the positions of the two nicks and have short 5' overhangs (Piganeau et al., 2013).

B. Indel formation at the ALK locus after a TAL^{ALK}, Cas9(ALK1) or nCas9(ALK1+ALK2) DSB. T7 endonuclease assay.

Number of experiments : n = 4 for TAL wt, n = 8 for het and mutant TAL (4 each for X4 and L4); n = 3 for Cas9 ; n = 4 for nCas9.



Figure 55, related to Figure 5. NPM-ALK cancer translocation junctions induced by TALENs, wild-type Cas9, and paired nCas.

A. <u>Nuclease expression is quantified after 48 h by Western blotting for the HA epitope for</u> TALEN or GFP expression for Cas9 and nCas9. The same blot was used to probe for both HA and tubulin.



<u>B. Translocation junction sequences from TALENs.</u> The TALEN recognition sequences at each DNA end are underlined; bps in italics represent the overhangs. The chromosome 5 end is in black and the chromosome 2 end is in red. Microhomologies (underlined), insertions (green) and lengths of deletions from each end are indicated.

Der5

HCT116 X4+/-

DNA ends: GC <u>TATATCCTCGAACTGCTA</u> CTGGGT <i>TCA(CCT C</i>		<i>a) atct</i> gatcacg <u>gtcggtccattgcataga</u> g
65		
GCTATATCCTCGAACTGCTACTGGGT <i>TCA</i>	AC	TCTGATCACGGTCGGTCCATTGCATAGAG
30 CCTATATCCTCCAACTCCTACTCCCTACTCA		GATCACCCTCCATTCCATACAC
57		
GCTATATCCTCGAACTGCTACTGGGTTC	с	TCTGATCACGGTCGGTCCATTGCATAGAG
60		
GCTATATCCTCGAACTGCTACTGGGTT	AG	.ATCTGATCACGGTCGGTCCATTGCATAGAG
55		
GCTATATCCTCGAACTGCTACTGGGTT	GGA	GATCACGGTCGGTCCATTGCATAGAG
61		
GCTATATCCTCGAACTGCTACTGGGT <i>TCA</i>	+40	AATCTGATCACGGTCGGTCCATTGCATAGAG
40bp: from chr11 (Homo sapiens genomi	ic DNA, chro	omosome 11q, clone:CMB9-3I4, complete
sequence)		
TTCCAGCGATCCTCCTACCTCGGCTTCCCAAAATGC1	rgag	
GCTATATCCTCGAACTGCTACTGGG=1	CTA	8TCGGTCCATTGCATAGAG
GCTATATCCTCCAAC -11	CCT	
63	cci	=INICACOGICOGICCATIGCATAGAG
GCTATATCCTCGAACTGCTA6	+13	11GTCCATTGCATAGAG
13bp: ACGCCAGCAACGC		
66		
GCTATATC		6GGTCGGTCCATTGCATAGAG
70		
GCTATATCCTCGAACTGCTACTGGGTT		25G
67		
27	A	TGATCACGGTCGGTCCATTGCATAGAG
71		

GCTATATCCTCGAACTGCTACTGGGTTCA		27
59		
GCTATA		25G
64		
		11GTCCATTGCATAGAG
53		
G41		49
52		
G43		85
72		
	TGTT	83
69		
	+14	GATCACGGTCGGTCCATTGCATAGAG
14bn: insertion of telomeric sequence:		
TTACCC_TTACCC_TA		
(Sequence unstream of insertion TACACA	ACCCA	
(Sequence upscream of insercion indada	Addd)	
50		
С		306

HCT116 L4+/-

GCTATATCCTCGAACTGCTACTGGGTTCA(CCT C	:	A)ATCTGATCACG <u>GTCGGTCCATTGCATAGA</u> G
27		
GCTATATCCTCGAACTGCTACTGGGTTCACCT		GATCACGGTCGGTCCATTGCATAGAG
44		
GCTATATCCTCGAACTGCTACTGGGT <i>TCA</i>		.ATCTGATCACGGTCGGTCCATTGCATAGAG
	TA A TA	1 8 9 6 8 6 6 6 9 6 6 6 9 6 6 6 6 6 6 6 6
26 32	IAAIA	=INICACOGICOGICCATIGCATAGAG
GCTATATCCTCGAACTGCTACTGGGT		2TCACGGTCGGTCCATTGCATAGAG
39		
GCTATATCCTCGAACTGCTA6	TGA	TGATCACGGTCGGTCCATTGCATAGAG
33		
GCTATATCCTCGAACTGCTA6		1ATCACGGTCGGTCCATTGCATAGAG
35		-
GCTATATCCTCGAACTGCTACTGGGTTC		/GTCGGTCCATTGCATAGAG
47 GCTATATCCTCGAACTGCTACTGG_2		-8 ФСССФССАФФССАФАСАС
48		=0ICGGICCAIIGCAIAGAG
GCTATATCCTCGA13		TGATCACGGTCGGTCCATTGCATAGAG
36		
GCTATATCCTCGAACTGCTACTGGGTTCACC.	CACCTC	19CATAGAG
25		
GCTATA <u>TCC</u>		15ATTGCATAGAG
31	_	
GCTATATCCTCG14	т	19CATAGAG
-24		-16
37		<u>1100</u> ATAGAG
GCTATAT		43
42		
Ст		54
34		
	CCC	70
38		
GCTATATCUTUGAACTGCTACTGGGTTCACCTC	AGCCTCTGAA	=183
41 CCTATATATCCTCCAACTCCTACTCCCTTCACC	+116	CARCACCCRCCCRCCARRCARACAC
116bp from TALEN expression vector	+110	GAICACGGICGGICCAIIGCAIAGAG
TIOND TIOW THINK CAPICOBION VECTOR		

CARTGGARAGTCCCTATTGGGGTACGTCATACGTCATTGACGTCAATGGGCGGGGGCGTCGTTGGGCGGTCAGCCAGG CGGGCCATTTACCGTAAGTTATGTAACGCG

HCT116 X4-/-

GC <u>TATATCCTCGAACTGCTA</u> CTGGGT <i>TCA(CCT</i>)	с	A) ATCTGATCACGGTCGGTCCATTGCATAGAG
0		
GCTATATCCTCGAACTGCTAC5	AG	4ACGGTCGGTCCATTGCATAGAG
9		
GCTATATCCTCGAACTGCTACTG3	С	9CGGTCCATTGCATAGAG
GCTATATCCTCGAACTGCTACTGGGTTCAC	AC	-43
6		
<u></u>		19
<u></u>		5 <u>CGGT</u> CGGTCCATTGCATAGAG
7		
<u>AG</u> 43		45
- 		
<u>TG</u> 41		65
56	AGT	
GCTATATCCTCGAA12 3	ATCCACTATAT	106
<u></u>		<u>GATC</u> ACGGTCGGTCCATTGCATAGAG
<u></u>		····-2 <u>TCA</u> CGGTCGGTCCATTGCATAGAG
GCTATATCCTCGAACT10		254
	+28	
8bp: TAGTAGC-TTTACACACAGTTTTCTGTA		
0bp from Homo sapiens BAC clone RP: TTACACACAGTTTTCTGTA	11-402C9 from 4	4, complete sequence
8		
<u>CC</u> 94		

HCT116 L4-/-

GCTATATCCTCGAACTGCTACTGGGTTCA(CCT C	A) ATCTGATCACGGTCGGTCCATTGCATAGAG
81 GCTATATCCTCGAACTGCTACTGGGT <u>TCAC</u>	6GGTCGGTCCATTGCATAGAG
56 GCTATATATCCTCGAACTGCTAC5	4ACGGTCGGTCCATTGCATAGAG
49,51 GCTATATCCTCGAACTGCTACTG <u>GGT</u> 70	9CGGTCCATTGCATAGAG
GCTATATCCTCGAACTGCTACTG3 CTACT 5 bp duplication from upstream of DSB on chr5	6GGTCGGTCCATTGCATAGAG
62 GCTATATCCTCGA <u>AC</u> 11	6GGTCGGTCCATTGCATAGAG

76		
GCT23 71	AA	••••••••••••••••••••••••••••••••••••••
GCTATATCC <u>TCG</u> 14		11GTCCATTGCATAGAG
GCTATATCCTCGAACTGC <u>TA</u> 6		43
GCTATATCCTCGAACTGCTACT <u>GG</u> -160		
36 58	TAA	15ATTGCATAGAG
<u></u> 58 54		4 <u>ACGG</u> TCGGTCCATTGCATAGAG
AGC28		39
	GTCA	54
<u>G</u> 68		50
<u>GGCTGG</u>		
<u>TGCC</u>		61
		····-2 <u>TC</u> ACGGTCGGTCCATTGCATAGAG
G128		48
		150
67 <u>GG</u> 94		107
<u>59</u> <u>CCTGGCC</u> 163		68
64 <u>CTG</u> 261		35
53 GCTATATCCTC15		
85100	A	309
77 GCT126		373
80 G36		468
79	GTG	
84 GCTATATCCTC	+10	-506
10bp insertion AAATCCACTA		
61		
	+19 clone RP11-3	66F6 map q28
73 GCT23	+69+2	35AG
Insertion 71bp		
695p insertion from 16 bp downstream of AGAACTACAGGCACACTCCACGCCTGGCTAATTTT	DSB on chr5 ITTTGTATATGI	GCAGATGGGGTCTCAGTATGTTCT
GCT	1+49+1	-37
Insertion 51bp		
49 bp insertion inverted from 37 bp down	nstream of DS	B on chr5
CCAACATACTGAGACCCCATCTGCACATATACAAAA	AAAAATTAGCCA	AGGCT
/ʊ	+516	18GCATAGAG

<u>C. Translocation junction sequences from wild-type Cas9.</u> The PAM sequence is underlined; bp in italics represents a possible overhang.

Der5

HCT116 L4+/-

DNA ends: GCAGTGATGTGATCATAGCTTGCTATAT <u>CCT</u> CGAA	GTAACCCTAATCTGATCACGGTCGGTCCATT
1,10,12,14,17,23,38,39,41,45,48 GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA	GTAACCCTAATCTGATCACGGTCGGTCCATT
42	
GCAGTGATGTGATCATAGCTTGCTATATCCTCGA.	GTAACCCTAATCTGATCACGGTCGGTCCATT
2,3,9	
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA	-1TAACCCTAATCTGATCACGGTCGGTCCATT
37,43	
GCAGTGATGTGATCATAGCTTGCTATATCCTCGA.	-2.AACCCTAATCTGATCACGGTCGGTCCATT
31	
GCAGTGATGTGATCATAGCTTGCTATATCCTCG <u>AA</u>	-4CCCTAATCTGATCACGGTCGGTCCATT
7	
GCAGTGATGTGATCATAGCTTG12.	GTAACCCTAATCTGATCACGGTCGGTCCATT
24	
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA	-14ATCACGGTCGGTCCATT
25	
GCAGTGATGTGATCATAG	GTAACCCTAATCTGATCACGGTCGGTCCATT
44	
GCAGTGATGTGATCATAGCTTGCTATATC5.	A -11CTGATCACGGTCGGTCCATT
8	
GCAGTGATGTGATCATAG	-1TAACCCTAATCTGATCACGGTCGGTCCATT
20	
GCAGTGATG <u>TGATCA</u>	-18CGGTCGGTCCATT
28	
GCAGTGATGTG <u>ATC</u>	-22GGTCCATT
47	
<u></u>	-20 <u>T</u> CGGTCCATT
32	
<u>GC</u> 151.	-215
11 240 20	107 106
145bp incertion with 107bp inverted dupliced	-120 DCD
AGGTAGAAGGCTGGAG-	LION FIOM CHES SOUP UPSTREAM DSB

 ${\tt tccagccgttataatgagactgtctttattaaaaataatttttaggccaggcacagtggcacacgtgataatactagcactttg agaggccaggcgtgcagatcacgttaaggccaggtcaag$

HCT116 L4-/-

DNA ends: GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA	GTAACCCTAATCTGATCACGGTCGGTCCATT
5 GCAGTGATG <u>TGATCA</u> 19.	-18CGGTCGGTCCATT
GCAGTG <u>AT</u> 26.	-11CTGATCACGGTCGGTCCATT

GCAGTGA <u>TG</u>	-14ATCACGGTCGGTCCATT
45	
4145.	-17 <u>ACGG</u> TCGGTCCATT
<u>GC</u> 32.	-80
<u>A</u> 81.	-36
TAG111.	-37
30	-5 <u>CC</u> TAATCTGATCACGGTCGGTCCATT
40	-5 <u>CC</u> TAATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATAT6. +74	-195
27bp inverted sequence on chr5 51bp downstream of D	SB
GACCGGAATGGCTGGGCAACATACTGAGACCCCATCTGCACATATACAAAA 7	AAAATTAGCCAGGCGTGGAGTGT
GCAGTGATGTGATCATAGCTTGCTATATC5. AC	-220
2 TCT 122.	-117
GCAGTG28. 42	-253
<u>GCAC</u> 231.	-54
<u>c</u> 81. 24	-230
<u>TT</u> 186.	-135
<u>GC</u> 279. 29	-80
57. 36	-317
GCAGTGATG25. +27 27bp inverted duplication on chr2 6bp downstream of	-369 DSB
CAATGGACCGACCGTGATCAGATTAG 21	
GC81.	-318
G <u>CAG</u> 30. 48	-419
<u>TC</u> 304.	-148
8 8 <u>TCACT</u>	-201
<u>GTT</u> 217. 46	-296
17bp: CTGCCCTTCTTTAAAAT	-331
	-449

TCCTCTATGCAATG

D. Translocation junction sequences from nCas9.

Der5

HCT116 L4+/-

DNA ends: CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAGAAC	GTAACCCTAATCTGATCACGGTCGGTCCATTGCATAG AGGAG
51 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCT5	-10TCTGATCACGGTCGGTCCATTGCATAGAGGAG
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAG <u>CT</u> 5	-13GATCACGGTCGGTCCATTGCATAGAGGAG
76 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTG13	-8AATCTGATCACGGTCGGTCCATTGCATAGAGGAG
21 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAGAAC	-24GTCCATTGCATACAGGAG
75 CTCGA ACTGCTAC33	-4CCCTAATCTGATCACGGTCGGTCCATTGCATAGAGGAG
9 CTCGA ACTGC36	-2AACCCTAATCTGATCACGGTCGGTCCATTGCATACAGGAG
25 CTCGAACTGCTACTGGGTTCACCTCAGCCTC15	-26CCATTGCATACAGGAG
31 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATA8	-26CCATTGCATAGAGGAG
14,43 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAG7	-37AGGAG
34 CTCGAACTGCTACTGGGTTCACCTCAG <u>CC</u> 17	-28ATTGCATAGAGGAG
3 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTA <u>GAAC</u>	-4710
45 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTA4+31	-436
31bp from chr2, from 5bp downstream of DSB CCTAATCTGATCACGGTCGGTCCATTGCATA (followed by a 6 b	p deletion)
CTCGAACTGCTACTGGGT28 -7	-20TCGGTCCATTGCATACAGGAG
6 CTCGAACTGCTACTGGGTTCACCTC <u>AG</u> 19	-37AGGAG
CTCGAACTGCTACTGGGTTC26	-31GCATAGAGGAG
7 bp: CGGTCCA 68	
CTCGAACTGCTACTGGGTTCACCTC	-392GAG
CTCGAACTGCTACTGGG <u>TT</u>	-31GCATAGAGGAG
CTCGAACTGCTACTGGGTTCACCTCA20 +9 TCAGCCTCT 60	-381GGAG
CTCGAACTGCTAC	-29ATTGCATAG AGGAG
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGG12+18 GTTCACCTCAGCCTCTGG	-4912
c4 45	-20GTCGGTCCATTGCATAGAGGAG
⁴⁹ <u>crc</u> -243	-27CATTGCATAGAGGAG
CTCGAACTGCTACT <u>GG</u>	-403.AG
20 CTCGAACTGCT <u>A</u>	-41 -4 G
CTCGA A40	-37AGGAG
CTCGAACTGCTACT <u>GGGT</u> 28	-6326
CTCGAACTGCTACTGGGTTCACCTCAGC18 +8 TGCTACTG 2	-7740
CTCGA <i>ACTGCTA<u>C</u></i>	-7033
	-7033
CTCGAACTG	-7639

30		
5	-5821	
13		
CATGC-1556	-5922	
53		
CTCGAACTGCTACTGGG29	-10063	
39		
<u>GTGA</u> -2667	-6831	
62		
TTG-1253	-9053	
70		
CTCGAACTGCTACTGGGT28	-11881	
52		
G1657	-9154	
46		
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAG7 T	-155118.	
18		
TCA-56	-6932	
11		
CTCGAACTGCTACTGGG29	-147117.	
55		
CA1960	-122	
2		
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAG7	-184147.	
41		
GCT-1051	-146109.	
10		
A75116	-13093	
15		
G1253	-212175.	
19		
CTCGA ACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAGAAC	-292255.	
5		
CTCGAACTGCTA	-264227.	
20		
277318	-31GCATAGAGGAG	
42	—	
CTCA-120161	-203166.	
5		
GCT-113154	-216179.	
50		
CTCGAACTGC	5-425	
165bp from Homo sapiens chr17, clone RP11-670E13,	complete sequence	
TGCAGCAGGCCGGAAAGGCCCCTGTCCAAGGCCAAAACATCTTACCATTG	TATTGCTGCTGCAGACAGGAAGGGGTGTATGCCTGTAGAAGTGCTGA	
GGATATGTAAAGGGAAGCCAGGGCCTGCGGTGTTGACACGGGGCAAGTCA	ACTCACATTAGAAATGAACCA	
29		
31	5-36	
205bp from Homo sapiens chr3, alternate assembly HuRef		

rge TCTCATCCAATTT 27

2' .-51 .-92+115-174. -137. +115bp from pCASD10A vector CGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGC AGTGCTGCCATAACCATG

HCT116 L4-/-

CTCGA ACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAGAAC	STAACCCTAATCTGATCACGGTCGGTCCATTGCATAG AGGAG
27,42	
CTCGAACTGCTACTG	-13ATCACGGTCGGTCCATTGCATAGAGGAG
30	
CTCGAACTG	-4912
32	
CTCGAACT	-375338.
22	
CTCGAACTGCTACTGGGTTCAC	-7134

23	
CT344	-7134
36	
<u>CATAGC</u> -1556	-5922
25	
C4	-494457.
28	
<u>G</u> 1/621/	-44/
40	-500 -463
21	-300403.
CT-10	-4811
41	
ATC-2061	-11679
39	
<u>CT</u> -3980	-272235.
14	
<u>CTGA</u> -2263	-5023
24	224
<u>C</u> 32/3	-324287.
G -31 -74	-400 -363
4	-400303.
ATTTT-68109	-137100.
19	
GCCCTGGCC-125166	-8144
6	
TGCC-86127	-7538
10	
166207	-174137. 2
<u>T</u> -212253	-375338.
-145 -187+105	-53 -16
106bp inverted insertion from chr5 117bp downstream	of DSB
TCTCACTTTGGGGAGGCCAAGGCGGACCACTTGAGGCCAAGAGTCCAATCT	GGCAACATACTGAGACCCCATCTGCACATATACAAAAAAAA
CCAGGCGTGTT	
12	

37

т

29

AGTCTATGCAATGCAAATGAGTCTATG

Figure S6, related to Figure 6. LIG3 deficiency is not required for translocation formation in human cells.

A. <u>LIG3 null cells: Western blot, PCR genotyping, ZFN expression, and indel formation. L3^{-/-} HCT116 cells are viable due to MtLIG1 expression, as seen in mouse cells (Simsek et al., 2011a).</u>

In the westerns, the same blot was used to probe for both epitopes.





Indel formation is not affected by LIG3 loss, as monitored by the T7endonuclease assay in L3^{+/-} and L3^{-/-} cells expressing MtLIG1 (abbreviated L3^{-/-}). Two loci were examined, p84/AAVS1 using ZFNP⁸⁴ and ALK using either wildtype Cas9 with gRNA ALK1 or nCas9 with gRNAs ALK1+ALK2. NT, not transfected (i.e., no nuclease).

B. Translocation junction sequences from ZFNs. The ZFN recognition sequences at each DNA end are underlined; bps in italics represent the overhangs. The chromosome 19 end is in black and the chromosome 22 end is in red. Microhomologies (underlined), insertions (green) and lengths of deletions from each end are indicated.

Der22

HCT116 Lig3+/-

ATCGT <u>TTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>	<i>GT</i> GGCC <u>CCACTGTGGGGT</u> GGAGGGGACAGAT
6,47,52,69,75	
ATCGT TTTTGGCCTCCCTATCA GTCAT <u>T</u>	GGCCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGG
2,29,36,45,54,68,73	
ATCGT TTTTGGCCTCCCTATCA GTCA <u>T</u> .	GGCCCCCCCTGTGGGGTGGAGGGGACAGAT
62	
ATCGTTTTTGGCCTCCCTATCAGTCA	GGCCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGG
20	
ATCGTTTTTGGCCTCCCTATCAGTCAT.	1GCCCCACTGTGGGGTGGAGGGGGACAGAT
64	
ATCGTTTTTGGCCTCCCTATCAGTCA	. – 1GCCCCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGG
11	

ATCGTTTTTGGCCTCCCTATCAG3	+94	8 TGTGGGGT GGAGGGGACAGAT
AGGTGATGGAGTTCTTCATGAAGGTGTACGGC 19	TACAGGGGAAAGCAG	CCTGGGCGGAAGCAGAAAGCCTGACGGCGCCA
78bp from ZFN plasmid	+78	. TGGCCCCCCTGTGGGGGTGGAGGGGACAGAT
<u>GAG</u> 36 32		158
<u>TA</u> 205 37		36
108 3		115
77 <u>A</u> 100		110
61	AG	60
22 <u>CA</u> 41		52
66 GTA45.	-	36
27	т	
A28 28		33
<u>A</u> 28 21		24
ATCGT TT 18 44		16GGAGGGGACAGAT
ATCGTTTTTG16 17 bp : ACCCCCGGCATCCCCAC 4	+17	ZUUCCACTGTGGGGTGGAGGGGGACAGAT
ATCGTTTTTGGCCTCCCTA7		11GGGGTGGAGGGGACAGAT
ATCGT TTTTGG<u>CC</u>13 8		6 ACTGTGGGGT GGAGGGGACAGAT
ATCGT TTTTGGCC 13 24		3.CCCACTGTGGGGTGGAGGGGGACAGAT
ATCGT TTTTGGC -14 15		.TGGCCCCACTGTGGGGTGGAGGGGGACAGAT
ATCGT TTTTGGCCTCCCT -8 33		. TGGCCCCCCCTGTGGGGGTGGAGGGGGACAGAT
ATCGTTTTTGGCCTCCCTATC5		. TGGCCCCACTGTGGGGGTGGAGGGGGACAGAT
ATCGT TTTTGGCCTCCCTATCA GTCAT.		5 CACTGTGGGGT GGAGGGGACAGAT
ATCGT TTTTGGCCTCCCTATCA GTCATT	CAT	4CCACTGTGGGGTGGAGGGGGACAGAT
ATCGTTTTTGGCCTCCCTATCAGTC-1.		. TGGCCCCACTGTGGGGTGGAGGGGACAGAT

94bp from Homo sapiens chromosome 10, alternate assembly HuRef GGACAATGCCTATGGAGAGATGGGAGAGAAATCAGAAAGAGGGCCCCTGGAAGACTGTAAGGGCCGGGACAGCCTGCAA CCTGGACATTATTT

HCT116 Lig3-/-

ATCGT <u>TTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>	<i>GT</i> GGCC <u>CCACTGTGGGGT</u> GGAGGGGACAGAT
3,8,37,39,43 ATCGT TTTTGGCCTCCCTATCA GTCA <u>T</u> .	GGCC CCACTGTGGGGT GGAGGGGACAGAT
9,31 ATCGT TTTTGGCCTCCCTATCA <u>GT</u> -2	GGCCCCACTGTGGGGTGGAGGGGACAGAT

48 ATCGT TTTTGGCCTCCCTATCA GTCA		1GCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
5		
ATCGTTTTTGGCCTCCCTATCAGT-2	G	GTGGCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGG
23		
4	10	ZCCCCCCTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
ATCGT TTTTGGCCTCCCTATCA 4		. TGGCCCCCCCTGTGGGGGTGGAGGGGGACAGAT
45		
ATCGTTTTTGGCCTCCCTATCAG3		1GCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
14		MCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
10		. IGGCCCCCCTGTGGGGGGGGGGGGGGGGGGACAGAT
ATCGTTTTTGGCCTCCCTATCAG3		3.CCCACTGTGGGGTGGAGGGGGACAGAT
30		
ATCGT TTTTGGCCTCCCTAT 6		GGCCCCCCTGTGGGGGTGGAGGGGGACAGAT
20		
ATCGTTTTTGGCCTCCCTATCAGTCAT.		5CACTGTGGGGGTGGAGGGGGACAGAT
ATCGT TTTTGGCCTCCCTATCA GTCAT.		9GTGGGGTGGAGGGGGACAGAT
11		
ATCGTTTTTGGCCTCCCTA7		2CCCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGG
18		
ATCGT TTTTGGCCTCCCT -8		. TGGCCCCACTGTGGGGTGGAGGGGGACAGAT
1/ A TOCT THE ACTION OF A CTOR		9 GTEGGGT CCACCCACACAT
12		
ATCGT TTTTGGCC -13		4CCACTGTGGGGTGGAGGGGGACAGAT
15		
		23 <u>ACA</u> GAT
36		27
AIC		=57
TGGG		69
6		
		19 <u>G</u> GGGACAGAT
38		
42		. TGGCCCCACTGTGGGGGTGGAGGGGACAGAT
42 -68		TGGCCCCACTGTGGGGGTGGAGGGGACAGAT
19		
164		29
29		
96		83
/ C 172		176
26		1/0
CAT199		141
16		
	+171	35

C. Junction sequences from Cas9. The PAM sequence is underlined; bp in italics represents a possible overhang.

Der5

<u>HCT116 Lig3+/-</u>

DNA ends: GCAGTGATGTGATCATAGCTTGCTATAT <u>CCT</u> CGAA		GTAACCCTAATCTGATCACGGTCGGTCCATT
1,9,12,14,19,31,32,35,47 GCAGGATGATGATCATAGCTTGCTATATCCTCGAA		GTAACCCTAATCTGATCACGGTCGGTCCATT
7,17,28,38,40 GCAGTGATGTGATCATAGCTTGCTATATCCTCGA.		GTAACCCTAATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 36	с	-1TAACCCTAATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 45		-8AATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 5		-19GGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCT3. 42		GTAACCCTAATCTGATCACGGTCGGTCCATT
$26 \\ \overline{C}$ -32		-47
39 TGTG		-87
6 C94.		-59
43 <u>AAA</u>		-354
41 GCAGTGATGTGATCATAGCTTGCTATATC <u>CT</u> 3.		-329
18 279.		-13GATCACGGTCGGTCCATT

HCT116 Lig3-/-

DNA ends: GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA		GTAACCCTAATCTGATCACGGTCGGTCCATT
12,15,20,22,24,32,37,39 GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 18		GTAACCCTAATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 16,28	с	GTAACCCTAATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 11	CT	GTAACCCTAATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 26		-1TAACCCTAATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 38		-3ACCCTAATCTGATCACGGTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTC <u>G</u> -1 29		-1TAACCCTAATCTGATCACGGTCGGTCCATT
GCAGTGAT <u>G</u> 25. 23		-20GTCGGTCCATT
GCAGTGATGTGATCATAGCTTGCTATATCCTCGAA 15		-138
106. 35		-21TCGGTCCATT
63. 25		-42
107.		-74
234.		GTAACCCTAATCTGATCACGGTCGGTCCATT
<u>A</u> 56.		-113

CAG55.	-337
31	
G <u>CAG</u> 32.	-419

D. Junction sequences from nCas9. The bps in italics represent overhangs.

Der5

HCT116 Lig3+/-

DNA ends: ctcgaactgctactgggttcacctcagcctctggaatagctagaac	GTAACCCTAATCTGATCACGGTCGGTCCATTGCATAG AGGAG
30 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAGAAAC	-28ATTGCATAGAGGAG
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAGAAC GT	-33ATAGAGGAG
34 CTCGAACTGCTACT	-1TAACCCTAATCTGATCACGGTCGGTCCATTGCATAG AGGAG
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGC6	-41 -4 G
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGA11	-36 G AGGAG
CTCGAACTGCTACTGGGTTCACC23 G	-26CCATTGCATAGAGGAG
CTCGAACTGCTACTGGGTTCACCTCAGC18	-33
CTCGAACTGCTACTGGG29	-22CGGTCCATTGCATAGAGGAG
CTCGAACTGCT	-33
CTCG-141	-27CATTGCATAGAGGAG
22. CTCGAAC39	-32CATAGAGGAG
35 CTCGA41	-34
18 CTCGAACTGCTACTG	-458
36 CTCGAACTGCT <u>ACTG</u> 31	-4912
5 CTCGAACTGCTACTGGGTTCACCTCAG <u>CCT</u> 16	-7740
27 1859	–37 <u>AG</u> GAG
32 <u>A</u> 9	-5013
31 <u>CATAGC</u> -1556	-5922
38 <u>CAT</u> -1859	-8447
6113154+26 26bp : CGGCACTTTACAACCTAAATGCCCATCAGT	-19GGTCGGTCCATTGCATAGAGAG
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGG12GTT	c -168 131 .
	-9861
<u>GA</u> -3677	-11376
	-35
11 CTCGAACTGCTACTG <u>GG</u> 29	-185148.
A-112153	-150113.
8 207248	-6427
21 TCGAACTGCTAC (TGGGTTCACCTCAGCCTCTGGAATAGCTAGA-2+304	GTAACCCTAATCTGATCACGGTCGGtCCATTGCATA) GAGGAG

HCT116 Lig3-/-

CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAGAAC	GTAACCCTAATCTGATCACGGTCGGTCCATTGCATAG AGGAG
35	
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAGAAC	-32CATAGAGGAG
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAATAGCTAG. 3- T 26	-34TAGAGGAG
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAA <u>TAG</u> 7	-37AGGAG
CTCGA ACTGCTACTGGGTTCACCTCAG <u>CC</u> 17 46	-28ATTGCATAGAGAG
C445+24 24bp repeat from chr2 directly downstream of DSB CCTAATCTGATCACGGTCGGTCCA	-1TAACCCTAATCTGATCACGGTCGGTCCATTGCATAG AGGAG
t1 CTCGAACTGCTACTGGGTTCA25	-23GGTCCATTGCATAGAGGAG
* CTCGA ACTGCTACTGGGTTCACCTCAGCCTCTGGAATA <u>G</u> 7 22	-447
CTCGAACTGCTACTGGGT <u>TC</u> 26	-27CATTGCATAG AGGAG
CTCGAACTGCTACTGGG29	-24GTCCATTGCATAGAGGAG
CTCGAACTGCTACT <u>GG</u> 30	-25TCCATTGCATAGAGAGAG
CTCGAACTGCTACTGGGTTCACCTCAGCCTCT14+24	-438
CTCGA41 C	-21TCGGTCCATTGCATAGAGGAG
CTCGA ACTG37	-33ATAGAGAGAG
CTCGAAC <u>TGC</u> 36	-33ATAGAGAG
CTCGA ACTGCTACT <u>GG</u> 30	-403. A G
CTCGA ACTGC36	-6124
CTCGAACT38 T	-6932
CTCGAACTGCTACTGGGTTCACCTCAGCCTCT <u>GG</u> 12	-9962
CTCGAACTGCTACTGGGTTCACCTCAGCCT <u>CTGG</u> 12	-9962
	$GTAACCCTAATCTGATCACGGTCGGTCCATTGCATAG {\bf AGGAG} {\bf G} {\bf G$
<u>CAT</u> -1859	-8447
$\frac{CA-19}{4}$	-10770
<u>CAC</u> -92133	-5417
C441	-159122.
<u>T</u>-64 104	-147110.
CTCGA ACTGCTAC <u>T</u> 32	-220183.
S CAC-103144	-12386
GTGGC-82123	-146108.
<u>TG</u> 90131	-159122.

