



# L'organisation 3D des chromosomes synthétiques de levure

Guillaume Mercy

## ► To cite this version:

Guillaume Mercy. L'organisation 3D des chromosomes synthétiques de levure. Génomique, Transcription et Protéomique [q-bio.GN]. Sorbonne Université, 2018. Français. NNT : 2018SORUS034 . tel-02111002

**HAL Id: tel-02111002**

**<https://theses.hal.science/tel-02111002>**

Submitted on 25 Apr 2019

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

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

# Université Pierre et Marie Curie

Ecole doctorale « Complexité du vivant »

*Institut Pasteur – Régulation Spatiale des Génomes*

## Thèse de doctorat

# L'organisation 3D des chromosomes synthétiques de levure

Présentée et soutenue publiquement le 29 janvier 2018

Présentée par Guillaume Mercy

Dirigée par Romain Koszul

Devant un jury composé de :

Frédéric Devaux

Cécile Fairhead

Joseph Schacherer

Gianni Liti

Président du jury

Rapporteur

Rapporteur

Examineur



# **Thèse de doctorat**

## **L'organisation 3D des chromosomes synthétiques de levure**

**Présentée par : Guillaume Mercy**

**dirigée par : Dr. R. Koszul**

Régulation Spatiale des Génomes  
Institut Pasteur

Pour obtenir le grade de :  
*Docteur de l'Université Pierre et Marie Curie*



*Je souhaite dédier ce manuscrit à ma grand-mère.  
Où que tu sois aujourd'hui, j'ai tenu ma promesse. . .*



# Remerciements

En premier lieu je tiens à remercier l'ensemble des membres de mon jury : Cécile Fairhead, Joseph Schacherer, Gianni Liti et Frédéric Deveaux, pour avoir accepté de me consacrer une partie de leur temps.

Un grand merci à Romain, mon directeur de thèse, pour l'opportunité que tu m'as offerte en me confiant ce sujet de thèse, mais également pour la confiance que tu m'as accordée, alors que nous nous connaissions à peine. J'espère avoir répondu à tes attentes. Pour moi, ces trois ans ont été très bénéfiques, aussi bien d'un point de vue professionnel, car j'ai énormément appris, que d'un point de vue personnel. Cette expérience m'a fait énormément grandir. Je ne pense pas être exactement la même personne à la fin de ce doctorat qu'au début. Alors merci !

Bien sûr, je souhaite également remercier tous les membres de l'équipe avec qui j'ai partagé ces trois ans. Un immense merci à Héloïse, pour tout ce que tu m'as appris, pour l'aide que tu m'as apportée dans la rédaction de certaines parties de ce manuscrit, mais, également, pour tous les bons moments partagés autour d'une bière, d'un café ou lors des conférences annuelles du consortium Sc2.0. Merci également à Julien, pour toutes les structures de génomes que tu as pu me générer et pour ton aide dans l'analyse des souches Sc2.0. *Grazie mille* Luciana et Vittore, pour toutes nos discussions et tout le travail que l'on a partagé. *Thank you Charlie, for your help correcting my english, all the discussion we shared and I will never forget those moments we spent learning RNAseq!* Merci à Martial et Agnès, vous avez été super, sans vous, je ne sais pas s'il aurait été aussi facile de préparer toutes ces banques Hi-C. Merci à tous les autres membres de l'équipe : Axel, Lyam, Théo, Aurèle et tous ceux dont je n'ai pas cité le nom.

*Gracias! Thank you!* Merci, Ingrid, mon épouse, qui a toujours cru en moi et qui m'a énormément supporté tout au long de mon doctorat et surtout ces derniers mois, dans la dernière ligne droite. Merci également à ma mère, sans qui je n'aurais jamais pu arriver jusque là. Merci à tout le reste de ma famille, en France et au Mexique, pour également m'avoir toujours supporté.

Merci à tout mes amis Enguerrand, Jon', Gautier, JP, Coralie, Louis, Tristan, Anto', Théo... et tous les autres qui m'ont toujours encouragé. Vous êtes ma deuxième famille, sans vous rien ne serait pareil. Une mention spéciale pour William, pour m'avoir aidé à maîtriser LaTeX !



# Résumé

Le projet international de synthèse des chromosomes de *Saccharomyces cerevisiae* (projet Sc2.0) a débuté il y a une dizaine d'années en suivant des principes établis par Jef Boeke, le coordinateur du projet. Les chromosomes synthétiques ont été conçus de manière à augmenter la stabilité du génome en supprimant toutes les séquences répétées (ARNt, éléments transposables, etc.), tout en y ajoutant un système d'évolution inductible dépendant du système Cré/LoxP (système SCRaMbLE). Ce système permet de générer rapidement de nombreux réarrangements chromosomiques, il peut ainsi être utilisé pour explorer très rapidement une large gamme de structures chromosomiques, issues de diverses conditions de croissance. Mon projet de thèse s'inscrit dans le cadre du projet européen IESY (Induced Evolution of Synthetic Yeast) et s'appuie sur le projet Sc2.0, qui vise analyser les génomes, les transcriptomes et les métabolomomes des souches Sc2.0 avant et après activation du système SCRaMbLE.

Bien que le design du projet Sc2.0 soit très conservateur en ce qui concerne le contenu des gènes, la suppression de plusieurs classes de séquences répétées peut affecter l'organisation du génome et potentiellement altérer les fonctions cellulaires. En utilisant la méthode de capture de conformation de chromosome couplée au séquençage de seconde génération (Hi-C), mon objectif a été de caractériser d'abord l'organisation 3D du génome des chromosomes parentaux Sc2.0, puis celle de génome de souches évoluées.

À ce jour, huit chromosomes (*synI*, *synII*, *synIII*, *synV*, *synVI*, *synIX-R*, *synX* et *synXII*) ont été entièrement assemblés séparément. En utilisant les souches contenant un ou plusieurs de ces chromosomes, nous avons pu montrer que leur organisation génomique n'est globalement pas affectée par leur présence. Quelques exceptions subsistent, avec *synIII* dont les cassettes HML et HMR ont été retirées, et *synXII* d'où l'ADNr a été retiré et déplacé vers un autre chromosome. À ce stade, nous concluons que l'ADN répétitif dispersé ne conduit pas la conformation moyenne globale de la chromatine chez *S. cerevisiae*. Nous avons également pu exploiter les cartes de contacts pour déterminer les réarrangements ayant pu se produire dans les souches SCRaMbLE.

En plus de ce projet, j'ai également pris part à trois autres projets : (1) l'analyse d'un chromosome synthétique qui a été conçu dans notre laboratoire dans le but d'améliorer la résolution des expériences de Hi-C ; (2) l'analyse de différents facteurs responsables de l'organisation des chromosomes de *S. cerevisiae* au cours des différentes phases du cycle cellulaire ; (3) La caractérisation de l'appariement des chromosomes homologues, en utilisant une levure hybride (*S. cerevisiae* x *Saccharomyces paradoxus*).





# Abstract

The international project Sc2.0 aims to build a fully synthetic genome of *S. cerevisiae*. Sc2.0 chromosomes have been designed following principles established by Jef Boeke, the coordinator of the project, 10 years ago. The main idea was to increase the genome stability by removing all repeated sequences (tRNA, transposable elements, etc.), and to implement SCRaMbLE (for Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution), an inducible, high-throughput chromosome rearrangement system. My PhD project was undertaken within the framework of the Induced Evolution of Synthetic Yeast (IESY) genome project (itself built upon the ongoing Sc2.0 project) which aims to explore synthetic strains before and after SCRaMbLE through genomic, transcriptomic and metabolomic studies.

The design of the synthetic yeast genome Sc2.0 is highly conservative with respect to gene content, the deletion of several classes of repeated sequences and the introduction of thousands of designer changes. However, this design may affect genome organization and potentially alter cellular functions. We investigated the Hi-C-determined three-dimensional (3D) conformations of Sc2.0 and scrambled chromosomes.

Currently, eight synthetic chromosomes (*synI*, *synII*, *synIII*, *synV*, *synVI*, *synIX-R*, *synX* et *synXII*) have been fully assembled. Using these strains we observed that the large-scale genomic organization is globally unaffected by the presence of synthetic chromosome(s). Two exceptions are *synIII*, which lacks the silent mating-type cassettes, and *synXII*, specifically when the ribosomal DNA is moved to another chromosome. We also exploited the contact maps to detect rearrangements induced in these SCRaMbLE strains.

In addition to this work, I collaborated on three other projects. : (1) we analyzed a synthetic design of a chromosomal region of *S. cerevisiae* aiming to improve the resolution of genomic 3C experiments ; (2) we studied the role of different factors involved in the chromosome architecture of *S. cerevisiae* during the cell cycle ; (3) we investigated vegetative homologous pairing using hybrid yeast (*S. cerevisiae* x *S. paradoxus*).



# Table des matières

|                                                                                    |             |
|------------------------------------------------------------------------------------|-------------|
| <b>Remerciements</b>                                                               | <b>v</b>    |
| <b>Résumé</b>                                                                      | <b>vii</b>  |
| <b>Abstract</b>                                                                    | <b>ix</b>   |
| <b>Table des figures</b>                                                           | <b>xv</b>   |
| <b>Liste des tableaux</b>                                                          | <b>xvii</b> |
| <br>                                                                               |             |
| <b>Introduction</b>                                                                | <b>3</b>    |
| <br>                                                                               |             |
| <b>1 L'étude de l'ADN et de l'organisation des génomes</b>                         | <b>3</b>    |
| 1.1 Les approches d'imageries . . . . .                                            | 4           |
| 1.2 Les approches génomiques . . . . .                                             | 6           |
| 1.3 La capture de conformation de chromosome (3C) . . . . .                        | 8           |
| 1.3.1 Le principe de la méthode . . . . .                                          | 8           |
| 1.3.2 Comprendre la matrice de contacts . . . . .                                  | 10          |
| 1.3.3 Les méthodes dérivées du 3C . . . . .                                        | 11          |
| <br>                                                                               |             |
| <b>2 Organisation des chromosomes de <i>S. cerevisiae</i></b>                      | <b>15</b>   |
| 2.1 Le nucléosome, l'unité de base de la fibre de chromatine . . . . .             | 16          |
| 2.2 L'organisation de la fibre de chromatine . . . . .                             | 17          |
| 2.3 Organisation 3D des chromosomes de <i>S. cerevisiae</i> . . . . .              | 20          |
| 2.3.1 Le regroupement des centromères au « spindle pole body » . . . . .           | 21          |
| 2.3.2 L'organisation des télomères . . . . .                                       | 23          |
| 2.3.3 Le Nucléole, compartimentation de la biosynthèse des ribosomes . . . . .     | 25          |
| 2.4 L'organisation fonctionnelle des chromosomes de <i>S. cerevisiae</i> . . . . . | 27          |
| 2.4.1 Le positionnement des ARNt dans l'espace nucléaire . . . . .                 | 27          |
| 2.4.2 L'enveloppe nucléaire inhibiteur de la transcription . . . . .               | 28          |

|          |                                                                                             |           |
|----------|---------------------------------------------------------------------------------------------|-----------|
| <b>3</b> | <b>La synthèse de génomes <i>de novo</i></b>                                                | <b>31</b> |
| 3.1      | Les débuts de la synthèse chimique d'ADN, du simple gène au génome complet .                | 32        |
| 3.1.1    | La synthèse de première base d'ADN, de 20 à 200 pb . . . . .                                | 32        |
| 3.1.2    | L'amélioration des méthodes de synthèse d'ADN pour passer le seuil des<br>1000 pb . . . . . | 34        |
| 3.1.3    | Des gènes aux génomes (viraux) . . . . .                                                    | 36        |
| 3.1.4    | <i>Mycoplasma mycoides</i> , le premier micro-organisme au génome synthétique               | 38        |
| 3.2      | L'édition de génome, une alternative à la synthèse <i>de novo</i> . . . . .                 | 39        |
| 3.2.1    | L'édition de gènes . . . . .                                                                | 39        |
| 3.2.2    | L'édition multiple de génome . . . . .                                                      | 40        |
| 3.2.3    | La réduction de génome . . . . .                                                            | 40        |
| 3.2.4    | <i>S. cerevisiae</i> , une plateforme d'assemblage <i>de novo</i> et d'édition de génome    | 42        |
| 3.3      | La conception et la construction d'un génome synthétique . . . . .                          | 44        |
| 3.3.1    | Design et modélisation . . . . .                                                            | 44        |
| 3.3.2    | Phase de construction . . . . .                                                             | 44        |
| 3.3.3    | Test du génome synthétique . . . . .                                                        | 45        |
| 3.4      | Les projets de synthèse de génomes eucaryotes, « GP-write » . . . . .                       | 45        |
| 3.5      | Les limites de la synthèse de génomes . . . . .                                             | 46        |
| <b>4</b> | <b>Projet de synthèse du génome de <i>S. cerevisiae</i>, projet Sc2.0</b>                   | <b>49</b> |
| 4.1      | Le design des chromosomes synthétiques . . . . .                                            | 50        |
| 4.1.1    | Les éléments supprimés . . . . .                                                            | 50        |
| 4.1.2    | Les éléments introduits . . . . .                                                           | 51        |
| 4.1.3    | Les éléments déplacés . . . . .                                                             | 51        |
| 4.1.4    | Éléments remplacés . . . . .                                                                | 53        |
| 4.1.5    | Les éléments non modifiés . . . . .                                                         | 53        |
| 4.2      | La synthèse des chromosomes synthétiques . . . . .                                          | 53        |
| 4.3      | La combinaison des chromosomes synthétiques . . . . .                                       | 56        |
| 4.4      | Le système d'évolution de SCRaMbLE . . . . .                                                | 57        |
|          | <b>Résultats et discussions</b>                                                             | <b>63</b> |
|          | <b>L'organisation 3D des chromosomes synthétiques</b>                                       | <b>63</b> |
|          | Première publication . . . . .                                                              | 63        |
|          | Résultats supplémentaires . . . . .                                                         | 73        |
| 4.4.1    | Introduction . . . . .                                                                      | 73        |
| 4.4.2    | L'organisation du nouveau chromosome regroupant les ADNt . . . . .                          | 73        |
| 4.4.3    | Regroupement des cassettes HML et HMR indépendamment des télomères                          | 74        |
| 4.4.4    | Organisation de la souche à trois chromosomes . . . . .                                     | 76        |

|                                                                                               |                                                                                                             |            |
|-----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|------------|
| 4.4.5                                                                                         | Analyse de la réplication de la souche à trois chromosomes . . . . .                                        | 78         |
| 4.4.6                                                                                         | Conclusion . . . . .                                                                                        | 82         |
| <b>Design d'un chromosome synthétique pour améliorer la qualité des données de 3C</b>         |                                                                                                             | <b>83</b>  |
| <b>Étude de l'interaction des chromosomes homologue d'une levure hybride</b>                  |                                                                                                             | <b>119</b> |
| 4.5                                                                                           | Introduction . . . . .                                                                                      | 119        |
| 4.6                                                                                           | Résultats et discussions . . . . .                                                                          | 121        |
| 4.6.1                                                                                         | Alignement des données haploïdes sur le génome hybride . . . . .                                            | 121        |
| 4.6.2                                                                                         | Filtrage de la matrice pour corriger les artefacts . . . . .                                                | 121        |
| 4.6.3                                                                                         | L'organisation du génome de l'hybride et les contacts inter-homologues<br>au cours de la phase G1 . . . . . | 124        |
| 4.7                                                                                           | Conclusion . . . . .                                                                                        | 126        |
| <b>L'organisation 3D des chromosomes de <i>S. cerevisiae</i> au cours du cycle cellulaire</b> |                                                                                                             | <b>127</b> |
| <b>Perspectives</b>                                                                           |                                                                                                             | <b>159</b> |
| <b>Références</b>                                                                             |                                                                                                             | <b>165</b> |
| <b>Bibliographie</b>                                                                          |                                                                                                             | <b>165</b> |
| <b>Annexes</b>                                                                                |                                                                                                             | <b>189</b> |



# Table des figures

|      |                                                                                                 |     |
|------|-------------------------------------------------------------------------------------------------|-----|
| 1.1  | Les premières observations des chromosomes et de la structure de l'ADN . . . . .                | 4   |
| 1.2  | Marquage des chromosomes par hybridation <i>in situ</i> . . . . .                               | 5   |
| 1.3  | Capture de conformation de chromosome (3C) . . . . .                                            | 9   |
| 1.4  | Normalisation de la matrice de contacts . . . . .                                               | 11  |
| 1.5  | Le 3C et ses dérivés . . . . .                                                                  | 12  |
| 2.1  | Modèles de chromatine de 30 nm . . . . .                                                        | 19  |
| 2.2  | Modèle Rabl de l'organisation des chromosomes . . . . .                                         | 21  |
| 2.3  | Regroupement des centromères . . . . .                                                          | 23  |
| 2.4  | Structure et organisation des régions télomériques . . . . .                                    | 24  |
| 2.5  | Répétition du rDNA . . . . .                                                                    | 26  |
| 3.1  | L'assemblage d'ADN synthétique par liaison d'oligonucléotides . . . . .                         | 33  |
| 3.2  | Assemblage par polymérisation . . . . .                                                         | 35  |
| 3.3  | Assemblage du génome de <i>M. genitalium</i> dans la levure . . . . .                           | 39  |
| 3.4  | MAGE . . . . .                                                                                  | 41  |
| 3.5  | Cycle d'édition de génome . . . . .                                                             | 43  |
| 4.1  | Les éléments déplacés du génome Sc2.0 . . . . .                                                 | 52  |
| 4.2  | Assemblage hiérarchique d'un chromosome synthétique de levure . . . . .                         | 55  |
| 4.3  | Combinaison de chromosomes synthétiques par endoréduplication . . . . .                         | 56  |
| 4.4  | Le système de SCRaMbLE . . . . .                                                                | 58  |
| 4.5  | Avancement du projet Sc2.0 . . . . .                                                            | 60  |
| 4.6  | Organisation du nouveau chromosome d'ARNt . . . . .                                             | 75  |
| 4.7  | L'organisation des cassettes HML et HMR dans le chromosome fusionné . . . . .                   | 77  |
| 4.8  | Organisation de la souche à trois chromosomes . . . . .                                         | 79  |
| 4.9  | Comparaison de la loi de distance génomique entre les chromosomes natifs et fusionnés . . . . . | 80  |
| 4.10 | Profil de répllication des chromosomes fusionnés . . . . .                                      | 81  |
| 4.11 | Matrice de similarité des génomes de <i>S. cerevisiae</i> et <i>S. paradoxus</i> . . . . .      | 120 |
| 4.12 | Matrice de contrôle de d'alignement du génome de l'hybride . . . . .                            | 122 |
| 4.13 | Orientation des paire de lectures . . . . .                                                     | 123 |

|                                                         |     |
|---------------------------------------------------------|-----|
| 4.14 Filtrage de la matrice de contacts . . . . .       | 124 |
| 4.15 Effet de Rabl ou appariement d'homologue . . . . . | 125 |
| 4.16 Tableau noir de Richard Feynman . . . . .          | 161 |



# Liste des tableaux

|     |                                                |    |
|-----|------------------------------------------------|----|
| 4.1 | Combinaison de chromosomes fusionnés . . . . . | 73 |
| 4.2 | Contacts du chromosome d'ADNt . . . . .        | 74 |



# **Introduction**

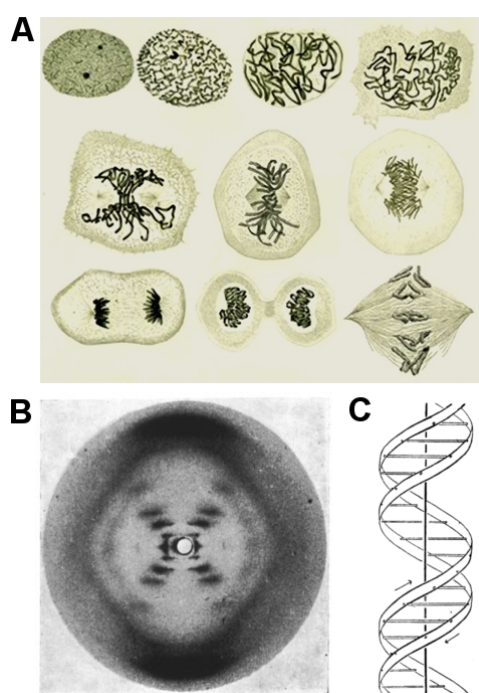


# Chapitre 1

## L'étude de l'ADN et de l'organisation des génomes

Les premières observations de chromosomes remontent aux travaux de Walther Flemming, en 1882, et de Carl Rabl, en 1885. En observant des cellules de salamandre au microscope optique après coloration à l'aniline, Flemming avait décrit et dessiné la répartition égale des chromosomes au cours de la division cellulaire (Figure 1.1), tandis que Rabl a décrit l'organisation des chromosomes (plus de détails sont donnés dans le chapitre 2) avec les centromères et les télomères regroupés à des pôles opposés du noyau (Rabl, 1885). En 1888, Theodor Boveri s'intéressa lui aussi à ces chromosomes, en observant des œufs d'*Ascaris megalocephala* (un ver nématode) au microscope, il décrit alors les chromosomes comme étant des unités très organisées. Puis en 1902, Theodor Boveri et Walter Sutton suggérèrent, dans des travaux indépendants, que les chromosomes pourraient être impliqués dans la transmission des caractères héréditaires. Un peu plus tard, en 1915, preuves cytologiques à l'appui, Thomas Hunt Morgan fut capable de combiner les lois Mendel, énoncées en 1866, à la théorie de Theodor Boveri et Walter Sutton afin d'affirmer que les chromosomes sont porteurs de l'information héréditaire. En 1944, Oswald T. Avery, Maclyn McCarty et Colin MacLeod confirmèrent que l'ADN est responsable de la transmission des caractères héréditaires en réussissant à transformer l'ADN d'une première souche de *Pneumococcus* dans une autre souche de *Pneumococcus*, lui conférant ainsi les caractéristiques de la première (Avery et al., 1944).

Au milieu du XX<sup>e</sup> siècle, bien que le rôle de l'ADN dans la cellule commence à être compris, sa structure reste inconnue et intrigue. En 1953, les analyses de la molécule d'ADN par diffraction au rayons-X (Franklin and Gosling, 1953; WILKINS et al., 1953) ont entraîné la découverte de la structure de la double hélice d'ADN (Figure 1.1) par Watson et Crick (WATSON and CRICK, 1953). Le repliement de cette molécule à l'intérieur du noyau des cellules reste quant à lui un mystère.



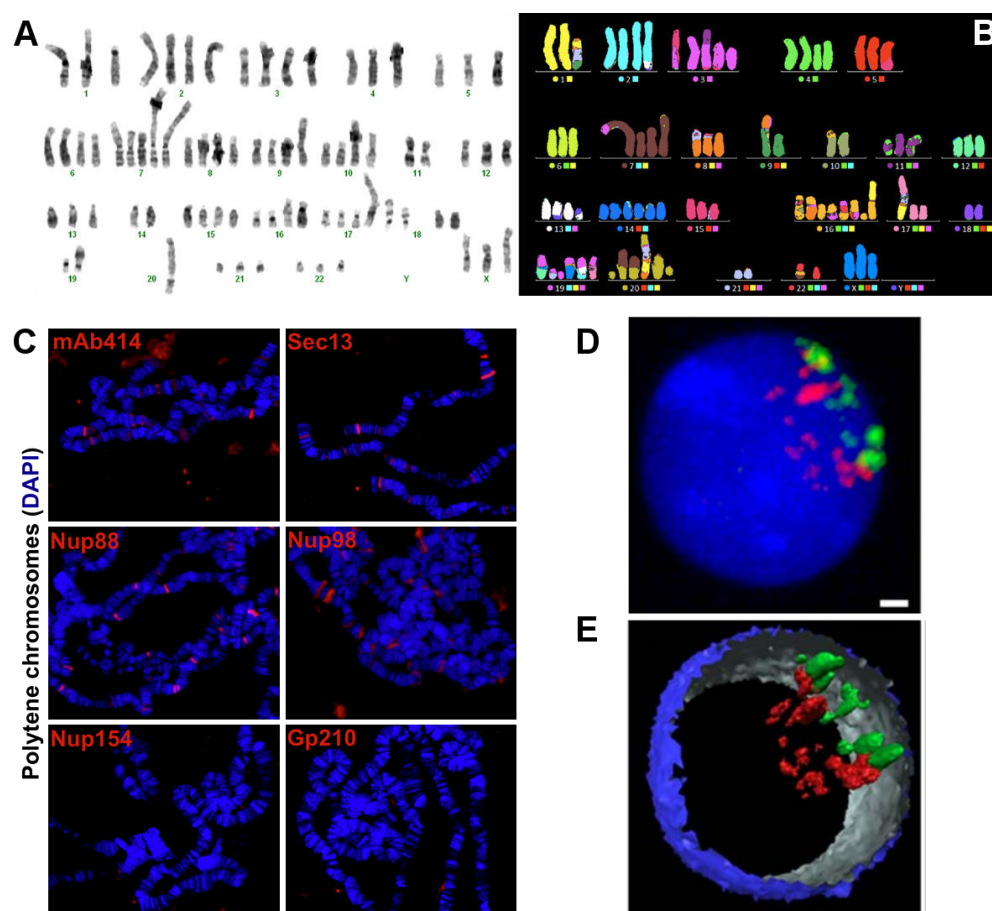
**FIGURE 1.1 Les premières observations des chromosomes et de la structure de l'ADN :** (A) Dessin de l'organisation des chromosomes d'*Ascaris megalocephala* au cours de la division cellulaire. (B) Observation de la double hélice d'ADN au rayon-X par Rosalind Franklin (Franklin and Gosling, 1953) (C) Première représentation de la double hélice d'ADN par Watson et Crick (WATSON and CRICK, 1953)

Aujourd'hui il est devenu très clair que la relation entre l'information génétique de la séquence de l'ADN (1D) et l'organisation spatiale (3D) de la molécule d'ADN est très importante, comme nous allons le voir dans le deuxième chapitre de ce manuscrit. Ce premier chapitre, quant à lui, rapporte une brève vue d'ensemble des méthodes généralement utilisées pour étudier l'organisation des chromosomes. Dans un premier temps, je présenterai les approches d'imageries qui ont pu être utilisées pour observer les chromosomes. Enfin, je présenterai les approches génomiques que j'ai utilisées au cours de mon doctorat et, notamment, l'approche de capture de conformation de chromosomes (3C), que je développerai un peu plus.

## 1.1 Les approches d'imageries

À la fin des années 1950, Tjio et Levan déterminent le compte correct de chromosomes par cellules humaines en utilisant des cultures cellulaires traitées à la colchicine<sup>1</sup> (TJIO and LEVAN, 1956). À la suite de cette découverte, Jérôme Lejeune et ses collègues ont pu mettre en évidence la première aberration chromosomale, en 1959, en observant qu'un chromosome supplémentaire était présent dans les cellules de patient atteint de trisomie (Lejeune et al., 1959). La simple microscopie optique, utilisée dans les années 1960, était suffisante pour permettre de détecter ce genre d'aberration, mais elle ne permettait pas d'identifier spécifiquement les chromosomes. Ce problème a été résolu grâce à des méthodes de coloration faisant apparaître des bandes sombres et claires alternées sur les chromosomes, permettant ainsi de les distinguer et de réaliser les premiers caryotypes. La coloration à la quinacrine moutarde est appelée « Q-banding » (Caspersson et al., 1972), et la coloration au Giemsa est appelée « G-banding » (Sumner et al., 1971) (Figure 1.2).

1. La colchicine perturbe les fuseaux mitotiques et bloque les cellules en métaphase de la mitose.



**FIGURE 1.2 Marquage des chromosomes par hybridation *in situ* :** (A) Caryotype obtenu par coloration au Giemsa, « G-banding ». (B) Caryotype obtenu par marquage FISH des chromosomes (Rondón-Lagos et al., 2014). (C) Marquage FISH de protéines (rouge) associées au chromosome (marquage au DAPI, bleu) sur des noyaux éclatés (Capelson et al., 2010). (D) À partir d'une image de noyau de lymphocyte humain marqué par des sonde de FISH, (E) une reconstruction 3D partielle peut être modélisée (Küpper et al., 2007).

Les méthodes de coloration en bandes permettent de distinguer les chromosomes les uns des autres dans des extraits cellulaires, mais elles ne permettent pas de visualiser ces derniers, ni de localiser des *loci* précis sur un chromosome à l'intérieur des noyaux. Des méthodes basées sur l'hybridation d'ADN avec des sondes ARN ou ADN radioactives se sont alors développées, permettant de marquer précisément la position de *loci* d'intérêts sur les chromosomes *in situ*, comme par exemple les ADN ribosomaux (Gall and Pardue, 1969; John et al., 1969). À la fin des années 1970 et au début des années 1980, les sondes radioactives ont été remplacées par des sondes fluorescentes, c'est le développement de la méthode d'hybridation *in situ* en fluorescence (FISH). Plusieurs types de sondes ont été développées : soit, des sondes ARN qui, une fois hybridées à l'ADN, sont reconnues par un anticorps couplé à un fluorochrome (Rudkin and Stollar, 1977), soit la sonde peut être directement couplée à un fluorochrome (Bauman et al., 1980; Cheung et al., 1977; Manning et al., 1975). Une fois couplés à l'ADN, les fluorochromes des

sondes peuvent être visualisés par microscopie électronique permettant ainsi de déterminer la position du *locus* d'intérêt sur le chromosome. La méthode de FISH est ensuite devenue l'une des méthodes standards de l'analyse de caryotype et de l'analyse de l'organisation de la structure des chromosomes. En 1985, elle a d'ailleurs permis de confirmer expérimentalement l'organisation territoriale des chromosomes suggérée par Carl Rabl (Rabl, 1885) et Theodor Boveri (Boveri, 1909) à la fin du XIX<sup>e</sup> siècle (Schardin et al., 1985). Grâce au développement du microscope confocal à balayage laser, des images de différentes sections du noyau marquées par des sondes de FISH ont pu être obtenues et être utilisées pour reconstruire une image 3D du noyau (Cremer and Cremer, 2010). La méthode de FISH a également permis de montrer l'association de certaines régions de chromosomes avec la membrane nucléaire (Capelson et al., 2010)(Figure 1.2). Comme beaucoup d'autres techniques, la méthode de FISH a ses limites. En effet, pour pouvoir être appliquée, il est nécessaire que les cellules soient préalablement fixées, elle ne permet donc pas de marquer le génome de cellules vivantes.

L'utilisation de système de répresseur/opéron fluorescents (FROS) a permis de dépasser cette limite et d'observer l'organisation de la chromatine *in vivo*. Cette méthode exploite la spécificité qui existe entre un répresseur et la séquence de l'opéron où il se lie (par exemple le couple LacI/LacO). En fusionnant une protéine fluorescente (comme la GFP) au répresseur et en intégrant la séquence de l'opérateur, en série, au niveau d'un *locus* d'intérêt, la position de ce dernier pourra être déterminée après expression du répresseur. Ce design permet de créer une accumulation de protéines fluorescentes en un même *locus* du génome qui peut ensuite être facilement visualisé par microscopie (Belmont and Straight, 1998; Robinett et al., 1996). Cette méthode a facilité l'étude de la dynamique de différents *loci* du génome de *S. cerevisiae* (Lassadi and Bystricky, 2011), mais elle reste très limitée par la faible résolution imposée par la diffraction de la lumière. Depuis une dizaine d'années, des méthodes de microscopie par fluorescence à super résolution ont été développées pour dépasser cette limite. Elles permettent d'atteindre une résolution de l'ordre du nanomètre, rendant possible la localisation de molécules fluorescentes individuelles (Betzig et al., 2006).

## 1.2 Les approches génomiques

Depuis le développement du séquençage Sanger en 1977 (Sanger et al., 1977), de nombreux projets de séquençage de génome ont été réalisés par la communauté scientifique, du phage phiX147, en passant par le génome de la levure, puis le génome humain. Tous ces efforts ont permis de considérablement améliorer la technologie de séquençage, jusqu'à la rendre plus accessible à la communauté scientifique aujourd'hui grâce au séquençage de nouvelle génération (NGS). Cette démocratisation des techniques de NGS a permis de déterminer la séquence de nombreux génomes de références, et leur nombre est en continuelle augmentation. Enfin, en traitant les génomes par différents procédés de biologie moléculaire avant de les séquencer, différents résultats peuvent être



obtenus. Ils permettent d'apporter de nouvelles perspectives à l'étude de ces génomes, notamment en étudiant la relation qui peut exister entre leur séquence, leurs fonctions et leurs structures.

Certaines nucléases, comme la DNase I, ont pendant longtemps été utilisées pour caractériser les sites d'interactions entre ADN et protéines, généralement, des sites de liaisons de facteurs de régulation du génome (amplificateurs, promoteurs, inhibiteurs). Lorsqu'un génome est digéré par la DNase I, les régions qui sont liées par des protéines (comme des facteurs de transcription) vont être protégées de l'action de la nucléase et peuvent être caractérisées. En comparant les résultats de deux digestions réalisées en présence et en absence du facteur de transcription étudié, sa spécificité peut être déterminée (Galas and Schmitz, 1978). En couplant cette méthode au séquençage NGS, DNase-seq, ces sites de liaison de protéine à l'ADN ont pu être cartographiés sur l'ensemble du génome (He et al., 2012; Ling et al., 2010; Voss et al., 2011). Une autre nucléase, la nucléase micrococcale (Mnase), coupe spécifiquement l'ADN libre qui n'est pas impliqué dans la formation de nucléosomes (Noll and Kornberg, 1977). Après digestion de la chromatine par la Mnase, plusieurs fragments d'ADN, dont les tailles correspondent à différentes combinaisons de nucléosomes (un seul nucléosome, deux nucléosomes ou plus), sont obtenus. En analysant, par séquençage haut-débit, les fragments d'ADN dont la taille correspond à un seul nucléosome, la position des nucléosomes sur le génome peut être cartographiée (Weiner et al., 2010). Ces cartographies de sites de liaisons de protéines à l'ADN (Farnham, 2009) ou des nucléosomes (Lee et al., 2007; Schones et al., 2008) ont également pu être déterminées en couplant des réactions d'immuno-précipitations de la chromatine au séquençage de seconde génération (ChIP-seq). Dans ce cas-là, les liaisons ADN-protéines sont d'abord fixées par réaction chimique, faisant généralement intervenir le formaldéhyde. Les protéines d'intérêts fixées à l'ADN peuvent ensuite être isolées à l'aide d'anticorps, avant de purifier l'ADN et de l'analyser.

Dans le paysage de la génomique, ces études ont permis de décrire différentes caractéristiques de la chromatine, comme son accessibilité ou la position de ses nucléosomes. Elles ont également permis d'annoter de nombreuses séquences, notamment celles impliquées dans les mécanismes de régulation génétiques. Cependant, elles ne permettent pas de déterminer l'organisation générale d'un génome dans le noyau des cellules étudiées. Cette question a pu être adressée par la technique de capture de conformation de chromosome (3C) développée par Job Dekker, en 2002, dans le laboratoire de Nancy Kleckner (Dekker et al., 2002). En effet, le 3C et ses variantes, dont la principale qui est, entre autres, couplée au séquençage de seconde génération (Hi-C) (Lieberman-Aiden et al., 2009), permettent d'étudier l'organisation génomique à l'échelle d'une population, pour de nombreux organismes (Crane et al., 2015; Duan et al., 2010; Le et al., 2013; Marbouty et al., 2015; Mizuguchi et al., 2014; Sexton et al., 2012).

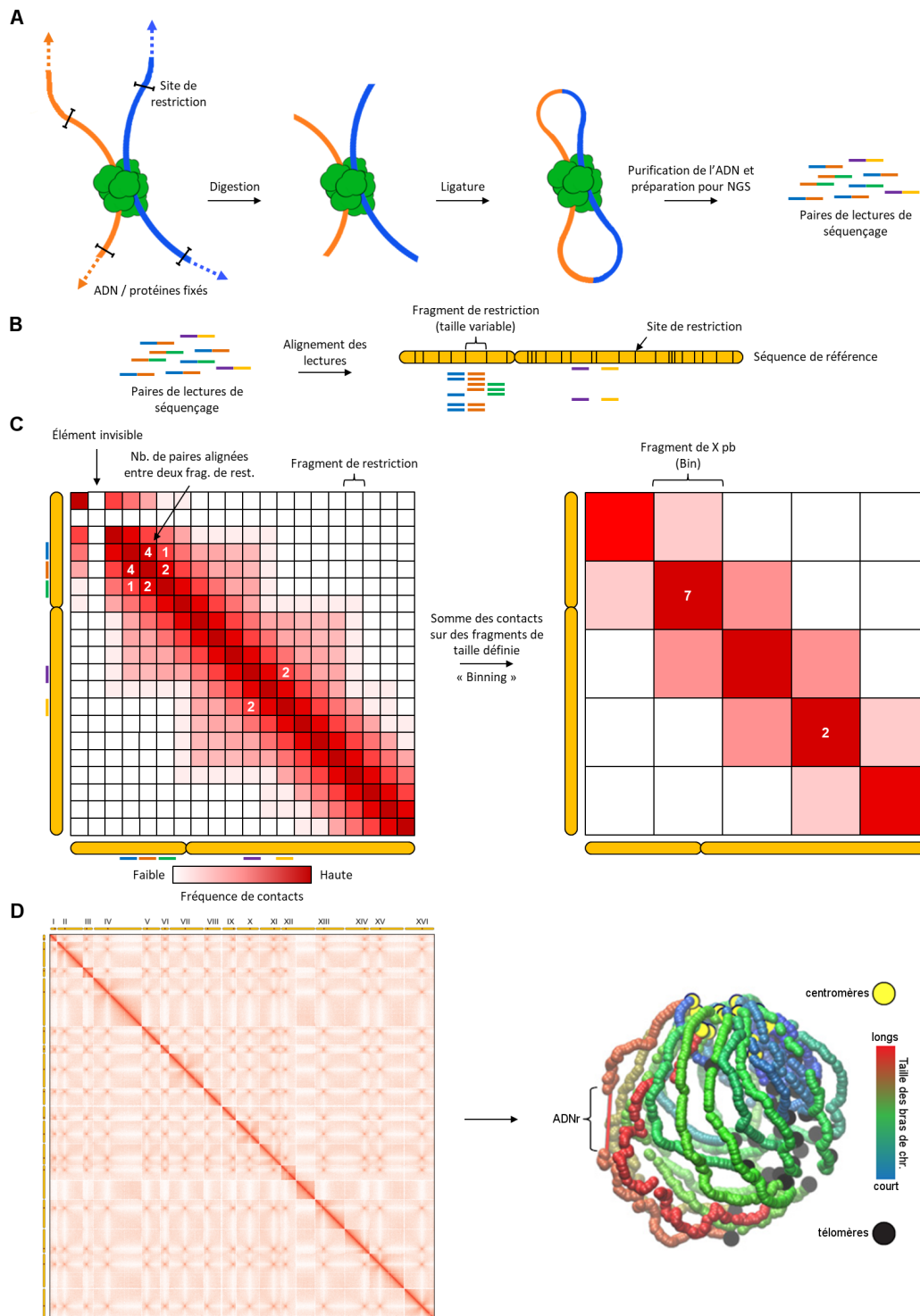
## 1.3 La capture de conformation de chromosome (3C)

### 1.3.1 Le principe de la méthode

La méthode de 3C est basée sur la capture d'interactions entre différents fragments du génome qui sont positionnés à courte comme à longue distances, les uns par rapport aux autres. La capture de ces interactions est réalisée en utilisant le formaldéhyde qui permet de créer des liaisons covalentes entre protéine-protéine et ADN-protéine. Après cette étape de fixation, les régions du génome qui étaient proches les unes des autres dans le noyau des cellules se retrouvent figées dans des complexes d'ADN et de protéines, ce sont ces complexes d'interactions que l'on va chercher à identifier. Pour cela, l'ADN est d'abord digéré à l'aide d'une enzyme de restriction, formant ainsi des extrémités libres au sein de ces complexes. Ces extrémités d'ADN libres pouvant appartenir à des régions éloignées du génome, dans la séquence 1D, sont ensuite assemblées par ligature, formant ainsi un ADN chimérique. Après une étape de purification de ces ADN, ils peuvent être caractérisés soit par PCR-quantitative (3C), ou bien par séquençage NGS (3C-seq, Hi-C ; figure 1.3). Lors de l'analyse des données, les fragments chimériques peuvent être identifiés et la fréquence des interactions entre les différentes régions du génome comptées. Une fois le résultat obtenu, il peut être représenté dans un tableau à double entrées sous forme de carte de contacts. La fréquence d'interaction est généralement représentée par une échelle de couleur, allant par exemple du blanc (fréquence d'interactions faible) au rouge (fréquence d'interactions élevée).

Les matrices de contacts peuvent être générées suivant différents niveaux de résolution. Le premier niveau de résolution, celui des fragments de restriction générés par la digestion du génome, va permettre de représenter les interactions entre ces différents fragments de restriction. En général, à ce niveau de résolution, la matrice est très bruitée à cause de la différence de taille des fragments et de couverture des lectures de séquençage qui peuvent varier d'un fragment à l'autre. Pour diminuer ce bruit dans la matrice, on peut sommer le total d'interactions présent sur des régions de taille préalablement définie (par exemple 5'000 pb). Autrement dit, au lieu de représenter les interactions entre différents fragments de restriction, on les représente entre sections de taille égale (figure 1.3). La méthode de 3C et ses dérivées sont dépendantes de la répartition des sites de coupure de l'enzyme de restriction utilisée lors de la préparation de l'échantillon. Si une région du génome ne contient pas de sites de restriction, alors cette région ne sera pas visible lors de l'analyse des données. De plus une enzyme de restriction qui coupe le génome plus fréquemment qu'une autre permettra d'atteindre une résolution plus fine qu'une enzyme coupant moins fréquemment.

Finalement les matrices de contacts peuvent être utilisées pour générer une structure 3D, qui représente l'organisation moyenne du génome de la population de cellules analysées (Lesne et al., 2014).



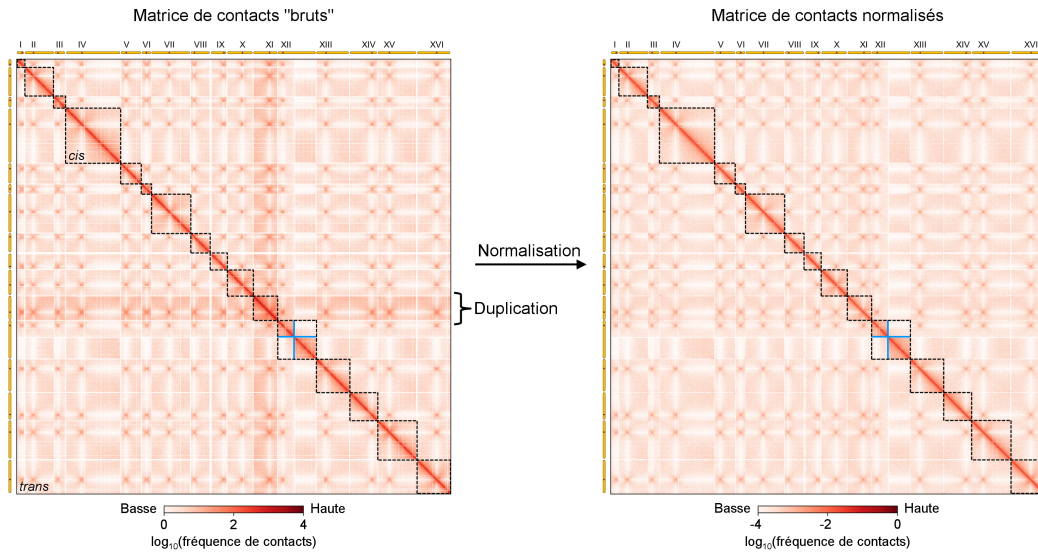
**FIGURE 1.3 Capture de conformation de chromosome (3C) :** (A) Après avoir été fixé au formaldéhyde, le génome est digéré avec une enzyme de restriction. Les extrémités d'ADN libre, pris dans les complexes protéine-ADN, sont ligaturées avant que l'ADN soit purifié pour être séquencé. (B) Les lectures de séquençage sont attribuées aux différents fragments de restrictions auxquels elles correspondent. (C) Le nombre d'interactions entre deux loci est compté à partir des paires de lectures qui leur ont été attribuées. Le résultat est représenté sous forme de carte de contacts (tableau à double entrées) et une couleur est attribuée à la valeur du contact : blanc lorsque le nombre de contacts est faible, et rouge si le nombre de contacts est élevé. Pour améliorer la qualité de la carte de contacts, les interactions peuvent être « binées » en sommant toutes les interactions d'une région de taille définie. (D) Pour faciliter sa compréhension, la carte de contacts peut être représentée sous forme de structure 3D (Lesne et al., 2014).

### 1.3.2 Comprendre la matrice de contacts

Au début de son développement, la méthode de 3C était restreinte à l'identification d'interactions intrachromosomiques (*cis*) entre paires de *loci* spécifiques (Dekker et al., 2002). À l'image des méthodes de DNase-seq ou de ChIP-seq, le couplage du 3C au séquençage NGS a permis d'étendre l'analyse à l'ensemble du génome (Simonis et al., 2006 ; Lieberman-Aiden et al., 2009). Dans ces cas-là, deux profils d'interactions peuvent être visualisés sur les cartes de contacts. Le premier motif, dont la fréquence d'interaction est en moyenne supérieure, correspond aux interactions de paires de *loci* qui résident sur le même chromosome (*cis*). Le second motifs correspond quant à lui aux interactions de *loci* qui résident sur des chromosomes différents (*trans*). Dans une matrice d'interaction à l'échelle du génome, les interactions *cis* apparaissent sous la forme de blocs carrés d'interactions élevées, centrées le long de la diagonale où chaque carré s'aligne avec un chromosome (figure 1.4). Cette tendance peut en partie s'expliquer par l'organisation territoriale des chromosomes, selon laquelle les chromosomes sont physiquement séparés et occupent un volume distinct dans le noyau.

Au sein d'un même chromosome, les différents *loci* n'interagissent pas tous de la même manière, une décroissance de la fréquence d'interaction dépendante de la distance est mesurée (figure 1.4). En d'autres termes, la fréquence d'interaction entre les *loci* en *cis* diminue lorsque leur distance génomique augmente. Dans la matrice de contacts, ce phénomène apparaît comme une diminution progressive de la fréquence d'interaction plus on s'éloigne de la diagonale, c'est la loi de distance génomique. Ce modèle peut s'expliquer par un mouvement aléatoire du chromosome, des *loci* qui sont proches dans le génome interagiront plus fréquemment s'ils se déplacent aléatoirement dans l'espace 3D. Bien sûr, en présence d'un phénomène biologique, comme par exemple le rapprochement des *loci* HMR et HML, localisés aux extrémités du chromosome III de *S. cerevisiae*, on pourra également observer une augmentation de la fréquence de contacts entre ces deux *loci* (Belton et al., 2015). Enfin, la loi de distance génomique est très utile pour identifier des phénomènes de translocation ou d'inversion de certaines régions d'un génome, ou bien même corriger un génome de référence qui n'avait pas été assemblé correctement en premier lieu (Marie-Nelly et al., 2014).