T73	7	
38 A-141 -182 -169 -132 187 -228 -90 -53 33 -228 -90 -53 CTCGAA -40 -309 -272 36 -272 -36 CTCGAACTCGCTACTCGGGTCACCTCAGCCTCTGGAA -10 +80-33 ATACAGGAG 80bp mixed repeats (chr5 and chr2) and insertions ATACAGGACCTCAGCCTCTGGAA-CACCGGTCGTCACCTCAGCCTCTGGCAACCACC CTGGGTTCACCTCAGAC-CACGCGTCCATGC-TGGCTAACCTCAGCCTCTGGCAACCACC -10 +80-33	T73114	-180143.
A-141	38	
28 187	A-141182	-169132.
	28	
33 CTCGA	187228	-9053
CTCGAA	33	
36 CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAA10 +80-33	CTCGAA40	-309272.
CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAA10 +80-33ATAGAGAG 80bp mixed repeats (chr5 and chr2) and insertions CTGGGTTCACCTCTAGCCTCTGGAA-CACGGTCCATGC-TGGCTTACCTCAGCCCTGGGAA-CCACGGTCGG-AACCAC	36	
80bp mixed repeats (chr5 and chr2) and insertions CTGGGTTCACCTCAGCCTCTGGAA-CACGGTCGGTCCATTGC-TGGGTTCACCTCAGCCTCTGGAA-CCACGGTCGG-AACCAC	CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAA10 +	80-33ATAGAGGAG
CTGGGTTCACCTCAGCCTCTGGAA-CACGGTCGGTCGATCGC-TGGGTTCACCTCAGCCTCTGGAA-CCACGGTCGG-AACCAC	80bp mixed repeats (chr5 and chr2) and insertions	
	CTGGGTTCACCTCAGCCTCTGGAA-CACGGTCGGTCCATTGC-TGGGTT	CACCTCAGCCTCTGGAA-CCACGGTCGG-AACCAC
24bp GGACCGACCGCCTCTGGACCGTGA	24bp GGACCGACCGCCTCTGGACCGTGA	
	1	
CTCGAACTG	CTCGAACTG	+41-123

45bp : 41bp inverted sequence 83bp downstream from DSB from ALK GAATAGTGGGTAGATTCTGTGTGTAAAGCCCAGCCCCCAACACA 44

134bp

31bp from NPM directly downstream of the DSB GCTACTGGGTTCACCTCAGCCTCTGGAATAG

103bp partly from CASD10A-coding plasmid

CTAGCCTTTTGCTCACATGTCATATGTAGAGAGGTACCTCGAGCGGCCCAAGCTTAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGAT AACGGACTA

14

(CTGCTACTGGGTTCACCTCAGC) (CTGCTACTGGGTTCACCTCAGCCTCTGGA ATAGCTAGAAC) (CTGCTACTGGGTTCACCTCAGCCTCTGGA-CCTCAGCCTCTGGAATAGCTAGAAC)

(CTGCTACTGGGTTCACCT)

25

CTCGAACTGCTACTGGGTTCACCTCAGCCTCTGGAAT.....-9+276-89.....-52.. +276 bp from CAS9 coding plasmid TACGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGACCGCCCAACGACCCCCGCCCA $\label{eq:constraint} TradeGrant And Grant TradeGrant And Grant TradeGrant Grant And Grant And$

Figure S7, related to Figure 7. Translocation junction sequences from ZFN^{EWS} and ZFN^{P84} expression after CtIP depletion.

A. Indel formation at the ZFN^{p84} site is minimally affected by CtIP knockdown, as monitored by the T7-endonuclease assay in either wild-type or XRCC4-deficient HCT116 cells. Western blot of CtIP knock-down is shown on the left. The same blot was used to probe for both CtIP and tubulin.

T7 assay p84



B. Translocation junction sequences from ZFNs in wild-type cells. The ZFN recognition sequences at each DNA end are underlined; bps in italics represent the overhangs. The chromosome 19 end is in black and the chromosome 22 end is in red. Microhomologies (underlined), insertions (green) and lengths of deletions from each end are indicated.

Der22

HCT116 wt-SiCONT

DNA ends: ATCGTTTTTGGCCTCCCTATCAGTCATT		GTGGCC<u>CCACTGTGGGGT</u>GGAGGGGACAGAT
3,5,25,27,31,40 ATCGTTTTTGGCCTCCCTATCAGTCATT		GGCCCCACTGTGGGGTGGAGGGGACAGAT
4,9,20,39 ATCG TTTTTGGCCTCCCTATCA GTCA <u>T</u> .		GGCCCCCCTGTGGGGTGGAGGGGACAGAT
ATCG TTTTTGGCCTCCCTATCA GTCA <u>T</u> .	AC	GGCCCCCCTGTGGGGTGGAGGGGACAGAT
ATCG TTTTTGGCCTCCCTATCA GTCAT.		3.CCCACTGTGGGGTGGAGGGGGACAGAT
ATCG TTTTTGGCCTCCCTATCA GT <u>C</u> -1. 23		4CCACTGTGGGGGTGGAGGGGGACAGAT
ATCGTTTTTGGCCTCCCTATCAG3 22		IGCCCCCCCTGTGGGGGTGGAGGGGGACAGAT
ATCG TTTTTGGCCTCCCTATCA GTCAT. 13		5CACTGTGGGGTGGAGGGGGACAGAT
ATCGTTTTTGGCCTCCCTATCAGTCAT. 34		11GGGGTGGAGGGGGACAGAT
ATCG TTTTTG 16 33		2CCCCACTGTGGGGTGGAGGGGGACAGAT
ATCGTTTTTGGCCTCCCTATCAG3. 42		12 GGGT GGAGGGGACAGAT
<u>CA</u> 41		52
<u>GTA</u> 45 29		36
<u>CAG</u> 45		47

HCT116 wt-SiCtIP

ATCGT <u>TTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>		GTGGCC <u>CCACTGTGGGGG</u> GGAGGGGACAGAT
65,71,85 ATCGT TTTTGGCCTCCCTATCA GTCA <i>TT</i>		GGCCCCACTGTGGGGTGGAGGGGACAGAT
81,84,88 ATCGTTTTTGGCCTCCCTATCAGTCAT.		GGCCCCACTGTGGGGTGGAGGGGGACAGAT
82 ATCGTTTTTGGCCTCCCTATCAGTCAT.	ATAC	GTGGCCCCACTGTGGGGTGGAGGGGGACAGAT
63,66 ATCG TTTTTGGCCTCCCTATCA GTCAT.		IGCCCCCCTGTGGGGTGGAGGGGACAGAT
72 ATCG TTTTTGGCCTCCCTATCA GTC-1.	CTC	2CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
54 ATCG TTTTTGGCCTCCCTATCA GTCA <i>T</i> .	с	2CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
59 ATCG TTTTTGGCCTCCCTATCA GTC-1.		2CCCCCACTGTGGGGGTGGAGGGGGACAGAT
56 ATCG TTTTTGGCCTCCCTATCA GT-2		GGCCCCCACTGTGGGGGTGGAGGGGGACAGAT
78 ATCG TTTTTGGCCTCCCTATCA 4		, TGGCCCCCACTGTGGGGGTGGAGGGGGACAGAT
55 ATCG TTTTTGGCCTCCCTATCA -4		
52 ATCGTTTTTGGCCTCCCT		. TGGCCCCC ACTGTGGGGGTGGAGGGGACAGAT
		-11 GGGGTGGAGGGGACAGAT
80		
95		. TGGCCCCACTGTGGGGGGGGGGGGGGGGGGGGGGGGGGG
ATCGTTTTTGGCCTCCCTATCAGTCAT.		=193

C. Translocation junction sequences from ZFNs in X4^{-/-} cells.

HCT116 X4-/-siCTRL

ATCGT <u>TTTTGGCCTCCCTATCA</u> GTCA <i>TT</i>		<i>GT</i> GGCC <u>CCACTGTGGGGT</u> GGAGGGGACAGAT	
15			
ATCGT TT<u>TTGGCC</u>13		4CCACTGTGGGGTGGAGGGGGACAGAT	
33		_	
31		4CCACTGTGGGGTGGAGGGGGACAGAT	
15			
ATCGT TTTTGGCC -13	A	25AGAT	
4			
	CTAGAG	4CCACTGTGGGGTGGAGGGGGACAGAT	
6			
ATCGTT	GTCAC	24CAGAT	
17			
AT24		2/	
22	C 222	50	
A23	GAA	=50	
-36		-42	
39			
AT	TC	41	

37		
<u>AGTA</u> 34		36
-38.		24
46	G	31
12 <u>AC</u> 26		37
AG36		41
42		54
11 <u>CAG</u> 40	G	41
<u>G</u> 53		49
		51
<u>TGG</u> 55		81
2 GAG36		49
<u>CAGA</u> 68		48
AGCC68		51
218		4CCACTGTGGGGTGGAGGGGGACAGAT
20 <u>AA</u> 86		164
32 <u>GGGGA</u> 52		253
27	т	18AGGGGACAGAT
40		32
5 AAGA162		319
16	ATCTG	26
26 TTGT46		375
29		_185
21		422
18		. <u>.</u> =432
41		408
45		194
<u>CAA</u> 27 31		426
108 bp inverted insertion from	+108 1321 bp upstream the	393 DSB on chr19

108 bp inverted insertion from 1321 bp upstream the DSB on chr19 GCATCGCCCCCCGCTGGCGGTGTCCCCAGGTCCTTAGGGTACCCCCCGGGGGTTATCCAACCACTTGGTGAGGCTGGTACCCTGC CCCCATTCCTGCACTGGCATGG

HCT116 X4-/-siCtIP

ATCGT<u>TTTTGGCCTCCCTATCA</u>GTCA*TT*

*GT*GGCC<u>CCACTGTGGGGT</u>GGAGGGGACAGAT

43		
ATCG TTTTTGGCCTCCCTAT<u>CA</u>4		7CTGTGGGGTGGAGGGGGACAGAT
ATCGT TTTTGGCCT<u>CCC</u>9		6ACTGTGGGGTGGAGGGGGACAGAT
38 ATCGT TTTTG<u>GCC</u>13		4CCACTGTGGGGTGGAGGGGGACAGAT
4 ATCGT TTTTGGCCTCCCTAT 6	GA	13
1,56,73		
2		
66	A	5CACTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
ATCGT TTTTGGCCTCCCTAT -6 85		24CAGAT
ATC23 93		7CTGTGGGGTGGAGGGGGACAGAT
ATCGT TTTT<u>GG</u> 15		15 T GGAGGGGACAGAT
<u>CA</u> 25		7CTGTGGGGTGGAGGGGACAGAT
25		6ACTGTGGGGTGGAGGGGGACAGAT
74 ATCGT TTTTGGCCTCCCTATC<u>AGT</u>-2		35
20 ATCGT TTTTGG 15		22GACAGAT
82 ATCGTTTTTGGCCTCCCTATCAG3.	AA	-37
35		21
18		31
<u>A</u> 28 59		24CAGAT
<u>CA</u> 25		40
34	GTCAAGT	31
-43		24 <u>CAG</u> AT
ATCGT T 20	GA	47
26 <u>CA</u>		10 <u>TGGGG</u> TGGAGGGGACAGAT
30 AGTA34		36
23 ATC	CTAGTAT	48
5 A TCC TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT		66
55		
10bp:ACAAAACATT	+10	. <u>.</u> -3/
70	AC	29
21		-41
52	c	272
85	G	3/3
23bp:CATTAAGGATTGGTGACAGAAAA	+23	61
7 AG40		49
78		-45
87		. <u>-</u>

28	GAGCTGT	80
13		
<u>G</u> 84		49
33		
<u>A</u> 28		132
62		
ATCGT TTT<u>T</u> -17		141
83		
<u></u>		. TG <u>GCCCCA</u> CTGTGGGGTGGAGGGGACAGAT
31		
<u></u>		. – 1 <u>GCCCCACTGTGGGGGT</u> GGAGGGGGACAGAT
40		
<u></u>		5 <u>C</u> ACTGTGGGGTGGAGGGGGACAGAT
61		
		• <u>•</u> =38•••••
84		147
42		· <u>·</u> -14/
42		172
10		=1/3
17		111
44		=111
-265		-5 састатасастососсососососо
54		
80		
ATCGT TTTTGGC		323
50		
CAG40		47
6		-
САА27		406
34		
ACATC24		436
58		
CAG299		312
51		
ATCGT TTTTG<u>GCC</u>13	+30	
30bp duplication from 215 bp upst	ream of DSB on	n chr22
ACTGCGCCCAGCCACGTTTGGAGTTTTTGAACA	GGGGGAATACTCT	TTGCCATTGTTTGCTTTTGGAATCCAGGA
37		
	+52	2CCCCACTGTGGGGGTGGAGGGGGACAGAT
52 bp from chr7 Homo sapiens BAC	clone RP11-31	0H4 from 7, complete sequence
TCCAGTGTTTCTTGTGGTTTGTGCTGAAACCTT	TGGTTCCCCTCCA	GGTCCT <u>CCCCA</u>
24		
ATCGT TTTT -17	+46	448
46bp inverted insertion from chr1	9, 333bp downs	stream of DSB
CTGGACAACCCCAAAGTACCCCGTCTCCCTGGC	TTTAGCCACCTCT	
Supplemental Experimental Procedures

Nucleases

All nucleases target intronic sequences. ZFN^{EWS}, ZFN^{FLI1}, and ZFN^{P84} have been described (Brunet et al., 2009; Piganeau et al., 2013). ZFN^{FLI-A} and ZFN^{FLI-B} were provided by Sangamo BioSciences, Inc. with an obligate heterodimer architecture through modification of the Fokl nuclease domain (right ZFN Fokl KK and left ZFN Fokl EL) (Miller et al., 2007) and have the following recognition sites and helices. All ZFNs have FLAG epitope tags.

ZFN Binding Sequence (underlined)	ZFN	Finger 1	Finger 2	Finger 3	Finger 4	Finger 5
CCTAAGCCCTTTCCTTCATTTGC <u>CAGGAGTAGAGAGGACA</u>	FLI-A-R	DNPNLNR	RSDDLSR	QRTHLTQ	RSANLAR	RSDNLRE
G <u>GATTCGGGAAAGGAAG</u> TAAACGGTCCTCATCTCTCCTGT	FLI-A-L	RSDNLST	QSSDLRR	RSDSLSV	QNQHRIN	QSGNLAR
GCCCCCTGGCCCAGGTGTCCCC <u>GATGAAAAGCAGGTT</u> A	FLI-B-R	HRSSLRR	RSDNLSE	RNANRIT	QSGNLAR	TSGNLTR
C <u>GGGGGACCGGGTCCA</u> CAGGGGCTACTTTTCGTCCAAT	FLI-B-L	RSDHLSR	RSDHLTQ	ERGTLAR	RSDHLTT	DQSTLRN

TALEN assembly and TAL^{ALK} and TAL^{NPM} have been described (Piganeau et al., 2013). The sequences of TAL^{P84} and TAL^{LAM} are shown below with the TAL DNA binding domains underlined. TAL^{ALK} and TAL^{P84} have HA epitope tags; TAL^{LAM}, and TAL^{NPM} have FLAG epitope tags.

TAL^{p64} subunits were assembled as described (Huang et al., 2011). 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 the laboratory of Bo Zhang (Beijing) and subcloned in the pCS2 vector containing the wild-type Fokl domain.

TAL^{p84} recognition sequence (underlined): TTTTCTGTCACCAATCCTGTCCCTAGTGGCCCCACTGTGGGGTGGAGGGGA

TAL^{p84-L}:

MAPKKKRKVYPYDVPDYAGYPYDVPDYAGSYPYDVPDYAAHGTVDLRTLGYSQQQQEKIKPKVR STVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQMSGARA LEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNL<u>TPEQVVAIASHD</u> GGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAI ASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQ AHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPV VLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQAL ETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQAL ETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQAL ETVQRLLPVLCQAHGLTPAQVVAIASNIGGKPALESIVALSIVALSIVA KQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKPALESIVAQLSRPDPALAALTNDHLVALA

TAL^{p84-R}:

MAPKKKRKVYPYDVPDYAGYPYDVPDYAGSYPYDVPDYAAHGTVDLRTLGYSQQQQEKIKPKVR STVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQMSGARA LEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASNG GGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAI ASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNDGGKQALETVQRLLPVLCQAHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNDGGKQALETVQRLLPVLCQAHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNDGGKQALETVQRLLPVLCQAHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNDGGKQALETVQRLLPVLCQAHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNDGGKQALETVQRLLPVLCQAHGLTPAQ VAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQ HGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQ RLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQAL ETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKPALESIVAQLSRPDPALAALTNDHLVALA CLGGRPALDAVKKGLPHAPALIKRTNRIPERTSIRVA

TAL^{LAM} cleaves the lamin locus on Chr1. Subunits were directly synthetized by Geneart as described (Miller et al., 2011) using the RVD: A (NI), T (NG), G (NK) and C (HD) and subcloned in the pVax vector containing the wild-type FokI domain.

TAL^{LAM} recognition sequence (underlined): TTGCTCCCGTTCTCTCTTTTTCCTCTTTAGCTCAGAGTAGCTA

TAL^{LAM-L}:

MDYKDHDGDYKDHDIDYKDDDKMAPKKRKRVGIHGVPMVDLRTLGYSQQQQEKIKPKVRSTVA QHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEAL LTVAGELRGPPGTLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLMLTPDQVVAIASNGGGKQ ALETVQRLLPVLCQDHGLTPEQVVAIASNKGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNG GGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGL TPEQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNKGGKQALETVQRLLPVLCQDHGL TPEQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNKGGKQALETVQRLLPVLCQ AHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQ HGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQ HGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQ AHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGRPALESIV AOLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVA

TAL

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPMVDLRTLGYSQQQQEKIKPKVRSTVA QHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEAL LTVAGELRGPPGTLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLMLTPDQVVAIASNIGGKQ ALETVQRLLPVLCQHGLTPEQVVAIASNKGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQHGLTPAQ VVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGL TPEQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNKGGKQALETVQRLLPVLCQ HGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNKGGKQALETVQRLLPVLCQ HGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNKGGKQALETVQRLLPVLCQ HGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNKGGKQALETVQRLLPVLCQ AHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNKGGKQALETVQRLLPVLCQ NCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNKGGKQALETVQRL VALASNKGGKQALETVQRLPVCQNFLPVLCQDHGLTPDQVVAIASNKGGKQALETVQR AHGLTPAQVVAIASNIGGKQALETVQRLPVCQDHGLTPDQVVAIASNKGGKQALETVQR NCCQDHGLTPEQVVAIASNGGGRPALESIV pCas9 GFP expressing wild-type Cas9 (Addgene plasmid 44719) and pCas9D10A GFP expressing nCas9 (Addgene plasmid 44720) allow simultaneous expression of GFP (Ding et al., 2013; Mali et al., 2013b). The gRNA expression vector was derived from Addgene plasmid 43860 MLM3636. Specific target sequences are underlined and PAM sequences are in bold:

NPM: CCT<u>CGAACTGCTACTGGGTTCAC</u>CTCA<u>GCCTCTGGAATAGCTAGAACTAC</u>AGG gRNA(NPM1): 5'-GTGAACCCAGTAGCAGTTCG-3' gRNA(NPM2): 5'-GCCTCTGGAATAGCTAGAACTAC-3' ALK: CCT<u>CAGGTAACCCTAATCTGATC</u>ACG<u>GTCGGTCCATTGCATAGAGG</u>AGG gRNA(ALK1): 5'-GATCAGATTAGGGTTACCTG-3' gRNA(ALK2): 5'-GTCGGTCCATTGCATAGAGG-3'

Cell lines

The human pre-B cell line NALM6 and its LIG4-defective derivative N114P2 (Grawunder et al., 1998) were maintained in RPMI 1640 medium supplemented with 10% heat inactivated fetal bovine serum (FBS). Human Dermal Fibroblasts, adult, HDFa (Life Technologies) and 411BR primary skin fibroblasts (O'Driscoll et al., 2001) were cultured in DMEM supplemented with 15% FBS and 1% pyruvate sulfate. The human wild-type HCT116 cell line and its derivatives generated by gene targeting were cultured in McCoy's 5A medium containing 10% FCS. LIG4 mutant (L4^{+/-}, L4^{-/-}) cells have been described (Oh et al., 2013). A manuscript detailing the derivation of XRCC4 mutant (X4^{+/-}, X4^{-/-}) cells is in preparation (B.R. and E.A.H.). X4^{-/-} cells were propagated in media containing G418 (1 mg/ml).

HCT116 LIG3^{flox/-} cells containing a conditional LIG3 allele and a deletion allele (Oh et al., 2014), here termed L3^{+/-}, were further engineered to express human LIG1 with a mitochondrial leader sequence (MtLIG1) as follows: a cDNA encoding the mitochondrial localization signal of human LIG3 (residues 1- 85) fused to human LIG1 (residues 233-918, which contains the catalytic core but not the nuclear localization signal) was constructed by ligating two PCR products. The resultant DNA fragment was linked with a cDNA encoding EYFP from Aequrea Victoria (Addgene plasmid 11180:pCAG-YFP), such that YFP was fused to the LIG1 C terminus, and then subcloned into the mammalian expression vector pCAGGS that confers resistance to neomycin (Addgene plasmid 31264:pGCGFP-G418). The final expression vector (pCAG-MLig1-YFP-neo) was confirmed by plow cytometry. To delete the remaining conditional LIG3 allele, cells were infected with an adenovirus type 5 (dE1/E3) virus encoding the Cre

recombinase, (Ad-CMV-Cre #1045, Vector Biolabs). After 24 h, cells were washed and then cultured in fresh medium containing 0.5 mg/ml G418. Single cells were isolated in 96-well plates using an SY3200 cell sorter. G418-resistant YFP-positive clones, i.e., expressing MtLIG1, that were also L3^{-/-}, were identified by Western blotting using antibodies to human LIG1 (Peng et al., 2012), GFP (Santa Cruz #8334) and LIG3 (GeneTEX #103172). The L3^{-/-} genotype was confirmed using primers Lig3 Exon 5 F1: 5'-AAA GCA ACC CTC CTG TCT TCT CCT GCA AGT-3' and Lig3 Exon 5 R1: 5'-TGG TAC CAG GGA TAG AGT CAC GGA CAA ACC AA-3'.

Nucleofection

Cells were transfected by Amaxa technology (Lonza) using a cell line nucleofector. We used 10⁶ HCT116 cells Kit V program D-032, 2 × 10⁶ NALM6 and N114 cells Kit L program C-005, and 5 × 10⁵ HDFa and 411BR cells Kit NHDF program P-022. For the Asel assay, HCT116 cells were nucleofected with 5 μ g each ZFN^{EWS} plasmid (i.e., ZFN^{EWS-R} and ZFN^{EWS-L}). For inducing the 3.2-kb deletion at FLII intron 4, we used 5 μ g each ZFN^{FLI1-A} and each ZFN^{FLI1-B} plasmid.

To induce t(19;22), cells were nucleofected with 3 to 5 μ g each ZFN^{EWS} plasmid and 1 to 2.5 each μ g ZFN^{P84} plasmid, depending on the cell line. To induce t(1;19), cells were nucleofected with 3 μ g each TAL^{LAM} and 1 μ g of TAL^{P84} plasmid. To induce t(2;5), cells were nucleofected with 2.5 μ g each TAL^{NPM} and 1 μ g each TAL^{ALK}, 3.5 μ g pCas9 GFP mixed with 3.5 μ g each gRNA plasmid (ALK1+NPM1), or 3.5 μ g pnCas9D10A GFP mixed with 3.5 μ g each gRNA plasmid (ALK1+ALK2+NPM1+NPM2).

Immunoblotting and knockdowns

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 cocktail protease inhibitor tablets (Complete, Roche). Typically 30 µg of protein extract from cells were electrophoresed on an 8% (w/v) Tris-HCl SDS PAGE gel, blotted, and then probed with antibody: anti-human LIG4 rabbit monoclonal antibody (Gene Tex), XRCC4 goat antibody clone C-20 (Santa Cruz Biotechnology), CtIP mouse monoclonal antibody 14-1 (Yu and Baer, 2000), ALK (31F12) mouse antibody (Cell Signal); anti-human LIG3 mouse monoclonal antibody (Becton Dickinson), FLAG mouse antibody (Sigma-Aldrich), HA rat antibody (Covance), alpha-tubulin (Sigma-Aldrich), and beta-actin (Sigma-Aldrich).

For CtIP knockdown experiments, cells were plated at a density of 2×10^5 cells per well in 6well plates one day prior the transfection, and transfected using LipofectamineTM RNAiMax (Invitrogen) with 40 nmol siRNA: siControl 5'-UGUGACUUAUCGGUGUGAC-3' or siCtIP 5'- GCUAAAACAGGAACGAAUC-3', following the manufacturer's protocol. Cells were incubated with the siRNA for 24 h, and media was removed prior to ZFN nucleofection (t0), as described above. The level of knock down was evaluated using Western blotting. Translocation frequency was determined at 48 h (t2).

Repair assays

T7 endonuclease I assays, based on the Surveyor assay (Guschin et al., 2010), have been described (Piganeau et al., 2013). Genomic DNA was isolated 48 h after transfection. Quantifications were made with Image J software. For the p84/AAVS1 locus, the uncut band is 353 bp; T7 cleavage gives ~233 bp and ~120 bp cut bands; for the ALK locus, the uncut band is 401 bp; T7 cleavage gives ~231 bp and ~170 bp cut bands. For intrachromosomal repair involving a single DSB (Asel assay), genomic DNA was isolated from cells 48 h after ZFN^{EWS} nucleofection. A ~724 bp PCR amplicon including the ZFN^{EWS} target site was generated using primers surrounding the EWS site, and then digested in vitro with Asel (New England BioLabs). Asel-resistant PCR products were gel purified with a gel purification kit (Cycle Pure Kit, EZNA) and reamplified. Products were cloned with a TOPO TA cloning kit (Invitrogen) and sequenced (GATC Biotech).

For intrachromosomal distal repair involving two DSBs, genomic DNA was isolated from cells 48 h after ZFN^{FLI-A} and ZFN^{FLI-B} transfection. A fragment corresponding to the 3.2-kb deletion was PCR amplified using primers flanking the two DSBs, and products were cloned and sequenced. Serial dilutions were performed with the same primer set.

Translocation frequency was calculated from a 96-well screen using small pools of cells and nested PCR to amplify translocation junctions 48 h after transfection, as described (Brunet et al., 2009; Piganeau et al., 2013). Frequencies were normalized to the number of viable cells 24 h after transfection. Cells were also counted 48 h after transfection; overall, mutant cells were reduced in number ~15%. Primers are listed below. Statistical analyses used a t test for frequency comparisons and Mann-Whitney for deletion and microhomology distributions.

Primers

	Intrachromosomal	Assays		
gene				
EWS	AseI-F- GCCACGTTTGGAGTTTTTGA	AseI-R- GGGCTGAGCTCCATAAATCA		
FLI1	FL1-NF- TCCACCCAGTTTTCCAGAGC	FL1-NR- TTTATTGAGGGCATTTTTGC		
	T7 Cleavage	9		
p84	p84-F-GTGTGTCACCAGATAAGGAATC	P84-R-TCTAGTCTGTGCTAGCTCTTCC		
ALK	ALK-F-AGATGGGCAGAGGCTTGAAAAG	ALK-R-TGAGGATGTTCTGGAAGGCAAA		
	Translocatio	ns		
	PCR(1)	Nested-PCR(2)		
t(1;19)				
Der1	DER1-F-CACCACGTGAGTGGTAGCC	DER1-NF- GCCTGGCCTTTCTTCTCTCT		
	DER1-R-GGGTTCCCTTTTCCTTCTCC	DER1-NR- CCTGTGCCATCTCTCGTTTC		
t(19;22)				
Der19	DER19-F-CCTAGGACGCACCATTCTCA	DER19-NF-CAAAGGGAGTTTTCCACACG		
	DER19-R-GGGGCTGAGCTCCATAAATC	DER19-NR-GAAATCCCCGTGGATAGAATG		
Der22	DER22-F-GCCTCCCGAGTAGCTGAGAT	DER22-NF-TGCCACTATGCCCAGCTATT		
	DER22-R-GGGTTCCCTTTTCCTTCTCC	DER22-NR-CCTGTGCCATCTCTCGTTTC		
t(2;5)				
Der5	DER5-F-CAGTTGCTTGGTTCCCAGTT	DER5-NF-GGGGAGAGGAAATCTTGCTG		
	DER5-R-AGGAATTGGCCTGCCTTAGT	DER5-NR-GCAGCTTCAGTGCAATCACA		

Supplemental References

- Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T., *et al.* (2013). A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell *12*, 238-251.
- Guschin, D.Y., Waite, A.J., Katibah, G.E., Miller, J.C., Holmes, M.C., and Rebar, E.J. (2010). A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol 649, 247-256.
- Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol 29, 699-700.
- Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25, 778-785.
- 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. Nat Biotechnol 29, 143-148.
- Peng, Z., Liao, Z., Dziegielewska, B., Matsumoto, Y., Thomas, S., Wan, Y., Yang, A., and Tomkinson, A.E. (2012). Phosphorylation of serine 51 regulates the interaction of human DNA ligase I with replication factor C and its participation in DNA replication and repair. J Biol Chem 287, 36711-36719.
- Simsek, D., Furda, A., Gao, Y., Artus, J., Brunet, E., Hadjantonakis, A.K., Van Houten, B., Shuman, S., McKinnon, P.J., and Jasin, M. (2011a). Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. Nature 471, 245-248.
- Yu, X., and Baer, R. (2000). Nuclear localization and cell cycle-specific expression of CtIP, a protein that associates with the BRCA1 tumor suppressor. J Biol Chem 275, 18541-18549.

2.1.1 Résultats supplémentaires : Knock down de XRCC4 par des ARN interférents

Nous avons voulu valider les résultats dans d'autres modèles cellulaires que ceux publiés. En plus d'avoir utilisé d'autres lignées KO ou exprimant une version mutée pour la LigIV, nous avons utilisé un siRNA dirigé contre la protéine XRCC4 dans les cellules HCT WT et cellules souches mésenchymateuses (MES) afin de confirmer le phénotype observé dans les cellules HCT KO pour le complexe de ligation. Les HCT wt et MES ont été lipofectées (RNAi MAX, Invitrogen) avec le siRNA dirigé contre XRCC4 et un siCTRL. Au bout de 48h, l'expression de XRCC4 est fortement diminuée dans les cellules HCT-wt siX4 et MES siX4 contrairement aux cellules non traitées ou traitées avec le siCTRL où l'expression de XRCC4 n'est pas affectée (Figure 22).



<u>Figure 22</u>: Détection de la protéine XRCC4 par Western blot dans les lignées HCTwt et MES. L'expression de la protéine XRCC4 est fortement diminuée dans les cellules traitées avec le siXRCC4 (siX4), en revanche son expression n'est pas affectée dans les cellules traitées avec le siCTRL.

À t=48h on introduit le couple de nucléase par nucléfection afin d'induire la translocation, et à t=72h les cellules sont récupérées afin d'examiner l'influence du *Knockdown* sur la nature des jonctions des translocations. Une PCR nichée est réalisée autour du site des jonctions de translocations, ensuite les puits positifs de chaque PCR ont été séquencés directement, en utilisant les primers de la PCR nichée du der22. Nous n'avons observé aucune différence significative entre les cellules traitées avec le siCTRL et les cellules traitées avec le siXRCC4, au niveau des jonctions. En effet, celles-ci présentent en majorité de courtes délétions (médiane= 7pb). Les jonctions présentent également majoritairement des courtes microhomologies de 1 à 2 pb (plus de 90% des jonctions) (Figure 23). Malgré une très forte inhibition de la protéine d'intérêt, nous nous retrouvons pas les résultats observés dans les cellules HCT XRCC4 KO. Une expression résiduelle de la protéine XRCC4 semble donc suffisante pour conserver une activité de la Ligase IV et assurer le bon déroulement du C-NHEJ. D'ailleurs, une étude réalisée sur des cellules irradiées et traitées avec des siRNA dirigés contre les protéines LigIV et LigIII, montre que même fortement ihnibées, ces ligases conservent une activité et les CDB sont religuées, mais avec des cinétiques de réparation plus longues (Windhofer et al., 2007).



<u>Figure 23</u> : Taille des délétions et répartition des microhomologies des lignées traitées avec le siXRCC4 (A) : Taille (bp) des délétions au niveau de la jonction der22 pour chacune des lignées cellulaires. Chaque point représente une jonction. (B) : Répartition du % des jonctions et de la longueur (pb) des microhomologies en fonction des lignées cellulaires. Random étant la répartition théorique de microhomologies pour une séquence aléatoire est montrée à titre comparatif.

DER22

HCTwt-siCTRL

TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.	.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTTCCT
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGG17	-97
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCA} \underline{{\tt G5}}$	-2CCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.	.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.	.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.	.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCC12	-13AGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.	-5CTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.	.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.	.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2	-5CTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT\overline{{\tt C3}}$.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC3}$.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2}$.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT\underline{T}$	-7AGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
${\tt TCTTTGGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT\underline{CA-2}$	-5CTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG5	.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2	-5CTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC3	.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCT
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2	-5CTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT\overline{CAT}.$	-15GATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTAGTCTC

HCTwt-siX4

TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC

TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -4..ACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 -2CCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC.-3 -3.CACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC.-3 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCATT -19......GGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC

TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. -11.....ACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC.-3.GCCACTAGGGACAGGACTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 -2CCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTA......-9 -3.CACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT, .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCA....-6 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCT......-10 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -10.....GACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -2CCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. -3.CACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC

DER22

TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCG<mark>TTTTTGGCCTCCCTATCA</mark>GTCATT GGCCAC<mark>TAGGGACAGGAT</mark>TGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC

Mes siCTRL

TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT4 GGCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC3} - 5 \dots {\tt CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCTCTC$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.G-12CAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTTCCT
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT4 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC3 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC3 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTC
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG5 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT\underline{CA}-2 -5 \dots CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCCTCCTAGTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCGTCGTCGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCGTCGTCGTCGTCGTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCGTCGTCGTCGTCGTCGTCGTCCTCCTCCTAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG$
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.} - 5 \dots {\tt CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTTCCT$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT4 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCTAT \underline{CAG} \dots -5 -15 \dots \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTATCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCTAGGCCTCCTAGTCTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTC} \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTC \\ {\tt GATTGGTGACAGAAAAGCCCCCATCGTTGTTTTTGGCCTCCTAGTC \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGTC \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTAGTC \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTCCTTCCTAGTC \\ {\tt GATTGGTGACAGAAAAGCCCCCATCCTTCCTTCCTAGTC \\ {\tt GATTGTTGGTGACAGAAAAGCCCCCATCCTTCCTTCCTTC$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTCC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.} - 5 \dots {\tt CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTTCCT$
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT.} - 5 \dots {\tt CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTTCCT$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG5 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCTATCAG\ldots -5 -10\ldots GACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCCTAGTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCGTCGTGGTGGTGGTGGTGGTGGTGGCCTCCTCCTAGTCGTGTGTGGTGGTGGTGGTGGTGGTGGCCTCCTCCTCTCCTAGTCGTGGTGGTGGTGGTGGTGGGTG$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC3 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT\underline{CA}-2 -5 \dots {\tt CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTTCCT$
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCC\underline{CTA}, \dots, -9 -8, \dots, {\tt GGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC}$
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT\underline{CA}-2 -5 \dots {\tt CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTCCTCCTAGTCTCTCTC$
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT\underline{C}3 -3.\\ {\tt CACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTCCTCCTCCTAGTCTCCTCCTAGTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTTCCT
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCTATCAG\ldots -5 + 4 - 17 \ldots \\ {\tt TTGGTGACAGAAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt TTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt TTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt TTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt TTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTC} \\ \\ {\tt TTGTTGGGGGAAGTTGAAAGCCCCCATCCTTAGGCCTCCTAGTC} \\ {\tt TTGTTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTC} \\ \\ {\tt TTGTTGGTGACAGAAAAGCCCCCATCCTTTTTGGCCTCCTAGTC} \\ \\ {\tt TTGTTGGTGACAGAAAAGCCCCCATCCTTCCTAGGCCTCCTAGTC} \\ \\ {\tt TTGTTGGTGACAGAAAAGCCCCCATCCTTTTTTTTGGCCTTCCTAGTC \\ \\ \\ {\tt TTGTTGGTGACAGAAAAGCCCCCATCCTTTTTTTTTTGGCCTTCCTAGTC \\ \\ \\ {\tt TTGTTGGTGACAGAAAAGCCCCCATCCTTTTTTTTTTGGCCTTCCTT$
GGAT
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCTATCAGTCAT.}-4ACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTCCTAGTCAGTC$
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG} {\tt5} - 2 {\tt CCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTTCCT$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG5 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGT\underline{CA}-2 -5 \dots CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCCTCCTAGTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCGTCGTCGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCGTCGTCGTCGTCGTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCGTCGTCGTCGTCGTCGTCGTCCTCCTCCTAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG$
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCA\underline{G} \dots - 5 - 2 \\ {\tt CCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCGTGTCGTCGTGTCGTCGTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCGCCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCGTGTGTGT$
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCTATC\ldots -7 + {\tt GG} {\tt G$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG5+3 -41TAGGCCTCCTCCTCCTAGTCTC
ACA
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCA\underline{G}\ldots -5} . \\ {\tt GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTC} \\ {\tt GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTGTC} \\ {\tt GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGGGACAGGATTGGTGACAGGAAAAAGCCCCCATCCTTAGGCCTCCTAGGGACAGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGGGACAGGACAGGAAGAAAAAGCCCCCATCCTTAGGCCTCCTAGGCACAGAAAAAAAA$
TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCTATCAGTC3 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGGCACAGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTCCTAGTCAGGACAGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGGGACAGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTAGTCAGTC
${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCTATCAG\ldots -5 -10\ldotsGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTAGTCTCCTAGTCTCCTCCTCCTCCTCCTAGTCTCCTCCTAGTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT$

Mes siX4b

TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5**GC**-11.....ACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG..-5**A**.GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC

AAT

TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTA.....-9 -8......GGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT, .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 -15......GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC ${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGG...-5} - 2 {\tt CCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt CCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGGCACAGGACAGGACTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt CCTTTGGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGGCACAGGACAGGACAGGACTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt CCTTTGGGGGAAGTTGTATGCAGGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt CCTTTGGGGGACGGACAGGACTGGTGACAGGACAGGCCCCCATCCTTAGGCCTCCTCCTAGTCTC} \\ {\tt CCTTTGGGGACAGGACTGGTGACAGGACAGGACAGGACTGGTGACAGACAGACAGACAGCCCCCATCCTCCTCCTCCTAGGCCCCCAGGACAGACACAGACACAGACACAGACAGACACAGACACAGACACAGACACAGACACAGACACAGACACAGACAGACACAGACAGACAGACACAGACACAGACAGACACAGACACAGACAGACAGACAGACACAGACAGACACAGACACAGACAGACACAGACAGACAGACACAGACACAGACACACACACAGACACACACACAGACACAG$ TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -3.CACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC.-3 -2CCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC.-3 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC ${\tt TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC-5 {\tt C} {\tt G} {\tt G} {\tt G} {\tt C} {\tt C} {\tt G} {\tt$ TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCAT. -3.CACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 -15......GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 -15......GATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTC.-3 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAG...-5 .GCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC TCTTTGGGGAAGTTGTATGCAGTGAGTAAATTCAACATCGTTTTTGGCCTCCCTATCAGTCA-2 -5...CTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCTCCTCCTAGTCTC

2.2 Article II : Établissement de modèles cellulaires porteurs de translocations

Cette partie est le fruit d'un travail collaboratif avec une autre étudiante en thèse Marion Piganeau, afin de mettre au point de nouveaux modèles cellulaires porteurs de translocations, notamment le sarcome d'Ewing et l'ALCL (la leucémie anaplasique à grandes cellules). De nombreuses études ont été menées afin de comprendre l'implication des translocations chromosomiques dans la tumorigenèse. Ces études sont principalement basées sur l'inhibition du gène de fusion par ARN interférence ou sur la surexpression du gène de fusion, et généralement dans des modèles murins. Cependant ces approches présentent certaines limitations. En effet, dans le cas de l'utilisation de l'ARN interférant, l'expression résiduelle de la protéine de fusion peut-être suffisante pour conserver une activité. Dans le cas de la surexpression du gène de fusion, celui-ci n'est pas exprimé sous son promoteur endogène et n'est pas régulé. Or de nombreux oncogènes connus sont soumis à un effet dose et peuvent avoir des effets opposés selon leur niveau d'expression.

Ce projet repose sur la génération de CDB sur des loci spécifiques à l'aide de nucléases artificielles : les ZFN et les TALEN. L'intérêt est de permettre l'échange de morceaux de chromosomes transloqués et pas de surexprimer le gène de fusion intégré dans le génome. Pour reproduire la translocation t(11,22)(q24;q12) trouvée dans le sarcome d'Ewing, deux paires de nucléases à doigt de zinc (ZFN^{EWS} et ZFN^{FLI1}) ont été développées, ciblant respectivement gènes EWS et FLI1 dans les cellules souches mésenchymateuses. Leurs sites de coupure sont situés entre les exons 7 et 8 de EWS et 5 et 6 de FLI1, points de jonctions principaux des translocations observées chez les patients. Les TALEN, TAL^{NPM} et TAL^{ALK} ciblent les points de cassures sur les gènes *NPM* sur le chromosome *5* et le gène *ALK* sur le chromosome 2 respectivement afin d'induire la formation de la translocation t(2;5)(p23;q35). Les séquences ciblées sont les mêmes que les points de cassures retrouvés dans les lignées cellulaires ALCL, SUP-M2 et SUDHL1. Après expression de ces deux nucléases simultanément, il est possible de former les chromosomes transloqués. La formation de la translocation conduit à l'expression du gène de fusion, oncogène, EWS-FLI1 pour le sarcome d'Ewing et NPM-ALK pour l'ALCL.

Nous avons montré que nous pouvions :

- Induire la translocation spécifiquement au niveau génomique dans les cellules humaines

109

(ADN : jonctions des translocations)

- Induire l'expression du gène de fusion (détection de l'ARNm de EWS-FLI1 et de NPM-ALK et de la protéine NPM-ALK)

 Récapituler très précisément les caractéristiques génomiques trouvées dans les cellules de patients (par analyse des séquences des jonctions retrouvées dans les cellules traitées par les nucléases).

Cancer translocations in human cells induced by zinc finger and TALE nucleases

Marion Piganeau,^{1,6} Hind Ghezraoui,^{1,6} Anne De Cian,¹ Lionel Guittat,² Mark Tomishima,³ Loic Perrouault,¹ Oliver René,¹ George E. Katibah,⁴ Lei Zhang,⁴ Michael C. Holmes,⁴ Yannick Doyon,⁴ Jean-Paul Concordet,⁵ Carine Giovannangeli,¹ Maria Jasin,^{3,7} and Erika Brunet^{1,7}

¹Museum National d'Histoire Naturelle, CNRS UMR7196, Inserm U565, 75005 Paris, France; ²INSERM U978, Université Paris 13, Sorbonne Paris Cité, Labex "Inflamex," 93017 Bobigny, France; ³Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA; ⁴Sangamo BioSciences, Inc., Point Richmond Tech Center, Richmond, California 94804, USA; ⁵Institut Cochin, Universite Paris Descartes, CNRS UMR 8104, Inserm U1016, 75014 Paris, France

Chromosomal translocations are signatures of numerous cancers and lead to expression of fusion genes that act as oncogenes. The wealth of genomic aberrations found in cancer, however, makes it challenging to assign a specific phenotypic change to a specific aberration. In this study, we set out to use genome editing with zinc finger (ZFN) and transcription activator-like effector (TALEN) nucleases to engineer, de novo, translocation-associated oncogenes at cognate endogenous loci in human cells. Using ZFNs and TALENs designed to cut precisely at relevant translocation breakpoints, we induced cancer-relevant t(II;22)(q24;qI2) and t(2;5)(p23;q35) translocations found in Ewing sarcoma and anaplastic large cell lymphoma (ALCL), respectively. We recovered both translocations with high efficiency, resulting in the expression of the *EWSRI–FLI1* and *NPMI–ALK* fusions. Breakpoint junctions recovered after ZFN cleavage in human embryonic stem (ES) cell-derived mesenchymal precursor cells fully recapitulated the genomic characteristics found in tumor cells from Ewing sarcoma patients. This approach with tailored nucleases demonstrates that expression of fusion genes found in cancer cells can be induced from the native promoter, allowing interrogation of both the underlying mechanisms and oncogenic consequences of tumor-related translocations in human cells. With an analogous strategy, the ALCL translocation was reverted in a patient cell line to restore the integrity of the two participating chromosomes, further expanding the repertoire of genomic rearrangements that can be engineered by tailored nucleases.

[Supplemental material is available for this article.]

Despite the wide range of recurrent chromosomal translocations identified in various cancers (more than 300 genes implicated) (Mitelman et al. 2007), the direct path from translocation formation to tumorigenesis is not always clear. In mouse and human cells, studies are mostly based on either ectopically expressing or silencing the fusion gene induced by the translocation. In the case of fusion protein expression from a cDNA (often randomly integrated into the genome), the choice of the fusion transgenic promoter is crucial because the level of fusion protein expression must often be tightly controlled to recapitulate endogenous levels or risk spurious results from overexpression. And in silencing strategies, even low levels of expression remaining for the fusion protein could mask to some extent the full cellular consequences of the translocation.

DNA double-strand breaks (DSBs) are considered to be causative lesions for many genomic rearrangements, including chromosomal translocations (Richardson and Jasin 2000; Mani and Chinnaiyan 2010). With the development of tailored endonucle-

lication date are at http://www.genome.org/cgi/doi/10.1101/gr.147314.112. Freely available online through the *Genome Research* Open Access option. ases like zinc finger nucleases (ZFNs) (Urnov et al. 2010; Carroll 2011) and more recently transcription activator-like effector nucleases (TALENs) (Doyon et al. 2011; Mussolino and Cathomen 2012), it is now possible to create a DSB in the genome of a human cell at any locus of interest for the purposes of gene correction and gene disruption. In addition, contemporaneous expression of two ZFNs targeting loci has led to the induction of translocations at model loci in human multipotent and stem cells (Brunet et al. 2009) and intrachromosomal rearrangements (e.g., deletions) in transformed cell lines (Lee et al. 2010, 2012). This approach to study translocation formation obviates the need for prior genetic manipulation or cloning of cells, significantly expanding the repertoire of human cells that can be interrogated for translocation.

In this study, we now investigate the formation of two specific translocations, one frequently observed in Ewing sarcoma and one found in anaplastic large cell lymphoma (ALCL) using both types of nucleases (ZFNs and TALENs). Ewing sarcoma is a prototype of a solid tumor carrying a specific chromosomal translocation; it enables the transcription of the EWSR1–FLI1 chimeric protein corresponding to the in-frame fusion of the EWSR1 amino terminus with the FLI1 carboxyl terminus. It is well accepted that the EWSR1–FLI1 fusion protein acts as a transcriptional factor, but target genes induced or repressed by the fusion protein are not fully identified yet

⁶These authors contributed equally to this work. ⁷Corresponding authors E-mail m-jasin@ski.mskcc.org E-mail ebrunet@mnhn.fr Article published online before print. Article, supplemental material, and pub-

(Chansky et al. 2004; Prieur et al. 2004; Smith et al. 2006; Riggi et al. 2010). ALCL is an aggressive T-cell non-Hodgkin lymphoma, accounting for as much as 10%–15% of children with the disease. About half of tumors exhibits the specific translocation t(2;5)(p23;q35) resulting in NPM1–ALK expression and constitutive ALK tyrosine kinase activity (Morris et al. 1994; Elmberger et al. 1995; Kuefer et al. 1997).

Results

Inducing Ewing sarcoma specific translocations with ZFNs

To target reported Ewing sarcoma breakpoints, two ZFN pairs were designed within the *EWSR1* and *FLI1* genes on chromosomes 22 and 11, respectively, to induce t(11;22)(q24;q12) translocations (Fig. 1A). ZFN^{EWS} targets *EWSR1* intron 7 and ZFN^{FL1} targets *FLI1*



Figure 1. Induction of t(11;22)(q24;q12) translocations in hES-MP cells with ZFNs. (*A*) Ewing sarcoma translocations involve breakpoints within the *EWSR1* and *FLI1* genes on chromosomes 22 and 11, respectively, creating an *EWSR1–FLI1* fusion gene on der(22). To induce t(11;22)(q24;q12), ZFNs are expressed in hES-MP cells to create DSBs (scissors) in both genes. (*B*) ZFN^{EWS} and ZFN^{FLI} cleavage sites are within *EWSR1 and FLI1* introns, respectively, relevant to the *EWSR1–FLI1* translocation. Zinc fingers in the ZFNs are designed to bind to the shaded sequences. (Arrows) The presumed DSB sites after FokI nuclease domain cleavage. ZFN cleavage activity in hES-MP cells is monitored by a T7-endonuclease assay (Guschin et al. 2010). The region around the ZFN site is amplified; the amplified product is then denatured, reannealed, and then subjected to T7 endonuclease cleavage. Insertions and deletions (indels) characteristic of imprecise DSB repair by NHEJ give rise to T7 endonuclease-cleavable DNA. (C) Nested PCR to detect derivative chromosomes der(11) and der(22) in hES-MP cells. Translocation breakpoint junctions are only detected after expression of both ZFN^{EWS} and ZFN^{FLI}. (*D*) RT-PCR detection of the *EWSR1–FLI1* fusion transcript after ZFN^{EWS} and ZFN^{FLI} expression in hES-MP cells. The forward primer overlaps the exon 2/3 junction of *EWSR1*, and the reverse primer is within exon 9 of *FLI1*, amplifying most of the *EWSR1–FLI1* coding sequence.

intron 5 (Fig. 1B), which contain breakpoints for the most common type of *EWSR1–FLI1* translocation (Plougastel et al. 1993). In particular, the ZFNs target sequences at breakpoint junctions reported in two tumors, T60 for *EWSR1* and T64 for *FLI1* (Supplemental Fig. S1; Zucman-Rossi et al. 1998).

We chose to test this system in human mesenchymal precursor cells, in particular, those derived from human embryonic stem cells (hES-MP) (Barberi et al. 2005), because of the presumed mesenchymal origin of this sarcoma (Tirode et al. 2007; Riggi et al. 2008). Both ZFNs efficiently generated DSBs at the respective target sites in these cells, as evidenced by the frequent formation of indels (Fig. 1B). To detect translocations, we used a nested PCR strategy on genomic DNA from cells transfected with the ZFN expression vectors (Fig. 1C). Translocation breakpoint junctions for both derivative chromosomes, der(11) and der(22), were detected upon expression of both ZFN^{EWS} and ZFN^{FLI} (Fig. 1C). Expression of a single ZFN, however, was not sufficient to give rise to translocations. We next asked whether the EWSR1-FLI1 fusion transcript was expressed from der(22). Strikingly, we could detect the EWSR1-FLI1 fusion transcript by RT-PCR of RNA extracted from the bulk hES-MP cell population treated with $\mbox{ZFN}^{\mbox{EWS}}$ and ZFN^{FLI}, but not from treatment with either ZFN^{EWS} or ZFN^{FLI} alone (Fig. 1D).

The detection of *EWSR1–FL11* fusion transcripts from the transfected cell population suggested efficient t(11;22)(q24;q12) translocation. To quantify translocations, we applied a high-throughput 96-well PCR screen for translocation junctions using small pools of cells (~800 cells per well) (Brunet et al. 2009). Translocation breakpoint junctions were detected in this small pool PCR at a frequency of $3.1 \pm 0.7 \times 10^{-4}$ for der(22) and $4.2 \pm 0.45 \times 10^{-4}$ for der(11). Reciprocal translocations were frequent: ~30% of wells with one derivative chromosome also contained the other derivative chromosome. The lack of a reciprocal translocation formation, as observed in a portion of tumors (Zucman-Rossi et al. 1998) or, in some cases, lack of amplification (Brunet et al. 2009).

Breakpoint junctions recovered after ZFN cleavage fully recapitulate the genomic characteristics found in Ewing tumor cells

The occurrence of translocation breakpoints within intronic sequences means that joining of DNA ends can occur with a variety of DNA end modifications without compromising the ability to express the EWSR1-FLI1 fusion protein. To gain insight into the joining mechanism, we analyzed 174 t(11;22)(q24;q12) breakpoint junctions detected by small pool PCR after ZFN expression and compared the joining characteristics (deletions, microhomology, insertions) with those from patient-derived translocations (Zucman-Rossi et al. 1998). Small deletions were observed in most ZFN-induced translocations: 74.5% of the derivative chromosomes had deletions of <30 bp total from the two DNA ends forming the junction (Fig. 2A; Supplemental Fig. S2). Exact breakpoints are not known for patient-derived events; however, for the 36 cases in which junctions from both derivative chromosomes were sequenced, the total deletion from each chromosome could be determined and was found to be similarly small, with 71% of deletions also <30 bp (Supplemental Fig. S3). Except in one tumor junction (T23 with a deletion of 8.3 kb), all deletions from patient-derived translocations were <200 bp. Thus, our PCR-based screening approach, which detects deletions

Breakpoint junctions can occur at microhomologies (short sequence identities) between DNA ends more frequently than that expected by chance. The involvement of microhomology for joining is considered to be more typical of the alternative pathway of NHEJ than of the canonical pathway (Weinstock et al. 2007; Yan et al. 2007; Simsek and Jasin 2010; Simsek et al. 2011). We determined the microhomology present at t(11;22)(q24;q12) junctions from both ZFN-induced and patient-derived events and compared the microhomology distributions with that expected from chance joining of two random sequences (Fig. 2B; Supplemental Figs. S2, S3). Microhomologies of \geq 3 bp were observed in only a fraction of junctions from both ZFN-induced events and tumors, although somewhat more frequently than that expected by chance. Only one particularly long microhomology of 8 bp was found in one tumor (T44) (Supplemental Fig. S3). These results imply that substantial preexisting microhomology is not required for translocation formation in human cells. Microhomology of 1 bp was often present in junctions involving the ZFN overhangs (e.g., asterisks, Supplemental Fig. S2), whereas for tumor translocation junctions, which derive from a variety of DNA breakpoint sequences and types of ends, the major class of events showed no microhomology at the junctions, as would be expected if microhomology was not driving these events. Despite these differences, overall the difference in microhomology length distributions between the ZFN-induced and patient-derived events was not significant (P = 0.076, Mann–Whitney test), supporting the ability of ZFN induction of translocations to model events seen in patients. They also point out that subtle design changes, i.e., avoiding sites where ZFN overhangs have the possibility to anneal at even 1 base, could even more faithfully recapitulate events in tumors.

As seen in some tumors from Ewing sarcoma patients (Zucman-Rossi et al. 1998), a portion of the ZFN-induced translocation breakpoints contained inserted sequences. Although many were only a few base pairs, several longer insertions were also detected that were derived from a variety of sources: nearby the DSB sites on chromosomes 11 and 22, other chromosomes, and exogenous DNA (i.e., the ZFN expression plasmids). For example, a patient der(11) junction was reported to have a 140-bp insertion from chromosome 1, with otherwise little modification to the chromosome 11 and 22 ends (T89) (Supplemental Fig. S3). Likewise, a ZFN-induced der(11) junction was identified with a 180-bp insertion from chromosome 5, also with little other modification to the DNA ends.

Furthermore, insertions were also identified in ZFN-induced translocations where different pieces of DNA were assembled and/ or duplicated in a more complex way (Fig. 2C; Supplemental Fig. S4). For duplications, der(22)-D duplicated 139 bp from chromosome 11 that is derived from 363 bp downstream from the DSB, and der(11)-A and der(11)-B duplicated chromosome 22 sequences derived from upstream and downstream of the DSB, respectively. A more complicated insertion is found in der(22)-B, which has a duplication of 153 bp from chromosome 11 derived from directly upstream of the DSB, but in inverted orientation; this type of insertion has previously been observed in patient translocations and has been termed a locally derived inverted sequence (LDIS)



Figure 2. Analysis of t(11;22)(q24;q12) breakpoint junctions. (*A*) Deletion lengths for der(11) and der(22) junctions from ZFN-induced translocations from hES-MP cells. Each value represents the combined deletion from both ends of an individual junction. The median deletion length is indicated by a bar on the graph, and the value is given *below* the graph. (*B*) Microhomology length distributions for junctions from ZFN-induced and tumor translocation (Zucman-Rossi et al. 1998). [Der(11) and der(22) are grouped.] Only junctions with simple deletions (i.e., without an insertion) are included. The probability that a junction will have X nucleotides of microhomology by chance assumes an unbiased base composition and is calculated as previously described (Roth et al. 1985). (C) Examples of complex breakpoint junctions with insertions derived from sequences near the DSB sites. Schematics of the derivative and parental chromosomes are shown. The DSB sites on the unrearranged chromosomes are represented by scissors. Segments that are duplicated, deleted, or added are represented as independent blocks and in the orientation relative to that found on the parental chromosomes (see Supplemental Fig. S4 for the sequences).

(Zucman-Rossi et al. 1998). Der(22)-B additionally contains an insertion from the ZFN plasmid. The longest insertion, in der(22)-C, is 994 bp and is derived from four DNA segments from three different chromosomes (Supplemental Fig. S4), suggestive of iterative processing of DNA ends prior to the ligation step (Simsek and Jasin 2010).

Inducing ALCL specific translocations with TALENs

With ZFNs, we managed to recapitulate Ewing sarcoma translocations in human stem cells, but a limitation of extending this method to other translocations is that ZFN design is not trivial. The recent development of TALENs provides a potentially more advantageous strategy to generate DSBs at endogenous loci in human cells, given the simpler code of base recognition (Miller et al. 2011). We designed TALENs to target DSBs to the *NPM1* and *ALK* genes on chromosomes 5 and 2, respectively, to induce t(2;5)(p23;q35) translocations found in ALCL patients (Fig. 3A; Morris et al. 1994). TAL^{NPM} targets *NPM1* intron 4 and TAL^{ALK} targets *ALK* intron 16, both of which contain breakpoints found in patient-derived *NPM1– ALK* translocations. In this manner, TALEN-induced chromosomal translocations would result in expression of the oncogenic *NPM1– ALK* fusion gene, preserving the N-terminal dimerization domain of NPM1 and the C-terminal cytoplasmic domain of ALK.

Two different human cell lines were tested in this system. We first expressed the TALENs in a human acute T-cell leukemia line, Jurkat, because ALCL is a lymphoma of T-cell origin (Gillis and Watson 1980). Jurkat cells are pseudodiploid with two normal NPM1 and ALK loci (Mathas et al. 2009). Both TALENs efficiently generated DSBs at the NPM1 and ALK target sites in these cells (Fig. 3B). Using a similar nested PCR strategy as for ZFN induction of EWSR1-FLI1 translocations, we detected translocation breakpoint junctions for both derivative chromosomes when the TALENs were coexpressed, while no PCR product was detected when a single TALEN was expressed (Fig. 3C). Applying the same high-throughput 96-well format used to detect EWSR1-FLI1 translocations, we detected NPM1–ALK1 translocations at a frequency of $\sim 1 \times 10^{-3}$ per cell (see Methods). Given this relatively high frequency of translocations, we attempted direct PCR detection of translocation junctions using only one round of PCR and found that der(2) and der(5) breakpoint junctions were readily apparent (Fig. 3E). In serial dilutions of genomic DNA, both breakpoint junctions were obtained with as little as 6.25 ng of DNA (representing \sim 1000 cells), corroborating that translocations arose at a frequency of $\sim 10^{-3}$ per cell.

Confirming the generation of translocations, the *NPM1–ALK* fusion transcript was detected by PCR from the bulk Jurkat cell population following coexpression of TAL^{NPM} and TAL^{ALK} (Fig. 3D), and the NPM1–ALK fusion protein was detected by Western blot analysis using an antibody directed against ALK (Fig. 3F). The NPM1–ALK fusion protein comigrated with that found in the ALCL tumor SUPM2 cells and was ~1/40th the level of protein seen in the SUPM2 cells (Fig. 3F). Expanding pools of cells, junctions were observed 25 d following TALEN expression (data not shown), indicating maintenance of the translocation.

We also performed similar experiments in retinal pigment epithelial cells, RPE-1, expressing hTERT to determine if the efficiency of t(2;5)(p23;q35) induced by TALENs is seen in other cells or may be related to the lymphocytic origin of the Jurkat cells. As in Jurkat cells, both TALENs efficiently generated DSBs at the target sites (Supplemental Fig. S5A). We recovered translocation junctions for der(2) and der(5) from RPE-1 cells coexpressing both TAL^{NPM} and TAL^{ALK} (Supplemental Fig. S5B); using one round of PCR on dilutions of genomic DNA from the transfected cells, translocation frequency was estimated at 0.5×10^{-2} to 1×10^{-2} (Supplemental Fig. S5C). The *NPM1–ALK* fusion transcript and fusion protein were detected from the bulk-transfected cell population (Supplemental Fig. S5D,E). Fluorescence in situ hybridization confirmed the presence and the high frequency of the translocation (two of 70 metaphases) (Fig. 3A). Junctions were observed at 8 and 26 d following TALEN expression (Fig. 3G), indicating maintenance of the translocation. These results demonstrate that TALENs can efficiently induce t(2;5)(p23;q35) in different human cell types and that the translocation can be propagated.