Comme toutes les méthodes basées sur le séquençage, le Hi-C rencontre un certain nombre de limites, notamment en ce qui concerne les séquences répétées. Lors de l'alignement, les lectures de séquençage correspondant à des séquences répétées (par exemple les régions subtélomériques, les transposons ou les séquences d'ADN ribosomal chez *S. cerevisiae*) ne pourront pas être attribuées à l'une plutôt qu'à l'autre. Éventuellement, les lectures s'alignant à cheval entre ces séquences répétées et leurs régions voisines pourront être comptabilisées, mais le nombre d'évènements restera très faible. Les séquences répétées ne seront pas visibles sur la matrices de contacts et une bande d'interactions nulles sera représentée (figure 1.3). Enfin, certaines régions du génome seront mieux couvertes que d'autres, mais ce biais de couverture peut être corrigé en normalisant

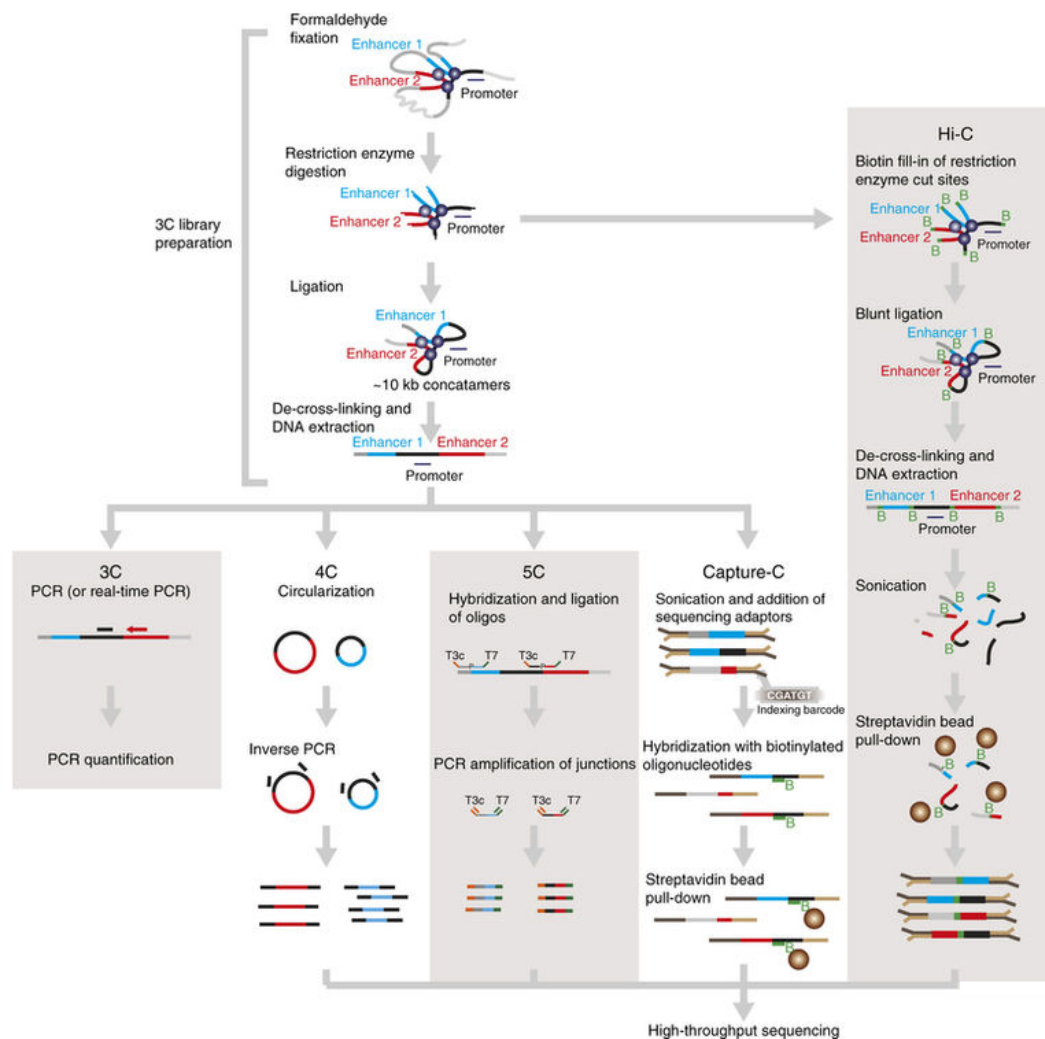


**FIGURE 1.4 Normalisation de la matrice de contacts :** La fréquence de contacts intrachromosomique, *cis*, étant plus élevée que celle des contacts interchromosomique, *trans*, les chromosomes sont facilement identifiable le long de la diagonale grâce à l'apparition de domaines carré. Si l'un des chromosomes de la souche est dupliqué, son signal est alors deux fois plus intense que les autres chromosomes (par exemple le chr. XI ici). Cette différence est éliminée lorsque la matrice est normalisée.

la matrice de contacts (Cournac et al., 2012; Imakaev et al., 2012). La normalisation est une procédure itérative réalisée de telle sorte que la somme de chaque ligne et de chaque colonne de la matrice soit égale à 1. Autrement dit, pour un fragment ou un bin de la matrice, la somme de l'ensemble de ses contacts sera égale à 1. Bien qu'elle permette d'éliminer certains biais de la matrice et qu'aucun détail ne soit perdu, cette procédure peut introduire des artefacts dans la matrice en diminuant le contraste général de la matrice. Par exemple, si une région du génome est dupliquée avant normalisation, le signal de ses contacts est deux fois plus élevé que celui d'une autre région non-dupliquée, alors qu'après la procédure de normalisation le signal de cette région sera au même niveau que celui du reste de la matrice. Après la procédure de normalisation, les duplications ne sont plus visibles (figure 1.4). En éliminant ces biais de couverture grâce à la normalisation, la comparaison entre plusieurs jeux de données est alors possible, et des matrices différentielles peuvent être générées. Ces matrices permettent de mettre en évidence les contacts qui vont être plus fréquents entre certaines régions du génome dans une condition plutôt qu'une autre.

### 1.3.3 Les méthodes dérivées du 3C

Depuis le développement de la méthode de 3C (Dekker et al., 2002), plusieurs dérivés ont été proposés : 4C, 5C, Capture-C et Hi-C (pour revue Davies et al., 2017 ; figure 1.5). Chacune de ces méthodes a été développée pour répondre à une problématique en particulier. Par exemple, à l'aide de sondes d'ARN, le Capture-C sert à enrichir une région de génome en particulier, ce qui permet d'augmenter la résolution de son signal. Dans le cadre de ce manuscrit nous avons utilisé



**FIGURE 1.5 Le 3C et ses dérivés :** Comparaison des différentes méthodes dérivées du 3C. Sauf la méthode de Hi-C pour laquelle une étape de biotinylation est ajoutée entre l'étape de digestion et de ligature, toutes les méthodes partagent la même préparation que pour le 3C. C'est le traitement de la « banque d'ADN 3C », à la fin qui varie en fonction de la méthode utilisée (Davies et al., 2017).

la méthode de Hi-C, à la différence du 3C, une base biotinylée est introduite au niveau des sites de restriction avant l'étape de ligature. En purifiant les fragments d'ADN ayant intégrés cette base biotinylée, nous pouvons ainsi séparer les fragments d'ADN correspondant à une interaction entre deux régions du génome des autres pour les séquencer.





## Chapitre 2

# Organisation des chromosomes de *S. cerevisiae*

Le dogme fondamental de la biologie moléculaire énoncé par Francis Crick en 1970 (Crick, 1970) décrit un flux séquentiel d'information de la molécule d'ADN vers la protéine, alors que le chemin inverse est impossible. Dans ce tableau, l'information génétique de la cellule est donc contenue dans un ou plusieurs chromosomes, représentant autant de molécules d'ADN. Pour autant ces molécules d'ADN ne sont pas aléatoirement organisées à l'intérieur du noyau de la cellule, résultant en une organisation de l'information génétique dans l'espace. Le génome levure de boulangerie *S. cerevisiae* est probablement l'un de ceux qui fut le plus étudié, pour des raisons historiques. Le chromosome III de levure fut le premier chromosome non viral séquencé (Oliver et al., 1992). Puis le génome entier de *S. cerevisiae* fut le premier génome eucaryote à être entièrement séquencé. Le génome haploïde comprend 16 chromosomes codant un peu plus de 6000 gènes, et mesure environ 12 méga paires de bases (Mpb) (Goffeau et al., 1996), le tout étant compacté à l'intérieur d'un noyau mesurant généralement 2 à 3  $\mu\text{m}$ . Ce repliement de l'ADN à l'intérieur du noyau ne se fait pas n'importe comment. Bien au contraire, l'ADN est replié de manière très spécifique et très organisée (Woodcock and Dimitrov, 2001), d'ailleurs sa structure tridimensionnelle (3D) est, depuis quelques années, considérée comme ayant une influence sur l'expression des gènes (Horn and Peterson, 2002). Ainsi, il est devenu une question intéressante d'essayer de trouver quelle influence réelle aurait cette organisation sur une variété de processus métaboliques, notamment la régulation de l'expression de l'information génétique, la condensation des chromosomes en mitose, la réparation des cassures d'ADN, etc.

Au sein d'un noyau de cellule eucaryote, on retrouve une grande quantité de protéines déposées sur les molécules d'ADN constituant les chromosomes. Le résultat de cette association consiste en la fibre de chromatine, elle-même structurée et organisée selon plusieurs niveaux de compaction. L'organisation du noyau peut donc s'analyser à différentes échelles, de la fibre d'ADN sous forme de chromatine à l'agencement de l'ensemble des chromosomes au sein du volume nucléaire. Dans

cette partie nous nous intéresseront à ces différents niveaux d'organisations et leur implication chez *S. cerevisiae*.

## 2.1 Le nucléosome, l'unité de base de la fibre de chromatine

Chez les organismes eucaryotes, la fibre de chromatine est le premier niveau de compaction de l'ADN qui compose leur génome. Elle est le produit d'une série de nucléosomes, l'unité structurale de base qui la compose (Oudet et al., 1975). Le nucléosome est obtenu par l'enroulement d'une séquence d'ADN de (en moyenne) 147 paires de base (pb) autour d'un octamère d'histone, formé à partir de quatre paires d'histones (H2A, H2B, H3 et H4) (Kornberg, 1974; Kornberg and Lorch, 1999; Luger et al., 1997a,b). Les nucléosomes sont séparés les uns des autres par un fragment d'ADN libre, dont la taille varie de 20 à 50 pb (Widom, 1992). Dans les années 1970, une série de nucléosomes compactant un long fragment d'ADN génomique a pu être observée pour la première fois par microscopie électronique, elle avait été décrite comme formant une structure en « collier de perles » (Olins and Olins, 1974), bien que l'agencement précis de cette organisation *in vivo* demeure relativement inconnu (Barbi et al., 2012). Cette structure primaire de l'ADN a été qualifiée de « fibre de 10 nm », par rapport à la taille du diamètre des nucléosomes qui la compose. Ce premier niveau de conformation permet de compacter l'ADN d'un facteur 6.

Les nucléosomes présentent des caractéristiques très intéressantes facilitant leur étude. Le segment d'ADN qui compose le nucléosome est protégé des coupures enzymatiques ou chimiques, alors que le fragment d'ADN connectant les nucléosomes est quant à lui relativement accessible (Zentner and Henikoff, 2014). Certaines nucléases, plus particulièrement la nucléase micrococcalle (Mnase), coupent l'ADN libre, qui n'est pas impliqué dans la formation des nucléosomes, ce qui est généralement le cas des fragments d'ADN liant les nucléosomes (Noll and Kornberg, 1977). En digérant la chromatine avec la Mnase, on obtient des fragments d'ADN de tailles correspondant à différentes combinaisons de nucléosomes (un seul nucléosome, deux nucléosomes ou plus). En analysant par séquençage haut-débit, les fragments d'ADN dont la taille correspond à un seul nucléosome, on peut cartographier la position des nucléosomes sur le génome (Weiner et al., 2010). Ces cartographie peuvent également être obtenues à la suite d'un fractionnement chimique (Brogaard et al., 2012; Voong et al., 2016). Elles ont permis d'identifier trois types de nucléosomes : (1) nucléosomes fixes ; (2) les nucléosomes variables ; et (3) les nucléosomes labiles (Hughes and Rando, 2014). Les nucléosomes fixes, comme leur nom l'indique, sont retrouvés toujours à la même position dans le génome de la population de cellules analysées. Ils sont très enrichis au niveau des frontières des éléments de régulation. Les nucléosomes variables sont en général situés à un peu plus de 1500 pb après les nucléosomes fixes, leur position précise variant entre les différents génomes de la population étudiée. Les nucléosomes labiles sont quant à eux très souvent localisés dans les régions régulatrices (Brogaard et al., 2012; Voong et al., 2016; Weiner et al., 2010; Xi et al., 2011). De nombreuses méthodes d'analyse génomique ont été développées pour permettre

d'identifier les caractéristiques de la structure primaire de la chromatine. Par exemple, les régions de chromatine sans nucléosome, qui sont généralement proches des régions de régulation tels que les promoteurs et les enhanceurs, peuvent être identifiées par FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements) (Giresi et al., 2007), un dérivé du DNase-Seq (Boyle et al., 2008; Crawford et al., 2006; Keene et al., 1981; Wu et al., 1979). Une fois fixées par le formaldéhyde, les régions peu chargées en nucléosomes sont solubles en solution aqueuse, ce qui n'est pas le cas des fragments d'ADN liés à des protéines (Nagy et al., 2003). Ces régions solubles peuvent alors être purifiées et analysées par séquençage.

Des études génétiques et biochimiques ont permis d'identifier les forces responsables de la position des nucléosomes (Kaplan et al., 2009; Korber and Hörz, 2004; Sekinger et al., 2005; van Bakel et al., 2013; Zhang and Pugh, 2011). Dans un premier temps, la flexibilité d'une molécule d'ADN enroulée autour des nucléosomes dépend de sa séquence. Les séquences riches en A/T sont ainsi très rigides et ne favorisent pas la formation de nucléosomes. Deuxièmement, l'encombrement stérique dû à la présence d'autres nucléosomes ou de protéines de liaison à l'ADN influence leurs positions, comme cela a été anticipé par des modèles statistiques (Kornberg and Stryer, 1988). Troisièmement, certaines familles de protéines modifiant la structure primaire de la chromatine, dont l'activité dépend de la présence d'ATP, comme SWI/SNF, ISWI et INO80, peuvent assembler, retirer ou déplacer les nucléosomes de la chromatine (Clapier and Cairns, 2009). Une étude biochimique récente, réalisée *in vitro*, a permis de décrire très précisément ces mécanismes (Krietenstein et al., 2016). Quatrièmement, Le passage de l'ADN polymérase II déstabilise le positionnement des nucléosomes, pour preuve la concentration en nucléosomes et plus faible et plus variable au niveau des gènes très transcrits (Tsankov et al., 2010). Finalement, les protéines structurelles, comme l'histone H1, peuvent limiter les effets d'encombrement stérique et assembler des structures qui ne sont pas stables sans l'intervention de ces protéines. En résumé, la combinaison de toutes ces forces délimite la position et les frontières des régions sans nucléosome, et guide le repliement des autres régions, donnant sa première dimension à la chromatine.

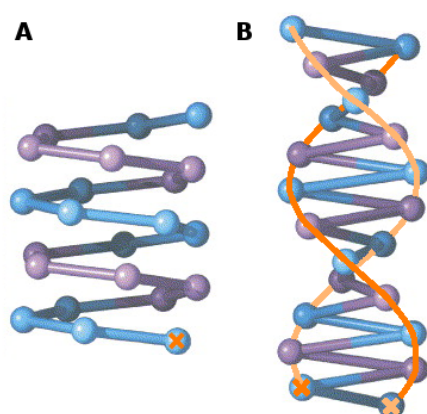
## 2.2 L'organisation de la fibre de chromatine

En 1976, lors d'une expérience *in vitro*, la compaction d'une chaîne de nucléosome en une fibre de 30 nm de diamètre a été mise en évidence (Finch and Klug, 1976; Marsden and Laemmli, 1979). Depuis cette découverte, la manière dont les nucléosomes interagissent les uns avec les autres et la façon dont ces chaînes de nucléosomes interagissent pour structurer la fibre de chromatine ont été beaucoup discutés (Fussner et al., 2011; Luger et al., 2012; Maeshima et al., 2010b; Tremethick, 2007). À partir d'analyses biochimiques combinées à la microscopie électronique et d'autres méthodes d'analyses biophysiques (diffraction par rayon X, pince optique/magnétique, ...), deux modèles de repliement de la fibre de chromatine ont été proposés pour décrire la fibre de 30 nm : (1) le modèle dit de solénoïde, ou hélice à un point de départ (figure 2.1A) (Finch and Klug,

1976; Robinson and Rhodes, 2006), et (2) le modèle d'hélice à deux points de départ (figure 2.1B) (Khorasanizadeh, 2004; Schalch et al., 2005). Dans le modèle de solénoïde, un réseau de nucléosomes s'organisent de sorte que les nucléosomes successifs sont adjacents les uns des autres dans une structure compacte et ils sont reliés par l'ADN de liaison, qui se replie vers l'intérieur de la fibre pour s'adapter aux longueurs variables d'ADN (Finch and Klug, 1976). Le modèle d'hélice à deux points de départ est basé sur un agencement en zigzag des nucléosomes, avec un ADN de liaison essentiellement rectiligne reliant les nucléosomes des côtés opposés de la fibre (Williams et al., 1986; Woodcock et al., 1993). Ces structures, en solénoïde ou en zigzag, sont dépendantes de la taille du fragment d'ADN espaçant les nucléosomes (Routh et al., 2008). Étant donné que la taille de ce fragment d'ADN varie au sein d'une même fibre de chromatine, il a été suggéré que les deux modèles, en solénoïde et en zigzag, puissent cohabiter sur cette même fibre (Grigoryev et al., 2009).

*In vivo*, la nature de la fibre de chromatine est plus complexe. En effet, les histones peuvent présenter des variations ou des modifications d'un nucléosome à l'autre ; la taille des fragments d'ADN espaçant les nucléosomes est également variable ; ou encore, des facteurs de transcription peuvent être présents dans certaines régions de la chromatine. Tous ces éléments, qui apportent de la variabilité et de la complexité dans la composition de la chromatine, rendent son étude très compliquée. Pour simplifier ces différentes variables expérimentales, la majorité des études réalisées se sont reposées sur l'utilisation de chromatine artificielle, basée en général sur des chaînes de nucléosomes obtenues à partir d'histones recombinants uniformes (Grigoryev et al., 2009; Lowary and Widom, 1998). En fonction des conditions expérimentales, la fibre de 30 nm pouvait être visualisée dans la conformation du modèle en zigzag ou bien du modèle de solénoïde. L'organisation en zigzag est préférée lorsque que la fibre est en présence de cations monovalents (Cui and Bustamante, 2000; Dorigo, 2004; Schalch et al., 2005). Alors que la structure en solénoïde est favorisée en présence de cations divalents (Kruithof et al., 2009) ou en présence de l'histone H1 (Robinson and Rhodes, 2006; Routh et al., 2008; Wong et al., 2007). D'autres analyses ont été effectuées sur une chromatine extraite du noyau de cellule (Grigoryev et al., 2016). Dans ce cas de figure, on a un modèle « pseudo-*in vivo* », avec une chromatine plus complexe. Ce modèle a permis d'identifier les deux organisations possibles sur la même fibre, dû à la variation de longueur du fragment d'ADN espaçant les nucléosomes.

De nombreux efforts ont été réalisés pour tenter d'observer cette structure de 30 nm de diamètre *in vivo*, mais aucune section nucléaire n'a permis de mettre la en évidence cette fibre de 30 nm (Chen et al., 2016; Dubochet et al., 1986; Fussner et al., 2011; Luger et al., 2012; Maeshima et al., 2016; Tremethick, 2007). L'hypothèse de son existence, *in vivo*, a été sérieusement remise en cause. Chez *S. cerevisiae*, certaines données obtenues par FISH, *in vivo*, avaient permis de suggérer que ces chromosomes puissent être compactés en une fibre de 30 nm (Bystricky et al., 2004), mais pouvant également soutenir d'autres degrés de compaction de l'ADN (Maeshima et al., 2010a).



**FIGURE 2.1 Modèles de chromatine de 30 nm. (A)** Modèle de repliement de la chromatine en hélice à un point de départ, le modèle dit de solénoïde. **(B)** Modèle d'hélice à deux points de départ. [Figure adaptée de (Robinson and Rhodes, 2006)]

Quelques années après, ces données ont été combinées à des données de capture de conformation des chromosomes (3C) et analysées de manière différente (Dekker, 2008). Cette nouvelle analyse suggère que l'organisation de la fibre de chromatine de *S. cerevisiae* est beaucoup plus proche de la fibre de 10 nm plutôt que de la fibre de 30 nm. De plus, il a été montré, dans les essais *in vitro*, que l'histone connecteur H1 (Hho1p chez *S. cerevisiae*), favorise la compaction de la chromatine (Carruthers et al., 1998; F. et al., 1979; Routh et al., 2008). Or, ce dernier est très peu présent sur les nucléosomes de la levure (Freidkin and Katcoff, 2001). Hho1p se lie très peu au nucléosome lorsque la machinerie de transcription est active, il aurait tendance à se lier aux histones quand l'expression des gènes est moins importante en phase stationnaire ou lors de l'entrée des cellules en quiescence (Schäfer et al., 2008). Hho1p joue un rôle de protection, inhibant la recombinaison homologue (Downs et al., 2003), lorsque le métabolisme de la cellule est ralenti, plutôt qu'un rôle de répression de la transcription. Les régions présentant de nombreux nucléosomes acétylés ont également tendance à être beaucoup plus décondensées que les régions peu acétylées (Robinson et al., 2008; Shogren-Knaak, 2006; Tumbar et al., 1999). Ce manque d'histone de connexion, peut donner le résultat d'une chromatine beaucoup plus étendue, décondensée. Toutes ces observations suggèrent que la structure de la chromatine de *S. cerevisiae* est moins compacte que les modèles d'organisations de la fibre de 30 nm proposés<sup>1</sup>.

Il est très probable que la fibre de 30 nm existe uniquement *in vitro*, dans des conditions où la chromatine est très diluée, favorisant ainsi les interactions entre nucléosomes d'un même segment d'ADN. Alors qu'au sein du noyau, où le nombre de nucléosomes libre, n'étant pas impliqués dans la formation d'une structure particulière, est bien supérieur, les interactions entre nucléosomes de différentes section d'ADN vont être favorisées. Ainsi, la possibilité de former les structures périodiques observées *in vitro* sera limitée (Luger et al., 2012; Maeshima et al., 2016).

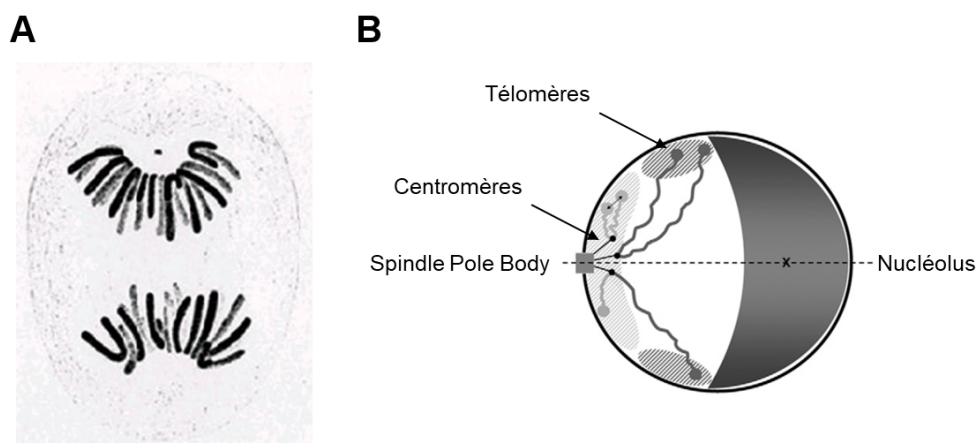
1. La chromatine de 30 nm a pu être observée dans des cellules présentant un très faible taux de transcription, un type spécifique d'histone de connexion H1, des nucléosomes à longue séquence d'ADN et très peu de protéines autre que les histones se liant à l'ADN. C'est le cas par exemple, chez les érythrocytes de poulet ou de necturus (Grigoryev et al., 2016; Scheffer et al., 2011; Williams et al., 1986).

Aucune analyse n'a permis de mettre évidence, *in vivo*, que la fibre de chromatine pouvait s'organiser en une fibre compacte de 30 nm, contrairement aux observations faites au cours des analyses *in vitro*. Au contraire, les condensines ont été retrouvées à l'intersection de ce qui semble être des fibres de chromatine de 10 nm (König et al., 2007). Une théorie alternative selon laquelle les condensines permettraient de structurer les chromosomes en fonctionnant comme des éléments de liaison stochastique entre plusieurs fibres de chromatine de 10 nm (Cuylen et al., 2011; Thadani et al., 2012), a récemment été proposée. Des simulations de dynamiques Brownienne de fibre de chromatine de la taille d'un chromosome de *S. cerevisiae*, soumis à interactions stochastiques entre différents sites de liaison de condensines, on permis de générer, *in silico*, un chromosome présentant les même caractéristiques physiques que celles ayant pu être déterminée *in vivo* (Cheng et al., 2015).

## 2.3 Organisation 3D des chromosomes de *S. cerevisiae*

Lorsque les chromosomes ont été découverts au XIX<sup>e</sup> siècle, leur observation s'est faite lors de la division cellulaire et plus particulièrement lors de la mitose, quand ils prennent une forme condensée (formant cette structure de filament sombre dans le noyau) qui leur permet de transférer les chromatides sœurs au deux cellules filles. Pendant l'interphase, qui regroupe les étapes du cycle cellulaire hors mitose, les chromosomes sont bien moins compacts. En observant des larves de salamandre sous un microscope juste avant la division cellulaire (début de la prophase) et juste après la division cellulaire (fin de la télophase), Carl Rabl a suggéré que les chromosomes, au cours de l'interphase, étaient organisés de manière non aléatoire et étirée (Rabl, 1885) conservant leur disposition anaphasique. Dans ce modèle d'organisation, les régions centrales (centromères) et terminales (télomères) des chromosomes sont regroupées à des pôles opposés du noyau (figure 2.2A). Cette organisation, qualifiée d'organisation Rabl, observée au cours de l'interphase est considérée comme étant le résultat de l'organisation des chromosomes au cours de l'anaphase de la mitose précédente.

Plusieurs espèces voient leur génome organisé selon cette organisation Rabl. C'est ainsi le cas des 16 chromosomes de *S. cerevisiae*, qui pendant toute la durée du cycle cellulaire sont ainsi polarisés. La polarisation des chromosomes est due à l'attachement de leurs centromères au niveau du centre organisateur de microtubules (ou spindle pole body, SPB), et à la position préférentielle des télomères à l'enveloppe du noyau. Chez *S. cerevisiae*, le regroupement des centromères n'est pas dépendant de leur organisation au cours de l'anaphase de la mitose précédente, et les télomères ne sont pas situés en vis-à-vis du SPB au niveau du pôle opposé du noyau. Pour ces différentes raisons on parle plutôt d'organisation « pseudo-Rabl » plutôt que Rabl (Jin et al., 2000)(figure 2.2B).