TALENs and ZFNs both cleave DNA using the FokI nuclease domain, but the left and right DNA binding domains have a different spacing from each other in the two types of nucleases, with the potential to influence the site(s) of DNA cleavage. For ZFN^{EWS} and ZFN^{FLI}, the spacers are 6 bp; cleavage occurs at two sites, adjacent to the DNA binding site and 1 bp into the spacer, leaving 5' overhangs of 6 and 4 bp, respectively (Fig. 1B; Smith et al. 2000), consistent with our junction analysis (Supplemental Fig. S2). In contrast, the spacers in TAL^{NPM} and TAL^{ALK} are 18 bp (Fig. 3B). To map the TALEN cleavage sites, TAL^{NPM} and TAL^{ALK} were expressed in vitro, and cleavage assays were performed on double-stranded DNA fragments labeled at either 5' end (Supplemental Fig. S6). For both TAL^{NPM} and TAL^{ALK}, we identified a region of cleavage spanning 6-7 bp that overlapped the center of the spacer (Fig. 3B). TALEN cleavage results primarily in 5' overhangs, potentially of different lengths, although a fraction of blunt ends are also possible.

To examine joining mechanisms from TALEN-induced translocations, we analyzed 161 breakpoint junctions arising from t(2;5)(p23;q35) in Jurkat cells (Supplemental Fig. S7A). Because the TALEN cleavage sites may vary in cells as in vitro, we considered the longest 5'-overhang encompassing the full set of in vitro cleavage sites. Consistent with the in vitro cleavage assays, bases predicted to be in an overhang were incorporated into some of the junctions, and deletions emanating from the predicted DSB sites were frequent. As seen in the ZFN-induced translocations, small deletions were most common in the TALEN-induced translocations: 81% of junctions had deletions <30 bp, while <5% of junctions had deletions of >100 bp.

One feature of TALEN-induced breakpoint junctions was the frequent presence of short insertions (Supplemental Fig. S7A). Almost half of breakpoint sequences showed insertions of <10 bp (47.5%), compared with 7% in the ZFN-induced translocations. A few insertions could be identified as duplications of sequences close to the DSB sites, but most of them were not readily evident as being templated. As shown for ZFN-induced translocations, some longer insertions were observed that were derived from a variety of sources: Duplications of regions located near the DSB sites on chromosomes 2 and 5 as well as sites more distant from the DSBs (Mb), and the TALEN expression plasmids. Complex insertions were also identified. For example, one der(5) junction had tandem duplications from both chromosomes 2 and 5 in addition to a deletion from each DSB and another der(5) junction had duplications from chromosome 5 from both upstream and downstream of the DSB (junctions 19 and 54, respectively) (Supplemental Fig. S7A).

The involvement of microhomology in the joining mechanism of TALEN-induced translocations is more problematic to analyze because of the large number of short insertions. Never-



Figure 3. Induction of t(2;5)(p23;q35) translocations with TALENs. (*A*) ALCL translocations have breakpoints within the *NPM1* and *ALK* genes on chromosomes 2 and 5, respectively, creating an *NPM1–ALK* fusion gene on der(5). To induce t(2;5)(p23;q35), TALENs are expressed to create DSBs (scissors) in both genes. FISH demonstrates the t(2;5)(p23;q35) translocation after TALEN expression in RPE-1 cells. Red and green signals are from an *ALK* probe that "breaks apart" upon translocation. The blue signal is from an *NPM1* probe. Of 70 metaphases screened, two exhibited translocations and three showed breaks with the *ALK* break-apart probe, likely due to remaining TALEN expression at this time. (*B*) TAL^{NPM} and TAL^{ALK} cleavage within *NPM1* and *ALK* introns, respectively, relevant to the *NPM1–ALK* translocation. DNA binding domains of TALENs are designed to bind the shaded sequences. (Arrows) DSB sites with different tail lengths representing the efficiency of cleavage in vitro, as assayed by in vitro expression of the TALENs (see Supplemental Fig. S6). TALEN cleavage activity in Jurkat cells. Translocation breakpoint junctions are only detected after expression of both TAL^{NPM} and TAL^{ALK}. (*D*) RT-PCR detection of the *NPM1-ALK* fusion transcript after TAL^{NPM} and TAL^{ALK} expression in Jurkat cells and in ALC cell line SUP-M2. The forward primer is within exon 2 of *NPM1*, and the reverse primer is within exon 29 of *ALK*, amplifying most of the *NPM1-ALK* coding sequence. (*E*) Single-round PCR to detect of the fragment marked (*) in C on serial dilutions of genomic DNA (50, 25, 12.5, 6.25, 3.125, and 1.56 ng). The number of times the PCR was positive for each dilution from six total experiments is indicated. The markers for 933 and 951 bp correspond to der(5) junctions, respectively, without end modification. The larger and smaller fragments seen in some of the lanes likely correspond to junctions with large insertions or deletions. (*F*) Detection of the NPM1-ALK fusion protein in Jurkat cells.

theless, when considering only sequences without insertions, microhomologies of 3–6 bp were observed in only a small fraction of sequences (6.2%), suggesting a minimal dependence on microhomology in formation of these translocations as found in ZFN-induced and tumor translocations.

All of the joining characteristics observed in Jurkat cells were recapitulated in translocations from hES-MP cells, including a large fraction of junctions with short insertions (24/38 total) (Supplemental Fig. S7B). Frequent small insertions were previously reported for intrachromosomal repair after expression of a single TALEN in hES cells (Hockemeyer et al. 2011), although not in another study using U2OS cells (Reyon et al. 2012).

The homodimerizing FokI domains in the TALENs leave open the possibility that newly formed translocation junctions could be recleaved by mixed TALENs (i.e., TAL^{NPM/ALK} and TAL^{ALK/NPM}). To circumvent this possibility, TALENs were also constructed with FokI heterodimerization domains (Doyon et al. 2011). Translocations were observed in Jurkat cells with the heterodimeric TALENs (Supplemental Fig. S8A). Junctions showed similar characteristics as with the homodimeric TALENs (Supplemental Fig. S8B), including frequent small insertions, indicating that cleavage of newly formed translocation junctions are not substantially affecting characteristics.

Reversion of the NPMI-ALK translocation

The ability to specifically induce translocations suggested that it should also be possible to reverse translocations with a similar approach (Fig. 4A). Genetic reversion permits the analysis of phenotypic consequences of fusion protein loss, such as whether tumor cells are "addicted" to the oncogene. To revert the t(2;5)(p23;q35) translocation, TAL^{NPM} and TAL^{ALK} were expressed in SUDHL-1 patient-derived cells, with efficient cleavage of both the NPM1 and ALK loci (Fig. 4B). Due to the position of the TALEN cleavage sites relative to the translocation breakpoint junctions, restoration of the intact chromosomes leaves a small segment of DNA from the partner chromosome that serves as a tag to differentiate the revertant chromosomes from the wild-type chromosomes 2 and 5 (Fig. 4C). By PCR, both revertant chromosomes were detected (Fig. 4C), and sequencing confirmed that the translocation had been reversed (Fig. 4D; Supplemental Fig. S9). We also tested the reversion strategy in SUPM2 cells, another ALCL patient-derived cell line, and also obtained revertant breakpoint junctions (Supplemental Fig. S9). As with translocations, many of breakpoint junctions for both revertant chromosomes from both ALCL cell lines had small insertions. Overall then, frequent small insertions were observed at four different junctions (two translocation and two reversion junctions), suggesting that they do not derive from a specific sequence context from the TALEN pair we used.

Discussion

In this report, we could reproduce cancer relevant translocations and revert them by targeting concurrent DSBs to intronic sequences in human cells using two types of designed nucleases, ZFNs and TALENs. With this approach, fusion genes created during translocation formation, which are potential oncogenes, are expressed from their endogenous promoters. In addition, translocation disrupts a copy of each participating gene, reproducing the heterozygosity (and potential haploinsufficiency) found in tumor cells. Therefore, this strategy provides a more accurate model for translocation-related tumor formation than ectopic expression of a fusion protein to study effects on proliferation and other cellular phenotypes of newly translocated cells and, further, provides an approach to revert the translocation.

Despite the fact that a ZFN induces cohesive 5'-overhangs that should be easily repaired by direct ligation or by polymerase fill-in followed by ligation, breakpoint junctions of ZFN-treated cells recapitulated the features found in Ewing sarcoma cells: deletions, short insertions, complex insertions from other chromosomes and/or nearby the break regions, and an absence of long microhomology. Overall ZFN-treated cells showed a comparable proportion of each type of junction modification as patient cells, with the exception of complex duplications, which are overrepresented in tumor cells (Zucman-Rossi et al. 1998). It is possible that these latter events in patients arise from multiple breaks in the same intronic regions (fragile sites). Translocation junctions obtained with TALENs frequently showed short deletions and other modifications seen in patient translocation junctions. Short insertions were overrepresented, although it is not clear if this is specific to the TALEN pair.

Relevant to Ewing sarcoma, the strategy presented here has the potential to resolve inconsistencies in previous studies that have attempted to identify target genes using ectopic expression or knockdown of the EWSR1-FLI1 fusion protein and also those in studies using mouse cells as a tumor model (for review, see Lessnick and Ladanyi 2012). Der(5) NPM1-ALK-driven ALCL tumorigenesis appears to be more defined, since the fusion protein is known to inhibit apoptosis and promote cell proliferation through the constitutively active ALK tyrosine kinase moiety (Kuefer et al. 1997; Greenland et al. 2001; Kasprzycka et al. 2006). Nonetheless, it must be noted that NPM1 is a multifunctional protein involved in many cellular processes (Colombo et al. 2006), including centrosome duplication through a C-terminal interaction with the CRM1 complex, such that der(2) expressing the NPM1 C terminus has the potential to participate in the lymphomagenesis of ALCL in addition to der(5) (Wang et al. 2005). And while normal levels of NPM1 induce senescence and/or apoptosis under UV damage conditions by binding TP53 (Kurki et al. 2004), NPM1 overexpression can lead to cellular transformation (Li et al. 2008), although paradoxically, loss of one allele of NPM1 leads to DNA instability allowing possible cellular transformation as well (Grisendi and Pandolfi 2005). Thus, formation of both derivative chromosomes with loss of at least one allele of NPM1 may all be critical to recapitulate the disease initiation state faithfully.

The present study provides a novel approach to understand translocation-related tumorigenesis and, more broadly, demonstrates that precise genome rearrangements can be achieved in relevant cell types with tailored nucleases. This study also expands the repertoire of genome engineering accomplished with these nucleases beyond gene mutation/correction and intrachromosomal rearrangements (Lee et al. 2010, 2012) and translocations in model systems (Brunet et al. 2009; Simsek et al. 2011). Unlike site-specific recombinases, endonucleases are designed to cleave endogenous sites, abrogating the need for prior genome manipulation. And because they introduce DSBs, endonucleases provide more accurate models for studying the etiology of genomic rearrangements and their potential for reversion, as well as providing more accurate cellular models for further functional investigation.



Figure 4. Reversion of the t(2;5)(p23;q35) translocation in SUDHL-1 cells. (*A*) Patient-derived SUDHL-1 cells carry the t(2;5)(p23;q35) translocation and express the *NPM1–ALK* fusion gene from der(5). To reverse the translocation, TAL^{NPM} and TAL^{ALK} are expressed to create DSBs in both fusion genes; repair between the derivative chromosomes restores intact chromosomes 5 and 2. (*B*) TAL^{NPM} and TAL^{ALK} efficiently cleave target loci in SUDHL-1 cells. Cleavage is monitored by the T7-endonuclease assay directed to the *NPM1 (left)* and *ALK (right)* loci. (C) PCR detection of revertant chromosomes 2 and 5. Reversion is only detected after expression of both TAL^{NPM} and TAL^{ALK}. Because the TALENs cleave to the side of the translocation breakpoint junctions in the SUDHL-1 cells, segments of the other chromosome remain to "tag" the revertant chromosomes 5 and 2, which restore the *NPM1* and *ALK* genes, respectively.

Methods

Cell lines, transfections, FISH

hES-MP cells, derived from hES strain H1, were provided by the SKI Stem Cell Research Facility (MSKCC), and cultivated in α -MEM supplemented with 10% FBS and 50 ng/mL recombinant human FGF (R&D Systems). Jurkat and SUDHL-1 and SUPM2 (DSMZ, Germany) cells were cultivated in RPMI-1640 medium supplemented with 10% heat inactivated FBS. The hTERTimmortalized retinal pigment epithelial cell line, RPE-1, was cultured in DMEM:F12 medium supplemented with 10% FBS.

Transfections were done as previously described (Brunet et al. 2009). Typically, 7.5×10^5 (for hES-MP and RPE-1) or 2×10^6 (for Jurkat, SUDHL-1, and SUPM2) cells were transfected by Amaxa technology (Lonza) using the cell line nucleofector kit V (program B-16 for hES-MP cells and X-001 for Jurkat, SUDHL-1, SUPM2, and RPE-1 cells) with 7.5 µg of each ZFN^{EWS} and 5 µg of each ZFN^{FL11}; or 5 µg of each TAL^{NPM} and 2 µg (for Jurkat, SUDHL-1, and SUPM2 cells) to 2.5 µg (for RPE-1 cells) of each TAL^{ALK}.

FISH was performed on metaphases of RPE-1 cells 48 h after transfection with TAL^{NPM} and TAL^{ALK} using *ALK* break apart probes (LPS019 Amplitech) and an *NPM1* BAC probe (RP11-117L6). Images were taken with the Zeiss Axio Observer.Z1 system.

PCR-based translocation detection, frequency determination, and sequencing of breakpoint junctions

Primer sequences are presented in Table 1. For detection of translocations in pooled cells, 100 ng of genomic DNA from cells transfected with ZFNs or TALENs was amplified by nested PCR 72 h post-transfection. Where noted, the transfected cell population was kept in culture for the indicated length of time to monitor maintenance of the translocation. For translocation frequency determination, a high-throughput 96-well screen using small pools of cells and nested PCR to amplify translocation junctions was used 48 h post-transfection, as described (Brunet et al. 2009). The number of PCR-positive and negative wells for translocation junctions was determined to calculate the translocation frequency. Amplified products from positive wells were directly sequenced to verify translocations and determine the junction sequences. For hES-MP cells, 96-well plates contained 8×10^4 transfected cells per plate, resulting in 17-33 positive wells per plate in four experiments for each derivative chromosome. For Jurkat cells, 96-well plates contained 2×10^5 transfected cells per plate. Centrifugation of plates with the nonadherent Jurkat cells, followed by media removal, resulted in cell loss, and thus cells were requantified prior to lysis, demonstrating $\sim 1 \times 10^5$ cells remaining. After transfection of the TALEN homodimer expression plasmids, two 96-well plates were analyzed for each derivative chromosome; 61 ± 2 positive wells were obtained per plate, which adjusted to a β distribution is estimated to represent about 95 translocations per plate, for a frequency of $\sim 1 \times 10^{-3}$. Sequencing of PCR products from positive wells showed mixed sequences in a single well from almost half of the wells corresponding to multiple junctions, consistent with the high frequency of translocations, and were removed from the junction analysis. For TALEN heterodimer experiments, the translocation frequency was half that of the homodimer experiments.

Translocation frequency was also determined for Jurkat cells and for RPE-1 cells by serial dilution of genomic DNA from cells transfected with the TALEN expression vectors followed by one round of PCR (i.e., non-nested) indicated in Figure 3C. For Jurkat cells, genomic DNA dilutions were 1:2, for 50, 25, 12.5, 6.25, 3.125, and 1.56 ng of DNA. Considering that one cell contains ~ 6 pg of DNA (Jurkat cells being a pseudodiploid human cell line), 6.25 ng of represents $\sim 10^3$ cells. PCR amplification with 6.25 ng of DNA was seen in half of the experiments with the homodimeric TALENs, corresponding to a frequency of $\sim 10^{-3}$ per cell. For the heterodimeric TALENs, junction detection was less efficient (one of three experiments with 6.25 ng of DNA and two of three experiments with 12.5 ng of DNA), implying about half the translocation efficiency as with homodimeric TALENs. Thus, translocation frequency determination by the 96-well format and by serial dilution of genomic DNA gave similar results. For RPE-1 cells, genomic DNA dilutions were 1:2, for 5, 2.5, 1.25, 0.625, and 0.31 ng of DNA. Amplification was detected at 1.25 ng of genomic DNA (two of three experiments) and 0.625 (one of three experiments), indicating a frequency of 0.5×10^{-2} to 1×10^{-2} per cell. The detection of two translocations by FISH from 70 total metaphases is in agreement with the high translocation frequency estimated for RPE-1 cells.

PCR for revertant chromosome detection and sequencing of breakpoint junctions

Both Chr(2) and Chr(5) revertant chromosomes were detected by PCR amplification 72 h post-transfection of the TALEN expression vectors. Chr(2)- and Chr(5)-amplified products were cloned with the TOPO-TA cloning system (Invitrogen) and sequenced.

T7 endonuclease I assay

Forty-eight hours after transfection, genomic DNA was extracted with QIAamp DNA Mini Kit (QIAGEN). The genomic region encompassing the ZFN and TALEN target sites was amplified with the primers listed below, typically using 50 ng of genomic DNA, melted, and annealed (5 min at 95°C, 95°C to 25°C at -0.5° C/30 sec, and 15 min at 4°C) to form heteroduplex DNA. The annealed DNA was treated with 1.5 units of T7 endonuclease I (New England BioLabs) for 10 min at 37°C and run on a 2.4% agarose gel.

RT-PCR and cDNA amplification

RNA was extracted from transfected cells with absolutely RNA Microprep Kit (Agilent Technologies) and RT-PCR was done following manufacturer directions (SuperScript II First-Strand Synthesis System; Invitrogen). Primers used for cDNA amplification of *EWSR1–FLI1* and *NPM1–ALK* are listed below.

Western blotting

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 cocktail protease inhibitor tablets (Complete, Roche). Typically 40 μ g of protein extract from TALEN-treated cells were run on an 8% (w/v) Tris-HCl SDS page gel, blotted, and then probed with ALK (31F12) Mouse Antibody (Cell Signal). Alpha-tubulin mouse antibody (Sigma-Aldrich) was used to detect those proteins as loading controls.

In vitro mapping of TALEN cleavage sites

The TNT SP6 Quick Coupled Transcription/Translation System (Promega) was used to express TAL^{NPM} and TAL^{ALK} (Mussolino et al. 2011). Sixty-eight-bp fragments containing target sites of TAL^{NPM} and TAL^{ALK} were 5'-radiolabeled on one strand.

Typically, 0.5 μ L of each TNT lysate containing one TALEN subunit was incubated overnight at 37°C with 10 nM target duplex in NEB4 buffer (New England Biolabs) supplemented

with 100 mM NaCl and 0.4 μg of salmon sperm (DNA competitor) in 10 $\mu L.$ The reaction was analyzed on a denaturing 12% PAGE gel.

Table 1. Primer sequences

A-Ewing system

Translocation

	PCR(1)		Nested-PCR (2)		
Der22	DER22-f DER22-b	AGGCTGGTCTCGAACTCCTG TCTGGAAAAGCCCGTTCAGT	DER22-Nf DER22-Nb	CAGCCACGTTTGGAGTTTTT GAATGAGGTGGAAGGCTCTG	
Der11	DER11-f DER11-b	ACCCCACATACCACAAGCTC AGGGGCTGAGCTCCATAAAT	DER11-Nf DER11-Nb	AAGGTCCCAGAGTTTGCTGA TTTTTCCCCCTTTCCTTTTC	
T7 cleavage EWSR1 gene EWS-T7-f EWS-T7-b FLI1 gene FLI1-T7-f FLI1-T7-b Fusion EWSR1- EWS-FL-RNA	GCCACGTTTGGAGTTTT GCCATTTTTCCCCTTAA ACCTTGACATCTCACCA GAATGAGGTGGAAGGC <i>FLI1</i> cDNA A-f CAGCAGGGCTACAGT	TGA TGG CCC TCTG TGCTTA			
EWS-FL-RNA B-ALCL syste	n-b GTTGGGGTTGGGGT	AGATTC			
D-ALCL Syste					

Translocation

	PCR(1)		Nested-PCR (2)	
Der5 Der2	DER5-f DER5-b DER2-f DER2-b	CAGTTGCTTGGTTCCCAGTT AGGAATTGGCCTGCCTTAGT TCCTTCAGTGTCCATCACGA GAACCTTGCTACCACCTCCA	DER5-Nf DER5-Nb DER2-Nf DER2-Nb	GGGGAGAGGAAATCTTGCTG GCAGCTTCAGTGCAATCACA CCCACCCTAGACGTCACT TTCACATCCTCCTCCTCTTCA
Reversion Revertant CH SUDHL-10 DHL1-cl SUPM2 ce M2-ch2 Revertant CH SUDHL-10 DHL-ch DHL-ch SUPM2 ce M2-ch5 M2-ch5	nr(2) cells hr2-f CAGAGACATGCC hr2-b GCCTAATGTTCA Ils -f CTGCCGGTAGAAGC -b CCACTTGAGGCCA/ rr(5) cells 5-f AAGGGGTTTCGTG/ 5-b CCCACCCTCTAGG Ils -f GTGCACCACCAAGT -b CCCACCCTCTAGG	CAGGACAG CTGCCAGGT GAGATG AGAGTCC ATGTTGG GTTGTCA CCAGTT GTTGTCA		
T7 cleavage NPM1 gene T7-NPM-fr ALK gene T7-ALK-f A T7-ALK-f A T7-ALK-b Fusion NPM1-A NP-AL-RNA-fr NP-AL-RNA-fr	ACCACCAAGTCCAGT ATCTCACTTTGGGGA AGATGGGCAGAGGCTT TGAGGATGTTCTGGAA 4LK cDNA f TTGTGAACTAAAGGC b CCAGGCTGGTTCAT(TGCGTTT GGCCAAG GGCAAA GGCAAA CGACAAA GCTATTCT		
C-Control cD	NA			
GAPDH-f TG		NGC		

GAPDH-b GGCATGGACTGTGGTCATGAG

Competing interest statement

G.E.K., L.Z., M.C.H., and Y.D. are past (G.E.K.) or present (L.Z., M.C.H., Y.D.) employees of Sangamo BioSciences, Inc.

Acknowledgments

We thank Francesca Cole and other members of the Jasin laboratory, Patrizia Alberti and Loic Ponger from the laboratory of "Regulation and Dynamics of Genomes" at MNHN, and Alexei Morozov in Malcolm Moore's laboratory at MSKCC for helpful discussions. We also thank Fyodor Urnov (Sangamo BioSciences, Inc.) for his help in initiating the project. M.P. is supported by a "Le Canceropole IDF" PhD grant and H.G. is supported by a "La Ligue Nationale contre le Cancer" PhD grant. This work was supported by CNRS, Inserm, MNHN, IBiSA (Infrastructures for Biology), and ANR-12-JSV6-0005-01 (E.B.) and by grant R01 NIHGM54668 (M.J.).

Author contributions: E.B. and M.J. designed the research. M.P., H.G., A.D.C., L.G., M.T., L.P., O.R., and E.B. performed the experiments. G.E.K., L.Z., M.H., and Y.D. designed new ZFN reagents. A.D.C., J.-P.C., and C.G. designed new TALEN reagents and did analytic analysis. E.B. and M.J. did statistical analysis of data. E.B. and M.J. wrote the paper.

References

- Barberi T, Willis LM, Socci ND, Studer L. 2005. Derivation of multipotent mesenchymal precursors from human embryonic stem cells. *PLoS Med* 2: e161.
- Brunet E, Simsek D, Tomishima M, DeKelver R, Choi VM, Gregory P, Urnov F, Weinstock DM, Jasin M. 2009. Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci* 106: 10620–10625.
- Carroll D. 2011. Genome engineering with zinc-finger nucleases. Genetics 188: 773–782.
- Chansky HA, Barahmand-Pour F, Mei Q, Kahn-Farooqi W, Zielinska-Kwiatkowska A, Blackburn M, Chansky K, Conrad EU III, Bruckner JD, Greenlee TK, et al. 2004. Targeting of EWS/FLI-1 by RNA interference attenuates the tumor phenotype of Ewing's sarcoma cells in vitro. *J Orthop Res* **22**: 910–917.
- Colombo Ê, Martinelli P, Zamponi R, Shing DC, Bonetti P, Luzi L, Volorio S, Bernard L, Pruneri G, Alcalay M, et al. 2006. Delocalization and destabilization of the Arf tumor suppressor by the leukemia-associated NPM mutant. *Cancer Res* **66**: 3044–3050.
- Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, Xia DF, Miller JC, Urnov FD, Gregory PD, Holmes MC. 2011. Enhancing zinc-fingernuclease activity with improved obligate heterodimeric architectures. *Nat Methods* 8: 74–79.
- Elmberger PG, Lozano MD, Weisenburger DD, Sanger W, Chan WC. 1995. Transcripts of the *npm–alk* fusion gene in anaplastic large cell lymphoma, Hodgkin's disease, and reactive lymphoid lesions. *Blood* **86:** 3517–3521.
- Gillis S, Watson J. 1980. Biochemical and biological characterization of lymphocyte regulatory molecules. V. Identification of an interleukin 2-producing human leukemia T cell line. *J Exp Med* **152**: 1709–1719.
- Greenland C, Touriol C, Chevillard G, Morris SW, Bai R, Duyster J, Delsol G, Allouche M. 2001. Expression of the oncogenic NPM–ALK chimeric protein in human lymphoid T-cells inhibits drug-induced, but not Fasinduced apoptosis. Oncogene 20: 7386–7397.
- induced apoptosis. *Oncogene* **20**: 7386–7397. Grisendi S, Pandolfi PP. 2005. NPM mutations in acute myelogenous leukemia. *N Engl J Med* **352**: 291–292.
- Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ. 2010. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* **649**: 247–256.
- Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, et al. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**: 731–734.
- Kasprzycka M, Marzec M, Liu X, Zhang Q, Wasik MA. 2006. Nucleophosmin/anaplastic lymphoma kinase (NPM/ALK) oncoprotein induces the T regulatory cell phenotype by activating STAT3. *Proc Natl Acad Sci* **103**: 9964–9969.

- Kuefer MU, Look AT, Pulford K, Behm FG, Pattengale PK, Mason DY, Morris SW. 1997. Retrovirus-mediated gene transfer of NPM–ALK causes lymphoid malignancy in mice. *Blood* **90**: 2901–2910.
- Kurki S, Peltonen K, Latonen L, Kiviharju TM, Ojala PM, Meek D, Laiho M. 2004. Nucleolar protein NPM interacts with HDM2 and protects tumor suppressor protein p53 from HDM2-mediated degradation. *Cancer Cell* 5: 465–475.
- Lee HJ, Kim E, Kim JS. 2010. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res* **20**: 81–89.
- Lee HJ, Kweon J, Kim E, Kim S, Kim JS. 2012. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res* 22: 539–548.
- Lessnick SL, Ladanyi M. 2012. Molecular pathogenesis of Ewing sarcoma: New therapeutic and transcriptional targets. Annu Rev Pathol 7: 145–159.
- Li Z, Boone D, Hann SR. 2008. Nucleophosonin interacts directly with c-Myc and controls c-Myc-induced hyperproliferation and transformation. *Proc Natl Acad Sci* 105: 18794–18799.
- Mani RS, Chinnaiyan AM. 2010. Triggers for genomic rearrangements: Insights into genomic, cellular and environmental influences. *Nat Rev Genet* 11: 819–829.
- Mathas S, Kreher S, Meaburn KJ, Johrens K, Lamprecht B, Assaf C, Sterry W, Kadin ME, 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. *Proc Natl Acad Sci* 106: 5831–5836.
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, et al. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**: 143–148.
- Mitelman F, Johansson B, Mertens F. 2007. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* **7**: 233–245.
- Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, Saltman DL, Look AT. 1994. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science 263: 1281–1284.
- Mussolino C, Cathomen T. 2012. TALE nucleases: Tailored genome engineering made easy. *Curr Opin Biotechnol* **23:** 644–650.
- Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39**: 9283–9293.
- Plougastel B, Zucman J, Peter M, Thomas G, Delattre O. 1993. Genomic structure of the EWS gene and its relationship to EWSR1, a site of tumorassociated chromosome translocation. *Genomics* 18: 609–615.
- Prieur A, Tirode F, Cohen P, Delattre O. 2004. EWS/FLI-1 silencing and gene profiling of Ewing cells reveal downstream oncogenic pathways and a crucial role for repression of insulin-like growth factor binding protein 3. *Mol Cell Biol* 24: 7275–7283.
- Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30: 460–465.
- Richardson C, Jasin M. 2000. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* **405**: 697–700.
- Riggi N, Suva ML, Suva D, Cironi L, Provero P, Tercier S, Joseph JM, Stehle JC, Baumer K, Kindler V, et al. 2008. EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in primary human mesenchymal stem cells. *Cancer Res* 68: 2176–2185.
- Riggi N, Suva ML, De Vito C, Provero P, Stehle JC, Baumer K, Cironi L, Janiszewska M, Petricevic T, Suva D, et al. 2010. EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes Dev* 24: 916–932.
- Roth DB, Porter TN, Wilson JH. 1985. Mechanisms of nonhomologous recombination in mammalian cells. *Mol Cell Biol* **5**: 2599–2607.
- Simsek D, Jasin M. 2010. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4–ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol* **17:** 410–416.
- Simsek D, Brunet E, Wong SY, Katyal S, Gao Y, McKinnon PJ, Lou J, Zhang L, Li J, Rebar EJ, et al. 2011. DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS Genet* 7: e1002080.
- Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, Carroll D. 2000. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* 28: 3361–3369.
- Smith R, Owen LA, Trem DJ, Wong JS, Whangbo JS, Golub TR, Lessnick SL. 2006. Expression profiling of EWS/FLI identifies NKX2.2 as a critical target gene in Ewing's sarcoma. *Cancer Cell* 9: 405–416.
- Tirode F, Laud-Duval K, Prieur A, Delorme B, Charbord P, Delattre O. 2007. Mesenchymal stem cell features of Ewing tumors. *Cancer Cell* 11: 421–429.
- Urnov FD, Rébar EJ, Holmes MC, Zhang HŠ, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11: 636–646.

- Wang W, Budhu A, Forgues M, Wang XW. 2005. Temporal and spatial control of nucleophosmin by the Ran–Crm1 complex in centrosome duplication. *Nat Cell Biol* 7: 823–830.
 Weinstock DM, Brunet E, Jasin M. 2007. Formation of NHEJ-derived
- Weinstock DM, Brunet E, Jasin M. 2007. Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. Nat Cell Biol 9: 978–981.
- Yan CT, Boboila C, Souza EK, Franco S, Hickernell TR, Murphy M, Gumaste S, Geyer M, Zarrin AA, Manis JP, et al. 2007. IgH class switching and

translocations use a robust non-classical end-joining pathway. *Nature* **449:** 478–482.

Zucman-Rossi J, Legoix P, Victor JM, Lopez B, Thomas G. 1998. Chromosome translocation based on illegitimate recombination in human tumors. *Proc Natl Acad Sci* 95: 11786–11791.

Received August 3, 2012; accepted in revised form April 3, 2013.

3 Discussion et perspectives

Durant ma thèse, j'ai étudié les translocations chromosomiques grâce à l'utilisation de nucléases artificielles (ZFN, TALEN, CRISPR/Cas9). Nous avons montré que nous pouvions induire tout type de translocations, même oncogéniques dans différents types cellulaires, même dans des cellules non transformées (article II).