**FIGURE 2.2 Modèle Rabl de l'organisation des chromosomes :** (A) Dessin de Carl Rabl représentant les chromosomes des larves de salamandres au cours de l'anaphase. Les centromères sont attachés aux microtubules entraînant le repliement des chromosomes sur eux-mêmes, et le positionnement des télomères du côté opposé aux centromères (adaptée de Rabl, 1885). (B) Organisation de « pseudo-Rabl » des chromosomes de *S. cerevisiae* maintenue pendant toute la durée de l'interphase. Les centromères sont regroupés au niveau du SPB, le nucléole est situé au pôle opposé au SPB et les télomères appartenant à des bras de chromosomes de tailles similaires forment des foyers au niveau de l'enveloppe nucléaire (adaptée de Therizols et al., 2010).

### 2.3.1 Le regroupement des centromères au « spindle pole body »

Le « spindle pole body » (SPB) est un important complexe protéique, composé d'au moins 18 protéines. Chacun des gènes codant ces protéines sont essentiels à la croissance de la cellule. Sa structure est dynamique, son diamètre peut varier de 80 nm en phase G1 à 110 nm en mitose, il varie aussi en fonction du nombre de chromosomes dans le noyau, 80 nm dans une cellule haploïde, 160 nm dans une cellule diploïde et encore le double pour un tétraploïde. L'augmentation du diamètre du SPB permet l'augmentation du nombre de microtubules associés à ce dernier (Bullitt et al., 1997; Byers and Goetsch, 1974). Chez *S. cerevisiae*, l'enveloppe nucléaire n'étant pas désassemblée au cours de la mitose (mitose fermée), le SPB y demeure ancré durant la totalité du cycle cellulaire. Créant une interface entre nucléoplasme et cytoplasme, le SPB organise les microtubules nucléaires et cytoplasmiques, jouant ainsi un rôle important dans la ségrégation des chromosomes et le positionnement du noyau dans la cellule, respectivement (Seybold and Schiebel, 2013).

Au début de chaque cycle cellulaire, le SPB est dupliqué au cours de la phase G1, et devient fonctionnel quand la réplication des chromosomes commence, au début de la phase S. Ainsi, à la fin du cycle, les cellules mère et fille obtiennent chacune une copie du SPB et du génome. Généralement, un problème de duplication du SPB ne sera pas détecté par la cellule avant la mitose. La cellule tentera tout de même de terminer sa mitose en effectuant une ségrégation des chromosomes avec un seul SPB, générant des aneuploidies (Chial and Winey, 1999). Pour cette raison, une duplication parfaite du SPB en G1 est essentielle au maintien de la stabilité génomique

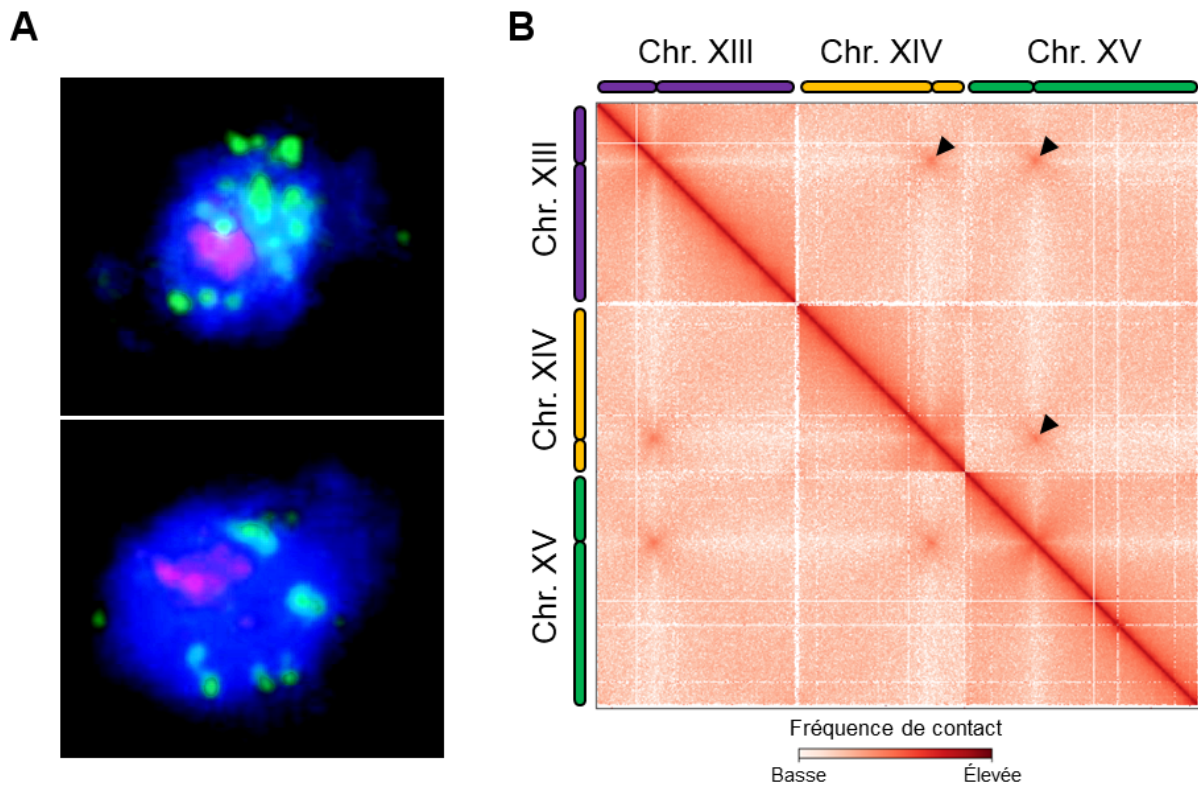
de la cellule. A la fin de la phase G1 les deux copies du SPB sont liées l'une à l'autre par un pont protéique, qui sera coupé durant la mitose, sous l'action de Cdc28 associée aux cyclines mitotiques (Clb1-4) (Mathias et al., 1996). Une fois les deux copies du SPB séparées, le noyau peut-être polarisé pour la ségrégation des chromosomes grâce à leur migration. La migration du SPB dépend des microtubules et de protéines moteurs Cin8 et Kip1 (Jacobs et al., 1988; Roof et al., 1992).

L'attachement des chromosomes au SPB se fait via un microtubule relié au kinétochore, ce dernier étant accroché à la région centromérique du chromosome (McAinsh et al., 2003). Chez *S. cerevisiae*, le centromère est une séquence d'environ 125 pb, où l'on retrouve trois régions très conservées entre chacun des centromères : CDEI, CDEII et CDEIII (Fitzgerald-Hayes et al., 1982). Cse4, l'histone centromérique de *S. cerevisiae*, se lie aux éléments CDEI et CDEII de chacun des centromères, formant ainsi un squelette sur lequel les protéines du kinétochore peuvent s'assembler (Cheeseman et al., 2002; Meluh et al., 1998). Une fois assemblé, le kinétochore se lie à l'extrémité positive d'un microtubule (Joglekar et al., 2009), permettant ainsi d'attacher les chromosomes au SPB. Cet attachement des centromères entraîne leur regroupement au niveau du SPB, synonyme de l'organisation en « pseudo-Rabl » des chromosomes. La suppression de la liaison kinétochore-microtubule entraîne la perturbation de ce regroupement des centromères (Jin et al., 2000).

Le regroupement des centromères au niveau du SPB a été mis en évidence par plusieurs méthodes. Notamment l'utilisation du système Cré/lox qui a permis de calculer des taux de recombinaison entre plusieurs régions du génome. Lorsque les sites loxP étaient situés dans régions adjacentes aux centromères, le taux de recombinaison calculé était plus important, suggérant que les centromères étaient plus proches les uns des autres (Burgess and Kleckner, 1999). Les centromères ont également été observés au microscope par méthode de FISH, ils ont alors été décrits comme se regroupant en forme de rosette autour du SPB (figure 2.3A) (Guacci et al., 1997; Jin et al., 1998). Ces résultats ont été confirmés plus récemment par l'utilisation de la méthode de 3C combinée au séquençage à haut débit (ou de seconde génération). Après analyse, la majorité de contacts interchromosomiques sont concentrés dans des fenêtres 20 kb au niveau des régions centromériques (figure 2.3B) (Duan et al., 2010; Guidi et al., 2015). Enfin, la distance entre les centromères et le SPB a pu être mesuré grâce à l'utilisation de différentes protéines rapportrices en microscopie (FROS) (Dorn et al., 2005). L'analyse de ces images a permis d'estimer la distance SPB-centromère entre 50 et 200 nm pour une cellule en phase G1 du cycle cellulaire.

Bien que la séquence des centromères soit très courte, elle est essentielle pour permettre la ségrégation précise des chromosomes au cours de la mitose (Clarke and Carbon, 1980; Cottarel et al., 1989). Ce mécanisme de ségrégation est rendue possible par la migration des SPBs au sein de l'enveloppe nucléaire. Cette migration dépend des microtubules et de protéines moteurs Cin8 et Kip1 (Jacobs et al., 1988; Roof et al., 1992). Pour *S. cerevisiae* qui complète la mitose dans un





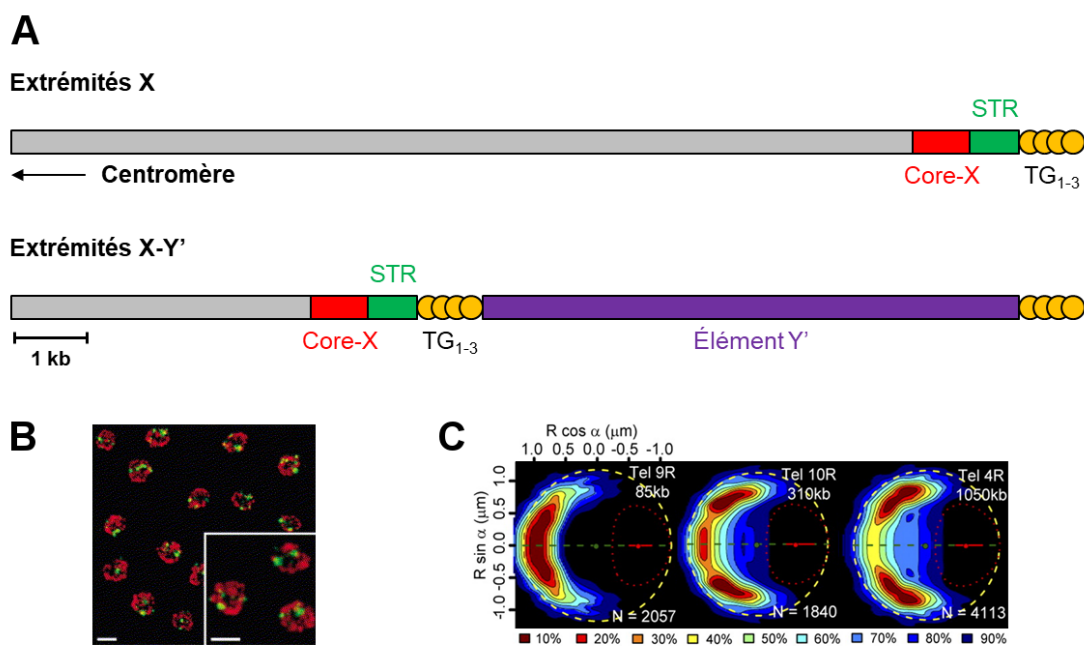
**FIGURE 2.3 Regroupement des centromères :** (A) Marquage FISH des régions centromérique (rouge) et des régions télomérique (vert) sur des noyaux de levures haploïdes (Jin et al., 1998). (B) Le regroupement des centromères observé sur une carte de contacts générée à partir de données de Hi-C (flèche noire).

noyau fermé, le SPB a un rôle particulier, car il influence l'organisation des chromosomes tout au long du cycle cellulaire.

### 2.3.2 L'organisation des télomères

Comme chez beaucoup d'autres organismes eucaryotes, les extrémités des chromosomes de *S. cerevisiae* sont protégées par un important complexe protéique, pour constituer les télomères. Les séquences télomériques se caractérisent par des domaines répétés d'environ 250 pb, principalement composées de répétitions de  $TG_{1-3}$ . En amont des télomères, on retrouve les subtélomères qui sont quant à eux des régions de quelques dizaines de kb ( $\approx 30$  en moyenne) souvent composées de nombreux gènes non-essentiels mais répétés. Chez tous les subtélomères, on retrouve une région de 500 pb très conservée, appelée core-X et de petites séquences répétées (STR, « Subtelomeric repeat sequences »). On retrouve également chez 17 des 32 subtélomères, l'élément Y' qui est une région de 4 à 8 kb dont la fonction reste inconnue (Liti and Louis, 2005; Louis, 1995).

Suite à des observations réalisées au microscope par méthode de FISH, les télomères et les subtélomères ont pu être localisés à la périphérie de la membrane nucléaire (figure 2.4). En plus de leur position, il a été observé que certains de ces télomères tendent à se rapprocher les uns des autres



**FIGURE 2.4 Structure et organisation des régions télomériques :** (A) Il existe deux types d'extrémités chromosomiques chez *S. cerevisiae*. Celles qui n'ont qu'un core X et celles qui ont à la fois un core X et un élément Y'. Entre les éléments X et les répétitions télomériques (TG<sub>1-3</sub> terminales), ou entre le core X et l'élément Y', se trouvent plusieurs petites séquences répétitives subtélomériques (STR) (Louis, 1995). (B) Marquage FISH des régions télomériques (vert) et des pores nucléaires (rouge) (Hediger et al., 2002). (C) Carte du noyau de levure représentant les probabilités de localisation pour un bras de chromosome court (Tel9R), moyen (Tel10R), ou long (Tel4R) (Therizols et al., 2010).

en foyers contenant de 3 à 8 télomères (Gotta et al., 1996; Hediger et al., 2002). De nombreuses protéines, enrichies à la périphérie nucléaire, participent à l'établissement de ces foyers. Les protéines Sir (Sir 4/Sir 2), recrutées par Rap1, font parties de cette famille de protéine (Hecht et al., 1995). Les complexes yKU (YKU70/YKU80) sont également nécessaires à l'organisation de ces télomères. La délétion d'une des protéines appartenant à ces complexes (SIR4, YKU70 ou YKU80) entraîne la désagrégation des foyers télomériques, mais n'altère la localisation périphérique des télomères (Bystricky et al., 2005; Hediger et al., 2002; Laroche et al., 1998; Tham et al., 2001). Certaines protéines de la membrane nucléaire Esc1 et Mps3 semblent jouer un rôle important dans cette localisation des foyers télomériques en interagissant, indépendamment l'une de l'autre, avec la protéine Sir4 (Andrulis et al., 2002; Bupp et al., 2007; Taddei et al., 2004). Lors de la délétion simultanée des complexes yKU et SIR, le nombre de télomères délocalisés augmente, suggérant que ces complexes ont des fonctions redondantes impliquées dans la position des télomères. Tous les mécanismes de localisation des télomères en périphérie ne sont pas encore totalement élucidés. Cependant, leurs positionnements sur la sphère que constitue la membrane nucléaire dépendent principalement des propriétés physiques des molécules d'ADN. Notamment, la taille des bras chromosomiques intervient dans la spécificité des télomères composant les différents foyers. Les bras courts ne semblent pas pouvoir spontanément explorer l'ensemble du noyau et leurs télomères ont tendance à être localisés en périphérie, proches des centromères, au contraire des bras plus

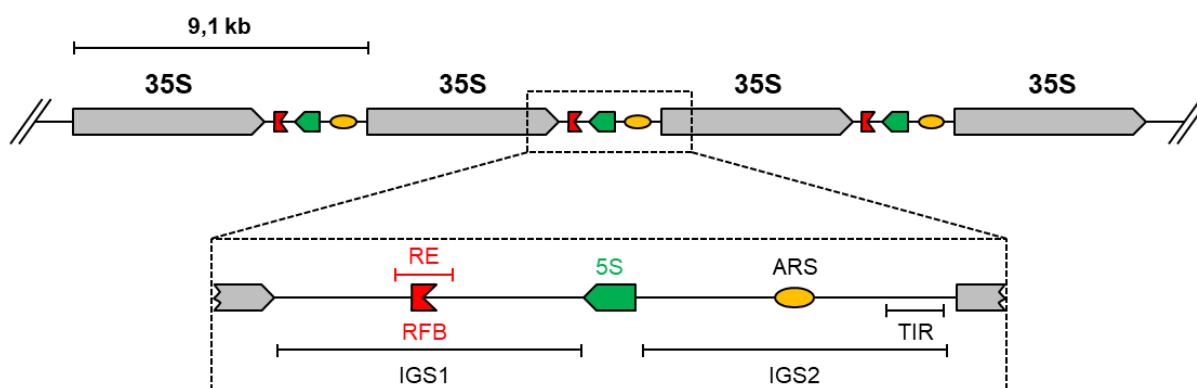
longs. Cette observation se vérifie par les fréquences d'interaction bien plus élevées entre les régions télomériques appartenant aux bras courts qu'avec le reste du génome (Cournac et al., 2012; Duan et al., 2010; Therizols et al., 2010). L'organisation des télomères est donc directement dépendante des caractéristiques physiques du chromosome, un bras court étant un facteur limitant ne permettant pas à son télomère d'explorer l'ensemble du noyau et d'établir des interactions avec des télomères issus de bras longs.

Ainsi, les foyers télomériques forment au sein du noyau des zones aux caractéristiques bien particulières. Ils contiennent des protéines qui leur sont spécifiques influençant leur organisation. Ils constituent un environnement local répressif pour la transcription pouvant influencer l'expression des gènes localisés à proximité.

### 2.3.3 Le Nucléole, compartimentation de la biosynthèse des ribosomes

Chez *S. cerevisiae*, les gènes codant les sous-unités ribosomales sont regroupés en un seul locus du génome, au milieu du bras droit du chromosome XII. Ce locus d'ADN ribosomal (ADNr) est organisé en une série de 100 à 200 séquences répétées de 9,1 kpb (Pasero and Marilley, 1993), il représente environ 10% du génome. Chacune des répétitions de l'ADNr est transcrite en ARNr 35S, un ARNr précurseur, par l'ARN polymérase I (ARN pol I) et en ARNr 5S par l'ARN polymérase III (ARN pol III) (figure 2.5). Cet ensemble d'ADNr et de protéines constitue le nucléole qui est un sous-compartiment du noyau où se fait la biosynthèse des ribosomes. Le nucléole est accolé à l'enveloppe nucléaire, en forme de croissant, du côté opposé au SPB. Dans des conditions normales de culture, il occupe approximativement un tiers du volume nucléaire (Bystricky et al., 2005; Léger-Silvestre et al., 1999; Yang et al., 1989). Sa morphologie est extrêmement influencée par le taux de croissance des cellules, probablement en réponse au taux de ribosomes nécessités par la cellule pour la production de protéines (Oakes et al., 1993; Powers and Walter, 1999).

Lors de l'analyse des caractéristiques du nucléole, l'utilisation artificielle de l'ARN polymérase II (ARN pol II) pour transcrire les gènes d'ADNr, à la place l'ARN pol I, a conduit à une importante destructuration du nucléole. Cette expérience a permis de mettre en évidence l'importance de l'ARN pol I, et de ses facteurs de transcriptions associés, dans l'organisation du nucléole (Oakes et al., 1993; Trumtel et al., 2000). Cette expérience a également permis de démontrer que la structure du nucléole n'est pas essentielle à la transcription des gènes d'ADNr puisque l'ARN pol II a pu maintenir leur transcription. Suggérant ainsi que le nucléole permet d'organiser et d'optimiser le processus transcriptionnel assuré par l'ARN pol I. Autrement dit, la structure du nucléole est intimement liée à sa transcription et son apparition est la conséquence de l'activité de l'ARN pol I (Albert et al., 2011; Sirri et al., 2008). Il a été proposé que la production des ARNr, par l'ARN pol I, conduit à un phénomène d'auto-assemblage du nucléole (Hernandez-Verdun et al., 2002; Misteli, 2001; Trumtel et al., 2000). Le repliement exacte des 2Mpb du nucléole est peu connu, mais il a été montré que les cohésines, les condensines et la protéine Sir2 jouent un



**FIGURE 2.5 Répétition du rDNA.** Chez *S. cerevisiae*, l'ARN ribosomal est codé dans 100 à 200 répétition placée en série. Chaque répétition a une taille de 9,1 kpb et permet de produire un précurseur d'ARNr 35S, transcrit par l'ARN pol I et un ARNr 5S transcrit par l'ARN pol III. Le gène du 5S est entouré de deux espaceur intergénique, IGS1 et IGS2. A l'intérieur de ces espaces, on retrouve une barrière de fourche de réplication (RFB) orientée, un enhanceur de recombinaison (RE), une région d'initiation de la transcription de l'ARN pol I (TIR) et une origine de réplication (ARS).

rôle important dans la compaction du nucléole au cours du cycle cellulaire (Guacci et al., 1994; Lavoie et al., 2002).

Une telle série de séquences répétées présente les caractéristiques idéales pour le fonctionnement de la recombinaison homologue, la voie de réparation de cassure double brin la plus utilisée par *S. cerevisiae*. La stabilité de ces séquences répétées est un facteur essentiel pour assurer la croissance et la survie des cellules, pour cette raison le nucléole doit être protégé de la machinerie de recombinaison. En effet, des événements de recombinaison de l'ADNr peuvent entraîner la génération d'ADNr circulaire (Sinclair and Guarente, 1997). Pour éviter de tels événements de recombinaison, *S. cerevisiae* a développé deux mécanismes de protection impliquant la protéine Sir2. Le premier mécanisme est corrélé avec une organisation locale des nucléosomes, la compaction de la chromatine, dépendante de Sir2 (Bryk et al., 1997; Fritze et al., 1997; Gottlieb and Esposito, 1989; Smith and Boeke, 1997). Le second concerne l'organisation plus globale de la structure de la chromatine avec l'ancrage du nucléole à l'enveloppe nucléaire (Chan et al., 2011; Mekhail et al., 2008). Ces deux mécanismes permettent ainsi d'empêcher la machinerie de recombinaison homologue d'accéder aux séquences répétées de l'ADNr. Cependant, en cas de cassures double brins dans ces séquences, elles peuvent être transitoirement délocalisées du nucléole vers le nucléoplasme pour être réparées par la machinerie de recombinaison homologue (Torres-Rosell et al., 2007).

Le nucléole participe directement à l'organisation du génome en séparant ces séquences répétées de l'ADNr du reste des chromosomes. Il influence également le positionnement de certaines régions chromosomiques telles que les télomères. En effet, la diminution du volume du nucléole a une conséquence directe sur l'organisation des régions télomériques qui viennent occuper l'espace

vide laissé par le nucléole (Therizols et al., 2010). Ainsi, la structure du génome et l'organisation du nucléole sont intimement liées.

## 2.4 L'organisation fonctionnelle des chromosomes de *S. cerevisiae*

L'organisation des chromosomes ne se fait pas aléatoirement, elle joue un rôle très important dans le métabolisme de l'ADN. Très souvent les altérations de cette organisation peuvent être associées à des perturbations de la réplication (Raghuraman and Brewer, 2010), de la transcription ou de la réparation de l'ADN. L'organisation du nucléole, décrite précédemment permet de parfaitement illustrer ce principe d'organisation fonctionnelle. En effet, bien que les machineries nucléolaires soient en échange permanent avec le nucléoplasme, le nucléole isole différentes étapes de la biogenèse des ribosomes. Il crée ainsi un domaine de répression pour des molécules qui sont normalement actives en dehors du nucléole (Sirri et al., 2008).

### 2.4.1 Le positionnement des ARNt dans l'espace nucléaire

Contrairement à l'ADNr qui est regroupé en un seul locus dans le génome de *S. cerevisiae*, les 274 gènes des ARN de transfert (ARNt), transcrits par l'ARN pol III, sont répartis le long de ses 16 chromosomes. À la suite d'observations microscopiques d'hybridation de fluorescence *in situ* (FISH), beaucoup de ces ARNt ont été identifiés formant des groupes dans l'espace nucléaire et certains de ces groupes ont été localisés à proximité du nucléole (Thompson et al., 2003; Wang et al., 2005). Cette association reflète probablement le fait que l'ARN 5S de l'ADNr et les ARNt soient transcrits simultanément par l'ARN pol III. Cette juxtaposition permet probablement l'activation simultanée des ARN pol I et pol III pour l'expression de la machinerie de synthèse des protéines. En utilisant du nocodazole, un agent biologique antimitotique permettant de dépolymériser les microtubules, la dissociation de certains groupes d'ARNt au nucléole a été observée, suggérant que leur localisation au nucléole dépend des microtubules (Haeusler et al., 2008). Bien que la position des groupes d'ARNt soit dépendante des microtubules, leur regroupement dépend de l'action des condensines. En effet, il a été montré qu'en supprimant certaines sous-unités de la condensines, le regroupement des ARNt est perdu (Haeusler et al., 2008). Lorsque les ARNt sont regroupés, l'expression des gènes qui sont à proximité est inhibée, ce phénomène est appelé « tRNA associated-silencing » (tgm, extinction de gènes associés au ARNt) (Wang et al., 2005). La suppression des condensines et par extension la perte du groupement des ARNt a permis de montrer que ce phénomène de tgm est dépendant de ce regroupement (Haeusler et al., 2008).