Nos travaux se sont focalisés sur la phase de ligation lors de la formation de translocations chromosomiques. Ces travaux, ont permis de mettre en évidence un rôle prédominant du complexe de ligation XRCC4/LigaseIV du C-NHEJ dans la formation des translocations chromosomiques dans les cellules humaines. Ce résultat inattendu est distinct de ce qui a été décrit pour les cellules murines (Simsek and Jasin, 2010; Simsek et al., 2011b), révélant ainsi une différence selon l'espèce dans la génération de ces événements oncogéniques. Comme mentionné dans les parties précédentes, notre stratégie consistait à exprimer des nucléases artificielles (ie les ZFN, TALEN, et CRISPR/Cas9) qui ciblent différents loci dans différentes lignées cellulaires humaines invalidées pour les protéines du complexe de ligation XRCC4/LigaseIV du C-NHEJ ou exprimant une version mutée de la LigIV. Ainsi, nous avons pu évaluer par screen PCR les variations de fréquence d'apparition des translocations ainsi que la nature des séquences des jonctions des chromosomes transloqués. Nous avons montré qu'en absence du complexe de ligation XRCC4/LigaseIV, la fréquence de translocation est souvent diminuée dans les mutants contrairement à ce qui a été observé dans les cellules murines déficientes pour ces mêmes protéines. Les séquences des jonctions des translocations sont également très différentes de celles obtenues dans les cellules sauvages avec l'apparition de longues délétions et de microhomologies, qui sont caractéristiques d'un mécanisme alt-NHEJ (article I Figure 2, 3 et 5). Nous avons également observé les mêmes résultats en réparation intrachromosomique en induisant une CDB ou deux CDB distantes de 3 kb (article I Figure 1). Nous avons enfin montré que la perte de certains composants du alt-NHEJ, tels que CtIP et la LigIII, n'affectait ni la fréquence de formation des translocations chromosomiques, ni la nature des jonctions de translocations, ce qui confirme que le alt-NHEJ joue un rôle mineur dans la formation de ces translocations dans les cellules humaines (article I Figure 6 et 7).

Pourquoi un mécanisme de formation des translocations chromosomiques différent entre les cellules humaines et les cellules murines ?

Nos résultats sont en contradiction avec les résultats décrits pour les cellules murines. En effet, alors que le C-NHEJ supprime l'instabilité génomique dans les cellules murines, celui-ci semble jouer un rôle prédominant dans la formation des translocations chromosomiques dans les cellules humaines.

Dans les modèles et lignées murins, le C-NHEJ protège contre la formation des translocations. En effet, plusieurs études ont démontré que ces évènements étaient dus à une réparation par le alt-NHEJ. Par exemple, il a été montré que les souris doublement déficientes pour les composants du C-NHEJ et p53, développaient des lymphomes agressifs induits par des coupures illégitimes par RAG sur les loci *lgh* et *c-myc* (Zhu et al., 2002). De même, les cellules B murines actives pour le CSR et déficientes pour les protéines du C-NHEJ montrent une augmentation dans la formation des translocations impliquant le gène *c-myc* (Boboila et al., 2010). Outre les cellules lymphoïdes, l'induction des translocations via des nucléases (ZFN et ISce-I) montrent une augmentation dans la fréquence d'apparition des translocations dans les cellules ES murines déficientes pour les protéines LIGIV, XRCC4 ou Ku70, par rapport aux cellules wt. Cependant dans toutes les lignées (wt ou déficientes) les jonctions des translocations sont inchangées et présentent de grandes délétions et un biais d'utilisation des microhomologies (Simsek and Jasin, 2010; Simsek et al., 2011a). Ces résultats démontrent que les translocations sont produites par le alt-NHEJ même en présence des facteurs de C-NHEJ.

Ces résultats tendent à montrer que le alt-NHEJ est une voie robuste chez la souris, qui est active même en présence du C-NHEJ. Pourquoi une telle différence entre l'homme et la souris ? Pour expliquer cette différence, on peut évoquer notamment des différences d'expression et d'activité qu'il y aurait au niveau des protéines du C-NHEJ entre les cellules murines et humaines. Bien que ces protéines soient conservées entre l'homme et la souris, l'activité DNA-PK serait beaucoup plus élevée dans les cellules humaines par rapport aux cellules de souris (Finnie et al., 1995). De plus, les cellules humaines déficientes pour Ku ne sont pas viables, car Ku jouerait un rôle crucial dans le maintien des télomères (Li et al., 2001; Wang et al., 2009). En revanche, l'absence de cette protéine ne semble avoir aucun impact sur la survie des cellules murines (Taccioli et al.,1994; Zhu et al., 1996). On pourrait

126

ainsi imaginer que le maintien de l'intégrité des extrémités des CDB dans les cellules murines serait moins fondamental, entrainant potentiellement leur dissociation ce qui peut conduire à la formation de translocations « tolérées » dans les cellules de souris. La voie alt-NHEJ serait ainsi plus robuste chez la souris et pourrait suppléer et renforcer la voie classique même quand celle-ci est active. De même, lorsque la voie classique est inhibée lors du CSR dans les cellules B murines, le alt-NHEJ est encore capable de produire 50% des immunoglobulines (Yan et al., 2007). De façon générale, on peut aussi remarquer que la durée de vie des souris est nettement inférieure à celle des hommes. On peut envisager que cela réduit considérablement le risque d'apparition et d'accumulation de mutations même si le alt-NHEJ est utilisé pour réparer les cassures. Par contre d'un point de vue évolutif, chez l'humain, il semble que la fidélité et la robustesse du C-NHEJ soit nécessaire pour diminuer la possibilité d'apparition de mutations (qui potentiellement sont les événements initiaux de cancers) au court de la vie d'un individu.

Pourquoi observons-nous une baisse dans la fréquence d'apparition des chromosomes dérivatifs dans les cellules mutantes ?

Dans les différents mutants LigIV (*ie* les cellules HCT 116 LigIV^{-/-}, X4^{-/-} et N114) nous avons montré une baisse dans la fréquence d'apparition des translocations par rapport aux cellules sauvages dans un certain nombre de conditions expérimentales. Nous nous sommes demandé si cette baisse pourrait être due au fait que ces cellules mutantes soient plus sensibles par rapport aux cellules wt. En effet, après transfection du couple de nucléases, les cellules porteuses de translocations peuvent ne pas survivre, entrainant ainsi une diminution de la fréquence de translocation. Pour vérifier cette hypothèse nous avons transfecté nos 5 lignées (HCT116 wt, LigIV^{+/-}, X4^{+/-}, LigIV^{-/-}, X4^{-/-}) avec les ZFN^{p84} et ZFN^{EWS} ciblant respectivement les chromosomes 19 et 22 pour induire la t(19 ;22). Nous avons ensuite compté le nombre de cellules à 24h, 48h et 72h après la transfection afin de suivre leur survie (Figure 24). Nous avons observé que le nombre de cellules mutantes est environ 15-20% plus bas que celui des cellules sauvages ou hétérozygotes. Néanmoins, cette différence du nombre de cellules après transfection ne peut pas, à elle seule, expliquer la grande différence dans la fréquence de translocations qui est dans les cellules mutantes 5 à 6 fois plus basse que dans les cellules sauvages.

Par la suite, nous avons voulu confirmer que cette baisse de fréquence n'était pas liée à la toxicité du couple de ZFN utilisé. Pour cela, nous avons refait les mêmes expériences mais cette fois-ci en transfectant nos lignées avec d'autres couples de nucléases. Nous avons utilisé la ZFN^{EWS} avec une TALEN, la TAL^{p84} qui coupe à proximité du site ciblé par la ZFN^{p84} pour induire la t(19 ;22). Cette même TALEN a également été utilisé avec une autre TALEN, la TAL^{LAM} qui cible le gène lamine sur le chromosome 1 pour induire cette fois-ci la t(1 ;19). Là aussi, une baisse de fréquence de translocation est observée dans les lignées mutantes avec ces deux autres paires de nucléases (article I Figure 2B). Ceci montre que la baisse de fréquence observée n'est pas seulement liée à la toxicité des ZFN^{P84} et ZFN^{EWS}.



<u>Figure 24</u>: Moyenne du nombre de cellules comptées à 24h et 72h après expression du couple de nucléases (ZFN^{EWS} et ZFN^{P84}) ou (TAL^{NPM} et TAL^{ALK}) dans les lignées wt, hétérozygotes et KO pour le complexe de ligation.

Cependant en utilisant des nucléases qui induisent des translocations oncogéniques, nous avons observé que la différence de fréquence de translocation n'était pas significative entre les lignées mutantes et sauvages (article I Figure 5A). Les nucléases (Cas9 et TALEN) ciblent les gènes *ALK* et *NPM*, et forment la translocation t(2 ;5) retrouvée dans le lymphome anaplasique à grandes cellules (ALCL). Les différents couples de nucléases utilisés créent des

cassures différentes. Ainsi, les TALEN créent des cassures avec des 5' overhangs, quand les nucléases de type CRISPR/Cas9 induisent des cassures franches pour la version sauvage ou des 5' overhangs d'une trentaine de nucléotides sur l'ADN pour la version nickase. En utilisant le couple de TALEN, la fréquence de translocation observée est de 5,1.10⁻⁴ contre $3,37.10^{-4}$ dans les lignées LigIV^{+/-} et LigIV^{-/-} respectivement, soit un facteur 1,5 de différence. En utilisant le système CRISPR/Cas9 la fréquence de translocation est de 13,2.10⁻⁴ contre 11,9.10⁻⁴ dans les lignées LigIV^{+/-}et LigIV^{-/-} pour la Cas9 sauvage. Enfin elle est de 6,18.10⁻⁴ dans les lignées hétérozygotes contre 1,8.10⁻⁴ dans les lignées KO lorsque l'on utilise la Cas9 nickase qui induit des extrémités avec des longs 5' overhangs. Ainsi en l'absence de la LigIV, la fréquence de translocation n'est pas réduite de façon significative pour des translocations oncogéniques. L'efficacité de coupure de ces nucléases étant élevée, le défaut de réparation dans les lignées mutantes pourrait favoriser et augmenter la formation des translocations. Il est également possible que la protéine de fusion NPM-ALK formée par la translocation donne un avantage prolifératif (Zhang et al., 2013). Il est à noter que lorsque l'on quantifie la survie globale des cellules traitées ou non avec les nucléases contre NPM et ALK nous n'observons pas de grande différence globale de la croissance (Figure 24). Cependant ceci n'exclut pas non plus une croissance spécifique des cellules transloquées.

L'autre hypothèse plausible pour expliquer le peu d'efficacité du alt-NHEJ pour induire des translocations dans les cellules humaines, serait la présence de l'hétérodimère Ku aux jonctions des cassures même dans nos cellules mutantes pour le complexe de ligation. Ku ayant une affinité très élevée aux CDB il pourrait rentrer en compétion avec PARP1 et l'empêcher de se lier à la cassure pour induire le alt-NHEJ (Cheng et al., 2011; Wang et al., 2006). En fixant ces extrémités, Ku pourrait être ainsi une barrière et entraver l'activité alt-NHEJ dans les lignées mutantes pour le complexe de ligation. En effet, Il a été montré que l'efficacité de ligation de plasmides épisomiques était inférieure à 5% dans les cellules KO pour le complexe de ligation. Cette efficacité est restaurée à plus de 70% lorsque Ku a été éteint (Fattah et al., 2010). Nous avons voulu voir si la fréquence de translocation pouvait être restaurée dans les lignées mutantes lorsque Ku était éteint. Pour se faire, nous nous sommes procurés la lignée HCT 116 KO conditionnel pour Ku : L4^{-/-}-Ku^{-/flox}, en collaborant avec le Pr E.Hendrickson. Dans ce système d'invalidation conditionnelle, l'induction de

l'activité recombinase par l'hydroxytamoxifène induit la délétion de la séquence comprise entre les deux sites LoxP, ce qui provoque l'invalidation du gène Ku86.

Nous avons traité cette lignée avec l'hydroxytamoxifène pendant 24h afin d'avoir une inhibition de Ku au moment de la transfection (Figure 25A). Ensuite à t=24h, nous avons transfecté le couple de nucléases afin d'induire la translocation. Pour cette expérience on a choisi d'utiliser le système CRISPR/Cas9, avec une Cas9 sauvage (*ie* coupures franches) et des gRNAs qui ciblent les gènes *NPM* et *ALK*. À t= 48 h, l'ADN génomique des cellules traitées et non traitées a été isolé puis analysé. Une PCR est réalisée afin de détecter la présence du chromosome dérivatif der5 (Figure 25B). En inhibant Ku en plus du complexe de ligation (L4^{-/-}-Ku^{-/-}), nos résultats montrent qu'aucun produit de PCR correspondant à la formation de la translocation n'est détecté (Figure 25B; puits 3). Au contraire, dans les cellules non traitées L4^{-/-}-Ku^{-/flox} (*ie* KO pour la ligaseIV et hétérozygote pour Ku) on détecte facilement la formation de der5 (Figure 25B; puits 2).

Un test T7 endocnuléase I, a également été réalisé afin d'évaluer le taux de mutations dans ces cellules. Pour le locus *ALK* et *NPM* un taux de mutation de 21% et de 14% a été relevé respectivement dans les lignées non traitées L4^{-/-}-Ku^{-/flox} (*ie* KO pour la ligaseIV et hétérozygote pour Ku), en revanche le taux de mutations n'est pas détectable aux loci *ALK* et *NPM* dans les lignées traitées (L4^{-/-}-Ku^{-/-}) (Figure25C). Ces résultats montrent que l'inhibition de Ku en plus du complexe de ligation ne restaure en aucun cas l'activité alt-NHEJ, au contraire la réparation décroit très fortement. Nous envisageons de refaire le même type d'expériences sur des cellules uniquement KO pour Ku pour voir quel est le rôle de Ku seul dans la réparation intra et interchromosomique dans les cellules humaines sur des sites endogènes (et non pas seulement sur des plasmides).



<u>Figure 25</u>: Analyse des cellules HCT L4^{-/-}Ku^{-/flox} traitées ou non à l'hydroxytamoxifène. (A) : Western blot pour vérifier le niveau d'expression de Ku. L'expression de la protéine Ku est totalement abolie dans les cellules L4^{-/-}Ku^{-/-} traitées à l'hydroxytamoxifène (à t=24h). Au contraire, les cellulaires non traitées L4^{-/-}Ku^{-/flox} ont un niveau d'expression identique aux cellules contrôles. (B) : Détection de la translocation t(2 ;5) dans les cellules L4^{-/-}Ku^{-/flox} traitées et non traitées. Une PCR nichée est réalisée pour détecter les jonctions du chromosome dérivatif 5 (Der5). (C) : Taux de mutations généré au locus *ALK* et *NPM* après expression des nucléases. Les séquences encadrant les sites de coupure sont amplifiées par PCR et soumises au test T7 endonuclease I. Les fragments non mutés suite aux coupures des gRNAs/Cas9 ne sont pas coupés par la T7 endonucléase (uncut). Les fragments mutés sont coupés (cut).

Les jonctions des translocations retrouvées chez les patients présentent peu de délétions et pas de réel biais d'utilisation de microhomologies

Le alt-NHEJ a été caractérisé comme étant la voie impliquée dans l'apparition des translocations chromosomiques dans les cellules murines, et cela même en présence du C-NHEJ (Simsek and Jasin, 2010; Weinstock et al., 2007). Les jonctions des translocations retrouvées dans les cellules murines montrent de larges délétions et un biais vers l'utilisation des microhomologies. D'autres études menées sur des cellules CHO (Guirouilh-Barbat et al., 2004, 2007), ou encore cellules B murines déficientes pour les protéines du C-

NHEJ montrent également les mêmes caractéristiques au niveau des jonctions de réparation (Difilippantonio et al., 2002; Zhu et al., 2002).

Ces résultats obtenus exclusivement dans les cellules murines ont été jusqu'à présent directement extrapolés aux cellules humaines, sans réelles études moléculaires. Afin de justifier ce raccourci, les auteurs se sont basés sur la nature des jonctions des translocations chromosomiques retrouvées chez les patients et reportées dans la littérature. Par exemple, dans la revue de (Boboila et al., 2012a), les auteurs ont écrit que les jonctions des translocations isolées de patients montrent la présence de grandes délétions, d'insertions et d'usage fréquent de microhomologies. Cependant, en examinant minutieusement ces articles cités dans cette revue et en re-analysant les jonctions des translocations reportées nous n'avons pas retrouvé ces résultats. Ainsi, dans l'article de (Kitada and Yamasaki, 2007), seules deux jonctions ont été séquencées, ce qui est loin d'être suffisant pour pouvoir tirer des conclusions. Dans l'article de (Mattarucchi et al., 2008), les auteurs ont séquencé des jonctions de translocations de patients atteints de Leucémie Myéloïde Chronique (t(9;22)). Ces derniers ont observé un biais d'utilisation de microhomologies. Cependant, en reanalysant ces séquences reportées dans cet article, nous avons constaté que plus de 50% des jonctions analysées possèdent des courtes microhomologies, de moins de 2 pb. Même constat pour les jonctions de translocations de 3 types de cancers liés à des translocations: t(8;21) (Zhang et al., 2002), t(4;11) (Gillert et al., 1999), et la t(9;11) (Langer et al., 2003), où aucun biais significatif d'utilisation des microhomologies n'est observé (Figure 26B). Concernant la présence de délétions aux jonctions, pour chacune des translocations : t(8;21) et la t(4;11), la médiane des délétions est de 5.5 et 14 pb respectivement ce qui est assez bas (Figure 26A). De même, nous n'avons observé aucune présence de délétions ni d'utilisation de microhomologies en analysant les jonctions des translocations dans nos modèles cellulaires porteurs de translocations caractéristiques du sarcome d'Ewing et de l'ALCL (Piganeau et al., 2013).



<u>Figure 26</u> : D'après (Gillert et al., 1999; Langer et al., 2003; Zhang et al., 2002). Analyse des jonctions de translocations retrouvées chez les patients. (A) : Taille des délétions et (B) utilisation des microhomologies ont été analysées aux jonctions des translocations t(8 ;22), t(4 ;11), et t(9 ;22) reportées dans ces articles.

Il est également à noter que les patients portant des mutations pour les protéines du C-NHEJ (*ie* LigIV, la DNA-PKcs, XLF, ou encore Artemis) présentent majoritairement un retard de croissance, un syndrome d'immunodéficience combiné sévère et une radiosensibilité. Ces patients ne sont pas réellement prédisposés à développer des cancers (Woodbine et al., 2014) (voir Table 1).

Le C-NHEJ est un mécanisme conservatif

En utilisant la nucléase CRISPR/Cas9 qui induit des cassures franches, nous avons observé que les extrémités des chromosomes impliquées dans le réarrangement chromosomique sont très conservées dans de nombreuses translocations. En effet, II y a peu ou pas de perte de part et d'autre de la jonction de réparation. D'ailleurs, la médiane du nombre de nucléotides perdus est de 1pb dans les lignées LigIV^{+/-}, alors que celle-ci est 283 fois plus élevée dans les lignées mutantes LigIV^{-/-}. Cela implique que le C-NHEJ est particulièrement efficace et indispensable pour le maintien de l'information aux extrémités franches. Cela

semble compatible avec le fait que le C-NHEJ est un mécanisme très conservatif lorsque les extrémités sont parfaitement complémentaires (Bétermier et al., 2014; Rass et al., 2012).

Rôle des autres facteurs du NHEJ dans la formation des translocations chromosomiques

La protéine XLF appartient au même complexe que XRCC4/LigaseIV lors du mécanisme de réparation par le C-NHEJ. Cette protéine forme un filament en interagissant directement avec XRCC4 afin de faciliter l'étape de ligation en favorisant la réparation d'extrémités noncohésives. Chez la souris, une déficience en LigIV ou XRCC4 entraine une létalité embryonnaire (pour revue : Deriano and Roth, 2013). XLF interagit avec ces deux protéines on pourrait donc imaginer qu'un KO XLF dans un modèle murin pourrait avoir les mêmes effets, or ce n'est pas le cas. En effet, les souris invalidées pour cette protéine, bien que sensibles aux IR ont un développement embryonnaire et lymphocytaire correct (Li et al., 2008; Vera et al., 2013). Ce qui laisserait imaginer que d'autres facteurs tels que ATM, H2AX ou 53BP1 pourraient compenser cette perte de fonction de XLF notamment au cours de la recombinaison V(D)J (Li et al., 2008; Liu et al., 2012; Oksenych et al., 2012; Zha et al., 2011). De même, dans les cellules humaines une compensation similaire pourrait avoir lieu, car les patients déficients pour XLF montrent une immunodéficience moins sévère que les individus déficients pour les autres protéines du C-NHEJ (Ahnesorg et al., 2006; Buck et al., 2006). En revanche, la réparation de cassures induites par des IR ou encore des agents génotoxiques est fortement affectée par l'absence de XLF comme démontré dans des fibroblastes humains (Buck et al., 2006), dans des lignées cellulaires humaines (Ahnesorg et al., 2006), ou encore dans les ES souris (Zha et al., 2007). Ce paradoxe suggérerait que le complexe de ligation (avec XLF) est indispensable en dehors du contexte de la recombinaison V(D)J. De même, des observations similaires que celles reportées pour les ES de souris ou encore les lignées de patients ont été relevées dans des cellules somatiques déficientes pour XLF (HCT116). Ces cellules montrent également une sensibilité aux IR et aux agents génotoxiques, et une diminution de l'efficacité de la recombinaison V(D)J (évaluée à l'aide de substrats extrachromosomiques) (Fattah et al., 2014). Cette lignée que nous avons à notre disposition est par conséquent validée comme modèle afin d'élucider les fonctions d'XLF. Nous envisageons donc de faire le même type d'expériences pour comprendre le rôle de la protéine XLF sur la formation des translocations chromosomiques. Dans les ES murines, il a été montré une augmentation de l'instabilité génomique et du taux de translocation. Néanmoins, cette augmentation est 2 fois moins importante que celle observée dans les cellules ES déficientes pour XRCC4 (Zha et al., 2007).

De même, PARP1 a été identifié comme étant impliqué dans la formation des translocations chromosomiques dans les cellules humaines. En effet, en inhibant cette protéine soit par des inhibiteurs soit par des siRNA, l'apparition des translocations est diminuée d'environ 20 fois dans les Jurkat, HEK293, ou encore dans les fibroblastes WI38 (Wray et al., 2013). Cependant une étude contredisant ces résultats a récemment été publiée, où aucun effet n'est observé lorsque PARP1 est inhibé dans des cellules HCT116 wt, par contre son inhibition dans les lignées HCT116 LigIV^{-/-} diminue les translocations d'un facteur 2 (Soni et al., 2014). Aux vues de ces résultats contradictoires, nous envisageons de refaire le même type d'expériences en inhibant PARP1 afin de définir le rôle exact de cette protéine. En effet nous utilisons différents types de nucléases (3 types) et nous pouvons quantifier précisément le taux de translocation contrairement à ce qui a été décrit dans l'étude de Wray et al où la détection des translocations a été faite en PCR non quantitative.

Enfin en collaboration avec le Dr. P. Calsou à Toulouse nous avons à notre disposition des cellules shDNA-PKcs/LigIV double déficientes. En effet, il a été montré que la LigIV a aussi une activité précoce qui permet l'autophosphorylation de la DNA-PK, qui agit très tôt dans la reconnaissance et la signalisation des CDB dans la voie C-NHEJ. L'étude de la formation des translocations dans ces cellules, nous permettra de savoir si le rôle de la LigIV dans ce mécanisme est un rôle précoce (directement en jouant sur la phosphorylation de la DNA-Pk comme montré par (Cottarel et al., 2013) ou tardif (juste au niveau de l'étape de ligation).

Quels sont les facteurs qui préviendraient la formation des translocations chromosomiques ?

Nos résultats montrent clairement un rôle majeur du C-NHEJ aussi bien dans la réparation intrachromosomique, que dans l'apparition des translocations chromosomiques. Ce mécanisme, qui a pour rôle de conserver l'intégrité du génome, conduit également à son instabilité en participant à la formation de structures interchromosomiques. Une des

135
questions majeures à laquelle il faudrait répondre est : quels seraient les facteurs qui pourraient prévenir et limiter l'apparition de ces réarrangements ? Parmi les candidats potentiels il y a les facteurs qui agissent dans la détection et la signalisation de ces CDB, notamment H2AX, 53BP1 ou encore ATM. En effet, il a été montré que ces protéines étaient nécessaires pour prévenir la formation de translocations suite aux CDB générées par AID entre le locus Igh et c-Myc (Franco et al., 2006a, 2006b; Ramiro et al., 2006). Les souris double déficientes pour H2AX et p53 développent des lymphomes de cellules T et B. Dans le cas des cellules B, les translocations les plus récurrentes ont lieu entre le locus *Igh* et c-Myc (Bassing et al., 2003; Celeste et al., 2003). L'histone H2AX est également impliqué dans la protection des extrémités lors de la recombinaison V(D)J en empêchant l'ouverture des extrémités codantes en épingle à cheveux par une autre nucléase qu'Artémis qui est spécifiquement associée au NHEJ pour cette étape (Helmink et al., 2011). La protéine 53BP1 semble influencer la dynamique chromatinienne pour rapprocher et faciliter la réparation des segments éloignés (Difilippantonio et al., 2008; Manis et al., 2004). Elle régule également la mobilité des extrémités de l'ADN, en particulier au niveau télomères déprotégés (de Lange, 2009). Ces résultats suggèrent que ces protéines permettraient la stabilisation des extrémités de la CDB afin d'empêcher leur dissociation et ainsi promouvoir une ligation appropriée, prévenant ainsi l'apparition de translocations (Bassing et al., 2003; Difilippantonio et al., 2008; Franco et al., 2006a). Des résultats récents montrent que BLM interagit avec la protéine 53BP1 et que le complexe 53BP1/BLM inhibe la résection des extrémités double brin de l'ADN et régule ainsi la réparation par la voie NHEJ (Grabarz et al., 2013). Analyser les translocations dont la formation dépend du C-NHEJ, dans un fond génétique déficient pour la protéine BLM, 53BP1, H2AX ou encore ATM permettra donc de comprendre précisément les mécanismes moléculaires qui induisent l'instabilité génomique.

Le NHEJ et la thérapie anticancéreuse

L'accumulation de CDB non réparées ou mal réparées sont hautement génotoxiques pour la cellule, ce concept est de ce fait exploité en thérapie anticancéreuse. Cette thérapie englobe la radiothérapie et la chimiothérapie qui génère de grandes quantités de CDB dans toutes les

cellules. Cependant, les cellules cancéreuses sont beaucoup plus sensibles, car celles-ci se divisent plus rapidement, elles sont souvent inactivées pour des facteurs de leurs machineries de réparation d'ADN (ex : BRCA1, BRCA2 mutées), mais sont également déréglées pour les points de contrôles du cycle cellulaire (inactivation de p53 par exemple).

À l'heure actuelle l'utilisation de la chimiothérapie conjointement avec la radiothérapie fait preuve d'une grande efficacité. Les agents radio-sensibilisants peuvent également être utilisés en combinaison avec la radiothérapie et la chimiothérapie ce qui peut grandement améliorer leur efficacité. Certains agents chimiotérapeutiques peuvent être utilisés en monothérapie lorsqu'ils induisent une "létalité synthétique". Celle-ci exploite le fait que de nombreuses cellules cancéreuses ont ou acquièrent des défauts dans l'une de leurs voies de réparation d'ADN, et deviennent ainsi dépendantes d'un autre mécanisme de réparation afin de survivre. L'inhibition de la voie de réparation compensatrice permettra de tuer sélectivement les cellules déficientes pour une réparation particulière (Farmer et al., 2005; The et al., 2005). D'ailleurs, ce principe de la létalité synthétique est appliqué pour traiter les cancers du sein, où la voie de recombinaison homologue est altérée et cela en utilisant des inhibiteurs de PARP afin d'altérer le NHEJ (Farmer et al., 2005; The et al., 2005).

Nous avons identifié le C-NHEJ comme étant la voie impliquée dans la formation des translocations chromosomiques dans les cellules humaines. Une inhibition sélective de cette voie pendant le traitement de la tumeur primaire pourrait prévenir l'apparition possible de leucémies secondaires (caractérisées par des translocations) à la suite d'une chimioradiothérapie. Cela pourrait également prévenir les mutations qui stimuleraient la croissance de ces cancers et la survie. L'inhibition de cette voie à long terme pourrait cependant entrainer des effets indésirables importants. En effet, cette voie est impliquée dans les processus physiologiques fondamentaux notamment le V(D)J et le CSR. Inhiber cette voie c'est également laisser place au alt-NHEJ qui est très mutagène, ce qui augmenterait l'instabilité génomique. Cibler ces deux voies au même temps pourrait s'avérer efficace et permettrait de prévenir la mutagène associée au alt-NHEJ. Ainsi, des inhibiteurs des composants du C-NHEJ peuvent être utilisés en combinaison avec des inhibiteurs de PARP. D'ailleurs, une étude a montré qu'en traitant les cellules murines et de hamster en combinant un inhibiteur de PARP (AG14361) avec un inhibiteur de la DNA-PK

137

augmenterait la sensibilité de ces cellules (Veuger et al.,2003). De même, des cellules déficientes pour la LigaseIV ont été radiosensibilisées après avoir été traitées par un inhibiteur de PARP (Löser et al., 2010).

4 Conclusions

Nos résultats ont mis en évidence un rôle majeur du complexe de ligation XRCC4/LigaseIV dans la détermination de la nature des séquences des jonctions des translocations et donc l'implication du C-NHEJ dans le mécanisme de formation des translocations. En effet, en son absence, la majorité des jonctions séquencées pour les cellules mutantes présentent de grandes délétions ce qui n'est pas le cas dans les cellules exprimant ce complexe. De plus, dans les cellules mutantes, il y a une prédominance de séquences qui présentent des microhomologies. Ce mécanisme est différent de celui retrouvé chez la souris. Ce projet nous a donc permis de mieux comprendre le mécanisme de formation des translocations dans les cellules humaines. A plus long terme, il sera nécessaire d'identifier et de comprendre le rôle de chacune des protéines nécessaires à la formation des translocations, en particulier dans les étapes de migration et de maintien des extrémités des chromosomes transloqués. Il sera important de déterminer comment les différents composants de la réponse aux dommages de l'ADN affectent les différentes étapes de la formation de ces structures mais également comment elles modulent la fréquence de translocations. De telles études sont utiles pour la conception de thérapies anticancéreuses plus efficaces, mais peuvent être également particulièrement utiles pour prévenir l'apparition de translocations initiatrices de tumeurs secondaires, qui sont un problème majeur lors de la chimioradiothérapie. Comprendre ce mécanisme permettra également d'expliquer pourquoi les modèles tumoraux murins ne corrèlent pas toujours avec l'apparition de tumeurs chez l'homme.

Annexe : Article méthodes

CHAPTER TWELVE

Creating Cancer Translocations in Human Cells Using Cas9 DSBs and nCas9 Paired Nicks

Benjamin Renouf^{*}, Marion Piganeau^{*}, Hind Ghezraoui^{*}, Maria Jasin^{†,1}, Erika Brunet^{*,1}

*Museum National d'Histoire Naturelle, INSERM U1154, CNRS 7196, Paris, France [†]Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, New York, USA ¹Corresponding authors: e-mail address: ebrunet@mnhn.fr; m-jasin@ski.mskcc.org

Contents

	254
2. Materials	254
2.1 Cas9, nCas9, and sgRNA expression plasmid preparation	254
2.2 Cell culture and transfection	254
2.3 T7 endonuclease I assay	255
2.4 PCR detection of translocations	255
2.5 PCR quantification of translocations	255
3. Methods to Induce and Detect Cancer Translocations in Human Cells	256
3.1 sgRNA design and expression plasmid construction	256
3.2 Cell transfections with sgRNA and Cas9 or nCas9 expression plasmid	s 260
3.3 T7 endonuclease I assay to estimate cleavage efficiency	262
3.4 PCR-based translocation detection	263
3.5 Quantification of potential off-target cleavage	264
3.6 Quantification of translocation frequency using a 96-well plate screen	n 266
3.7 Translocation frequency determination by serial dilution	268
4. Conclusions	269
Acknowledgments	269
References	269

Abstract

Recurrent chromosomal translocations are found in numerous tumor types, often leading to the formation and expression of fusion genes with oncogenic potential. Creating chromosomal translocations at the relevant endogenous loci, rather than ectopically expressing the fusion genes, opens new possibilities for better characterizing molecular mechanisms driving tumor formation. In this chapter, we describe methods to create cancer translocations in human cells. DSBs or paired nicks generated by either wild-type Cas9 or the Cas9 nickase, respectively, are used to induce translocations at the relevant loci. Using different PCR-based methods, we also explain how to quantify translocation frequency and to analyze breakpoint junctions in the cells of interest. In addition, PCR detection of translocations is used as a very sensitive method to detect off-target effects, which has general utility.

1. INTRODUCTION

The discovery that a chromosomal translocation was associated with oncogenesis was a watershed event in tumor biology research (Chandra et al., 2011; Rowley, 1973). Hundreds of recurrent reciprocal translocations have now been found in a variety of human cancers, including hematological malignancies, sarcomas, and epithelial tumors (Mani & Chinnaiyan, 2010; Mitelman, Johansson, & Mertens, 2007). These translocations are considered primary causes of many cancers and have been important for the development of targeted therapies. A typical consequence of a chromosomal translocation is the formation of a gene fusion that leads to expression of a novel protein with oncogenic potential; alternatively, translocations can lead to a gene coming under the control of a strong promoter, such that overexpression confers an oncogenic property. The cellular consequences of fusion gene expression have been widely investigated using ectopic expression in cells without the translocation or gene silencing in cells carrying the translocation. However, these studies are not optimal for several reasons. Ectopic expression does not recapitulate the genetics of the disease and may lead to nonphysiological levels of fusion gene expression, and silencing the fusion gene in tumor cells does not take into account the numerous other mutations that the tumor cells have acquired. Thus, having methods to induce translocations at will in a variety of cell types would provide a significant advantage to cancer researchers.

Translocations involve chromosome breakage and aberrant joining of DNA ends. Two concurrent DNA double-strand breaks (DSBs), one on each chromosome, have been shown to induce translocations (Richardson & Jasin, 2000). Joining of the DNA ends typically involves some type of nonhomologous end-joining (NHEJ) repair (Weinstock, Elliott, & Jasin, 2006). The simplest system for inducing concurrent DSBs is the expression of a nuclease which cleaves specific sites. Initial studies performed in model systems used the I-SceI endonuclease, a homing endonuclease from yeast, in which I-SceI sites were introduced at specific loci.

However, the development of designer nucleases (ZFNs and TALENs) has allowed DSBs to be introduced into genomes without prior modification (Gaj, Sirk, & Barbas, 2014; Urnov, Rebar, Holmes, Zhang, & Gregory, 2010). As a result, chromosomal translocations could be readily induced in human and mouse cells at endogenous loci (Brunet et al., 2009; Piganeau et al., 2013; Simsek et al., 2011), including translocations associated with human tumors. EWS–FLI1 translocations associated with Ewing sarcoma were induced by ZFNs directed to the EWS and FLI1 loci in mesenchymal cells (Piganeau et al., 2013). NPM–ALK translocations associated with anaplastic large cell lymphoma (ALCL) were induced by TALENs directed to the NPM and ALK loci in Jurkat cells (Piganeau et al., 2013). Of note, the ALCL translocation could also be reversed in patient cell lines carrying the translocation, by TALENs directed to the NPM–ALK and ALK–NPM loci. The versatility of TALENs suggests that designer nucleases could be used to induce translocations involving any loci.

The most recently developed designer nucleases are RNA-guided, which as such are easiest to design. Currently, the most commonly used nuclease is Cas9 from S. pyogenes (Cong et al., 2013; Hsu, Lander, & Zhang, 2014; Mali, Yang, et al., 2013). The guide RNA (sgRNA) has ~ 20 nucleotides of sequence complementary to a target site, followed by a Protospacer Adjacent Motif (PAM) sequence (NGG) which is critical for binding to Cas9. When both Cas9 and the sgRNA are expressed in cells, the target site is cleaved on both strands a few nucleotides away from the PAM, creating a DSB (Jinek et al., 2012). Because Cas9 has two active sites, each cleaving a defined strand, Cas9 can be converted to a nickase by mutation of one active site (nickase Cas9 or nCas9). For example, Cas9 D10A only cleaves the DNA strand complementary to the sgRNA. When two sgRNAs are provided which bind opposite DNA strands in close proximity, however, paired nicks can be introduced, also creating a DSB but with a long overhang (Mali, Yang, et al., 2013; Ran et al., 2013). Paired nicks are considered to have fewer potential off-target sites, since two distinct sgRNAs are required for double-strand cleavage. Following the principles established with TALENs and ZFNs (Piganeau et al., 2013), Cas9 has also been used to induce NPM-ALK and other tumor-relevant translocations (Choi & Meyerson, 2014; Ghezraoui et al., 2014; Torres et al., 2014). More recently, nCas9-induced paired nicks have also been used to induce translocations (Ghezraoui et al., 2014). In this chapter, we detail methods for the induction of translocations by Cas9 and nCas9, using a PCR screen for translocation junctions (Brunet et al., 2009).

2. MATERIALS

2.1. Cas9, nCas9, and sgRNA expression plasmid preparation

- Expression plasmids can be obtained from Addgene (https://www. addgene.org/CRISPR/). We use pCas9_GFP (Addgene plasmid 44719) to induce DSBs and pCas9D10A_GFP to induce paired nicks (nCAS9) (Addgene plasmid 44720), together with sgRNA expression plasmids derived from MLM3636 (Addgene plasmid 43860).
- 2. Competent bacteria, e.g., DH5α.
- 3. LB agar plates with antibiotic (ampicillin for the plasmids above).
- 4. LB medium.
- 5. PureLink[®] HiPure Plasmid Filter Maxiprep Kit (Invitrogen).
- 6. NanoDrop 2000c (Thermo Scientific).

2.2. Cell culture and transfection

- 1. RPE1 (hTert-RPE1) cells or mesenchymal stem cells (MSC) are used in this chapter, although the approach is applicable to any other cell type that can be transfected.
- Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F-12, Life Technologies) with 10% Fetal Bovine Serum (FBS) for RPE1 cells; alpha-Minimum Essential Eagle Medium (αMEM, Life Technologies), supplemented with 10% FBS and 2 ng/mL bFGF (Recombinant Human FGF basic (146 aa) 233-FB-025 R&D systems) for MSC.
- **3.** T150 flasks, 150 cm².
- 4. Cas9, nCas9, and sgRNA expression plasmids, at concentrations $\geq 2 \,\mu g/\mu L$.
- **5.** 6-well plates.
- 6. 48-well plates.
- 7. 96-well plates.
- 8. 0.05% Trypsin–EDTA (1×).
- 9. Dulbecco's Phosphate-Buffered Saline (DPBS, Life Technologies).
- 10. Nucleofector II device (Lonza).
- 11. Cell Line Nucleofector Kit V with cuvettes (Lonza).

2.3. T7 endonuclease I assay

- 1. QIAamp DNA Mini Kit (Qiagen).
- Primers: Primers may be designed using a variety of programs such as Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/ primer3plus.cgi/). Settings are chosen to yield 22 bp primers with melting temperatures ~62 °C.
- 3. Phusion High-Fidelity DNA Polymerase (Thermo Scientific).
- 4. T7 endonuclease I (New England Biolabs).
- 5. NEBuffer 2.1 (New England Biolabs).
- 6. $2 \times$ T7 loading buffer: 50% sucrose, bromophenol blue, 260 µg/mL proteinase K.
- 7. 2.4% agarose gel with Ethidium Bromide (EtBr).
- 8. $0.5 \times$ TBE running buffer (Life Technologies).
- 9. UV station.

2.4. PCR detection of translocations

- Primers for nested PCR (two sets of primers). Settings are chosen to yield 20 bp primers with melting temperatures ~60 °C.
- 2. FastStart Taq DNA polymerase (Roche).
- 3. 1% agarose gel with EtBr.
- **4.** $0.5 \times$ TBE running buffer.
- 5. UV station.

2.5. PCR quantification of translocations

- 10 × Lysis buffer: 100 mM Tris-HCl (pH 8), 4.5% NP40, 4.5% Tween20.
- 2. Proteinase K (New England Biolabs).
- 3. 2× Master Mix 1:1× GC-RICH solution (Roche FastStart Taq), 2× PCR Buffer with 20 mM MgCl₂ (Roche FastStart Taq), 400 μM dNTP mix (Roche FastStart Taq), 4% DMSO, 0.01% Tween20, 0.01% NP40. Store at 4 °C for up to 2 weeks.
- 4. 2× Master Mix 2:1× GC-RICH solution (Roche FastStart Taq), 2× PCR Buffer with 20 mM MgCl₂ (Roche FastStart Taq), 400 μM dNTP mix (Roche FastStart Taq), 4% DMSO, 0.01% Tween20, 0.01% NP40, 0.25 μL of SYBR[®] Green I (10,000× in DMSO, Sigma–Aldrich) diluted in 2 mL of 20% DMSO and Rox (concentrations depend on

the real-time PCR machine; 30 nM of Rox is used with MX3005P from Agilent). Store at 4 °C for up to 2 weeks.

- 5. Primers for nested PCR.
- 6. FastStart Taq DNA polymerase (Roche).
- 7. Real-time quantitative PCR machine (MX3005P, Agilent).

3. METHODS TO INDUCE AND DETECT CANCER TRANSLOCATIONS IN HUMAN CELLS

Both Cas9 and nCas9 can be used to generate translocations. We have found that Cas9 is somewhat more efficient (Ghezraoui et al., 2014), but nCas9 has fewer off-target concerns. Two translocations are used to illustrate the methods described in this chapter: NPM–ALK found in ALCL (Elmberger, Lozano, Weisenburger, Sanger, & Chan, 1995; Kuefer et al., 1997; Morris et al., 1994) (Fig. 12.1A) and EWS–FLI1 found in Ewing sarcoma (May et al., 1993) (Fig. 12.2A).

3.1. sgRNA design and expression plasmid construction

1. For the translocation of interest, a good starting point for designing sgRNA target sequences is close to reported breakpoint junctions in patients, if available. If the translocation generates a fusion gene, target sites within introns involved in the translocation are requisite. When using wild-type Cas9, design two sgRNAs, one to each chromosome; when using nCas9 to induce paired nicks, design two pairs of sgRNAs to each chromosome. The ~ 20 bp target sequence is located upstream of the requisite NGG PAM sequence. For a 20 bp target sequence, follow the simple rule N₂₀NGG for Cas9 and CCNN₂₀-spacer-N₂₀NGG for nCas9. The spacer can be a few bp or longer (Mali, Aach, et al., 2013; Ran et al., 2013). Specific Websites can be used to minimize the number of off-target sites of the sgRNAs (e.g., http://crispr.mit.edu/, http://zifit.partners.org/ZiFiT/Disclaimer.aspx, https://chopchop.rc.fas. harvard.edu/; see also Montague, Cruz, Gagnon, Church, & Valen, 2014; Xie, Shen, Zhang, Huang, & Zhang, 2014). The U6 promoter, which prefers G as the transcription start site, drives sgRNA expression in MLM3636; thus, either the target sequence should begin with G or a G should be inserted before the target sequence.



Figure 12.1 Induction of the NPM–ALK translocation by Cas9 DSBs or nCas9 paired nicks. (A) Induction of DSBs or paired nicks at the *NPM* and *ALK* loci can lead to chromosomal translocations and expression of the NPM–ALK fusion protein from Der5. (B) Position of the sgRNAs. With wild-type Cas9, DSBs are formed using sgRNAs NPM1 and ALK1. With nCas9, paired nicks are formed using NPM1+NPM2 (37 bp apart) and ALK1+ALK2 (41 bp apart), resulting in 5'-overhangs if the strands are separated. (C) Translocations are only detected when two DSBs or two paired nicks are introduced. PCR is performed with primers indicated in (A). (D) Estimate of cleavage efficiency at the ALK locus using wild-type Cas9 and either the ALK1 or ALK2 sgRNA and nCas9 with both the ALK1 and ALK2 sgRNAs. The T7 assay detects about 60% indel formation in all three instances.



Figure 12.2 sgRNA off-target analysis and estimation of translocation frequency by serial dilution. (A) The translocation of interest is EWS–FL11. The EWS sgRNA has a potential off-target site that contains two mismatches on another chromosome. (B) Using the T7 assay, the EWS sgRNA with Cas9 gives detectable indel formation at the EWS locus, while no cleavage is observed at the off-target site. Using the translocation assay, the EWS and FL11 sgRNAs with Cas9 give a detectable signal in all four samples (150 ng DNA each); however, a signal is also observed for translocation between the off-target site and the EWS locus. (C) Estimate of translocation frequency using serial dilution of genomic DNA from RPE1 cells using Cas9 and sgRNAs EWS and FL11. In this example, the frequency of Der22 formation is $\geq 0.375 \times 10^{-3}$, determined for three positive wells from (((4 total wells $\times 12.5$ ng DNA)/6.25 ng) $\times 10^3$ cells).

NPM-ALK

Cas9-generated DSBs with sgRNAs NPM1 and ALK1; nCas9-generated paired nicks with sgRNAs NPM1+NPM2 and ALK1+ALK2 (Fig. 12.1B). NPM genomic sequence followed by sgRNA sequences:

5'-**CCT**<u>CGAACTGCTACTGGGTTCAC</u>CTCA<u>GCCTCTGGAA-TAGCTAGAACTAC</u>**AGG**-3' sgRNA NPM1: 5'-GTGAACCCAGTAGCAGTTCG-3' sgRNA NPM2: 5'-GCCTCTGGAATAGCTAGAACTAC-3' ALK genomic sequence followed by sgRNA sequences: 5'-**CCT**<u>CAGGTAACCCTAATCTGATCACGGTCGGTCCATT-GCATAGAGG</u>**AGG**-3' sgRNA ALK1: 5'-GATCAGATTAGGGTTACCTG-3' sgRNA ALK2: 5'-GTCGGTCCATTGCATAGAGG-3'

 For each sgRNA, order the synthesis of two DNA oligonucleotides (sense and antisense) that would anneal to BsmBI-linearized plasmid MLM3636. (http://zifit.partners.org/ZiFiT/CSquare9GetOligos.aspx) Resuspend at 100 μM in annealing buffer (10 mM Tris (pH 7.5), 1 mM EDTA, 50 mM NaCl).

sgRNA NPM1 oligonucleotides:
5'-ACACC <u>GTGAACCCAGTAGCAGTTCG</u> G-3'
5'-AAAACCGAACTGCTACTGGGTTCACG-3'
sgRNA ALK1 oligonucleotides:
5'-ACACCGATCAGATTAGGGTTACCTGG-3'
5'-AAAACCAGGTAACCCTAATCTGATCG-3'

- 3. Anneal the two oligonucleotides to generate a duplex. Add 10 μ L each of the sense and antisense oligonucleotides to 80 μ L of annealing buffer. Place tube in a standard heating block at 95 °C for 5 min. Remove the heating block from the apparatus and allow to cool to room temperature on the workbench. Slow cooling to room temperature should take 45–60 min. Store on ice or at 4 °C until ready to use.
- 4. Ligate the duplex into BsmBI-linearized sgRNA plasmid MLM3636 vector.
- **5.** Transform into competent bacteria. Spread on a LB agar plate and incubate overnight at 37 °C.
- 6. Proceed to plasmid isolation with PureLink[®] Quick Plasmid *Miniprep* Kit (Invitrogen). Because the BsmBI overhangs are distinct (GTGT

and TTTT), a single duplex oligonucleotide should ligate in the correct orientation. Sequencing is recommended to confirm.

- 7. For a positive clone, proceed to plasmid isolation with PureLink[®] HiPure Plasmid Filter *Maxiprep* Kit (Invitrogen). Measure the DNA concentration (e.g., using a NanoDrop 2000c spectrophotometer). The final DNA plasmid concentration should be $>2 \ \mu g/\mu L$.
- 8. Note: If the translocation rate is very low, sgRNAs can be cloned directly in the same plasmid as Cas9 (pX330-U6-Chimeric_BB-CBh-hSpCas9, Addgene plasmid 42230) or nCas9 (pX335-U6-Chimeric_BB-CBh-hSpCas9n(D10A), Addgene plasmid 42335). Other sgRNA plasmids contain markers such as GFP (Addgene plasmid PX458) for FACS sorting or the puromycin resistance gene (Addgene plasmid PX459) for selection to increase the recovery of transfected cells.

3.2. Cell transfections with sgRNA and Cas9 or nCas9 expression plasmids

The protocol has been optimized for RPE1 and MSC cells, but any other cell line of interest can be used.

Transfection can induce high cell lethality, which can be explained by (i) inappropriate cell culture conditions, (ii) inappropriate transfection program, (iii) lack of sgRNA specificity, and (iv) induction of a lethal translocation. Check these parameters to reduce mortality.

- RPE1 and MSC cells are cultured in a T150 flask in 20 mL of DMEM/ F-12 medium and supplemented αMEM medium, respectively, at 37 °C and 5% CO₂. Passage every 2–3 days at a split ratio of 1:5 to 1:10, keeping them at ≤80% confluency.
- 2. The day before transfection, pass the cells without antibiotics (to decrease lethality during electroporation) to reach \sim 70–80% confluency on the day of transfection.
- 3. On the day of transfection, prepare Eppendorf tubes containing 3.5 μ g of each sgRNA plasmid and 3.5 μ g of the Cas9 or nCas9 expression plasmid. As controls, transfect one of the two sgRNA expression plasmids for Cas9 or two of the four sgRNA expression plasmids for nCas9 and the same quantity of the sgRNA expression plasmid in which no sgRNA was cloned. For nCas9, you can transfect one sgRNA each targeting the two chromosomes to exclude nick-induced translocations (Fig. 12.1C). The total volume of transfected DNA should not exceed 10 μ L.
- **4.** Prefill 2–3 wells of a 6-well plate with 1 mL medium and prewarm to 37 °C.

- 5. Prefill three 96-well plates with 50 μ L medium per well and prewarm to 37 °C.
- 6. Trypsinize the RPE1 or MSC cells and resuspend in prewarm supplemented medium DMEM/F-12 or α MEM, respectively (typically 10 mL). Do not use 4 °C media at any time.
- 7. Count the cells and for each transfection put 7.5×10^5 cells in a 15-mL Falcon tube. The program and the number of cells must be adjusted for each cell line (refer to Lonza Nucleofector Protocols). Conditions described here are optimized for MSC and RPE1 cells.
- 8. Centrifuge 10 min at $90 \times g$ at room temperature.
- **9.** Carefully remove the medium without aspirating the cells. FBS from the medium may inhibit transfection. A supplemental wash with DPBS may increase the transfection efficiency.
- 10. Carefully resuspend cells in $100 \ \mu L$ of Cell Line Nucleofector Kit V (Lonza) solution. To prevent cell lethality, avoid all unnecessary cell agitation.
- **11.** Transfer the cells into the tube containing DNA and then transfer the cell/DNA mix into the Amaxa DNA cuvettes.
- Electroporate using the Nucleofector II system (Lonza) on program B-016 for MSCs and X-001 for RPE1 cells.
- 13. Transfer transfected cells to 5 mL of prewarmed medium.
- 14. Dilute the cells as follows to a final volume of 6 mL: 1.2 mL (1/5), 600 μ L (1/10), and 300 μ L (1/20). Plate 50 μ L of each cell dilution into each well of a prewarmed 96-well plate. In addition, transfer 1 mL of the 5 mL of cells to a well of a 48-well plate for cell counts (see step 3.6.1).
- 15. Transfer the rest of the cells into wells of a 6-well plate (step 3.2.4) for further DNA, protein, or chromosome analysis. If the translocation frequency is high enough, translocations can be directly detected by Western blotting using antibodies directed against the fusion protein (frequency $\geq 10^{-3}$) or by microscopy using FISH probes (frequency $\geq 10^{-2}$) (Piganeau et al., 2013). Prepare supplementary wells during the transfection if you plan these additional analyses.

Induction of the *NPM*–*ALK* translocation is frequent enough that expression of the NPM–ALK fusion protein can be detected in the bulk population of transfected cells (Fig. 12.1A).

16. Incubate cells at 37 $^{\circ}$ C, 5% CO₂.

17. Note: To estimate transfection efficiency, our experiments utilize pCas9_GFP and pCasD10A9_GFP in which GFP is expressed as a 2A fusion. The percent GFP-positive cells is determined by flow cytometry 48 h after transfection. Poor transfection can result in low translocation efficiency. Test several programs to optimize the transfection efficiency for each cell line.

3.3. T7 endonuclease I assay to estimate cleavage efficiency

This assay estimates the efficiency of sgRNAs to direct cleavage by quantifying insertions and deletions (indels) resulting from DSB repair via NHEJ (Guschin et al., 2010). For nCas9 paired nicks, the estimate is done using each of the sgRNAs separately with wild-type Cas9 as well as with the two sgRNAs together with nCas9 (Fig. 12.1D).

Several sgRNAs can be designed to target the same locus, such that sgRNAs that are found to efficiently induce indels can be used to generate translocations.

- 1. Five days after transfection, trypsinize cells in the 6-well plate (from step 3.2.15).
- **2.** Centrifuge at $200 \times g$ at $4 \degree C$ for 5 min.
- 3. Remove supernatant and wash the pellet with 1 mL of DPBS.
- 4. Repeat centrifugation.
- 5. Remove supernatant. Cell pellets can be stored at -80 °C.
- 6. Extract genomic DNA with the QIAamp DNA Mini Kit (Qiagen). Genomic DNA can be stored at -20 °C.
- 7. Design a set of primers amplifying a 300–500 bp region encompassing the target sites on nontranslocated chromosomes. Cleavage sites should not be located in the middle of the amplicon, in order to obtain two distinct bands after T7 endonuclease I cleavage.

ALK locus primers generate a 401-bp fragment from the wild-type locus and smaller fragments from the modified locus (Fig. 12.1D):

ALK-F 5'-AGATGGGCAGAGGCTTGAAAAG-3' ALK-R 5'-TGAGGATGTTCTGGAAGGCAAA-3'

- 8. Perform a 35-cycle PCR on 50 ng of genomic DNA in a total of 25 μ L to amplify regions encompassing the target sites. We typically use Phusion High-Fidelity DNA Polymerase (Thermo Scientific).
- **9.** Verify the quality of the amplification by running 5 μ L of the PCR reaction on a 1% agarose gel with EtBr in 0.5 × TBE. Only one band should be amplified.

- **10.** Mix 10 μ L PCR reaction with 10 μ L 2 × NEBuffer 2.1 in two different PCR tubes.
- 11. Melt and reanneal amplicons by incubating as follows: $95 \,^{\circ}C$ for 5 min, $95 \,^{\circ}C$ to $25 \,^{\circ}C$ at $-0.5 \,^{\circ}C/s$, and 15 min at 4 $^{\circ}C$. This step converts fragments with and without indels into mismatched heteroduplex DNA. Add 1.5 U of T7 endonuclease I in one of the two tubes. Add the same volume of buffer in the second one as a control.
- **12.** Incubate at 37 °C for 20 min. This step allows T7 endonuclease I to cleave mismatched DNA duplexes.
- 13. Add 10 μ L of 2 × T7 loading buffer containing Proteinase K to 10 μ L DNA. Incubate at room temperature for 5 min. In this step, T7 endonuclease I is degraded by Proteinase K.
- 14. Load the PCR product on a 2.4% agarose gel with EtBr and run at 100 V for 20 min in $0.5 \times \text{TBE}$ buffer.
- **15.** Capture the gel image with a UV imaging station (Fig. 12.1D). The amount of cleaved DNA estimates the Cas9-induced mutation rate. Each sgRNA should induce detectable levels of indels when tested with Cas9. If the T7 endonuclease I assay reveals a low indel efficiency for a chosen sgRNA, increase the amount of the sgRNA expression vector or redesign the sgRNA. If using the nCas9 approach, the four sgRNAs must induce indels separately.

3.4. PCR-based translocation detection

1. With a low translocation frequency, a nested PCR approach is typically used. Design two sets of primers on each side of the translocation junction, the second set located within the first product (see Section 2.4). For a frequency $> 10^{-4}$, however, translocations can be detected with a single set of PCR primers. The amplified product should be 600–1000 bp to recover potentially long deletions from the DNA ends during translocation formation.

Der5, corresponding to the NPM-ALK fusion:
External primers:
Der5-F 5'-CAGTTGCTTGGTTCCCAGTT-3'
Der5-R 5'-AGGAATTGGCCTGCCTTAGT-3'
Internal primers:
Der5-NF 5'-GGGGAGAGGAAATCTTGCTG-3'
Der5-NR 5'-GCAGCTTCAGTGCAATCACA-3'

- Perform a 23-cycle PCR on 100–150 ng genomic DNA (extracted in step 3.3.6) with the external set of primers. We classically use FastStart Taq DNA polymerase (Roche).
- 3. Perform a 40-cycle PCR on $0.5-1 \ \mu$ L of the first PCR reaction with the internal set of primers. Nested PCR is a highly sensitive method, presenting contamination risks. Work carefully in a dedicated place, wear gloves, take care not to contaminate a tube with already amplified PCR products, and clean up your bench and your material after each PCR.
- 4. Load the PCR product on a 1% agarose gel with EtBr in $0.5 \times$ TBE buffer.
- 5. Capture the gel image with a UV imaging station (Fig. 12.1C).
- 6. Note: Translocations require DSBs or paired nicks on both chromosomes (Fig. 12.1C). PCR amplification of translocation junctions may lead to larger or smaller products than expected due to indels generated during translocations formation (see below Fig. 12.2C), which can be confirmed by sequencing.

3.5. Quantification of potential off-target cleavage

When using Cas9, potential off-target sites should be determined during sgRNA design, because mismatches to the ~ 20 nt target sequence do not necessary abrogate cleavage. Sequences containing up to five mismatches compared to the target sequence, especially in the PAM-distal region (Hsu et al., 2013), should be considered as potential sites for cleavage by Cas9. Potential off-target effects from nCas9 and single sgRNAs are significantly reduced since nicks are generated. However, off-target paired nicks are still possible with nCas9 if two sgRNAs bind by chance relatively close to each other. Thus, all combinations of sgRNAs should be considered for potential off-target binding.

EWS-FLI1

EWS genomic sequence followed by sgRNA sequence: 5'-GGAATCCAGGACACATCTTTAGG-3' sgRNA EWS: 5'-GGAATCCAGGACACATCTTT
Potential off-target site for the sgRNA EWS on Chr9 (two mismatches) (Fig. 12.2A): 5'-GGAT*TCCAGGACC*CATCTTTGGG-3'
FLI1 genomic sequence followed by sgRNA sequence: 5'-CCTCCCATGGCTGTCTCTAGACC-3' sgRNA FLI1: 5'-GGTCTAGAGACAGCCATGGG After determining potential sgRNA off-target sites, design a set of primers amplifying a 300–500 bp region encompassing the potential off-target site. Then perform the T7 endonuclease I assay (Section 3.3).

EWS locus primers generate a 354-bp fragment from the wild-type locus and smaller fragments from the modified locus (Fig. 12.2B): EWS-F 5'-CCTCAGCCACCCAGAGTGTT-3' EWS-R 5'-TAGCTGCCTCCCCACTTTACAT-3' EWS off-target locus primers generate a 465-bp fragment from the wild-type locus and smaller fragments from the modified locus EWS-OFF-F 5'-ACTCACCTGGTTGGGTTGTCTT-3' EWS-OFF-R 5'-GTCCGTACTATGAAGGGGTCGT-3'

2. Design primers for detection of translocations between one of the target chromosomes and the potential off-target site. Then proceed to nested PCR to detect translocations (Section 3.4)

Der22, corresponding to the EWS-FLI1 fusion (Fig. 12.2B):
External primers:
Der22-F 5'-ATCCTACAGCCAAGCTCCAA-3'
Der22-R 5'-GGCCTCATTGTTTCTGGCTA-3'
Internal primers:
Der22-NF 5'-CTACGGGCAGCAGAGTGAGT-3'
Der22-NR 5'-TTCCTCAAGGCTCTGGAAAA-3'
Der9, corresponding to the off-target-FLI1 fusion (Fig. 12.2B):
External primers:
Der9-F 5'-TAGTGGGGGAAGGAGAGACA-3'
Der9-B 5'-GCCAGGTTTCTTAGGGCTTT-3'
Internal primers:
Der9-NF 5'-GAAAAGGCTCCATTTCATGC-3'
Der9-NB 5'-GGGCTGAGCTCCATAAATCA-3'

A positive signal in one of these two assays (presence of indels after T7 assay or translocation detection by PCR) will reveal off-target cleavage of the tested sgRNA. However, the T7 assay is less sensitive than the translocation assay. The T7 assay sensitivity is ~1–2%, while translocations can be easily detected at a frequency $\geq 10^{-5}$. For example, in Fig. 12.2B a translocation between the target site and the off-target site can be amplified when no signal can be detect in the T7 assay. Thus, PCR detection of translocation formation is a highly sensitive assay to detect off-target cleavage and can be used as a general method in other applications to evaluate off-target sites.

3.6. Quantification of translocation frequency using a 96-well plate screen

Translocation frequencies can be accurately determined using a 96-well plate format (Brunet et al., 2009; Piganeau et al., 2013). The basic strategy is that small pools of cells are placed in each well, such that most wells on a plate contain <1 translocation. The translocation frequency can be determined from the number of negative and positive wells, with correction for the number of wells with two or more translocations. Having a single event in most positive wells also means that unique translocation junctions can be scored.

- 48-well plate (step 3.2.14): Trypsinize cells in the well ≤24 h after transfection, before cells have had a chance to divide. Use a small volume of trypsin (typically 100 µL) and count the cells contained in this well. This number represents 1/5th the transfected cells surviving transfection in each of the 96-well plates. Do not exceed this time frame to ensure an accurate cell number to calculate translocation frequency.
- 96-well plates (step 3.2.14): Remove all of the culture medium 48 h to 5 days after transfection, depending on the growth of the cells. Plates can be stored wrapped in Parafilm at −80 °C and thawed for later use.
- 3. Add Proteinase K to 2.5 mL of $1 \times$ lysis solution at final concentration of 100 µg/mL. Place 25 µL Proteinase K solution into each well of the thawed 96-well plate.
- 4. Incubate at 55 $^{\circ}$ C for 120 min in a humid chamber, such as a closed plastic container with moist towels at the bottom.
- 5. Transfer the lysates into 96 tubes on PCR tube strips.
- 6. Incubate at 95 °C for 10 min to inactivate Proteinase K.
- 7. Add 200 nM of the external primers (same primers used in Section 3.4) to 5 mL of $1 \times$ PCR Master Mix 1, and add 25 μ L of FastStart Taq polymerase. Perform the first round of PCR on 4–7 μ L cell lysate in a total volume of 50 μ L.
- 8. Add 200 nM of the internal primers (same primers used in Section 3.4) to 2 mL of $1 \times PCR$ Master Mix 2 (containing the SYBR Green), and add 10 μ L of FastStart Taq polymerase. Perform the second round of PCR on 0.5–1 μ L of the first PCR in a total volume of 20 μ L. The PCR program has to contain a denaturation curve cycle.
- **9.** Count the number of wells with a positive denaturation curve. Since nested PCR fragments corresponding to translocation junctions are

designed to amplify typically 600–900 bp, all wells with a $T_{\rm m} > 85$ °C are considered to be positive, corresponding to fragments >100–150 bp.