À la suite d'analyses de données de 3C, le regroupement et le positionnement des ARNt ont été discutés (Cournac et al., 2012; Duan et al., 2010). Dans une première analyse, les observations faites en microscopie ont été confirmées (Duan et al., 2010). Cependant les données n'étaient pas normalisées, ce qui signifie qu'un locus dont la couverture est supérieure à un autre, pour

différentes raisons, aura une tendance à interagir plus avec le reste du génome. Par extension, il donne l'impression d'interagir plus avec les ADN<sub>r</sub> qu'avec un autre locus du génome ayant une couverture inférieur. Dans une seconde analyse, où cette fois les données étaient normalisées, permettant ainsi d'éliminer ce biais de couverture, ces observations n'ont pas pu être confirmées (Cournac et al., 2012). Pour autant, cette seconde analyse ne remet pas en question les observations faites en microscopie, mais elle suggère qu'il faudra certainement des données de meilleure qualité pour que ce regroupement et ce positionnement des ARNt puissent être détectés par les méthodes de 3C.

## 2.4.2 L'enveloppe nucléaire inhibiteur de la transcription

Le regroupement des ARNt n'est pas le seul mécanisme permettant la répression de l'expression génique. Les gènes situés dans les régions subtélomériques des chromosomes, en périphérie du noyau, voient également leur expression inhibée. En positionnant artificiellement des gènes rapporteurs dont l'expression dépend de l'ARN pol II dans des régions télomériques, il a été montré que la périphérie du noyau jouait un rôle d'inhibiteur sur l'activité de transcription de l'ARN pol II. (Gottschling et al., 1990; Halme et al., 2004; Pryde and Louis, 1999; Renauld et al., 1993). Cette activité d'inhibition a pu être reliée à la présence de protéines du complexe SIR (Sir2, Sir3 et Sir4) à la membrane nucléaire (Andrulis et al., 1998). Plus précisément, *in vivo*, lorsqu'un télomère n'est pas réprimé, il s'associe à l'enveloppe nucléaire grâce à son interaction avec YKU, le rapprochant des autres télomères. Les zones de foyer de télomères étant enrichies en protéines du complexe SIR (Martino et al., 2009), ce rapprochement va permettre d'assurer la répression de ce nouveau télomère par la voie d'ancrage de la protéine Sir4. Il a été également démontré que la protéine Sir4 est essentielle à l'inhibition de la chromatine et elle permet de faire le lien entre le point d'ancrage et le domaine d'ADN inhibé (Roy et al., 2004). Cette double fonction apporte un mécanisme d'auto-organisation aux compartiments réprimés (Taddei et al., 2010).

Ce principe d'organisation fonctionnelle de la périphérie du noyau peut également être illustré par le mécanisme de répression des cassettes HML $\alpha$  et HMRA nécessaires à la modification de type sexuel de *S. cerevisiae*. Sous sa forme haploïde, *S. cerevisiae* peut avoir deux types sexuels :  $\alpha$  ou a. Trois différents locus du chromosome III contiennent l'information nécessaire pour déterminer ce type sexuel : HML $\alpha$ , HMRA et MAT. Les loci HML $\alpha$  et HMRA contiennent les copies des gènes  $\alpha$  et a, respectivement. Ils sont réprimés par les protéines du complexe SIR, alors que les gènes du locus MAT sont exprimés pour déterminer le type sexuel de la levure. Le type sexuel peut être modifié en remplaçant l'allèle du locus MAT par la copie de l'allèle  $\alpha$  ou a des loci HML $\alpha$  ou HMRA. Afin de maintenir le type sexuel de la levure, les cassettes HML $\alpha$  et HMRA sont donc inhibées et localisées proche de l'enveloppe nucléaire (Haber, 2012; Mekhail and Moazed, 2010).

En positionnant artificiellement le locus HMRA sur un plasmide, dans une souche n'exprimant pas Yku70 et Esc1 (responsables de l'ancrage de l'ADN à l'enveloppe nucléaire), il a été montré

que l'inhibition de son expression était indépendante de sa présence à la membrane nucléaire et pouvait être maintenue tant que des protéines du complexe SIR étaient présentes dans le noyau des cellules (Gartenberg et al., 2004). De plus, yKu peut interagir directement avec l'extrémité des télomères en l'absence des protéines Sir, ainsi les télomères peuvent être ancrés à l'enveloppe nucléaire sans que l'expression des gènes des régions subtélomériques ne soit inhibée (Maillet et al., 2001; Taddei et al., 2004). Par conséquent, les phénomènes d'ancrage et d'inhibition peuvent être dissociés l'un de l'autre (Gartenberg et al., 2004). Cependant la combinaison de ces deux phénomènes est essentielle à la compartimentation des domaines réprimés. Cette compartimentation permet non seulement de concentrer les protéines du complexe SIR, mais elle permet également de les éloigner d'autres sites où elles pourraient interférer avec d'autres types de régulations (Taddei et al., 2009). Cette séquestration des protéines du complexe SIR permet également d'éviter l'expansion de chromatine réprimée (Renauld et al., 1993). En conclusion, l'efficacité avec laquelle la chromatine est réprimée est issue de considération spatiale, ainsi que de la concentration locale en protéines Sir.

Alors qu'en règle générale la membrane nucléaire est un environnement défavorable à la transcription de gène par l'ARN pol II, dans certains cas elle semble tout de même être perméable à cette transcription. Par exemple, l'expression de certains gènes est favorisée lorsqu'ils sont relocalisés à la membrane nucléaire au niveau de pore nucléaires. Certains gènes inductible, comme GAL1-10, GAL2, INO1, HSP104 et HKX1, sont déplacés du nucléoplasme vers des pores nucléaires lorsque la cellule les signaux de régulation d'activation de leur transcription (Brickner et al., 2007; Brickner and Walter, 2004; Cabal et al., 2006; Dieppois et al., 2006; Schmid et al., 2006). Des expériences d'immuno-précipitations ont également permis de montrer que les protéines qui composent les pores nucléaires ont une tendance à interagir avec des régions du génome très transcrites (Casolari et al., 2004; Luthra et al., 2007).

En résumé, le génome de *S. cerevisiae* est très organisé dans le noyau, aussi bien dans le temps que dans l'espace, et cette organisation peut affecter certains processus biologiques comme la transcription, la réplication et la réparation de l'ADN (Taddei et al., 2010). Son organisation, qui suit le un modèle de pseudo-Rabl (Jin et al., 1998), peut se résumer en quatre points :

- 1) L'attachement et le regroupement des centromères au SPB ;
- 2) La position des télomères en petits foyers dynamiques accrochés à l'enveloppe nucléaire ;
- 3) Le regroupement préférentiel des bras des chromosomes en fonction de leur taille ;
- 4) L'isolement du nucléole des autres chromosomes du côté opposé au SPB.

Cette organisation persiste en grande partie pendant la durée du cycle cellulaire (Jin et al., 1998) et même au cours de la phase de quiescence (Guidi et al., 2015).





## Chapitre 3

# La synthèse de génomes *de novo*, ou génomique synthétique

Les chimistes ont commencé à étudier la structure de la matière grâce à différentes techniques d'analyses, leur permettant ainsi de décrire ce qu'ils pouvaient percevoir. C'est ensuite en cherchant à synthétiser ces molécules qu'ils ne faisaient qu'observer qu'ils ont pu apprendre bien plus de ces composés chimiques. De la même manière, le domaine de la biologie est aujourd'hui en pleine transition de l'ère du déchiffrement de la séquence d'ADN à l'ère de la synthèse de génome ; et cette transition nécessite un tout nouveau niveau de compréhension de la biologie, qu'il nous manque toujours. Cette transition peut être associée à une nouvelle discipline, la biologie synthétique, dont le premier meeting a été tenu en 2004 au MIT. Alors que la biologie synthétique a de nombreux aspects, l'un de ces sous-domaines appelé génomique synthétique s'emploie à construire une large variété de génomes. L'un des buts très importants de la génomique synthétique est de développer de nouvelles plateformes pour les biotechnologies ainsi que de trouver de nouvelles solutions aux différents problèmes sociétaux tels que la crise énergétique annoncée ou et la bioremédiation en utilisant des organismes synthétiques.

Dans ce chapitre, je vais vous détailler l'évolution de ce qui fait aujourd'hui la génomique synthétique. Pour cela nous allons commencer par la toute première synthèse d'un gène synthétique en 1970, en passant par la première synthèse d'un génome viral en 2002, puis celle d'une bactérie en 2010. Enfin nous terminerons par les projets de synthèses de génomes eucaryotes.

## 3.1 Les débuts de la synthèse chimique d'ADN, du simple gène au génome complet

### 3.1.1 La synthèse de première base d'ADN, de 20 à 200 pb

La synthèse chimique de molécule d'ADN double brin, et plus généralement de gène, trouve sa source dans les travaux de l'équipe du professeur Har Gobind Khorana<sup>1</sup>. En 1970, ils décrivent la synthèse du gène de l'ARN de transfert de l'alanine (ARNt<sub>ALA</sub>) (Agarwal et al., 1970; Khorana et al., 1972), c'est la première synthèse chimique de gène. La séquence de 77 paires de bases (pb) de ARNt<sub>ALA</sub> était la seule séquence d'acide ribonucléique connue, elle avait été élucidée, en 1965, par l'équipe du professeur Robert William Holley<sup>2</sup> (Holley et al., 1965).

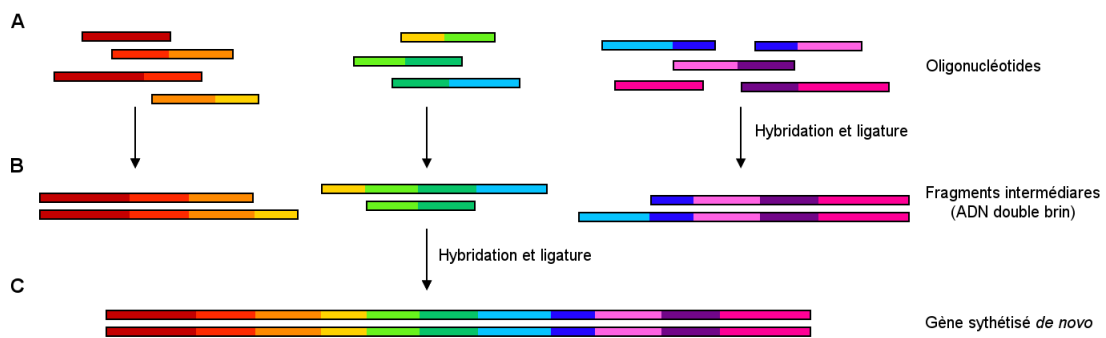
Pour synthétiser la séquence de l'ARNt<sub>ALA</sub>, l'équipe du Pr. Khorana a développé un processus d'assemblage d'ADN par liaison d'oligonucléotides, en trois étapes (Khorana et al., 1972). (1) La séquence de l'ARNt<sub>ALA</sub> est divisée en 15 petits fragments d'ADN (5 à 20 nucléotides), se chevauchant les uns par rapport aux autres. Ces fragments d'ADN sont ensuite synthétisés sous forme d'oligonucléotides. (2) Les oligonucléotides sont combinés par hybridation des séquences chevauchantes, puis ils sont liés les uns aux autres sous l'action de la T4 ligase. Trois fragments d'ADN double brins de tailles intermédiaires (20 pb, 33 pb et 31pb) sont alors obtenus. (3) Les extrémités cohésives de ces trois fragments leurs permettent ensuite d'être assemblés dans l'ordre souhaité à la suite d'une seconde réaction catalysée par la T4 ligase. Une fois assemblée, la séquence du fragment d'ADN synthétique doit être caractérisée<sup>3</sup>. La synthèse du gène de l'ARNt<sub>ALA</sub> a été la preuve de concept permettant démontrer qu'il était possible de synthétiser un fragment d'ADN double brin à partir d'une séquence prédéfinie. L'équipe du Pr. Khorana s'était arrêtée à la synthèse du gène de l'ARNt<sub>ALA</sub>, ils ne l'avaient pas exprimé, mais ils suggèrent qu'il sera intéressant d'étudier la réplication et l'expression d'un tel gène synthétique.

Une fois cette première synthèse décrite, la synthèse chimique d'ADN a commencé à intéresser d'autres équipes de recherche et par extension le domaine a commencé à se développer. En 1975, le Dr. Keiichi Itakura membre de l'équipe du Pr. Arthur Riggs, a utilisé la méthode de liaison d'oligonucléotides pour synthétiser l'opéron lactose (lacO) (Itakura et al., 1975). Une fois la synthèse terminée, il ont montré que ce LacO synthétique était fonctionnel *in vivo* (Heyneker et al., 1976). Enfin, en utilisant ce LacO synthétique et en utilisant toujours la même méthode

1. En 1968, Har Gobind Khorana, Robert W. Holley et Marshall Nirenberg obtiennent le Prix Nobel de physiologie ou médecine pour leur interprétation du code génétique et de sa fonction dans la synthèse des protéines (site du Prix Nobel).

2. En utilisant deux ribonucléases pancréatique capables de fragmenter les molécules d'ARN, à la position de nucléotides spécifiques, puis en déterminant l'ordre des fragments obtenus et leur composition par chromatographie, l'équipe du professeur Holley (Holley et al., 1965) a pu déterminer la séquence et la structure de l'ARNt<sub>ALA</sub>. Cette découverte a été l'une des clés essentielles à l'explication du système de synthèse de protéine à partir d'ARN messager

3. En 1970, le séquençage de l'ADN n'avait pas encore été mis au point, l'assemblage du gène de l'ARNt<sub>ALA</sub> avait donc été vérifié par analyse d'isotope et par chromatographie.



**FIGURE 3.1 L'assemblage d'ADN synthétique par liaison d'oligonucléotides :** (A) Les oligonucléotides synthétisés chimiquement sont hybridés puis ligaturés sous l'action de la T4 ligase. (B) Suite à cette première réaction, des fragments d'ADN double brins avec des extrémités cohésives sont obtenus. Ils peuvent être assemblés par ligature, (C) pour former le gène synthétisé *de novo*.

d'assemblage, l'équipe du Pr. Riggs a synthétisé et exprimé les gènes de la somatostatine<sup>4</sup> et de l'insuline chez *Escherichia coli* (Crea and Kraszewski, 1978; Itakura et al., 1977). Après les avoir exprimé chez *E. coli*, ils ont pu montrer que ces deux hormones étaient toujours fonctionnelles (Goeddel et al., 1979; Itakura et al., 1977).

Contrairement à l'exemple du gène de l'ARNt<sub>ALA</sub>, l'équipe du Pr. Riggs connaissait uniquement les séquences protéiques de la somatostatine et de l'insuline, ils n'avaient pas idée de leur séquence nucléique. Ils se sont alors servis du code génétique pour déterminer les séquences nucléotidiques à synthétiser. Afin de maximiser leur probabilité de succès dans la réécriture de ces séquences nucléotidiques, ils ont pris parti de suivre les trois principes suivants : (1) utiliser les codons identifiés pour être favorisés par *E. coli* (Ref) ; (2) les oligonucléotides à synthétiser ont été choisis de manière à ne pas générer d'assemblage indésirable ; (3) éviter la succession de séquences riches en AT et riches en GC qui pourraient entraîner l'arrêt de la transcription. Une fois les gènes synthétiques obtenus, ils ont été clonés dans un plasmide sous le contrôle d'un promoteur inducible, transformés dans *E. coli*, avant d'être isolés pour être vérifiés grâce au séquençage par dégradation chimique sélective (Maxam and Gilbert, 1977). Au travers de cette synthèse, l'équipe du Pr. Riggs a introduit la notion d'optimisation de codon pour la synthèse de gènes d'une espèce destinés à être exprimés par une autre espèce, dans ce cas là des gènes humains exprimés chez *E. coli*.

A cette époque, la synthèse des oligonucléotides pouvait prendre plusieurs mois. Il faut donc prendre beaucoup de recul pour mesurer les prouesses réalisées par les équipes du Pr. Khorana et du Pr. Riggs. Il aura fallu attendre le milieu des années 1980, pour que les premiers synthétiseurs, semi-automatisés, fassent leurs apparitions dans certains laboratoires de biologie moléculaire, permettant ainsi de rendre un peu plus accessible la synthèse d'oligonucléotides.

4. La somatostatine est une hormone humaine permettant d'inhiber la sécrétion d'hormone de croissance, d'insuline et de glucagon.

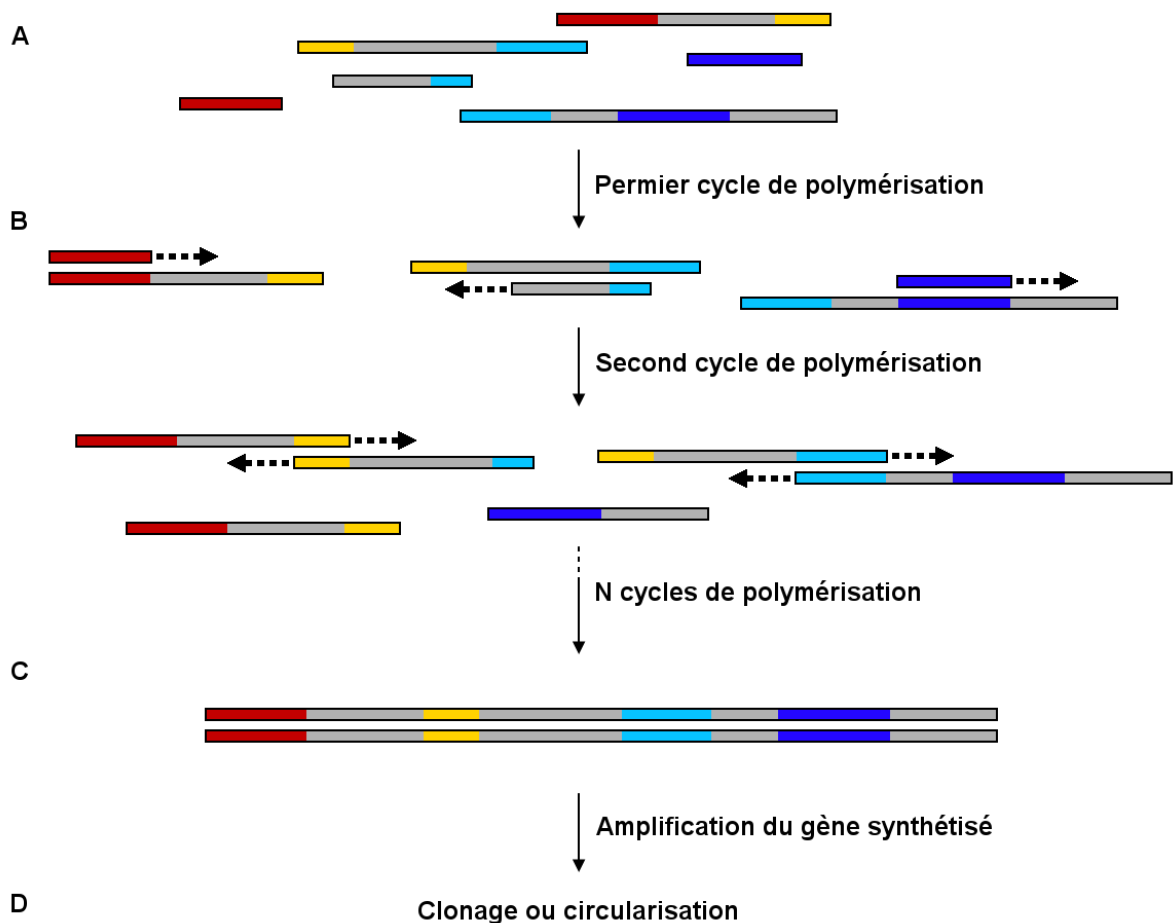
### 3.1.2 L'amélioration des méthodes de synthèse d'ADN pour passer le seuil des 1000 pb

A partir de la fin des années 1980, la démocratisation de la synthèse d'oligonucléotide a permis de développer nouvelles techniques de biologie moléculaire, en particulier le séquençage Sanger (Sanger et al., 1977) et l'amplification de fragments d'ADN génomique par réaction de polymérisation en chaîne (PCR) (Saiki et al., 1988, 1985). Cette démocratisation de la synthèse d'oligonucléotides a également permis d'augmenter la taille, la quantité et la qualité des molécules produites. Elle a également permis au domaine de la synthèse d'oligonucléotides de franchir un nouveau palier. En 1988, la synthèse d'un fragment d'ADN de plus de 1000 pb a été réussie pour la première fois : Wlodek Mandecki et T. J. Bolling décrivent en effet pour la première fois la synthèse chimique d'un plasmide de 2,1 kb (Mandecki and Boiling, 1988; Mandecki et al., 1990). Pour obtenir ce résultat, ils ont utilisé la méthode d'assemblage d'ADN par chevauchement des extrémités d'oligonucléotides précédemment développée (Khorana et al., 1972) et couplée à une méthode de clonage basée sur l'utilisation de l'enzyme de restriction FokI.

Dans le même temps, en 1990, Rosen et Dillon développèrent une nouvelle méthode d'assemblage d'ADN, toujours basée sur l'utilisation d'oligonucléotides, mais cette fois-ci en utilisant une ADN polymérase comme catalyseur de la réaction (et plus la T4 ligase). Pour cette méthode, le chevauchement des oligonucléotides n'est plus nécessairement intégral, et une vingtaine de nucléotides suffisent à jouer le rôle d'amorces pour l'ADN polymérase qui complète alors les nucléotides manquant (Dillon and Rosen, 1990). Grâce à cette méthode d'assemblage par cycle de polymérisation (PCA), Rosen et Dillon ont réussi à synthétiser le gène Rev d'HIV-2 (325 pb) en utilisant seulement quatre oligonucléotides (60 à 125 nucléotides) se chevauchant de 15 à 30 nucléotides (Dillon and Rosen, 1990).

En 1992, Prodromou et Pearl proposent une amélioration du protocole de Rosen et Dillon. La nouveauté de leur approche concerne les séquences chevauchantes des oligonucléotides. Lors du dessin de celles-ci, ils proposent de faire en sorte que leurs températures d'hybridation soient comprises entre 52-56°C. Cette uniformisation de la température d'hybridation leur permet ainsi de s'assurer de leur spécificité, ainsi que d'homogénéiser les conditions de la réaction de polymérisation. En utilisant une dizaine d'oligonucléotides (54 à 86 nucléotides), ils purent ainsi assembler le gène du lysozyme humain (522 pb) (Prodromou and Pearl, 1992).

Bien que les méthodes de synthèse d'oligonucléotides se soient considérablement améliorées, la limitation reste alors la synthèse de longs oligonucléotides. En effet, cette étape nécessite la purification des oligonucléotides sur gel d'acrylamide. La taille de ces oligonucléotides présente donc un frein au développement de la méthode de PCA. Pour pallier ce problème, Stemmer et al., proposent d'initier la réaction d'assemblage par PCR avec un mélange d'oligonucléotides de 40 nucléotides. Pour illustrer leur méthode, ils décrivent la synthèse d'un plasmide d'environ 3 kb



**FIGURE 3.2 L'assemblage d'oligonucléotides par polymérisation (PCA)** se fait en quatre étapes : **(A)** la synthèse d'oligonucléotides ; **(B)** ces oligonucléotides vont d'abord s'hybrider et s'allonger à partir de leur extrémités 3' grâce à l' $ADN_{pol}$ . Au cycle suivant, ces mêmes fragments d'ADN vont continuer à être étirés en utilisant l'extrémité 3' d'un autre fragment comme nouvelle amorce, et ainsi de suite jusqu'à ce que le fragment synthétique soit obtenu (la T4 ligase n'est plus requise) ; **(C)** À la fin de la réaction, le fragment d'ADN synthétisé peut-être amplifié par PCR en utilisant des amorces s'hybridant à chacune des extrémité de la séquence synthétique désirée (cette étape peut introduire des mutations) ; **(D)** Le fragment d'ADN est ensuite cloné ou directement circularisé s'il contient une origine de réplication permettant son maintien dans une cellule receveuse.

qu'ils ont assemblé en une réaction à partir de 134 oligonucléotides de 40 nucléotides (Stemmer et al., 1995).

Le développement techniques de synthèse d'oligonucléotides, de la méthode PCR et les améliorations apportées aux protocoles d'assemblage d'ADN synthétique, ont permis à ce domaine de faire un important bon en avant et les possibilités de synthèse d'ADN semblent bien plus importante qu'au début de son développement.

### 3.1.3 Des gènes aux génomes (viraux)

Passer de la synthèse de plasmides de quelques milliers de paires de bases aux petits génomes de phage ou de virus (< 10 kb), n'avait rien d'insurmontable du moment que la séquence de ces génomes viraux était caractérisée. En 1977, le génome du bactériophage  $\phi$ X174 est le premier à avoir été entièrement séquencé (Sanger et al., 1977). D'autres séquences suivirent, comme le génome du virus de la polio type 1 (Mahoney) (PV1(M)), en 1981 (Kitamura et al., 1981; Racaniello and Baltimore, 1981), ou encore celui du bactériophage T7, en 1983 (Dunn et al., 1983). En 2002, la séquence du PV1(M) servira de modèle, à Cello et ses collègues pour synthétiser un ADN complémentaire (ADNc) de ce virus (Cello, 2002). Cette étude est un jalon important dans le domaine de la génomique synthétique car ce fut la première fois qu'un génome fut entièrement reconstruit à partir d'oligonucléotides obtenus par synthèse chimique.

Le génome du virus de la polio est constitué d'ARN simple brin (+), pour le synthétiser, Cello et al. ont commencé par synthétiser un ADNc précédé du promoteur T7 pour qu'il puisse ensuite être transcrit en ARN. Le génome du PV1(M) est plus grand ( $\approx$  7500 pb) que les autres plasmides ou fragments d'ADN synthétisés jusqu'alors. Pour assembler cet ADNc, la séquence du PV1(M) a été divisée en trois fragments de tailles intermédiaires (3026 pb, 1895 pb et 2682 pb), qui sont eux même le résultat de l'assemblage de fragments plus petite taille (400 à 600 pb). Ces petits fragments ont été obtenus par la méthode d'assemblage par liaison (Khorana et al., 1972) ou par PCA (Dillon and Rosen, 1990; Stemmer et al., 1995), à partir d'oligonucléotides d'environ 69 pb. Une fois synthétisés, ces fragments furent clonés dans des plasmides pour être amplifiés par *E. coli*. A cette étape, et plus particulièrement lors de la réaction de polymérisation, des erreurs peuvent être introduites dans la séquence de ces fragments, qui doivent être séquencées pour contrôle. Les erreurs de synthèse furent alors corrigées par mutagenèse dirigée. Une fois les séquences vérifiées, tous ces fragments ont été assemblés pour former l'ADNc du PV1(M). En mélangeant leur ADNc synthétique à un milieu composé d'extrait cellulaire, les chercheurs obtinrent des particules virales fonctionnelles.

En 2003, Smith et al., assemblèrent le génome du bactériophage  $\phi$ X174 (5386 pb) à partir d'un mélange d'oligonucléotides synthétisés chimiquement (Smith et al., 2003). Tout comme Cello et ses collègues en 2002, Smith et ses collègues commencent par assembler leurs oligonucléotides

par hybridation et liaison. Mais au lieu de cloner les fragments obtenus suite à cette première réaction pour les assembler, ils les mélanges pour finir l'assemblage par polymérisation, en PCA. Une fois l'étape de PCA terminée ils purifièrent sur gel l'ADN correspondant à la taille du génome du bactériophage  $\phi$ X174. Ces molécules furent alors circularisées et transformées dans *E. coli* pour faire une sélection biologique. Contrairement à la méthode publiée un an plus tôt (Cello, 2002), il semblerait que la combinaison de l'assemblage d'oligonucléotides par liaison avec la PCA puissent accélérer considérablement le processus de synthèse de génome (Smith et al., 2003).

En 2005, Chan et al. ont partiellement réécrit les 39'937 pb du génome du bactériophage T7. Ils ont remplacé 11'515 pb de la partie 5' par 12'179 pb d'ADN modifié en utilisant à la fois des fragments d'ADN synthétiques et des fragments du génome du bactériophage T7 amplifiés par PCR. Ce génome partiellement synthétique leur a permis d'obtenir un bactériophage viable dont les caractéristiques clés de l'original ont été maintenues, tout en leur ayant permis de faciliter les manipulations génétiques grâce à l'élimination et l'ajout de nouveaux sites de restrictions. Grâce à ces modifications ils peuvent ainsi facilement modifier ou ajouter de nouveaux éléments génétiques (promoteurs, séquences codantes ...), pour étudier leur fonction (Chan et al., 2005).

Malgré tout, la synthèse de fragments d'ADN par PCA peut présenter certaines limites dues au taux de mutations résultant de la réaction de polymérisation. En général si l'on cherche à produire des fragments d'ADN de 500 pb il faudra analyser  $\approx 1,5$  clones pour trouver la séquence sans mutation, alors que pour des fragments quatre fois plus grand, de 2000 pb, il faudrait analyser 15 fois plus de clones (soit 30 clones) pour identifier un clone correcte (Kodumal and Santi, 2004). Kodumal et ses collègues, ont alors proposé une nouvelle approche en deux étapes pour faciliter la synthèse de génomes. Dans un premiers temps, par PCA, ils construisent des fragments d'ADN de 500 pb (appelés « synthons ») qu'ils clonent dans des plasmides pour vérifier leurs séquences. Ces synthons peuvent ensuite être assemblés par succession de sous-clonage jusqu'à obtenir le fragment d'ADN désiré, en ayant considérablement réduit le taux d'erreurs dans la séquence (Kodumal and Santi, 2004).

Les méthodes d'assemblage d'ADN se sont considérablement améliorées depuis la synthèse de l'ARNt<sub>ALA</sub> en 1970 (Agarwal et al., 1970). Le développement des nouvelles méthodes de synthèse d'oligonucléotides et plus particulièrement les technologies permettant d'en synthétiser plusieurs milliers sur des micro-plaques ont permis de réduire considérablement leur coût de production (Kosuri et al., 2010). De plus, le temps nécessaire à l'assemblage d'un d'un gène s'est considérablement accéléré, en passant de quelques mois à plusieurs jours. Ainsi, la technologie de synthèse d'ADN semble prête à surmonter la tâche de fabriquer des chromosomes artificiels et même de s'attaquer au projet de synthèse d'un génome minimal (Hutchison et al., 1999).

### 3.1.4 *Mycoplasma mycoïdes*, le premier micro-organisme au génome synthétique

Les méthodes d'assemblage *in vitro*, illustrées par les exemples précédents, permettent de synthétiser des fragments d'ADN de quelques milliers de paires de base, mais elles restent très laborieuses et limitées pour permettre de synthétiser un génome bactérien de plusieurs millions de paires de bases. Pour pallier cette limite technologique, au cours de ces dernières années, une voie d'assemblage de génome synthétique basée sur le mécanisme naturel de recombinaison homologue chez la levure *S. cerevisiae* a émergé (Gibson et al., 2008). Elle permet d'assembler de plus large fragment d'ADN, plus rapidement, à partir de fragments d'ADN de 1 à 15 kb, assemblés *in vitro*, en les introduisant dans le noyau de *S. cerevisiae*, par électrolocation. Ces fragments serviront alors de matrice à la machinerie de recombinaison homologue qui pourra les insérer dans le génome de *S. cerevisiae* ou les assembler en un nouveau chromosome (Gibson et al., 2008; Larionov et al., 1996; Mézard et al., 1992; Szostak et al., 1983).

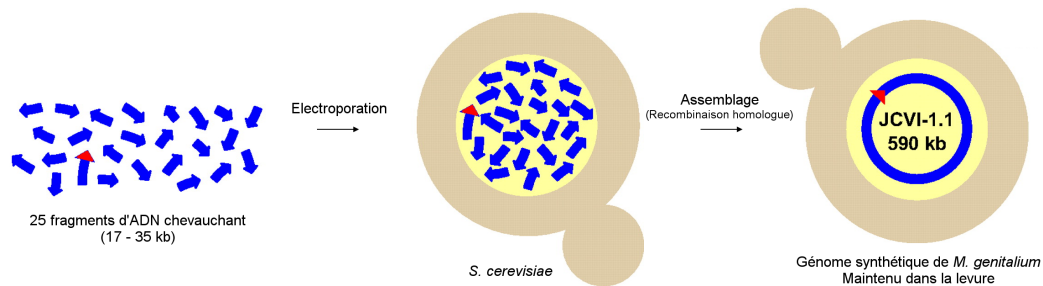
En 2008, grâce à cette méthode, une équipe de l'institut J. Craig Venter a reconstruit le génome de *Mycoplasma genitalium* (583 kb). C'est le premier génome bactérien entièrement assemblé à partir d'oligonucléotides synthétisés chimiquement (Gibson et al., 2008). Ce génome ayant été assemblé dans le noyau de *S. cerevisiae*, des étapes supplémentaires sont nécessaires pour le transférer dans les cellules de *Mycoplasma* afin qu'il puisse être transcrit et répliqué. Ainsi, Lartigue et al., ont développé une méthode de « transplantation génomique » (Lartigue et al., 2007), leur permettant de remplacer le génome natif d'une cellule receveuse, d'une certaine espèce de *Mycoplasma* (par exemple *Mycoplasma mycoïdes*), par celui d'une autre espèce de *Mycoplasma* (par exemple *Mycoplasma capricolum*)<sup>5</sup>. En 2010, cette même équipe assemble, dans la levure, le génome de *M. mycoïdes* (1,08 Mb), qui est presque deux fois plus grand que celui de *M. genitalium* (Gibson et al., 2010). Cette fois-ci, ils transplantèrent ce génome de *M. mycoïdes* synthétique dans des cellules de *M. capricolum*, produisant ainsi la toute première souche bactérienne avec un génome synthétique. Cette nouvelle souche est baptisée *JCVI-Syn1.0* (Gibson et al., 2010).

La stratégie d'assemblage par recombinaison homologue est une méthode qui peut s'avérer très chronophage. Cependant, suivant la finalité du projet, elle permet d'intégrer une infinité de nouvelles caractéristiques et modifications dans la séquence du génome de référence. Chez certains organismes la recombinaison homologue n'est pas très efficace, comme dans l'exemple précédent, le chromosome synthétique peut être assemblé dans le noyau de *S. cerevisiae*, puis être transplanté dans l'organisme cible. La transplantation de génome reste l'étape la plus limitante de la synthèse de génome. La méthode de transplantation qui fonctionne pour les cellules de

---

5. La transplantation de génome est un remplacement du génome d'une espèce bactérienne par celui d'une autre espèce bactérienne. Pour qu'un génome puisse être transplanté, il est essentiel que la cellule receveuse ait la machinerie nécessaire à la réplication du génome transplanté. Par exemple, les cellules de *M. capricolum* peuvent recevoir le génome de *M. mycoïdes*, mais l'inverse n'est pas possible, les cellules de *M. mycoïdes* ne peuvent pas recevoir le génome de *M. capricolum* (Lartigue et al., 2007).





**FIGURE 3.3 Assemblage du génome de *M. genitalium* dans la levure :** Le génome de *M. genitalium* est découpé en 25 fragments d'ADN de 17 à 35 kpb. Ils sont tous transformés dans le noyau de *S. cerevisiae* en même temps lors d'une réaction d'électroporation. Grâce à un fragment contenant la séquence d'un centromère de levure, tous ces fragments peuvent être assemblés par recombinaison homologue, pour former un chromosome qui sera répliqué et maintenu par la levure (Gibson, 2014).

*Mycoplasma*, s'est avérée totalement inefficace pour des cellules humaines alors que ces dernières, comme *Mycoplasma*, n'ont pas de paroi cellulaire (Lartigue et al., 2007). Pour chaque nouveau génome synthétisé il faudra probablement développer une nouvelle méthode d'insertion dans les cellules receveuses.

## 3.2 L'édition de génome, une alternative à la synthèse *de novo*

La diminution des coûts de la synthèse chimique d'ADN et le développement des différentes méthodes d'assemblages ont permis de rendre possible la synthèse de grands génome, mais ces méthodes, dites « bottom-up », sont loin d'être utilisées en routine dans les laboratoires de recherche ou l'industrie. En effet, elles restent très coûteuses et très longues à mettre en place. En parallèle de la synthèse de génome, des méthodes d'édition de génomes se sont développées. Ces approches, dites de « top-down », sont moins coûteuses et plus rapides à mettre en place. En général, elles s'appuient sur des méthodes de biologie moléculaire permettant de modifier ou de réduire le génome de l'organisme utilisé, de manière à ce qu'il acquière de nouvelles caractéristiques.

### 3.2.1 L'édition de gènes

Ces dernières années l'édition de gène c'est démocratiser au sein des laboratoires de recherches. Les nucléases telles que les TALENs (transcription activator-like effector nucleases) (Christian et al., 2010) et le système CRISPR-Cas9 (clustered regularly interspaced short palindromic repeat/CRISPR associated systems) (Jinek et al., 2012), très faciles à utiliser et permettant d'introduire des coupures doubles brins (CDB) à des loci bien précis du génome, ont grandement contribué à ce phénomène. En introduisant un CDB dans la séquence d'un gène, ce dernier peut être réparé soit par « Non-Homologous End Joining » (NHEJ) ou par recombinaison homologue. Très souvent, lors de la réparation par NHEJ une insertion ou une délétion de base se produit au niveau du site de la réparation, permettant de disrupter le gène. Dans le cas de la recombinaison

homologue, des ADN donneurs peuvent être utilisés pour remplacer la séquence du gène (pour revue voir (Gaj et al., 2013)).

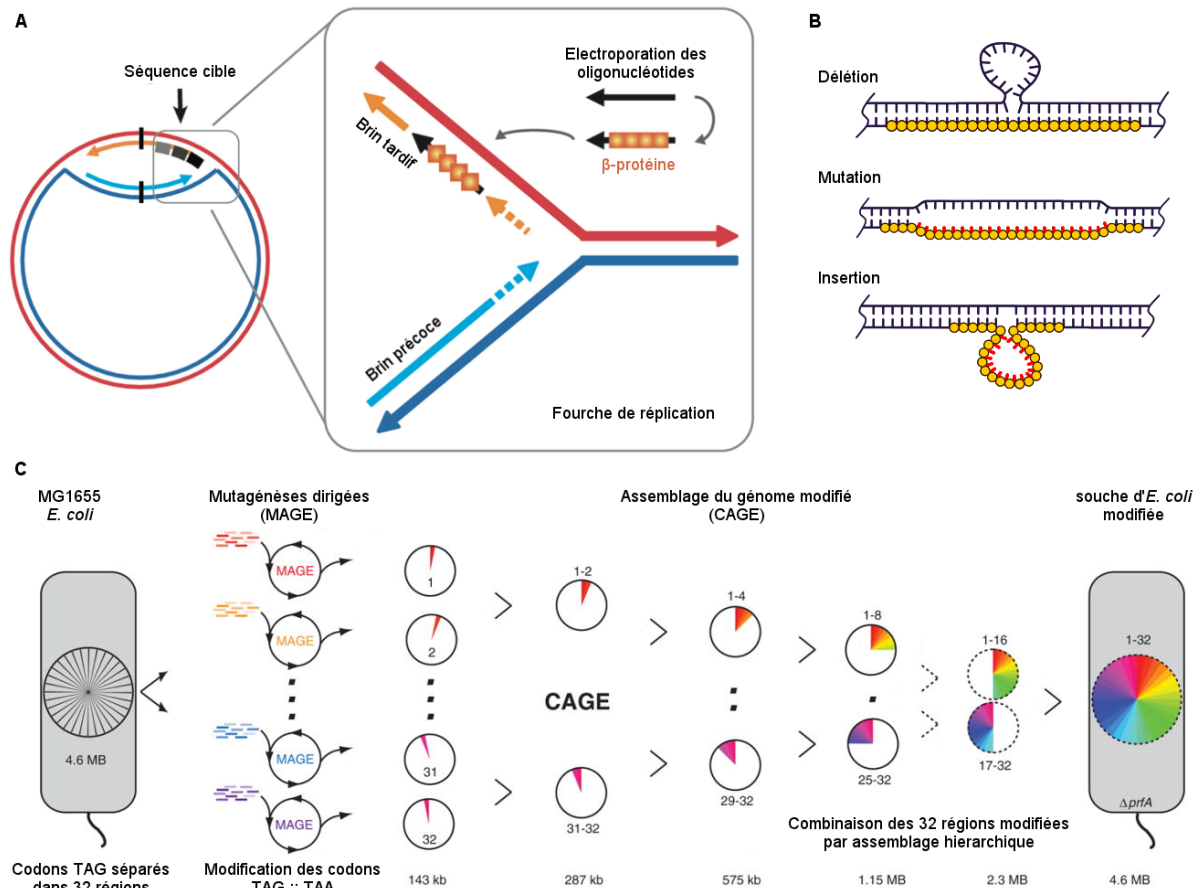
Bien que ces outils soient très facile à utiliser, ils ne permettent que l'édition d'un seul gène à la fois. Pour cette raison un projet de construction de génome synthétique ne dépendra pas uniquement de leur utilisation, mais ils pourront être utilisés pour effectuer des corrections dans la séquence synthétique, si nécessaire.

### 3.2.2 L'édition multiple de génome

En 2009, l'équipe de George Church a créé une méthode d'ingénierie génomique permettant d'introduire simultanément une multitude de modifications (substitutions, insertions ou délétions) dans le génome d'*E. coli* à partir d'oligonucléotides (figure 3.4). Cette méthode est baptisée MAGE, pour « Multiplex Automated Genome Engineering » (Wang et al., 2009). En utilisant plusieurs oligonucléotides à la fois, MAGE peut modifier plusieurs loci simultanément, ou elle peut être aussi utilisée pour introduire plusieurs mutations au niveau d'un seul locus, pour générer une diversité génomique. En utilisant cette méthode, l'équipe de George Church a introduit des mutations dans 24 gènes impliqués dans la voie métabolique DXP, afin d'obtenir une souche d'*E. coli* capable de produire 5 fois plus de lycopène que la souche initiale, en seulement trois jours (Wang et al., 2009). En 2011, ils combinent la méthode de MAGE à la conjugaison bactérienne pour remplacer les 314 codons STOP « TAG » d'*E. coli*, par le codon STOP « TAA » (Isaacs et al., 2011). Cette approche dérivée de MAGE est baptisée CAGE, pour « Conjugative Assembly Genome Engineering ». La méthode de MAGE est optimale pour introduire une trentaine de modifications dans la séquence d'un génome, tandis que CAGE est conçu combiner plusieurs portions de génome préalablement éditées. De manière général, ces deux méthodes sont plus avantageuses que la synthèse de génome *de novo*, si les mutations à intégrer sont dispersées au travers de la séquence du chromosome et que la similarité entre les chromosomes natif et synthétique est supérieure à 90% (Isaacs et al., 2011).

### 3.2.3 La réduction de génome

Avec le développement de la génomique synthétique, beaucoup d'équipes ont cherché à réduire les génomes au strict minimum, de manière à déterminer quel pourrait être le génome minimal d'une espèce. De telles réductions de génomes permettraient de simplifier les organismes facilitant leur étude et leur modification. Le concept de génome minimal semble simple à première vue, mais il s'avère être un problème bien plus complexe dans la pratique. En plus des gènes essentiels et non-essentiels, il existe de nombreux gènes « quasi-essentiels », c'est-à-dire qu'ils ne sont pas critiques pour la viabilité des cellules mais qu'ils sont néanmoins indispensables à leur croissance. Par conséquent, au cours du processus de minimisation du génome, un compromis doit être trouvé entre la taille du génome et le taux de croissance de l'organisme (Hutchison et al., 2016). Ces



**FIGURE 3.4 MAGE.** MAGE (Multiplex Automated Genome Engineering) est utilisé dans une souche d'*E. coli* dont la machinerie de correction est supprimée et qui exprime la bêta-protéine, une protéine de liaison à l'ADN simple brin de bactériophage. **(A)** Une fois les oligonucléotides introduits par électroporation, ils vont être stabilisés par la beta-protéine et à la séquence d'ADN dont ils sont complémentaires, au cours de la réplication. À la fin de la division cellulaire, dans l'une des cellules filles, au niveau du site de mutation il y a un mésappariement entre les deux brins d'ADN. L'un des deux brins contient la mutation désirée et l'autre non. Après la division cellulaire suivante, ce mésappariement sera corrigé, donnant naissance à deux cellules filles, l'une ayant une copie du chromosome natif et l'autre un chromosome muté. Après 6 à 8 cycles d'électroporation, la mutation est introduite dans 99% des génomes de la population de cellules. **(B)** En fonction du design de l'oligonucléotide, MAGE peut permettre d'introduire des substitutions, des délétions ou des insertions dans le génome d'*E. coli*. **(C)** La méthode de CAGE permet d'étendre le MAGE à l'ensemble du génome, en combinant séquentiellement les différentes régions préalablement modifiées par la méthode de MAGE.

génomés minimaux sont en général très discutés, un organisme sauvage a en général la capacité de s'adapter à différents environnements ; en réduisant son génome, cet organisme devient sensible aux modifications de son environnement (Danchin and Fang, 2016). Autrement dit un génome sera minimal dans des conditions de culture bien précises.

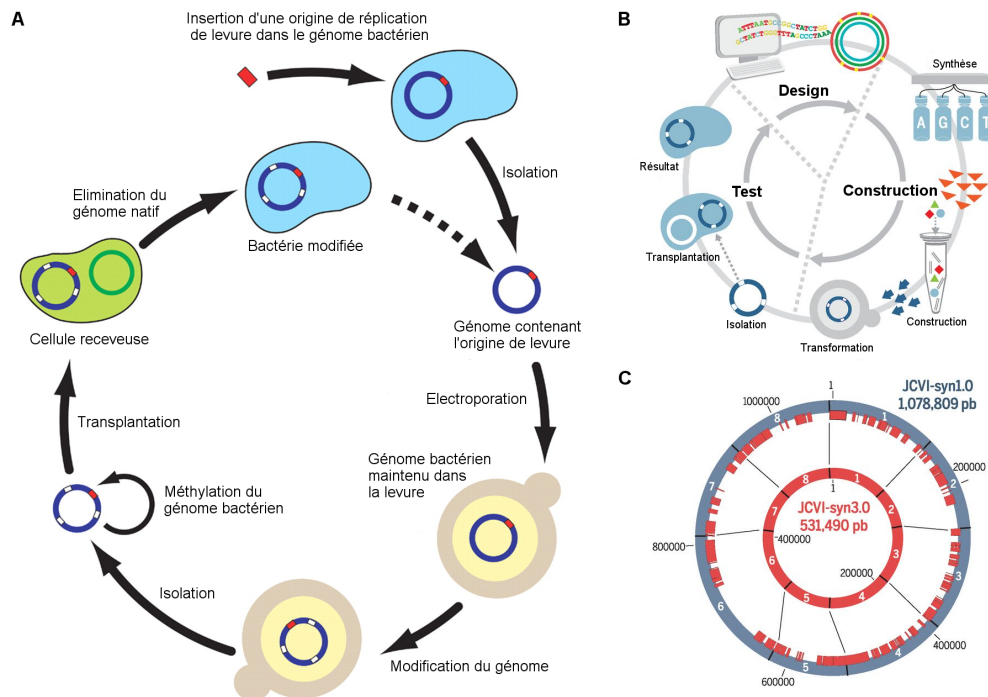
Plusieurs équipes ont réduit les génomes de nombreux organismes, comme par exemple *E. coli* (Pósfai et al., 2006; Yu et al., 2002), *Bacillus subtilis* (Ara et al., 2007; Tanaka et al., 2013) ou *Schizosaccharomyces pombe* (Giga-Hama et al., 2007). En utilisant le système de transposon T5n (Jorgensen et al., 1979), une équipe Coréenne (Yu et al., 2002) a intégré deux sites LoxP dans le génome d'*E. coli*. Ensuite, en exprimant la recombinase Cre, ils ont supprimé environ 300 kb, situé entre ces deux sites loxP, sans que la croissance de la souche ne soit affectée dans des conditions standard de culture. En 2006, une équipe Américaine (Pósfai et al., 2006) a utilisé le système de recombinaison du phage  $\lambda$ -red couplé à des cassette de résistance à un antibiotique (Baba et al., 2006; Datsenko and Wanner, 2000), toujours pour réduire le génome d'*E. coli*. Après avoir supprimé environ 700 gènes du génome d'*E. coli*, ils ont pu observer que la croissance de la souche réduite était similaire à la celle de la souche parentale. Ils ont également constaté que l'efficacité de transformation de la souche réduite avait considérablement augmenté par rapport à la souche parentale. Bien que les souches réduites soient en général plus sensible aux facteurs environnementaux que les souche parentales dont elles sont issues, il est possible que la réduction de leurs génomes puisse leur apporter des avantages insoupçonnés (Pósfai et al., 2006).

Tous ces résultats de réductions de génomes représentent des sources d'informations considérables pour la conception future de génomes synthétiques, ils peuvent permettre de déterminer quelles séquences d'ADN garder ou éliminer en fonctions du phénotype recherché.