- **10.** $\leq 12-14$ positive wells per 96-well plate: Each well is estimated to contain no more than one translocation and the translocation frequency is (*p*), the number positive wells divided by the number of plated cells (determined in step 3.6.1).
- 11. >12–14 positive wells per 96-well plate: The translocation frequency is determined using a correction following a beta cumulative distribution function k(x,a,b), where k = number of translocations per well, p = total number of positive wells per number of plated cells, n = number of cells per well, and x = 1 p, a = n k, b = k + 1. If the number of wells with two or more translocations is high to interfere with precise frequency measurements, calculate from a plate with fewer cells per well (step 3.2.14).

PCR products from positive wells can be sent for sequencing to analyze breakpoint junctions (e.g., Der5, Fig. 12.3 for Cas9 and Fig. 12.4 for nCas9).



Der5 translocation junction

Figure 12.3 Cas9 cleavage at the *NPM* and *ALK* loci and example of a translocation junction sequence. (A) Cas9 cleavage at the *NPM* and *ALK* loci using the NPM1 and ALK1 sgRNAs, respectively, leads to DSBs. Only the part of the sgRNA that binds DNA is shown. PAM sequence, boxed; arrows, cleavage sites. (B) DNA ends from Cas9 cleavage that are relevant to Der5 formation are shown. NHEJ leads to a variety of Der5 translocation junctions, but a common junction (shown) is direct ligation of the DNA ends, presumably involving fill-in of the 1-base 5' overhang. For additional junctions, see Ghezraoui et al. (2014).



Figure 12.4 nCas9 cleavage at the *NPM* and *ALK* loci and example of a translocation junction sequence. (A) nCas9 cleavage at the *NPM* and *ALK* loci using the NPM1 + NPM2 and ALK1 + ALK2 sgRNAs, respectively, leads to paired nicks. The relative position of the two sgRNAs at each locus has been termed "PAMs out." Only the part of the sgRNA that binds DNA is shown. PAM sequence, boxed; arrows, cleavage sites. (B) The two DNA ends from nCas9 cleavage that are relevant to Der5 formation are shown. Long overhangs are predicted. NHEJ leads to a variety of Der5 translocation junctions, an example of which is shown. In this typical junction, deletions occur primarily in the overhangs at both DNA ends, although in other junctions deletions extend well into the double-stranded region on one or both sides. For additional junctions, see Ghezraoui et al. (2014) (µhom=microhomology).

At the local sequence level, very different junctions are obtained for Cas9 and nCas9 due to the different DNA end structures. The small pool PCR means that with a unique translocation in a well, the reciprocal translocation junction can also be determined. Cells with a reciprocal translocation can be enriched by sib selection (Brunet et al., 2009).

3.7. Translocation frequency determination by serial dilution

The protocol described in Section 3.6 requires removal of medium from the 96-well plate. Thus, with nonadherent cells, cell loss is a potential problem. Here, we suggest an alternative method for translocation frequency assessment when frequencies are $>10^{-4}$, which is not atypical in our experience.

- 1. Perform serial dilutions on genomic DNA in quadruplicates starting from 100 to 1.56 ng DNA (extraction in step 3.3.6).
- 2. Perform PCR on each dilution as in step 3.4.3.
- 3. Load on a 1% agarose gel with EtBr.
- Capture the gel image with a UV imaging station. Considering that one human diploid cell contains 6 pg of DNA, 6.25 ng represent about 10³ cells. Consider all the quadruplicates to determine the translocation frequency (Fig. 12.2C for Der22).

4. CONCLUSIONS

In this chapter, we described a method to induce cancer translocations using Cas9 DSBs or nCas9 paired nicks in human cells. The ease at which potential target sites can be found at any locus of interest makes this approach readily adaptable for expanding the repertoire of possible cancer translocations that can be induced in any cell type of interest. Concerns about offtarget effects are diminished by the use of nCas9. In addition, nCas9 will better recapitulate some types of breakpoints junctions found in patient cells (Ghezraoui et al., 2014). Cells carrying the chosen translocations provide a more accurate model to understand mechanisms of tumor initiation than standard ectopic expression of a fusion protein. Results from such studies may assist the design of targeted therapies, for example, to avoid therapyinduced secondary tumors with specific translocations (Cowell et al., 2012). Our approach can also be used to decipher the repair mechanisms that are involved in translocation formation, because PCR screening of translocation junctions allows accurate frequency determination in parallel with the precise determination of breakpoint junctions (Ghezraoui et al., 2014). Finally, translocation assays provide a highly sensitive method to detect off-target cleavage for any application, and so has general utility.

ACKNOWLEDGMENTS

We thank Carine Giovannangeli, Anne De Cian and Jean-Paul Concordet (MNHN) and Matt Krawczyk (MSKCC) for their helpful discussions. This work was supported by ANR-12-JSV6-0005 (B.R.), Le Canceropole IDF (M.P.), La Ligue Nationale contre le Cancer (H.G.), and NIH grant GM054668 (M.J.).

REFERENCES

Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V. M., Gregory, P., et al. (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.

- Chandra, H. S., Heisterkamp, N. C., Hungerford, A., Morrissette, J. J., Nowell, P. C., Rowley, J. D., et al. (2011). Philadelphia chromosome symposium: Commemoration of the 50th anniversary of the discovery of the Ph chromosome. *Cancer Genetics*, 204, 171–179.
- Choi, P. S., & Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. *Nature Communications*, 5, 3728.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819–823.
- Cowell, I. G., Sondka, Z., Smith, K., Lee, K. C., Manville, C. M., Sidorczuk-Lesthuruge, M., et al. (2012). Model for MLL translocations in therapy-related leukemia involving topoisomerase IIbeta-mediated DNA strand breaks and gene proximity. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 8989–8994.
- Elmberger, P. G., Lozano, M. D., Weisenburger, D. D., Sanger, W., & Chan, W. C. (1995). Transcripts of the npm-alk fusion gene in anaplastic large cell lymphoma, Hodgkin's disease, and reactive lymphoid lesions. *Blood*, *86*, 3517–3521.
- Gaj, T., Sirk, S. J., & Barbas, C. F., 3rd. (2014). Expanding the scope of site-specific recombinases for genetic and metabolic engineering. *Biotechnology and Bioengineering*, 111, 1–15.
- Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.-B., Aallmyr, A., Ruis, B., et al. (2014) Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Molecular Cell*, 55, 829–842.
- Guschin, D. Y., Waite, A. J., Katibah, G. E., Miller, J. C., Holmes, M. C., & Rebar, E. J. (2010). A rapid and general assay for monitoring endogenous gene modification. *Methods* in Molecular Biology, 649, 247–256.
- Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157, 1262–1278.
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31, 827–832.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816–821.
- Kuefer, M. U., Look, A. T., Pulford, K., Behm, F. G., Pattengale, P. K., Mason, D. Y., et al. (1997). Retrovirus-mediated gene transfer of NPM-ALK causes lymphoid malignancy in mice. *Blood*, *90*, 2901–2910.
- Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., et al. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, 31, 833–838.
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNAguided human genome engineering via Cas9. Science, 339, 823–826.
- Mani, R. S., & Chinnaiyan, A. M. (2010). Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nature Reviews Genetics*, 11, 819–829.
- May, W. A., Gishizky, M. L., Lessnick, S. L., Lunsford, L. B., Lewis, B. C., Delattre, O., et al. (1993). Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 5752–5756.
- Mitelman, F., Johansson, B., & Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, 7, 233–245.
- Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOP-CHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, 42, W401–W407.

- Morris, S. W., Kirstein, M. N., Valentine, M. B., Dittmer, K. G., Shapiro, D. N., Saltman, D. L., et al. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science*, 263, 1281–1284.
- Piganeau, M., Ghezraoui, H., De Cian, A., Guittat, L., Tomishima, M., Perrouault, L., et al. (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Research*, 23, 1182–1193.
- Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154, 1380–1389.
- Richardson, C., & Jasin, M. (2000). Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature*, 405, 697–700.
- Rowley, J. D. (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, 243, 290–293.
- Simsek, D., Brunet, E., Wong, S. Y., Katyal, S., Gao, Y., McKinnon, P. J., et al. (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS Genetics*, 7, e1002080.
- Torres, R., Martin, M. C., Garcia, A., Cigudosa, J. C., Ramirez, J. C., & Rodriguez-Perales, S. (2014). Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nature Communications*, 5, 3964.
- Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 11, 636–646.
- Weinstock, D. M., Elliott, B., & Jasin, M. (2006). A model of oncogenic rearrangements: Differences between chromosomal translocation mechanisms and simple double-strand break repair. *Blood*, 107, 777–780.
- Xie, S., Shen, B., Zhang, C., Huang, X., & Zhang, Y. (2014). sgRNAcas9: A software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS One*, 9, e100448.

6 Bibliographie

Adamo, A., Collis, S.J., Adelman, C. a, Silva, N., Horejsi, Z., Ward, J.D., Martinez-Perez, E., Boulton, S.J., and La Volpe, A. (2010). Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia. Mol. Cell *39*, 25–35.

Ahnesorg, P., Smith, P., and Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. Cell *124*, 301–313.

Akopiants, K., Zhou, R.-Z., Mohapatra, S., Valerie, K., Lees-Miller, S.P., Lee, K.-J., Chen, D.J., Revy, P., de Villartay, J.-P., and Povirk, L.F. (2009). Requirement for XLF/Cernunnos in alignment-based gap filling by DNA polymerases lambda and mu for nonhomologous end joining in human whole-cell extracts. Nucleic Acids Res. *37*, 4055–4062.

Alt, F.W., and Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. Proc. Natl. Acad. Sci. U. S. A. *79*, 4118–4122.

Alt, F.W., Zhang, Y., Meng, F.-L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell *152*, 417–429.

Amé, J.-C., Spenlehauer, C., and de Murcia, G. (2004). The PARP superfamily. Bioessays 26, 882–893.

Andres, S.N., Modesti, M., Tsai, C.J., Chu, G., and Junop, M.S. (2007). Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. Mol. Cell 28, 1093–1101.

Audebert, M., Salles, B., and Calsou, P. (2004). Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J. Biol. Chem. *279*, 55117–55126.

Averbeck, N.B., Ringel, O., Herrlitz, M., Jakob, B., Durante, M., and Taucher-Scholz, G. (2014). DNA end resection is needed for the repair of complex lesions in G1-phase human cells. Cell Cycle *13*, 2509–2516.

Aylon, Y., Liefshitz, B., and Kupiec, M. (2004). The CDK regulates repair of double-strand breaks by homologous recombination during the cell. *23*, 4868–4875.

Bakkenist, C.J., and Kastan, M.B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature *421*, 499–506.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. a, and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Barton, O., Naumann, S.C., Diemer-Biehs, R., Kunzel, J., Steinlage, M., Conrad, S., Makharashvili, N., Wang, J., Feng, L., Lopez, B.S., et al. (2014). Polo-like kinase 3 regulates CtIP during DNA double-strand break repair in G1. J. Cell Biol. *206*, 877–894.

Bassing, C.H., Suh, H., Ferguson, D.O., Chua, K.F., Manis, J., Eckersdorff, M., Gleason, M., Bronson, R., Lee, C., and Alt, F.W. (2003). Histone H2AX : A Dosage-Dependent Suppressor of Oncogenic Translocations and Tumors. *114*, 359–370.

Beamish, H.J., Jessberger, R., Riballo, E., Priestley, a, Blunt, T., Kysela, B., and Jeggo, P. a (2000). The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity. Nucleic Acids Res. *28*, 1506–1513.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J., and Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. Curr. Opin. Biotechnol. *32*, 76–84.

Bennardo, N., Cheng, A., Huang, N., and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet. *4*, e1000110.

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

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

Biosciences, P.S. (2014). Sangamo BioSciences Announces Presentation At ICAAC of New Clinical Data Demonstrating Sustained Functional Control of Viremia in Multiple HIV-Infected Subjects Treated with SB-728-T Control of Viral Load During Treatment Interruption of Antiretroviral Th. 8–11.

Bitinaite, J., Wah, D. a, Aggarwal, a K., and Schildkraut, I. (1998). Fokl dimerization is required for DNA cleavage. Proc. Natl. Acad. Sci. U. S. A. *95*, 10570–10575.

Blier, P.R., Griffith, A.J., Craft, J., and Hardin, A. (1993). Binding of Ku Protein to DNA. *268*, 7594–7601.

Boboila, C., Jankovic, M., Yan, C.T., Wang, J.H., Wesemann, D.R., Zhang, T., Fazeli, A., Feldman, L., Nussenzweig, A., Nussenzweig, M., et al. (2010). Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. Proc. Natl. Acad. Sci. U. S. A. *107*, 3034–3039.

Boboila, C., Alt, F.W., and Schwer, B. (2012a). Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. (Elsevier Inc.).

Boboila, C., Oksenych, V., Gostissa, M., Wang, J.H., Zha, S., Zhang, Y., Chai, H., Lee, C.-S., Jankovic, M., Saez, L.-M.A., et al. (2012b). Robust chromosomal DNA repair via alternative end-joining in the absence of X-ray repair cross-complementing protein 1 (XRCC1). Proc. Natl. Acad. Sci. U. S. A. *109*, 2473–2478.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509–1512.

Boubakour-Azzouz, I., Bertrand, P., Claes, A., Lopez, B.S., and Rougeon, F. (2012). Terminal deoxynucleotidyl transferase requires KU80 and XRCC4 to promote N-addition at non-V(D)J chromosomal breaks in non-lymphoid cells. Nucleic Acids Res. *40*, 8381–8391.

Boulton, S.J., and Jackson, S.P. (1996). Saccharomyces cerevisiae Ku7O potentiates pathways. *15*, 5093–5103.

Bouwman, P., Aly, A., Escandell, J.M., Pieterse, M., Bartkova, J., van der Gulden, H., Hiddingh, S., Thanasoula, M., Kulkarni, A., Yang, Q., et al. (2010). 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. Nat. Struct. Mol. Biol. *17*, 688–695.

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. *106*.

Bryans, M., Valenzano, M.C., and Stamato, T.D. (1999). Absence of DNA ligase IV protein in XR-1 cells: evidence for stabilization by XRCC4. Mutat. Res. *433*, 53–58.

Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondanèche, M.-C., Sanal, O., Plebani, A., Stéphan, J.-L., Hufnagel, M., le Deist, F., et al. (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. Cell *124*, 287–299.

Buisson, R., Dion-Côté, A.-M., Coulombe, Y., Launay, H., Cai, H., Stasiak, A.Z., Stasiak, A., Xia, B., and Masson, J.-Y. (2010). Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination. Nat. Struct. Mol. Biol. *17*, 1247–1254.

Bunting, S.F., Callén, E., Wong, N., Chen, H.-T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., et al. (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell *141*, 243–254.

Burhans, W.C., and Weinberger, M. (2007). DNA replication stress, genome instability and aging. Nucleic Acids Res. *35*, 7545–7556.

Caldecott, K.W. (2003). XRCC1 and DNA strand break repair. DNA Repair (Amst). 2, 955–969.

Caldecott, K.W., McKeown, C.K., Tucker, J.D., Ljungquist, S., and Thompson, L.H. (1994). An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. Mol. Cell. Biol. *14*, 68–76.

Caldecott, K.W., Tucker, J.D., Stanker, L.H., and Thompson, L.H. (1995). Characterization of the XRCC1-DNA ligase III complex in vitro and its absence from mutant hamster cells. Nucleic Acids Res. *23*, 4836–4843.

Celeste, A., Difilippantonio, S., Difilippantonio, M.J., Fernandez-Capetillo, O., Pilch, D.R., Sedelnikova, O. a., Eckhaus, M., Ried, T., Bonner, W.M., and Nussenzweig, A. (2003). H2AX Haploinsufficiency Modifies Genomic Stability and Tumor Susceptibility. Cell *114*, 371–383.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. a, Somia, N. V, Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. *39*, e82.

Chandra, H.S., Heisterkamp, N.C., Hungerford, A., Morrissette, J.J.D., Nowell, P.C., Rowley, J.D., and Testa, J.R. (2011). Philadelphia Chromosome Symposium: commemoration of the 50th anniversary of the discovery of the Ph chromosome. Cancer Genet. *204*, 171–179.

Chapman, J.R., Taylor, M.R.G., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. Mol. Cell 47, 497–510.

Chapman, J.R., Barral, P., Vannier, J.-B., Borel, V., Steger, M., Tomas-Loba, A., Sartori, A. a, Adams, I.R., Batista, F.D., and Boulton, S.J. (2013). RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. Mol. Cell *49*, 858–871.

Chappell, C., Hanakahi, L. a, Karimi-Busheri, F., Weinfeld, M., and West, S.C. (2002). Involvement of human polynucleotide kinase in double-strand break repair by nonhomologous end joining. EMBO J. *21*, 2827–2832.

Chen, S., Oikonomou, G., Chiu, C.N., Niles, B.J., Liu, J., Lee, D. a, Antoshechkin, I., and Prober, D. a (2013). A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. Nucleic Acids Res. *41*, 2769–2778.

Cheng, Q., Barboule, N., Frit, P., Gomez, D., Bombarde, O., Couderc, B., Ren, G.-S., Salles, B., and Calsou, P. (2011). Ku counteracts mobilization of PARP1 and MRN in chromatin damaged with DNA double-strand breaks. Nucleic Acids Res. *39*, 9605–9619.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.-J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107–119.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Mol. Cell 40, 179–204.

Cimprich, K. a, and Cortez, D. (2008). ATR: an essential regulator of genome integrity. Nat. Rev. Mol. Cell Biol. *9*, 616–627.

Corneo, B., Wendland, R.L., Deriano, L., Cui, X., Klein, I. a, Wong, S.-Y., Arnal, S., Holub, A.J., Weller, G.R., Pancake, B. a, et al. (2007). Rag mutations reveal robust alternative end joining. Nature *449*, 483–486.

Cottarel, J., Frit, P., Bombarde, O., Salles, B., Négrel, A., Bernard, S., Jeggo, P. a, Lieber, M.R., Modesti, M., and Calsou, P. (2013). A noncatalytic function of the ligation complex during nonhomologous end joining. J. Cell Biol. *200*, 173–186.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. *2*, 292–301.

Cuneo, M.J., Gabel, S. a, Krahn, J.M., Ricker, M. a, and London, R.E. (2011). The structural basis for partitioning of the XRCC1/DNA ligase III- α BRCT-mediated dimer complexes. Nucleic Acids Res. *39*, 7816–7827.

Davis, A.J., Chen, B.P.C., and Chen, D.J. (2014). DNA-PK: A dynamic enzyme in a versatile DSB repair pathway. DNA Repair (Amst). *17*, 21–29.

DeFazio, L.G., Stansel, R.M., Griffith, J.D., and Chu, G. (2002). Synapsis of DNA ends by DNAdependent protein kinase. EMBO J. *21*, 3192–3200.

Della-Maria, J., Zhou, Y., Tsai, M.-S., Kuhnlein, J., Carney, J.P., Paull, T.T., and Tomkinson, A.E. (2011). Human Mre11/human Rad50/Nbs1 and DNA ligase IIIalpha/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway. J. Biol. Chem. *286*, 33845–33853.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z. a, Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y., and Yan, N. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science *335*, 720–723.

Deriano, L., and Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu. Rev. Genet. 47, 433–455.

Difilippantonio, M.J., Petersen, S., Chen, H.T., Johnson, R., Jasin, M., Kanaar, R., Ried, T., and Nussenzweig, a. (2002). Evidence for Replicative Repair of DNA Double-Strand Breaks Leading to Oncogenic Translocation and Gene Amplification. J. Exp. Med. *196*, 469–480.

Difilippantonio, S., Gapud, E., Wong, N., Huang, C.-Y., Mahowald, G., Chen, H.T., Kruhlak, M.J., Callen, E., Livak, F., Nussenzweig, M.C., et al. (2008). 53BP1 facilitates long-range DNA end-joining during V(D)J recombination. Nature *456*, 529–533.

Dobbs, T. a, Tainer, J. a, and Lees-Miller, S.P. (2010). A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. DNA Repair (Amst). *9*, 1307–1314.

Donzelli, M., and Draetta, G.F. (2003). Regulating mammalian checkpoints through Cdc25 inactivation. EMBO Rep. *4*, 671–677.

Downs, J. a, and Jackson, S.P. (2004). A means to a DNA end: the many roles of Ku. Nat. Rev. Mol. Cell Biol. *5*, 367–378.

Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia, D.F., Miller, J.C., Urnov, F.D., Gregory, P.D., and Holmes, M.C. (2011). Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat. Methods *8*, 74–79.

Dray, E., Etchin, J., Wiese, C., Saro, D., Williams, G.J., Hammel, M., Yu, X., Galkin, V.E., Liu, D., Tsai, M.-S., et al. (2010). Enhancement of RAD51 recombinase activity by the tumor suppressor PALB2. Nat. Struct. Mol. Biol. *17*, 1255–1259.

Ellenberger, T., and Tomkinson, A.E. (2008). Eukaryotic DNA ligases: structural and functional insights. Annu. Rev. Biochem. 77, 313–338.

Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T.F., Tkáč, J., Cook, M. a, Rosebrock, A.P., Munro, M., Canny, M.D., et al. (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. Mol. Cell *49*, 872–883.

Farmer, H., Mccabe, N., and Lord, C.J. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *239*, 236–239.

Fattah, F., Lee, E.H., Weisensel, N., Wang, Y., Lichter, N., and Hendrickson, E. a (2010). Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. PLoS Genet. *6*, e1000855.

Fattah, F.J., Kweon, J., Wang, Y., Lee, E.H., Kan, Y., Lichter, N., Weisensel, N., and Hendrickson, E. a (2014). A role for XLF in DNA repair and recombination in human somatic cells. DNA Repair (Amst). *15*, 39–53.

Ferretti, L.P., Lafranchi, L., and Sartori, A. a (2013). Controlling DNA-end resection: a new task for CDKs. Front. Genet. *4*, 99.

Finnie, N.J., Gottlieb, T.M., Bluntt, T., Jeggot, P.A., and Jackson, S.P. (1995). activity xrs-6 site-specific. *92*, 320–324.

Franco, S., Gostissa, M., Zha, S., Lombard, D.B., Murphy, M.M., Zarrin, A. a, Yan, C., Tepsuporn, S., Morales, J.C., Adams, M.M., et al. (2006a). H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. Mol. Cell *21*, 201–214.

Franco, S., Alt, F.W., and Manis, J.P. (2006b). Pathways that suppress programmed DNA breaks from progressing to chromosomal breaks and translocations. DNA Repair (Amst). *5*, 1030–1041.

Fraser, J., Ramsden, D.A., and Oettinger, A. (1995). Cleavage at a V (D) J Recombination Signal Requires Only RAG1 and RAG2 Proteins and Occurs in Two Steps mutation. *83*, 387–395.

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. Nat. Biotechnol. *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. Nat. Biotechnol. *32*, 279–284.

Gaj, T., Gersbach, C. a, and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. *31*, 397–405.

Gauss, G.H., and Lieber, M.R. (1996). Mechanistic Constraints on Diversity in Human V (D) J Recombination. *16*, 258–269.

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 1–14.

Gillert, E., Leis, T., Repp, R., Reichel, M., Ho, A., Breitenlohner, I., Angermu, S., Borkhardt, A., Harbott, J., Lampert, F., et al. (1999). A DNA damage repair mechanism is involved in the origin of chromosomal translocations t (4; 11) in primary leukemic cells.

Goodarzi, A. a, and Jeggo, P. a (2013). The repair and signaling responses to DNA double-strand breaks. (Elsevier Inc.).

Gottlieb, T.M., and Jackson, P. (1993). The DNA-Dependent Protein Ki for DNA Ends and AssocWbn with Ku Ant. 72, 131–142.

Grabarz, A., Guirouilh-Barbat, J., Barascu, A., Pennarun, G., Genet, D., Rass, E., Germann, S.M., Bertrand, P., Hickson, I.D., and Lopez, B.S. (2013). A Role for BLM in Double-Strand Break Repair Pathway Choice: Prevention of CtIP/Mre11-Mediated Alternative Nonhomologous End-Joining. Cell Rep. *5*, 21–28.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., and Lieber, M.R. (1997). letters to nature Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *388*, 3–6.

Guirouilh-Barbat, J., Huck, S., Bertrand, P., Pirzio, L., Desmaze, C., Sabatier, L., and Lopez, B.S. (2004). Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. Mol. Cell *14*, 611–623.

Guirouilh-Barbat, J., Rass, E., Plo, I., Bertrand, P., and Lopez, B.S. (2007). Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. Proc. Natl. Acad. Sci. U. S. A. *104*, 20902–20907.

Haber, J.E., and Leung, W.Y. (1996). Lack of chromosome territoriality in yeast: promiscuous rejoining of broken chromosome ends. Proc. Natl. Acad. Sci. U. S. A. *93*, 13949–13954.

Haince, J.-F., McDonald, D., Rodrigue, A., Déry, U., Masson, J.-Y., Hendzel, M.J., and Poirier, G.G. (2008). PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. J. Biol. Chem. *283*, 1197–1208.

Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.-R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C., et al. (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature *484*, 69–74.

Han, L., Mao, W., and Yu, K. (2012). X-ray repair cross-complementing protein 1 (XRCC1) deficiency enhances class switch recombination and is permissive for alternative end joining. Proc. Natl. Acad. Sci. U. S. A. *109*, 4604–4608.

Harper, J.W., and Elledge, S.J. (2007). The DNA damage response: ten years after. Mol. Cell 28, 739–745.

Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J. a (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science *329*, 1355–1358.

Helmink, B.A., Tubbs, A.T., Dorsett, Y., Bednarski, J.J., Walker, L.M., Feng, Z., Sharma, G.G., Mckinnon, P.J., Zhang, J., Bassing, C.H., et al. (2011). and aberrant repair in G1-phase lymphocytes. Nature *469*, 245–249.

Heyer, W.-D., Li, X., Rolfsmeier, M., and Zhang, X.-P. (2006). Rad54: the Swiss Army knife of homologous recombination? Nucleic Acids Res. *34*, 4115–4125.

Hoeijmakers, J.H.J. (2009). DNA damage, aging, and cancer. N. Engl. J. Med. 361, 1475–1485.

Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G.M., Kohn, D.B., Gregory, P.D., Holmes, M.C., et al. (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nat. Biotechnol. *28*, 839–847.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science 327, 167–170.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. Nat. Biotechnol. *29*, 699–700.

Huertas, P., and Jackson, S.P. (2009). Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. J. Biol. Chem. *284*, 9558–9565.

Huertas, P., Cortés-Ledesma, F., Sartori, A. a, Aguilera, A., and Jackson, S.P. (2008). CDK targets Sae2 to control DNA-end resection and homologous recombination. Nature *455*, 689–692.

Ira, G., Pellicioli, A., Balijja, A., Wang, X., Fiorani, S., Carotenuto, W., Liberi, G., Bressan, D., Wan, L., Hollingsworth, N.M., et al. (2004). DNA end resection , homologous recombination and DNA damage checkpoint activation require CDK1. *431*, 1011–1017.
Jazayeri, A., Falck, J., Lukas, C., Bartek, J., Smith, G.C.M., Lukas, J., and Jackson, S.P. (2006). ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. Nat. Cell Biol. *8*, 37–45.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. a (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. *31*, 233–239.

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 doublestrand break repair pathway in mammalian cells. EMBO J. *19*, 3398–3407.

Junop, M.S., Modesti, M., Guarné, a, Ghirlando, R., Gellert, M., and Yang, W. (2000). Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. EMBO J. *19*, 5962–5970.

Kabotyanski, E.B., Gomelsky, L., Han, J., Stamato, T.D., and Roth, D.B. (1998). Double-strand break repair in Ku86- and. *26*, 27–31.

Kass, E.M., and Jasin, M. (2010). Collaboration and competition between DNA double-strand break repair pathways. FEBS Lett. *584*, 3703–3708.

Kim, H., and Kim, J.-S. (2014). A guide to genome engineering with programmable nucleases. Nat. Rev. Genet. *15*, 321–334.

Kim, Y.G., and Chandrasegaran, S. (1994). Chimeric restriction endonuclease. Proc. Natl. Acad. Sci. U. S. A. *91*, 883–887.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. U. S. A. *93*, 1156–1160.

Kitada, K., and Yamasaki, T. (2007). The MDR1/ABCB1 regional amplification in large inverted repeats with asymmetric sequences and microhomologies at the junction sites. Cancer Genet. Cytogenet. *178*, 120–127.

Klein, H.L., and Symington, L.S. (2009). Breaking up just got easier to do. Cell 138, 20–22.

Klein, I. a, Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95–106.

Lakshmipathy, U., and Campbell, C. (1999). The Human DNA Ligase III Gene Encodes Nuclear and Mitochondrial Proteins The Human DNA Ligase III Gene Encodes Nuclear and Mitochondrial Proteins. *19*.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

De Lange, T. (2009). How telomeres solve the end-protection problem. Science 326, 948–952.

Langer, T., Metzler, M., Reinhardt, D., Viehmann, S., Borkhardt, A., Reichel, M., Stanulla, M., Schrappe, M., Creutzig, U., Ritter, J., et al. (2003). Analysis of t(9;11) chromosomal breakpoint sequences in childhood acute leukemia: almost identical MLL breakpoints in therapy-related AML after treatment without etoposides. Genes. Chromosomes Cancer *36*, 393–401.

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, H.J., Kim, E., and Kim, J. (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. 81–89.

Lee, H.J., Kweon, J., Kim, E., Kim, S., and Kim, J. (2012). Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. 539–548.

Lee-Theilen, M., Matthews, A.J., Kelly, D., Zheng, S., and Chaudhuri, J. (2011). CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. Nat. Struct. Mol. Biol. *18*, 75–79.

Leppard, J.B., Dong, Z., Mackey, Z.B., and Tomkinson, A.E. (2003). Physical and Functional Interaction between DNA Ligase III α and Poly (ADP-Ribose) Polymerase 1 in DNA Single-Strand Break Repair Physical and Functional Interaction between DNA Ligase III $_{\Box}$ and Poly (ADP-Ribose) Polymerase 1 in DNA Single-Strand Brea.

Levin, D.S., McKenna, A.E., Motycka, T. a., Matsumoto, Y., and Tomkinson, A.E. (2000). Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. Curr. Biol. *10*, 919–922.

Li, X., and Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. Cell Res. *18*, 99–113.

Li, G., Nelsen, C., and Hendrickson, E.A. (2001). Ku86 is essential in human somatic cells. 2001.

Li, G., Alt, F.W., Cheng, H.-L., Brush, J.W., Goff, P.H., Murphy, M.M., Franco, S., Zhang, Y., and Zha, S. (2008). Lymphocyte-specific compensation for XLF/cernunnos end-joining functions in V(D)J recombination. Mol. Cell *31*, 631–640.

Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S.Y., Bhagwat, A.S., Malani, N., Anguela, X.M., Sharma, R., Ivanciu, L., et al. (2011a). In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature *475*, 217–221.