### 3.2.4 *S. cerevisiae*, une plateforme d'assemblage *de novo* et d'édition de génome

De nombreuses méthodes d'édition de génomes ont été développées chez les organismes modèles comme *S. cerevisiae*, *E. coli*, *B. subtilis* ou encore les lignées cellulaires humaines, mais elles sont parfois très limitées pour éditer le génome d'espèce comme *Mycoplasma* ou des génomes viraux. Pour résoudre ce problème, l'équipe de Craig Venter, a mis au point une méthode permettant d'utiliser *S. cerevisiae* comme une plateforme d'édition de génome de *Mycoplasma* (Lartigue et al., 2009). Après avoir introduit un centromère de levure dans le génome de *Mycoplasma*, ce dernier peut être transféré et maintenu dans le noyau de *S. cerevisiae*, où il pourra être édité (Figure 3.5).

En 2016, l'équipe de Craig Venter a minimisé le génome synthétique de *Mycoplasma*, JCVI-Syn1.0 (1079 kb) (Gibson et al., 2010), pour obtenir un génome a été considérablement réduit, JCVI-Syn3.0 (531 kb) (Hutchison et al., 2016). Après quatre cycles de design, d'édition et de test (Figure 3.5), au cours desquels ils ont utilisé l'efficacité de la recombinaison et des différentes



**FIGURE 3.5 Cycle d'édition de génome.** (A) Cycle d'édition d'un génome bactérien dans la levure (Lartigue et al., 2009). (B) Une fois le design d'un chromosome synthétique déterminé *in silico*, les séquences d'ADN nécessaire à sa construction peuvent être synthétisées et assemblées. Enfin une fois assemblé ce chromosome peut être testé pour évaluer son design. Si des problèmes sont identifiés, la région problématique doit être redesignée, resynthétisée et rassemblée. (C) Cartes des génomes JCVI-syn1.0 (bleu) et JCVI-syn3.0 (rouge) (Hutchison et al., 2016).

méthodes d'édition de génome chez *S. cerevisiae*, ils ont réussi à supprimer plusieurs centaines de gènes du génome initial. *JCVI-syn3.0* est une approximation de génome cellulaire minimal, c'est un compromis entre la taille d'un petit génome et un taux de croissance suffisant pour l'expérimentation. Il conserve presque tous les gènes impliqués dans la synthèse et le traitement des macromolécules. De manière inattendue, il contient également 149 gènes avec des fonctions biologiques inconnues, suggérant la présence de fonctions non découvertes qui sont essentielles à la vie (Figure 3.5).

Récemment, une autre équipe de l'institut Craig Venter a assemblé le génome du virus herpès simplex de type 1 (HSV-1) dans la levure, à partir de fragments directement amplifié du génome parental (Oldfield et al., 2017). Dans un premier temps ils ont divisé le génome de HSV-1 en 11 fragments qu'ils ont assemblés séparément par recombinaison dans des vecteurs pouvant faire la navette entre *S. cerevisiae* et *E. coli*. Ainsi une fois assemblés, des mutations peuvent être introduites dans ces différents fragments grâce au système CRISPR-Cas9 dans la levure. Ils peuvent ensuite être transférés dans *E. coli* où ils pourront être amplifiés et isolés en grande quantité. Finalement la version complète du génome de HSV-1, natif ou modifié, pourra être assemblé en combinant les différents fragments dans la levure. Cette méthode permet à Oldfield et ses collaborateurs de modifier très rapidement le génome du virus HSV-1 et d'obtenir une

collection de différents variants (Oldfield et al., 2017). En infectant des cellules en culture<sup>6</sup> avec ces variants ils peuvent identifier le rôle des différents gènes qui composent le génome de HSV-1.

Dans le cas des génomes synthétiques assemblés dans la levure, pendant que le temps qu'ils y restent, ils peuvent être facilement modifiés grâce à l'efficacité de la recombinaison et aux différentes méthodes d'édition de génome chez *S. cerevisiae*. Le noyau de la levure représente une sorte « d'atelier » pour les génomes qui y sont assemblés (Mitchell and Ellis, 2017).

### 3.3 La conception et la construction d'un génome synthétique

Modifier de façon contrôlée un chromosome, donc un génotype dans le but d'obtenir un phénotype désiré, est un challenge qui peut s'avérer compliqué en fonction du niveau d'altérations requises sur le génome initial à disposition. Ce processus de construction de génome *de novo* peut se décomposer en 3 phases : (1) une première phase de conceptualisation, de modélisation et de « design » du nouveau génome, qui est faite *in silico* ; (2) une phase de synthèse et d'assemblage du génome ; et (3) une phase de contrôle qui va permettre de déterminer si le génome construit permet d'obtenir le phénotype recherché ou si d'éventuelles modifications sont à apporter au design initial.

#### 3.3.1 Design et modélisation

Comme pour tous projets de construction, la première étape précédant la construction d'un génome synthétique est le design de la séquence à synthétiser. Dans un premier temps la taille des génomes étudiés a nécessité le développement de nouveaux outils informatiques capables d'assister l'édition de séquence à grande échelle. Parmi ces outils, on peut citer J5 (Hillson et al., 2012), Clotho (Xia et al., 2011), Genome compiler (<http://www.genomecompiler.com/>), ou BioEdit, développé dans le cadre du projet de synthèse du génome de *S. cerevisiae* (projet Sc2.0, détaillé dans le chapitre 4 de ce manuscrit). Afin de revenir sur des modifications délétères voire létales pour l'organisme synthétique, il est important que ces logiciels puissent garder la trace des changements effectués dans le génome initial. Dans un second temps, le choix des modifications à effectuer, comme la délétion ou l'ajout de certains gènes, peut être assisté par une série de modélisations de voies métaboliques (Karr et al., 2012; Lewis et al., 2012).

#### 3.3.2 Phase de construction

En fonction de l'organisme choisi la stratégie d'assemblage des chromosomes varie. Les synthèses des génomes de *Mycoplasma* (Gibson et al., 2008, 2010; Hutchison et al., 2016), dont le génome peut être transplanté dans la cellule receveuse une fois l'assemblage terminé (Lartigue

---

6. L'infection a été réalisée sur des cellules mammifères Vero, une lignée cellulaire isolée à partir de cellules épithéliales de rein extraites d'un singe vert africain.

et al., 2009) a été assemblé dans la levure en utilisant son système de recombinaison. Dans le cadre du projet Sc2.0 (voir chapitre 4 de ce manuscrit), l'assemblage des chromosomes synthétiques est également effectué, *in vivo*, directement dans la levure (Annaluru et al., 2014; Dymond et al., 2011; Richardson et al., 2017). D'autres exemples de génomes synthétiques ont été assemblés par réactions enzymatique *in vitro*, comme ça a pu être le cas du génome du virus PV1(M) (Cello et al., 2002) ou celui du phage  $\phi$ X174 (Smith et al., 2003).

### 3.3.3 Test du génome synthétique

Afin de déterminer si le chromosome synthétique de la nouvelle souche possède les caractéristiques attendues, une étape de séquençage est nécessaire. Alors, on peut donc caractériser les fonctions génome, pour identifier les changements phénotypiques introduits lors de la synthèse. Par exemple, on peut étudier la viabilité de l'organisme synthétique, dans différentes conditions de milieux ou de température. Des analyses de transcriptome ou de production de métabolites peuvent également être effectuées. Dans le cas où les résultats des différents tests effectués ne sont pas satisfaisants, le design initial et le génome peuvent être modifiés jusqu'à ce qu'ils satisfassent les objectifs fixés. Par exemple, lors de synthèse du génome de *M. mycoïdes*, la délétion d'une seule base dans la séquence du gène de la protéine dnaA a empêché l'obtention de cellule viable. C'est seulement après avoir remplacé différents fragments d'ADN synthétiques par des fragments d'ADN natifs que cette erreur a pu être identifiée (Gibson et al., 2010).

## 3.4 Les projets de synthèse de génomes eucaryotes, « GP-write »

Le projet Sc2.0 (voir chapitre 4 de ce manuscrit) est le premier projet dont l'objectif est d'aboutir à la synthèse complète de génome eucaryote. Avec à peu près 35% du génome déjà synthétisé (Richardson et al., 2017) et le temps nécessaire pour synthétiser un chromosome qui s'est considérablement réduit, ce projet touche à sa fin. Ce projet semble précurseur d'une vague de synthèse de génomes eucaryote : ainsi, plusieurs chercheurs du monde entier, dont les Pr. Jef Boeke et George Church, ont créé un consortium international, « GP-Write » (Genome Project-write), qui vise à synthétiser différents génomes eucaryotes allant des cellules animales aux cellules végétales. Le but de ce projet restant la synthèse de régions génomiques de cellules de lignées cellulaires humaines (Boeke et al., 2016). Les intérêts scientifiques d'une telle approche sont nombreux, par exemple, en éliminant certaines séquences dont la fonction n'est pas connue dans une portion de génome, on pourrait éventuellement découvrir leur rôle dans le fonctionnement de la cellule. Comme le projet de séquençage du génome humain ou la recherche sur les cellules souches, auparavant, ce projet de synthèse du génome humain soulève évidemment de nombreuses questions éthiques et rencontre beaucoup d'opposants. Pour être développé dans les meilleures conditions, le consortium « GP-write » tente également de définir le cadre éthique et normatif dans lequel le projet devra se dérouler (Boeke et al., 2016).

La réussite du projet « GP-Write » dépend du développement de nouvelles technologies qui vont permettre d'assembler, d'éditer, d'introduire et de maintenir de larges constructions d'ADN dans les cellules humaines. Les outils d'édition de génome comme les TALENs ou le système CRISPR/Cas9 ne suffiront pas pour permettre d'assembler un génome. Comme pour le génome de *Mycoplasma*, la levure servira de plateforme pour l'assemblage de fragments de génomes synthétiques humains. Alina Chan, membre du laboratoire du Dr. Pamela Silver à l'université d'Harvard, a développé en collaboration avec Daniel Gibson, de l'institut Craig Venter, des chromosomes artificiels humains (HAC) pouvant être répliqué par des cellules humaine, mais également *S. cerevisiae* et *E. coli* (communication lors du congrès SB7.0, Singapour, 2017). Ces HACs vont ainsi pouvoir être assemblés et édités dans les cellules de *S. cerevisiae* avant d'être introduits dans les cellules humaines où ils pourront être maintenus. Pour transférer les HACs du noyau de *S. cerevisiae* aux noyaux des cellules humaine, les sphéroplastes de levure peuvent être directement fusionnés aux cellules humaine (Brown et al., 2017). Lors de la transplantation du génome de *Mycoplasma*, assemblé dans des cellules de levure (Lartigue et al., 2009), une étape de méthylation avait été nécessaire pour qu'il soit accepté par la cellule receveuse. Dans le cas de HACs, il sera peut-être essentiel que la levure servant à les assembler soit capable d'exprimer les protéines d'histones, pour que la chromatine humaine soit déjà formée avant que le chromosome soit transféré.

### 3.5 Les limites de la synthèse de génomes

Le facteur limitant de l'assemblage d'ADN *de novo* est l'introduction du chromosome ou du génome synthétique dans la cellule receveuse, mais mise à part cette étape qui peut être limitante, aujourd'hui, la technologie ne semble plus être un obstacle à la synthèse de génome. Cependant, dans certains cas la synthèse de génome de certains organismes peut soulever un certain nombre de problèmes. Par exemple, le fait que Oldfield et ses collaborateurs (Oldfield et al., 2017) aient choisi de modifier un pathogène humain soulèvent de nombreuses questions concernant la potentielle double utilité des approches de génomique synthétique et de la balance bénéfices / risques du développement d'une telle méthode de modification de virus infectieux. De manière importante, le fait qu'il n'y ait pas de synthèse d'ADN dans leur procédure permet de contourner l'effort de biosécurité du consortium international de synthèse de gènes ([www.genesynthesisconsortium.org](http://www.genesynthesisconsortium.org)), qui surveille les commandes de synthèse de gènes de manière à limiter la production de séquences d'ADN pouvant être associées à de dangereux pathogènes. Bien que ce mécanisme de surveillance doive être capable d'éviter la synthèse *de novo* de génomes viraux, ce système n'est pas encore infaillible. Récemment le génome de la variole du cheval ( $\approx 212$  kb) a été synthétisé à partir d'oligonucléotides, commandés de manières classique (Koblentz, 2017). Alors que le virus de la variole du cheval est connu pour être dangereux pour l'homme ou d'autres espèces agricoles, son homologue, le virus de la variole, qui a été éradiqué en 1980, peut entraîner de sévères maladies ; une telle stratégie de synthèse *de novo* pourrait être employée pour synthétiser le virus de la variole,



pour un coût d'environ 100'000 USD, même si cela nécessite une certaine expertise scientifique. Ces travaux montrent qu'il existe plusieurs méthodes permettant de synthétiser des virus, éteints ou pas...



# Chapitre 4

## Projet de synthèse du génome de *S. cerevisiae*, projet Sc2.0

En 2005, les Professeurs Srinivasan Chandrasegaran et Jef Boeke, de l'Université Johns Hopkins, ont commencé à réfléchir à l'idée de concevoir et de synthétiser un chromosome eucaryote, plus précisément de réécrire entièrement le génome de *S. cerevisiae* en une version synthétique. Avant de se lancer dans cet ambitieux projet, ils ont commencé par synthétiser le bras droit du chromosome IX (*synIX-R*), qui est le plus petit bras des chromosomes de *S. cerevisiae*. La synthèse de *synIX-R* a été terminée en 2011 (Dymond et al., 2011), ce premier succès a lancé définitivement le projet de synthèse du génome de *S. cerevisiae* (projet Sc2.0).

Le projet Sc2.0 est d'autant plus complexe qu'il est le premier projet de synthèse d'un génome eucaryote. Très peu d'informations sont disponibles sur les modifications pouvant être effectuées dans la séquence des chromosomes synthétiques. En effet, parmi les génomes qui ont été synthétisés (Cello, 2002; Gibson et al., 2008, 2010) aucune modification majeure n'a été faite par rapport à leur séquence de référence. Cependant, l'étape de conceptualisation du génome synthétique, qui est une étape cruciale pour garantir que la nouvelle séquence puisse être fonctionnelle *in vivo*, peut s'appuyer sur les informations de génomiques disponibles concernant *S. cerevisiae*. La conception des chromosomes synthétiques a donc été réalisée par rapport à la séquence de référence de la souche S288C (Saccharomyces Genome Database). Les modifications apportées à la séquence native ont été introduites sous la forme d'une série de suppressions, d'insertions et de substitutions de bases, en suivant trois règles. (1) Quelques soient les modifications réalisées dans le génome de *S. cerevisiae*, le phénotype de la souche synthétique devra rester proche du phénotype sauvage. La composition et l'ordre des gènes seront conservés dans la séquence des chromosomes synthétiques. (2) Les séquences synthétiques devront être stables, elles ne devront pas recombiner spontanément. Pour cela, les séquences répétées (entre autres les éléments transposable et les répétitions subtelomériques) seront éliminées des chromosomes synthétiques et tous les gènes codant pour des ARNt seront déplacés sur un nouveau chromosome spécialement conçu à cet effet. (3) Enfin, un système d'évolution inductible dépendant du système Cré/LoxP (SCRaMbLE,

« Synthetic Chromosome Rearrangement and Modification by LoxPsym-mediated Evolution ») sera intégré à la séquence des chromosomes synthétiques. Ce système de SCRaMbLE permet de générer rapidement de nombreux réarrangements chromosomiques, il pourra ainsi être utilisé pour explorer une large gamme de structures chromosomiques, issues de diverses conditions de croissance.

Dans ce chapitre, je vais vous présenter le design des chromosomes synthétiques du projet Sc2.0 et la manière dont ils sont synthétisés. Chacun des chromosomes étant synthétisé séparément, je vais également vous présenter la méthode de croisement et d'endoréduplication induite qui est utilisée pour les combiner au sein d'un même noyau. Enfin, nous verrons l'intérêt d'avoir introduit le système de SCRaMbLE dans la séquence des chromosomes synthétiques.

## 4.1 Le design des chromosomes synthétiques

### 4.1.1 Les éléments supprimés

L'un des principaux objectifs du design des chromosomes synthétiques est d'éliminer toutes sources d'instabilité du génome, de manière à ce que les réarrangements qui puissent intervenir dans les chromosomes synthétiques soient le résultat de l'activation du système SCRaMbLE. Les éléments transposables qui ne sont pas essentiels à la croissance de *S. cerevisiae* et qui sont connus pour contribuer à l'instabilité de son génome ont donc été éliminés de la séquence des chromosomes synthétiques (Lemoine et al., 2005; Xu and Boeke, 1987). Il est possible que lors de la diminution des nombres de copies de transposons, les transposons Ty1 restants soient activés pour se mobiliser, le gène SPT3, requis pour la transcription du transposon Ty1 (Boeke et al., 1986), peut être supprimé préalablement pour éviter ce genre de problème.

Dans le design des chromosomes synthétiques, les télomères ont été remplacés par des télomères artificiels composés de répétitions de la séquence  $(TG_{1-3})_n$ . Pour ce qui est des séquences répétées des subtélomères, une seule copie des gènes répétés a été gardée, et les éléments Y', dont la fonction est inconnue, ont été supprimés des chromosomes synthétiques. En revanche, les cores X, qui jouent un rôle essentiel dans le silencing des télomères et la ségrégation des chromosomes (Louis, 1995) et dont la séquence est variable, ont été remplacés par un core X consensus.

Les cassettes HML et HMR, du chromosome III, n'étant pas essentielles à la croissance des cellules dans les conditions de laboratoire, ont été retirées du design de la séquence du chromosome III synthétique (*synIII*). Étant donné que les souches de levures utilisées ne codent plus le gène HO qui est à l'origine de la conversion du type sexuel des souches de levures sauvages, le type sexuel de *synIII* pourra être décidé modifié par le manipulateur, rendant les cassettes HML et HMR inutiles dans son design.

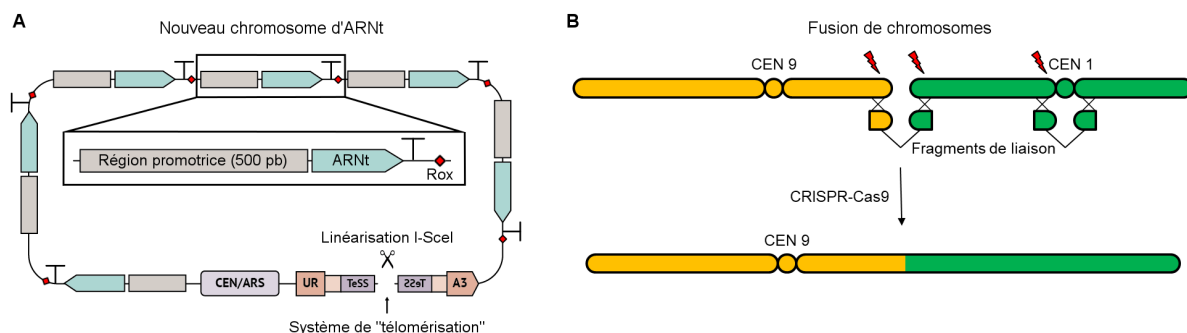
Le génome de *S. cerevisiae* contient environ 285 introns, pour la majorité d'entre eux leur délétion n'affecte pas la croissance de la levure (Parenteau et al., 2008). Afin de se débarrasser de la machinerie d'épissage, le design des chromosomes synthétiques prévoit de supprimer la majorité des introns. En effet, il pourra s'avérer compliqué d'éliminer tous les introns car la délétion de certains d'entre eux et notamment la délétion des introns des gènes codant pour des protéines ribosomiques peut affecter la croissance de la cellule (Parenteau et al., 2011). Dans ce cas-là, il est prévu que leur séquence soit conservée dans la séquence du chromosome synthétique. La suppression des introns codant pour de petits ARN nucléolaires (ARNsno), qui participent à la maturation des ARNr, peut aussi avoir une influence sur la croissance des cellules. Il est prévu que ces introns soient tout de même supprimés, mais que les séquences de leurs ARNsno soient exprimées de manière ectopique dans les chromosomes synthétiques.

### 4.1.2 Les éléments introduits

Chez *S. cerevisiae*, comme chez d'autres organismes, certains gènes ne sont pas essentiels à la survie de la cellule. En général, s'ils sont supprimés un par un, la survie de la cellule ne sera pas affectée, ces gènes sont qualifiés de gènes non-essentiels. Dans le but de créer un système d'évolution rapide, des sites loxP symétriques (Hoess et al., 1986) ont été insérés 3 pb après le codon STOP de tous ces gènes non-essentiels. Contrairement aux sites loxP canoniques, les sites LoxP-sym n'ont pas d'orientation spécifique, ils peuvent donc s'aligner selon deux orientations. En conséquence, lorsque la recombinase Cre sera exprimée dans le noyau des levures synthétique des inversions et des délétions pourront se produire à probabilité égale (Dymond et al., 2011). Ce système d'évolution, basé sur l'utilisation des sites loxP-sym et de la recombinase Cre, a été baptisé système SCRaMbLE (« Synthetic Chromosome Rearrangement and Modification by LoxP-sym-mediated Evolution »).

### 4.1.3 Les éléments déplacés

Avec 275 gènes codant pour 42 ARNt différents, chaque chromosome natif de *S. cerevisiae* contient un ou plusieurs gènes d'ARNt (Percudani et al., 1997). Ces gènes sont essentiels à la croissance de *S. cerevisiae* puisqu'ils sont directement impliqués dans la synthèse des protéines, mais ils sont également connus pour être une importante source d'instabilité du génome, ils sont des cibles privilégiées de la transposition Ty1 (Admire et al., 2006; Bridier-Nahmias and Lesage, 2012; Ji et al., 1993). Pour minimiser la probabilité de transposition dans les chromosomes synthétiques, il a été prévu que tous leurs ADNt soient replacer sur un nouveau chromosome contenant uniquement les séquences de ces derniers (figure 4.1). Les chromosomes synthétiques devraient ainsi être beaucoup plus stables, alors qu'au contraire, on peut imaginer que ce nouveau chromosome sera quant à lui très instable. Pour pouvoir induire des modifications dans ce chromosome et explorer les effets d'un changement du nombre d'ARNt, un dérivé du système de SCRaMbLE, basé sur le système Dre/Rox, a été inclus dans le design de ce nouveau chromosome. Les séquences d'ADNt



**FIGURE 4.1 Les éléments déplacés du génome Sc2.0 :** (A) Design du nouveau chromosome d'ADNT. Les séquences d'ADNT sont placées en série, entre une région promotrice et une séquence de terminateur. Un site de recombinaison Rox, du système Dre/Rox est placé après chaque gène d'ARNt. (B) Pour fusionner deux chromosomes, le système CRISPR-Cas9 peut être utilisé pour induire une CDB aux niveaux des régions subtélomériques à fusionner et du centromère à éliminer. Les coupures sont réparées par recombinaison homologue en utilisant des fragments d'ADN faisant la liaison comme modèle.

pourront ainsi recombiner entre elles au sein de ce nouveau chromosome sans être déplacées dans des régions des chromosomes synthétiques, lors de l'induction de la protéine Dre. De même que l'activation du système SCRaMbLE original, n'entraînera pas de recombinaison au sein de ce chromosome (figure 4.1). Enfin, ce nouveau chromosome est actuellement construit sous forme de chromosome circulaire mais la séquence d'un système de télomérisation<sup>1</sup> (Mitchell and Boeke, 2014) a été incluse dans le design du chromosome pour pouvoir le linéariser (figure 4.1A).

Il est prévu que le chromosome I synthétique (*synI*) mesure environ 180 kb, une fois toutes les modifications appliquées, faisant de lui un chromosome extrêmement petit. Avec la présence du nouveau chromosome contenant tous les ADNT, la souche synthétique finale aurait 17 chromosomes. Pour éviter cela et maintenir la souche à 16 chromosome, il est prévu que *synI* soit fusionné à un autre chromosome. En introduisant des coupures doubles brins (CDB) aux niveaux des deux régions télomérique à fusionner (par exemple *synIX-R* et *synI-L*), et en utilisant un fragment d'ADN contenant un marqueurs de sélection et des séquences d'homologies aux deux régions coupées, ces dernières pourront être réparées en fusionnant ces deux chromosomes. Dans le même temps, l'un des deux centromères doit être éliminé (par exemple CEN-*synI*), ce qui peut être fait en utilisant la même méthode, en introduisant cette fois-ci des coupures doubles brins autour du centromère à éliminer (figure 4.1B).

1. le système de télomérisation, ou « *telomerator system* », est un élément synthétique, qui a été développé chez *S. cerevisiae* pour linéariser un chromosome initialement circulaire, tout en permettant de le conserver au cours des cycles mitotiques suivant grâce à des télomères fonctionnels. En plus des séquences télomériques, il contient le site de reconnaissance de l'endonuclease I-SceI au centre d'un intron placé artificiellement dans la séquence d'un marqueur de sélection auxotrophique (Mitchell and Boeke, 2014).