Li, L., Wu, L.P., and Chandrasegaran, S. (1992). Functional domains in Fok I restriction endonuclease. Proc. Natl. Acad. Sci. U. S. A. *89*, 4275–4279.

Li, T., Huang, S., Zhao, X., Wright, D. a, Carpenter, S., Spalding, M.H., Weeks, D.P., and Yang, B. (2011b). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res. *39*, 6315–6325.

Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. J. Biol. Chem. 283, 1–5.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu. Rev. Biochem. *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.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B., 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.

Liu, X., Jiang, W., Dubois, R.L., Yamamoto, K., Wolner, Z., and Zha, S. (2012). Overlapping functions between XLF repair protein and 53BP1 DNA damage response factor in end joining and lymphocyte development. Proc. Natl. Acad. Sci. *109*, 3903–3908.

Liyanage, M., Weaver, Z., Barlow, C., Coleman, a, Pankratz, D.G., Anderson, S., Wynshaw-Boris, a, and Ried, T. (2000). Abnormal rearrangement within the alpha/delta T-cell receptor locus in lymphomas from Atm-deficient mice. Blood *96*, 1940–1946.

Löser, D. a, Shibata, A., Shibata, A.K., Woodbine, L.J., Jeggo, P. a, and Chalmers, A.J. (2010). Sensitization to radiation and alkylating agents by inhibitors of poly(ADP-ribose) polymerase is enhanced in cells deficient in DNA double-strand break repair. Mol. Cancer Ther. *9*, 1775–1787.

Lou, Z., Minter-Dykhouse, K., Franco, S., Gostissa, M., Rivera, M. a, Celeste, A., Manis, J.P., van Deursen, J., Nussenzweig, A., Paull, T.T., et al. (2006). MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. Mol. Cell *21*, 187–200.

Ma, J., Kim, E.M., Haber, J.E., and Lee, S.E. (2003). Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism To Repair Double-Strand Breaks Lacking Overlapping End Sequences Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism To Repair Double-Strand Breaks Lacking Overlapping End Sequence.

Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M.R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell *108*, 781–794.

Mahowald, G.K., Baron, J.M., and Sleckman, B.P. (2008). Collateral damage from antigen receptor gene diversification. Cell *135*, 1009–1012.

Mak, A.N.-S., Bradley, P., Cernadas, R. a, Bogdanove, A.J., and Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science *335*, 716–719.

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. Nat. Biotechnol. *31*, 833–838.

Manis, J.P., Morales, J.C., Xia, Z., Kutok, J.L., Alt, F.W., and Carpenter, P.B. (2004). 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. Nat. Immunol. *5*, 481–487.

Mansour, W.Y., Rhein, T., and Dahm-Daphi, J. (2010). The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. Nucleic Acids Res. *38*, 6065–6077.

Marculescu, R., Le, T., Simon, P., Jaeger, U., and Nadel, B. (2002). V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. J. Exp. Med. *195*, 85–98.

Mari, P.-O., Florea, B.I., Persengiev, S.P., Verkaik, N.S., Brüggenwirth, H.T., Modesti, M., Giglia-Mari, G., Bezstarosti, K., Demmers, J. a a, Luider, T.M., et al. (2006). Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. Proc. Natl. Acad. Sci. U. S. A. *103*, 18597–18602.

Masson, M., Niedergang, C., Schreiber, V., Murcia, J.M., Murcia, G. De, Schreiber, R.I.E., and Muller, S. (1998). XRCC1 Is Specifically Associated with Poly (ADP-Ribose) Polymerase and Negatively Regulates Its Activity following DNA Damage XRCC1 Is Specifically Associated with Poly (ADP-Ribose) Polymerase and Negatively Regulates Its Activity following DNA Damage.

Mattarucchi, E., Guerini, V., Rambaldi, A., Campiotti, L., Venco, A., Pasquali, F., Curto, F. Lo, and Porta, G. (2008). Microhomologies and Interspersed Repeat Elements at Genomic Breakpoints in Chronic Myeloid Leukemia. *632*, 625–632.

Mccarty, M. (1943). (From the Hospital of The Rockefeller Institute for Medical Research).

McClendon, A.K., and Osheroff, N. (2007). DNA topoisomerase II, genotoxicity, and cancer. Mutat. Res. *623*, 83–97.

Meaburn, K.J., Misteli, T., and Soutoglou, E. (2007). Spatial genome organization in the formation of chromosomal translocations. Semin. Cancer Biol. *17*, 80–90.

Meek, K., Kienker, L., Dallas, C., Wang, W., Dark, M.J., Venta, P.J., Huie, M.L., Hirschhorn, R., and Bell, T. (2001). SCID in Jack Russell Terriers: A New Animal Model of DNA-PKcs Deficiency. J. Immunol. *167*, 2142–2150.

Meek, K., Dang, V., and Lees-Miller, S.P. (2008). DNA-PK: the means to justify the ends? Adv. Immunol. *99*, 33–58.

Miller, J., McLachlan, a D., and Klug, a (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. *4*, 1609–1614.

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.-L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K. a, et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nat. Biotechnol. *25*, 778–785.

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. Nat. Biotechnol. *29*, 143–148.

Mimitou, E.P., and Symington, L.S. (2008). Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. Nature 455, 770–774.

Mimitou, E.P., and Symington, L.S. (2009). Nucleases and helicases take center stage in homologous recombination. Trends Biochem. Sci. *34*, 264–272.

Mimori, T., Akizuki, M., Yamagata, H., Inada, S., Yoshida, S., and Homma, M. (1981). Characterization of a high molecular weight acidic nuclear protein recognized by autoantibodies in sera from patients with polymyositis-scleroderma overlap. J. Clin. Invest. *68*, 611–620.

Mimori, T., Hardin, J. a, and Steitz, J. a (1986). Characterization of the DNA-binding protein antigen Ku recognized by autoantibodies from patients with rheumatic disorders. J. Biol. Chem. *261*, 2274–2278.

Mimoris, T., and Hardin, J.A. (1986). Mechanism of Interaction between Ku Protein and DNA*.

Misteli, T., and Soutoglou, E. (2009). The emerging role of nuclear architecture in DNA repair and genome maintenance. Nat. Rev. Mol. Cell Biol. *10*, 243–254.

Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. Nat. Rev. Cancer 7, 233–245.

Moscou, M., and Bogdanove, A. (2009). Recognition by TAL Effectors. Science 326, 1501.

Moshous, D., Callebaut, I., Chasseval, D., Corneo, B., Cavazzana-calvo, M., Deist, L., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., et al. (2001). Artemis , a Novel DNA Double-Strand Break Repair / V (D) J Recombination Protein , Is Mutated in Human Severe Combined Immune Deficiency. *105*, 177–186.

Mussolino, C., Morbitzer, R., Lütge, 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 Res. *39*, 9283–9293.

Nakamura, K., Kogame, T., Oshiumi, H., Shinohara, A., Sumitomo, Y., Agama, K., Pommier, Y., Tsutsui, K.M., Tsutsui, K., Hartsuiker, E., et al. (2010). Collaborative action of Brca1 and CtIP in elimination of covalent modifications from double-strand breaks to facilitate subsequent break repair. PLoS Genet. *6*, e1000828.

Neal, J. a, and Meek, K. (2011). Choosing the right path: does DNA-PK help make the decision? Mutat. Res. 711, 73–86.

Neale, M.J., and Keeney, S. (2006). Clarifying the mechanics of DNA strand exchange in meiotic recombination. Nature *442*, 153–158.

Neves, H., Ramos, C., da Silva, M.G., Parreira, a, and Parreira, L. (1999). The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. Blood *93*, 1197–1207.

Nimonkar, A. V, Ozsoy, a Z., Genschel, J., Modrich, P., and Kowalczykowski, S.C. (2008). Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. Proc. Natl. Acad. Sci. U. S. A. *105*, 16906–16911.

Nowsheen, S., and Yang, E.S. (2012). The intersection between DNA damage response and cell death pathways. Exp. Oncol. *34*, 243–254.

Numata, M., Saito, S., and Nagata, K. (2010). Biochemical and Biophysical Research Communications RAG-dependent recombination at cryptic RSSs within TEL – AML1 t (12;21)(p13;q22) chromosomal translocation region. Biochem. Biophys. Res. Commun. *402*, 718–724.

Ochi, T., Blackford, A.N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V.M., Robinson, C. V, et al. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair.

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 242, 1517-1523.

Oksenych, V., Alt, F.W., Kumar, V., Schwer, B., Wesemann, D.R., Hansen, E., Patel, H., Su, a., and Guo, C. (2012). Functional redundancy between repair factor XLF and damage response mediator 53BP1 in V(D)J recombination and DNA repair. Proc. Natl. Acad. Sci. *109*, 2455–2460.

Orthwein, A., Fradet-Turcotte, A., Noordermeer, S.M., Canny, M.D., Brun, C.M., Strecker, J., Escribano-Diaz, C., and Durocher, D. (2014). Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. Science *344*, 189–193.

Panier, S., and Boulton, S.J. (2014). Double-strand break repair: 53BP1 comes into focus. Nat. Rev. Mol. Cell Biol. *15*, 7–18.

Parada, L. a, McQueen, P.G., Munson, P.J., and Misteli, T. (2002). Conservation of relative chromosome positioning in normal and cancer cells. Curr. Biol. *12*, 1692–1697.

Parada, L. a, McQueen, P.G., and Misteli, T. (2004). Tissue-specific spatial organization of genomes. Genome Biol. *5*, R44.

Park, J.-Y., Zhang, F., and Andreassen, P.R. (2014). PALB2: The hub of a network of tumor suppressors involved in DNA damage responses. Biochim. Biophys. Acta *1846*, 263–275.

Parsons, J.L., Dianova, I.I., Finch, D., Tait, P.S., Ström, C.E., Helleday, T., and Dianov, G.L. (2010). XRCC1 phosphorylation by CK2 is required for its stability and efficient DNA repair. DNA Repair (Amst). *9*, 835–841.

Paul, K., Wang, M., Mladenov, E., Bencsik-Theilen, A., Bednar, T., Wu, W., Arakawa, H., and Iliakis, G. (2013). DNA ligases I and III cooperate in alternative non-homologous end-joining in vertebrates. PLoS One *8*, e59505.

Paull, T.T. (2010). Making the best of the loose ends: Mre11/Rad50 complexes and Sae2 promote DNA double-strand break resection. DNA Repair (Amst). *9*, 1283–1291.

Pauwels, K., Podevin, N., Breyer, D., Carroll, D., and Herman, P. (2014). Engineering nucleases for gene targeting: safety and regulatory considerations. N. Biotechnol. *31*, 18–27.

Pavletich, N.P., and Pabo, C. (1991). Zinc Finger-DNA Recognition : Crystal Structure of a Zif268-DNA Complex at 2 . 1. 809–817.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K. a, Liu, O., Wang, N., Lee, G., Bartsevich, V. V, Lee, Y.-L., et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biotechnol. *26*, 808–816.

Perez-Pinera, P., Ousterout, D.G., and Gersbach, C. a (2012). Advances in targeted genome editing. Curr. Opin. Chem. Biol. *16*, 268–277.

Petermann, E., and Helleday, T. (2010). Pathways of mammalian replication fork restart. Nat. Rev. Mol. Cell Biol. *11*, 683–687.

Piganeau, M., Ghezraoui, H., Cian, A. De, Guittat, L., Katibah, G.E., Zhang, L., Tomishima, M., Perrouault, L., Rene, O., Holmes, M.C., et al. (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. 1182–1193.

Poehlmann, A., and Roessner, A. (2010). Importance of DNA damage checkpoints in the pathogenesis of human cancers. Pathol. Res. Pract. *206*, 591–601.

Pommier, Y., Leo, E., Zhang, H., and Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. Chem. Biol. *17*, 421–433.

Price, B.D., and D'Andrea, A.D. (2013). Chromatin remodeling at DNA double-strand breaks. Cell *152*, 1344–1354.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.-T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R. a, Lowe, S.W., et al. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature *440*, 105–109.

Ramsden, D. a (2011). Polymerases in nonhomologous end joining: building a bridge over broken chromosomes. Antioxid. Redox Signal. *14*, 2509–2519.

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.

Raoul Tan, T.L., Kanaar, R., and Wyman, C. (2003). Rad54, a Jack of all trades in homologous recombination. DNA Repair (Amst). *2*, 787–794.

Rass, E., Grabarz, A., Plo, I., Gautier, J., Bertrand, P., and Lopez, B.S. (2009). Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. Nat. Struct. Mol. Biol. *16*, 819–824.

Rass, E., Grabarz, a, Bertrand, P., and Lopez, B.-S. (2012). [Double strand break repair, one mechanism can hide another: alternative non-homologous end joining]. Cancer Radiother. *16*, 1–10.

Reczek, C.R., Szabolcs, M., Stark, J.M., Ludwig, T., and Baer, R. (2013). The interaction between CtIP and BRCA1 is not essential for resection-mediated DNA repair or tumor suppression. J. Cell Biol. *201*, 693–707.

Reinhardt, H.C., and Yaffe, M.B. (2009). Kinases that control the cell cycle in response to DNA damage: Chk1, Chk2, and MK2. Curr. Opin. Cell Biol. *21*, 245–255.

Reynolds, P., Anderson, J. a, Harper, J. V, Hill, M. a, Botchway, S.W., Parker, A.W., and O'Neill, P. (2012). The dynamics of Ku70/80 and DNA-PKcs at DSBs induced by ionizing radiation is dependent on the complexity of damage. Nucleic Acids Res. *40*, 10821–10831.

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

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/lgH translocations. Cell *135*, 1028–1038.

Roberts, S. a, Strande, N., Burkhalter, M.D., Strom, C., Havener, J.M., Hasty, P., and Ramsden, D. a (2010). Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. Nature *464*, 1214–1217.

Rocha, P.P., Micsinai, M., Kim, J.R., Hewitt, S.L., Souza, P.P., Trimarchi, T., Strino, F., Parisi, F., Kluger, Y., and Skok, J. a (2012). Close proximity to Igh is a contributing factor to AID-mediated translocations. Mol. Cell *47*, 873–885.

Rogakou, E.P., Pilch, D.R., Ann, H., Ivanova, V.S., William, M., Orr, A.H., and Bonner, W.M. (1998). CELL BIOLOGY AND METABOLISM : DNA Double-stranded Breaks Induce DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139 *. 1–12.

Rogakou, E.P., Boon, C., Redon, C., and Bonner, W.M. (1999). Megabase Chromatin Domains Involved in DNA Double-Strand Breaks In Vivo. *146*, 905–915.

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. Nat. Genet. *34*, 287–291.

Rooney, S., Chaudhuri, J., and Alt, F.W. (2004). The role of the non-homologous end-joining pathway in lymphocyte development. Immunol. Rev. *200*, 115–131.

Roos, W.P., and Kaina, B. (2013). DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. Cancer Lett. *332*, 237–248.

Roukos, V., and Misteli, T. (2014). The biogenesis of chromosome translocations. Nat. Cell Biol. *16*, 293–300.

Roukos, V., Burman, B., and Misteli, T. (2013a). The cellular etiology of chromosome translocations. Curr. Opin. Cell Biol. 25, 357–364.

Roukos, V., Voss, T.C., Schmidt, C.K., Lee, S., Wangsa, D., and Misteli, T. (2013b). Spatial dynamics of chromosome translocations in living cells. Science *341*, 660–664.

Rowley, J.D. (2008). ASH 50th anniversary review Chromosomal translocations : revisited yet again. *112*, 2183–2189.

Roy, R., Chun, J., and Powell, S.N. (2012). BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat. Rev. Cancer *12*, 68–78.

Ryan, K.M., Ernst, M.K., Rice, N.R., and Vousden, K.H. (2000). Role of NF-kappaB in p53-mediated programmed cell death. Nature *404*, 892–897.

Saintigny, Y., Delacôte, F., Varès, G., Petitot, F., Lambert, S., Averbeck, D., and Lopez, B.S. (2001). Characterization of homologous recombination induced by replication inhibition in mammalian cells. EMBO J. *20*, 3861–3870.

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. Annu. Rev. Biochem. 77, 229–257.

Sartori, A. a, Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. Nature *450*, 509–514.

Sax, K. (1938). Good- hanson.

Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. Annu. Rev. Genet. 45, 167–202.

Schlegel, B.P., Jodelka, F.M., and Nunez, R. (2006). BRCA1 promotes induction of ssDNA by ionizing radiation. Cancer Res. *66*, 5181–5189.

Schreiber, V., Dantzer, F., Ame, J.-C., and de Murcia, G. (2006). Poly(ADP-ribose): novel functions for an old molecule. Nat. Rev. Mol. Cell Biol. 7, 517–528.

Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F.Z., et al. (2011). In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. Stem Cells *29*, 1717–1726.

Shell, B.K., Nash, R.A., Schär, P., Wei, Y., Robins, P., Carter, K., Caldecott, K., and Pappin, D.J.C. (1995). Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. Molecular Cloning and Expression of Human cDNAs Encoding a Novel DNA Ligase IV and DNA Ligase III, an E.

Sibanda, B.L., Critchlow, S.E., Begun, J., Pei, X.Y., Jackson, S.P., Blundell, T.L., and Pellegrini, L. (2001). Crystal structure of an Xrcc4-DNA ligase IV complex. Nat. Struct. Biol. *8*, 1015–1019.

Simsek, D., and Jasin, M. (2010). Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. Nat. Struct. Mol. Biol. *17*, 410–416.

Simsek, D., Furda, A., Gao, Y., Artus, J., Brunet, E., Hadjantonakis, A.-K., Van Houten, B., Shuman, S., McKinnon, P.J., and Jasin, M. (2011a). Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. Nature *471*, 245–248.

Simsek, D., Brunet, E., Wong, S.Y.-W., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., et al. (2011b). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet. *7*, e1002080.

Singleton, B.K., Torres-Arzayus, M.I., Rottinghaus, S.T., Taccioli, G.E., and Jeggo, P. a (1999). The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. Mol. Cell. Biol. *19*, 3267–3277.

Smith, J., Berg, J.M., and Chandrasegaran, S. (1999). A detailed study of the substrate specificity of a chimeric restriction enzyme. Nucleic Acids Res. *27*, 674–681.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318–331.

Soni, A., Siemann, M., Grabos, M., Murmann, T., Pantelias, G.E., and Iliakis, G. (2014). Requirement for Parp-1 and DNA ligases 1 or 3 but not of Xrcc1 in chromosomal translocation formation by backup end joining. Nucleic Acids Res. *42*, 6380–6392.

Soulas-Sprauel, P., Le Guyader, G., Rivera-Munoz, P., Abramowski, V., Olivier-Martin, C., Goujet-Zalc, C., Charneau, P., and de Villartay, J.-P. (2007). Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. J. Exp. Med. *204*, 1717–1727.

Soutoglou, E., and Misteli, T. (2008). On the contribution of spatial genome organization to cancerous chromosome translocations. J. Natl. Cancer Inst. Monogr. *20892*, 16–19.

Stark, J.M., Pierce, A.J., Oh, J., Pastink, A., and Jasin, M. (2004). Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences.

Stavnezer, J., Guikema, J.E.J., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. *26*, 261–292.

Sun, J., Lee, K.-J., Davis, A.J., and Chen, D.J. (2012). Human Ku70/80 protein blocks exonuclease 1-mediated DNA resection in the presence of human Mre11 or Mre11/Rad50 protein complex. J. Biol. Chem. *287*, 4936–4945.

Sung, P. (1994). Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. Science *265*, 1241–1243.

Sy, S.M.-H., Huen, M.S.Y., Zhu, Y., and Chen, J. (2009a). PALB2 regulates recombinational repair through chromatin association and oligomerization. J. Biol. Chem. *284*, 18302–18310.

Sy, S.M.H., Huen, M.S.Y., and Chen, J. (2009b). PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proc. Natl. Acad. Sci. U. S. A. *106*, 7155–7160.

Symington, L.S. (2014). End Resection at Double-Strand Breaks: Mechanism and Regulation. Cold Spring Harb. Perspect. Biol. 6.

Symington, L.S., and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. Annu. Rev. Genet. 45, 247–271.

Szczepek, M., Brondani, V., Büchel, J., Serrano, L., Segal, D.J., and Cathomen, T. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat. Biotechnol. *25*, 786–793.

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.

Taccioli, G.E., Gottilieb, T.M., Blunt, T., Priestey, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., and Jeggo, P. A, (1994). Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. Science *265*, 1442–1445.

Takeda, S., Nakamura, K., Taniguchi, Y., and Paull, T.T. (2007). Ctp1/CtIP and the MRN complex collaborate in the initial steps of homologous recombination. Mol. Cell 28, 351–352.

Terasawa, M., Shinohara, A., and Shinohara, M. (2014). Canonical Non-Homologous End Joining in Mitosis Induces Genome Instability and Is Suppressed by M-phase-Specific Phosphorylation of XRCC4. PLoS Genet. *10*, e1004563.

The, A., Kaklamanis, L., Arnaouti, M., Foukas, P., Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., et al. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly (ADP-ribose) polymerase. *7*, 913–918.

Tomkinson, A.E., and Sallmyr, A. (2013). Structure and function of the DNA ligases encoded by the mammalian LIG3 gene. Gene *531*, 150–157.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature.

Tsai, A.G., and Lieber, M.R. (2010). Mechanisms of chromosomal rearrangement in the human genome. BMC Genomics *11 Suppl 1*, S1.

Tsai, A.G., Lu, H., Raghavan, S.C., Muschen, M., Hsieh, C.-L., and Lieber, M.R. (2008). Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell *135*, 1130–1142.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J. a, Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat. Biotechnol. *32*, 569–576.

Tupler, R., Perini, G., and Green, M.R. (2001). Expressing the human genome. Nature 409, 832–833.

Urnov, F.D., Miller, J.C., Lee, Y.-L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646–651.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C. a, Holt, R. a, et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Vera, G., Rivera-Munoz, P., Abramowski, V., Malivert, L., Lim, A., Bole-Feysot, C., Martin, C., Florkin, B., Latour, S., Revy, P., et al. (2013). Cernunnos deficiency reduces thymocyte life span and alters the T cell repertoire in mice and humans. Mol. Cell. Biol. *33*, 701–711.

Veuger, S.J., Curtin, N.J., Richardson, C.J., Smith, C. M., and Durkacz, B.W. (2003). Radiosensitization and DNA Repair Inhibition by the Combined Use of Novel Inhibitors of DNA-dependent Protein Kinase and. Cancer Research 63, 6008-6015.

Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A.D., Feldhahn, N., Resch, W., Oliveira, T.Y., Chait, B.T., et al. (2013). Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. Science *339*, 711–715.

Walker, J.R., Corpina, R. a, and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature *412*, 607–614.

Wang, B., Matsuoka, S., Ballif, B. a, Zhang, D., Smogorzewska, A., Gygi, S.P., and Elledge, S.J. (2007). Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. Science *316*, 1194–1198.

Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H., and Iliakis, G. (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res. *34*, 6170–6182.

Wang, Y., Ghosh, G., and Hendrickson, E. a (2009). Ku86 represses lethal telomere deletion events in human somatic cells. Proc. Natl. Acad. Sci. U. S. A. *106*, 12430–12435.

Weinstock, D.M., Richardson, C. a, Elliott, B., and Jasin, M. (2006). Modeling oncogenic translocations: distinct roles for double-strand break repair pathways in translocation formation in mammalian cells. DNA Repair (Amst). *5*, 1065–1074.

Weinstock, D.M., Brunet, E., and Jasin, M. (2007). Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. Nat. Cell Biol. *9*, 978–981.

West, S.C. (2003). Molecular views of recombination proteins and their control. Nat. Rev. Mol. Cell Biol. 4, 435–445.

Windhofer, F., Wu, W., and Iliakis, G. (2007). Low Levels of DNA Ligases III and IV Sufficient for Effective NHEJ. 475–483.

Wold, M.S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu. Rev. Biochem. *66*, 61–92.

Woodbine, L., Gennery, A.R., and Jeggo, P. a (2014). The clinical impact of deficiency in DNA non-homologous end-joining. DNA Repair (Amst). *16C*, 84–96.

Wray, J., Williamson, E.A., Singh, S.B., Wu, Y., Cogle, C.R., Weinstock, D.M., Zhang, Y., Lee, S., Zhou, D., Shao, L., et al. (2013). PARP1 is required for chromosomal translocations. *121*, 4359–4365.

Wu, Q., Ochi, T., Matak-Vinkovic, D., Robinson, C. V, Chirgadze, D.Y., and Blundell, T.L. (2011). Non-homologous end-joining partners in a helical dance: structural studies of XLF-XRCC4 interactions. Biochem. Soc. Trans. *39*, 1387–1392, suppl 2 p following 1392.

Xia, B., Sheng, Q., Nakanishi, K., Ohashi, A., Wu, J., Christ, N., Liu, X., Jasin, M., Couch, F.J., and Livingston, D.M. (2006). Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol. Cell *22*, 719–729.

Xia, B., Dorsman, J.C., Ameziane, N., de Vries, Y., Rooimans, M. a, Sheng, Q., Pals, G., Errami, A., Gluckman, E., Llera, J., et al. (2007). Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nat. Genet. *39*, 159–161.

Xie, A., Kwok, A., and Scully, R. (2009). Role of mammalian Mre11 in classical and alternative nonhomologous end joining. Nat. Publ. Gr. *16*, 814–818.

Xing, M., Yang, M., Huo, W., Feng, F., Wei, L., Jiang, W., Ning, S., Yan, Z., Li, W., Wang, Q., et al. (2015). Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. Nat. Commun. *6*, 1–12.

Xu, Y., Ayrapetov, M.K., Xu, C., Gursoy-Yuzugullu, O., Hu, Y., and Price, B.D. (2012). Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair. Mol. Cell *48*, 723–733.

Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A. a, Manis, J.P., et al. (2007). IgH class switching and translocations use a robust non-classical end-joining pathway. Nature *449*, 478–482.

Yaneva, M., Kowalewski, T., and Lieber, M.R. (1997). Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. EMBO J. *16*, 5098–5112.

Yano, K., and Chen, D.J. (2008). Live cell imaging of XLF and XRCC4 reveals a novel view of protein assembly in the non-homologous end-joining pathway ND ES SC RIB ND ES SC Dynamic Movement of Nuclear Proteins. 1321–1325.

Yoo, S., and Dynan, W.S. (1999). Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. Nucleic Acids Res. *27*, 4679–4686.

Yu, X., and Chen, J. (2004). DNA Damage-Induced Cell Cycle Checkpoint Control Requires CtIP, a Phosphorylation-Dependent Binding Partner of BRCA1 C-Terminal Domains DNA Damage-Induced Cell Cycle Checkpoint Control Requires CtIP, a Phosphorylation-Dependent Binding Partner of BRCA1.

Yun, M.H., and Hiom, K. (2009). CtIP-BRCA1 modulates the choice of DNA double-strandbreak repair pathway throughout the cell cycle. Nature *459*, 460–463.

Zeman, M.K., and Cimprich, K. a (2014). Causes and consequences of replication stress. Nat. Cell Biol. *16*, 2–9.

Zha, S., Alt, F.W., Cheng, H.-L., Brush, J.W., and Li, G. (2007). Defective DNA repair and increased genomic instability in Cernunnos-XLF-deficient murine ES cells. Proc. Natl. Acad. Sci. U. S. A. *104*, 4518–4523.

Zha, S., Guo, C., Boboila, C., Oksenych, V., Cheng, H.-L., Zhang, Y., Wesemann, D.R., Yuen, G., Patel, H., Goff, P.H., et al. (2011). ATM damage response and XLF repair factor are functionally redundant in joining DNA breaks. Nature *469*, 250–254.

Zhang, Y., and Jasin, M. (2011). An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat. Struct. Mol. Biol. *18*, 80–84.

Zhang, F., Fan, Q., Ren, K., and Andreassen, P.R. (2009). PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. Mol. Cancer Res. 7, 1110–1118.

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. (2013). The Potent Oncogene NPM-ALK Mediates Malignant Transformation of Normal Human CD4+ T Lymphocytes. Am. J. Pathol. *183*, 1971–1980.

Zhang, Y., Strissel, P., Strick, R., Chen, J., Nucifora, G., Beau, M.M. Le, Larson, R.A., and Rowley, J.D. (2002). Genomic DNA breakpoints in AML1⁻RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t (8; 21) leukemia. 8–13.

Zhu, C., Bogue, M.A., Lim, D., Hasty, P., and Roth, D.B. (1996). Ku86-Deficient Mice Exhibit Severe Combined Immunodeficiency and Defective Processing of V (D) J Recombination Intermediates. *86*, 379–389.

Zhu, C., Mills, K.D., Ferguson, D.O., Lee, C., Manis, J., Fleming, J., Gao, Y., Morton, C.C., and Alt, F.W. (2002). Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. Cell *109*, 811–821.

Zhu, Z., Chung, W.-H., Shim, E.Y., Lee, S.E., and Ira, G. (2008). Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. Cell *134*, 981–994.

Zhuang, J., Jiang, G., Willers, H., and Xia, F. (2009). Exonuclease function of human Mre11 promotes deletional nonhomologous end joining. J. Biol. Chem. *284*, 30565–30573.

Résumé/Abstract

Les translocations chromosomiques qui consistent en l'échange de morceaux de chromosomes sont une des caractéristiques génétiques de nombreux cancers. Les séquences des jonctions des chromosomes transloqués chez les patients correspondent à une réparation par NHEJ. Nous avons étudié le rôle du complexe de ligation XRCC4/LigaseIV du C-NHEJ dans la formation de ces réarrangements chromosomiques dans les cellules humaines. Nous avons utilisé différentes nucléases artificielles (ZFN, TALEN, et CRISPR/Cas9) afin d'introduire deux CDB sur deux chromosomes et nous avons ainsi réussi à générer différentes translocations. Des lignées sauvages et mutantes pour ce complexe de ligation ont été utilisées et la fréquence formation de translocations a été quantifiée par PCR. Nous avons pu observer que celle-ci est souvent diminuée dans les différentes lignées mutantes. Les jonctions des translocations obtenues par séquençage sont modifiées dans des cellules déficientes pour ce complexe. En effet, elles présentent de longues délétions et un biais d'utilisation de microhomologies, indiquant l'utilisation d'un mécanisme alt-NHEJ. Une altération de cette voie dans les cellules humaines n'affecte d'ailleurs pas la formation de ces réarrangements chromosomiques. Ainsi, contrairement aux cellules de souris, les translocations dans les cellules humaines sont générées par le C-NHEJ.

Chromosomal translocations involve the exchange of chromosome pieces and are often associated with oncogenesis. It has been shown that breakpoint junctions of translocated chromosomes found in patients are typical of a repair by NHEJ. Here we investigated the specific role of XRCC4/LigaseIV, the ligation complex of C-NHEJ, on chromosomal translocation formation in human cells. Using different nucleases (ZFN, TALEN, et CRISPR/Cas9) targeting two chromosomes, we studied the induction of translocation in wt and KO human cells, expressing or not the XRCC4/LigaseIV complex. We found that translocation frequency was mostly reduced in XRCC4/LigaseIV deficient cells when we quantified the induction of translocation by PCR. In addition, we analyzed the breakpoint junctions by sequencing. Strikingly, we found that junctions of translocations show large deletions, and a bias towards the use of longer microhomologies only in XRCC4/LigaseIV KO cells, signature of the alt-NHEJ activity. In contrast, translocation formation was not affected in alt-NHEJ deficient cells. Thus conflicting with results obtained in rodent cells where alt-NHEJ promotes translocation formation, translocations in human cells are generated by the C-NHEJ.