### 4.1.4 Éléments remplacés

Dans le génome de *S. cerevisiae*, le codon « TAG » est le codon STOP le moins fréquent. Le design des chromosomes synthétiques prévoit de le remplacer par le codon STOP « TAA ». L'élimination du codon stop « TAG » du génome synthétique permettra de manipuler, dans le futur, le code génétique de la levure synthétique. En effet, une fois libre, le codon « TAG » pourra être utilisé pour incorporer des acides aminés artificiels (Isaacs et al., 2011; Park et al., 2011). Une telle modification du génome permet d'ajouter un niveau supplémentaire de contrôle de la levure synthétique. Par exemple, le codon « TAG » peut être placé dans des gènes essentiels pour spécifier à un ARNt synthétique d'utiliser un acide aminé non génétiquement codé, apporté dans le milieu de culture. Ainsi la levure synthétique est dépendante de l'Homme et le risque de propagation dans la nature est limité.

Pour faciliter l'assemblage *in vitro* des fragments d'ADN nécessaires à la construction du génome synthétique, certaines séquences courtes (1 – 10 pb) ont été recodées pour introduire ou éliminer des sites de restriction uniques dans les séquences synthétiques (Dymond et al., 2011). D'autres séquences de moins de 30 pb ont également été recodées pour générer des marqueurs de PCR. Ces modifications permettent de créer des séquences synonymes, qui peuvent être utilisées comme base pour la conception d'amorce de PCR. Elles permettent ensuite d'amplifier spécifiquement des fragments homologues des génomes sauvage ou synthétique. Ainsi en fonction du produit de PCR obtenu, l'intégration du fragment d'ADN synthétique peut être validée ou non. Aucune autre optimisation de codon n'a été effectuée, pour éviter d'altérer l'expression des gènes. D'ailleurs si, au cours de la synthèse, l'un des marqueurs de PCR altère l'expression du gène où il se trouve, la séquence du gène natif est rétablie.

### 4.1.5 Les éléments non modifiés

L'ordre des gènes de la souche sauvage a été préservé dans la levure synthétique. De même que les régions non codantes, sauf indication contraire (comme par exemple les éléments transposables ou les éléments Y' des régions subtélomériques), n'ont pas été modifiées. En effet, pour que le phénotype de la levure synthétique soit équivalent à celui de la souche sauvage, les modifications faites dans les séquences non codantes ont été limitées au maximum. Le risque de perturbation d'éléments critiques dans une région non-annotée du génome est très important.

## 4.2 La synthèse des chromosomes synthétiques

Une fois que la séquence synthétique du chromosome a été déterminée, il avait été initialement prévu que les chromosomes soient assemblés selon un plan d'assemblage hiérarchique. À partir d'oligonucléotides (60 pb), commandés auprès d'entreprises spécialisées dans la synthèse d'ADN, de petits fragments d'ADN double brins ( $\approx 750$  pb) étaient assemblés par PCA (voir chapitre 2

de ce manuscrit). Ils sont ensuite utilisés pour former des minichunks ( $\approx 3$  kb) par clonage. Ces derniers sont ensuite assemblés en chunks ( $\approx 10$  kb), qui sont une fois de plus utilisés pour être assemblés en mégachunks (30 à 50 kb). L'assemblage de ces minichunks en chunks, ou de ces chunks en mégachunks est réalisé *in vitro*. Une fois les mégachunks assemblés ils sont ensuite transformés dans la levure, qui peut les introduire dans son génome par recombinaison homologue. Pour cela, les extrémités du fragment transformé doivent être identiques à la séquence du chromosome à remplacer, sur une centaine de paires de bases (100 à 200 pb). La séquence native correspondante est ainsi remplacée au fur et à mesure par les fragments d'ADN synthétique (figure 4.2).

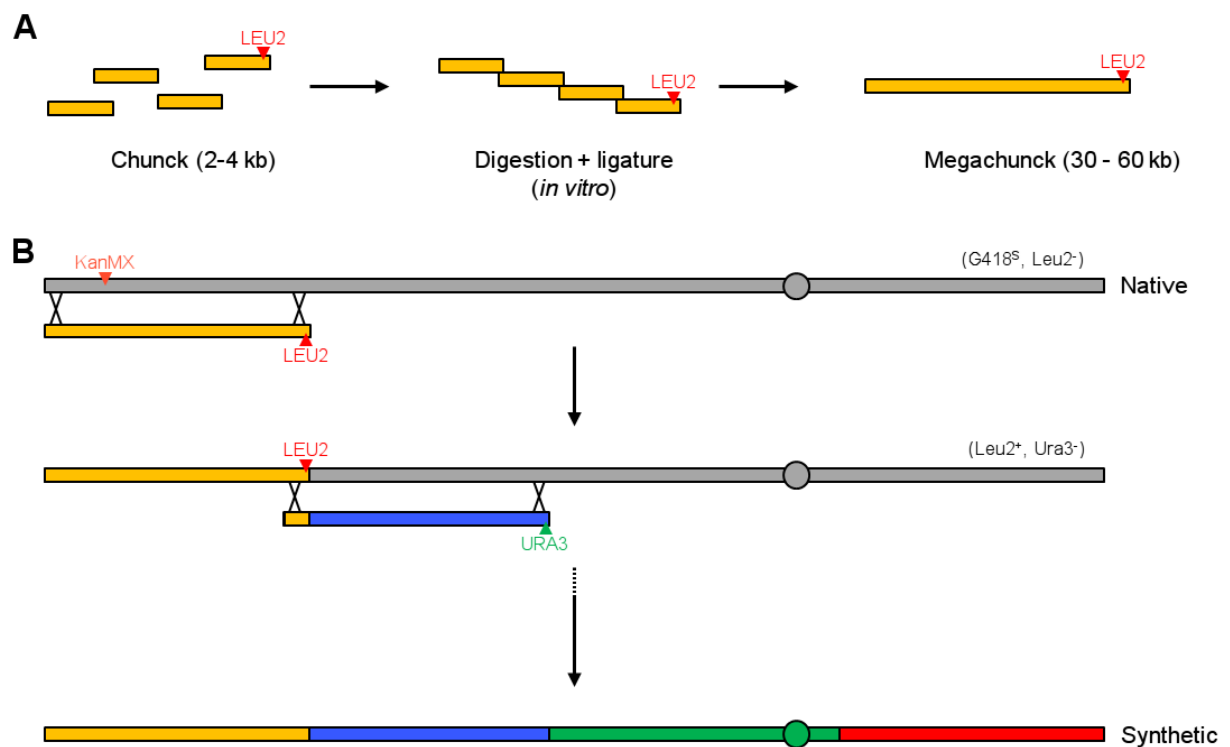
Pour sélectionner les levures ayant intégré le mégachunk à leur génome, un marqueur de sélection (LEU2 ou URA3) est ajouté à l'extrémité droite du mégachunk. Préalablement à l'intégration du premier mégachunk, une cassette kanMX est introduite dans la séquence native du chromosome à des fins de sélection négative de la première intégration. Ensuite, lors de l'intégration du second mégachunk, le premier marqueur d'auxotrophie (LEU2) est éliminé par recombinaison, et la sélection est imposée par un autre marqueur d'auxotrophie (URA3) introduit à l'extrémité droite du nouveau mégachunk (figure 4.2). De cette manière, le chromosome synthétique peut être assemblé, *in vivo*, avec une sélection alternée pour les marqueurs URA3 et LEU2. Cette méthode a été baptisée méthode de « SwAP-In », pour « switching auxotrophies progressively by integration » (Dymond et al., 2011; Richardson et al., 2017).

Aujourd'hui, la diminution des coûts de la synthèse de fragments d'ADN ( $\approx 0,09$  € / pb) a permis de simplifier la méthode d'assemblage hiérarchique. Les minichunks ( $\approx 3$  kb), voir même les chunks ( $\approx 10$  kb), pouvant être directement commandés auprès des entreprises spécialisées, les étapes d'assemblage *in vitro* ont pu être raccourcies, voir même évitées. En effet, plusieurs minichunks ou chunks, peuvent être directement transformés et intégrés au génome de la levure. Ainsi la synthèse du mégachunk, *in vitro*, n'est plus nécessaire (Richardson et al., 2017).

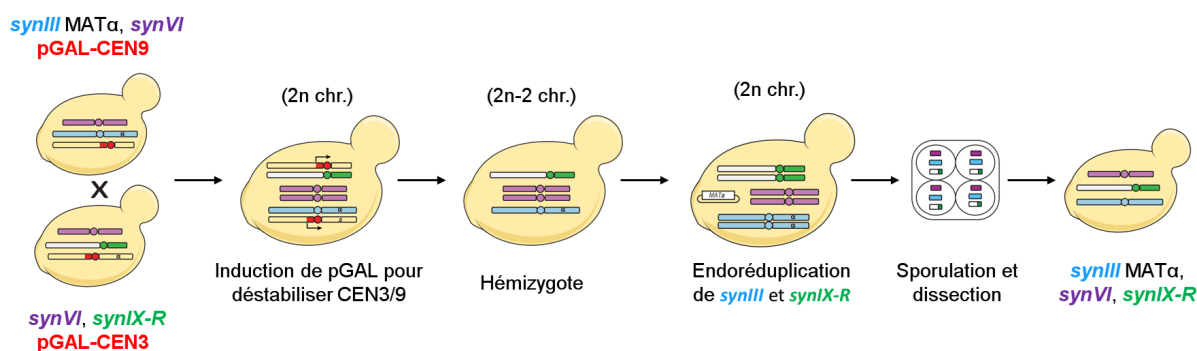
L'assemblage hiérarchique des chromosomes synthétiques permet entre autres de confirmer l'intégration de l'ADN synthétique et la perte de la séquence native, grâce aux marqueurs de sélection et aux marqueurs de PCR, qui ont été intégrés dans la séquence du chromosome synthétique. Elle permet également d'identifier tout segment susceptible de conférer une croissance lente ou d'être létale à la levure. Si un défaut dans le phénotype est identifié, la séquence du dernier segment intégré peut être modifiée de manière à corriger ce phénotype. Généralement, la correction la plus simple est de remplacer la séquence synthétique qui pose problème par la séquence native correspondante.

Le projet Sc2.0 s'est organisé en un consortium international regroupant des équipes de : Johns Hopkins University (États-Unis), New York University (États-Unis), Beijing Genomics Institute





**FIGURE 4.2 Assemblage hiérarchique d'un chromosome synthétique de levure :** (A) Les chunks sont assemblés *in vitro* en mégachunk (B) Les mégachunks peuvent ensuite être transformés dans le levure qui pourra les utiliser pour remplacer son ADN natif par recombinaison homologue. Grâce à des marqueurs de sélection les levures ayant intégré le mégachunk peuvent être sélectionnées et isolées.



**FIGURE 4.3 Combinaison de chromosomes synthétiques par endoréduplication**

(Chine), Tianjin University (Chine), Tsinghua University (Chine), Nat'I University of Singapore (Singapour), Edinburgh University (Royaume-Unis), Imperial College London (Royaume-Unis), Australian Wine Research Institute (Australie) et Macquarie University (Australie). La synthèse des 16 chromosomes de *S. cerevisiae* se faisant séparément, chacun des chromosomes a pu être attribué aux différentes équipes du consortium permettant ainsi de paralléliser la synthèse du génome Sc2.0 (figure 4.5). En plus d'accélérer le processus de synthèse, cette séparation de la synthèse des chromosomes permet de les caractériser et de les tester séparément avant qu'ils ne soient combinés. Une modification qui aurait un effet négatif sur la croissance ou la survie de la cellule, pourra ainsi être identifiée et corrigée plus facilement. Enfin, temps que le nouveau chromosome contenant les ADNt n'est pas terminé, les chromosomes synthétiques ne peuvent pas être combinés. En effet, les ADNt sont essentiels à la survie des cellules, bien qu'ils soient parfois présent en plusieurs copies sur différents chromosomes (275 gènes codant pour 42 ARNt), certaines combinaisons de chromosomes synthétiques sont impossibles tandis que d'autres le sont (par exemple *synII* et *synXII* ; *synIII* et *synXII* ; *synIII*, *synVI* et *synIX-R* ; *synV* et *synX*).

### 4.3 La combinaison des chromosomes synthétiques

Une fois que deux chromosomes sont entièrement synthétisés, dans deux souches de levures différentes, ils peuvent être combinés en croisant ces deux souches de manière à obtenir une cellule diploïde contenant les deux chromosomes synthétiques et leurs deux homologues natifs. La partie la plus délicate restant l'isolement d'une cellule haploïde qui contienne les deux chromosomes synthétiques. Si l'on cherche à combiner seulement deux chromosomes synthétiques, on peut envisager de faire sporuler la cellule diploïde et chercher à isoler une spore contenant les deux chromosomes synthétiques intacts. Cependant, à mesure que le nombre de chromosomes synthétiques augmente, les possibilités de croisements et de recombinaisons méiotiques rendent cette tâche quasi impossible. Pour éviter tout problème de crossing-over entre l'un des chromosomes synthétiques et son homologue natif, l'idée est d'éliminer le chromosome natif et d'avoir une cellule diploïde homozygote pour le chromosome synthétique, avant l'entrée en méiose.

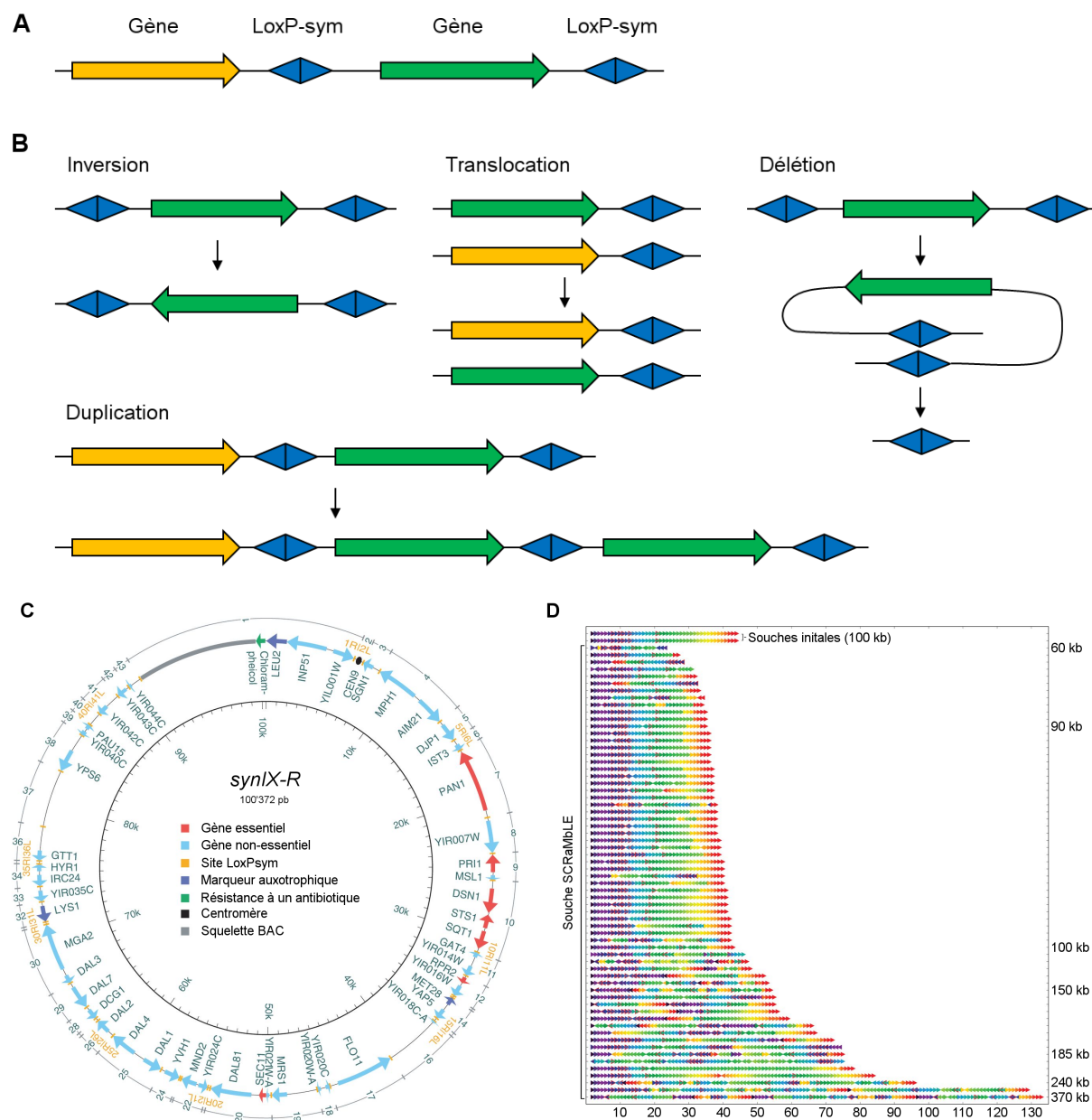
Au cours de la croissance végétative de la levure diploïde hétérozygote, les chromosomes natifs peuvent être éliminés en plaçant des promoteurs inductibles pGAL1 autour de leurs centromères. L'activation de ces promoteurs, va déstabiliser les centromères qu'ils entourent et les détacher du SPB, entraînant alors la perte de ces chromosomes, générant une souche à «  $2n-2$  » chromosomes à la fin de la mitose (Hill and Bloom, 1987). Des hémizygotes sont ainsi générés et la cellule peut endoredupliquer les chromosomes synthétiques, qui ont perdus leur homologues natifs, pour régénérer un état à «  $2n$  » chromosomes (Reid et al., 2008). Ainsi, une cellule diploïde homozygote pour les chromosomes synthétiques est obtenue. Ensuite il ne reste plus qu'à faire sporuler ce diploïde et isoler une spore contenant une copie de chaque chromosome synthétique, pour obtenir une cellule haploïde contenant plusieurs chromosomes synthétiques (figure 4.3).

## 4.4 Le système d'évolution de SCRaMbLE

Le SCRaMbLE est un système d'évolution rapide permettant de générer une multitude de génomes différents à partir du génome synthétique Sc2.0. Il est basé sur la recombinaison entre les différentes séquences LoxP-sym positionnées tout le long des chromosomes synthétiques. Ces séquences mesurent 34 pb et proviennent du bactériophage P1, elles sont les substrats de la protéine Cre qui catalyse les événements de recombinaison entre deux de ces séquences. Pour garder le contrôle de ce système, la protéine Cre est apportée aux levures synthétiques sur un plasmide, où elle est placée sous le contrôle d'un promoteur inductible à œstradiol (Lindstrom and Gottschling, 2009). Ainsi le système de SCRaMbLE ne sera activé que si ce plasmide est transformé dans les levures synthétiques et si de l'œstradiol est ajouté au milieu de culture.

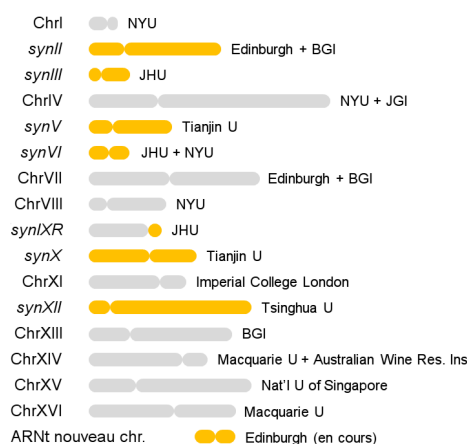
Le système SCRaMbLE a été conçu pour permettre d'introduire un maximum de réarrangements rapides et stochastiques au sein des chromosomes synthétiques, générant ainsi une importante diversité génétique et phénotypique au sein de la population de levure synthétique étudiée. En appliquant une pression de sélection, lors de l'activation du système, la souche synthétique devrait être capable de rapidement s'adapter à son nouvel environnement. Enfin, les séquences LoxP-sym n'ont pas été introduites en aval des gènes essentiels car leur délétion au cours de l'activation du système de SCRaMbLE pourrait s'avérer bien trop critique pour la survie de la cellule. Cependant, il est peu probable mais pas impossible qu'une série de gènes, dont des gènes essentiels, soit éliminée au cours de l'activation du SCRaMbLE, si elle est située entre deux sites loxP-sym.

Récemment, Shen et al., des membres du consortium Sc2.0, ont analysé 64 clones d'une souche contenant une version circulaire de *synIX-R* après induction du système SCRaMbLE (Dymond et al., 2011; Shen et al., 2016). Dans un premier temps ils ont pu constater la spécificité du système SCRaMbLE, puisque 99% des réarrangements observés se sont produits dans la séquence du *synIX-R*. Ils ont ensuite pu mettre en évidence, parmi ces 64 souches, 156 délétions, 89 inversions, 94 duplications, et 55 autres réarrangements trop complexes pour qu'ils soient clairement identifiés.



**FIGURE 4.4 Le système de SCRaMbLE :** (A) et (B) le système de SCRaMbLE permet de générer des inversions, des translocations, des délétions ou des duplications dans les chromosomes synthétiques. (C) et (D) 64 différentes versions du chromosome *synIX-R* ont été obtenues après l'induction du système SCRaMbLE.

Enfin, alors que la taille initiale de *synIX-R* est d'environ 100 kb, après l'induction du SCRaMbLE la taille des chromosomes analysés est extrêmement variable allant de 60 kb pour le plus petit jusqu'à 370 kb pour le plus grand (figure 4.4). Aux vues de ce résultat, alors que *synIX-R* ne contenait que 43 sites LoxP-sym, une fois la souche Sc2.0 complète, qui devrait compter 3932 LoxP-sym, on peut imaginer générer une infinité de génomes grâce à ce système de SCRaMbLE.



**FIGURE 4.5 État de l'art du projet Sc2.0**

Jusqu'à présent, 6.5 des 16 chromosomes de *S. cerevisiae* ont été synthétisés (*synII*, *synIII*, *synV*, *synVI*, *synIX-R*, *synX* et *synXII*). Parmi ces chromosomes, le *synIII* existe en deux versions, MAT $\alpha$  ou MAT $\alpha$ , et le chromosome XII, qui existe avec ou sans rDNA, qui a été relocalisé soit sur un plasmide ou au milieu du bras droit du chromosome III. La délétion des ARNt de la séquence des chromosomes synthétiques limite la possibilité de combinaison des chromosomes, jusqu'à ce que le nouveau chromosome contenant tous les ARNt soit terminé. Certaines combinaisons restent tout de même possible, nous avons ainsi pu analyser plusieurs souches combinant deux ou trois chromosomes synthétiques : *synII* et *synXII* ; *synIII* et *synXII* ; *synV* et *synX* ; *synIII* et *synIX-R* ; ou *synIII*, *synVI* et *synXI-R*.

## **Résultats et discussions**





# L'organisation 3D des chromosomes synthétiques

## Première publication

Comme il a été présenté dans le chapitre 2, quatre caractéristiques majeures peuvent être retenues de l'organisation des chromosomes de *S. cerevisiae* : (1) l'attachement et le regroupement des centromères au SPB ; (2) la position des télomères en petits foyers dynamique à la l'enveloppe nucléaire ; (3) le regroupement préférentiel des bras des chromosomes en fonction de leur taille ; (4) et l'isolement du nucléole des autres chromosomes du côté opposé au SPB. Dans le cadre de ce premier projet, nous avons cherché à déterminer si les chromosomes synthétiques, avec les modifications qu'on leur connaît, conservent la trajectoire et l'organisation de leurs homologues natifs.

Une fois les chromosomes synthétiques terminés, les différents collaborateurs du projet Sc2.0 nous les ont fait parvenir pour que l'on puisse analyser leur organisation, en utilisant la technique de Hi-C. Nous avons alors comparé l'organisation spatiale de ces chromosomes synthétiques par rapport à leurs homologues natifs. Nous avons ainsi montré que l'organisation d'un chromosome synthétique ne présentait pas de différences majeures par rapport à son homologue natif, sauf certains cas précis. De plus, du fait de l'absence de séquences répétées dans les chromosomes synthétiques, la qualité de l'alignement des lectures de séquençage sur leurs génomes de références est améliorée, les cartes de contacts des chromosomes synthétiques présentent un signal continu, sans région invisible. Ces observations mettent en évidence l'importance de l'un des concepts de la synthèse de génome *de novo*, qui vise à affiner les systèmes expérimentaux pour faciliter la résolution des questions complexes.

Dans le cas des chromosomes *synIII* et *synXII*, nous avons tout de même observer quelques différences. D'abord la délétion des cassettes HML et HMR du chromosome *synIII* nous ont permis de mettre en évidence la perte de la boucle habituellement formée par le chromosome III due à leur interaction spécifique. Le locus des répétitions d'ADNr rapproché des centromères sur le bras droit du chromosome III entraîne d'importante modification dans l'organisation générale

du génome. Le bras droit du chromosome III est coupé en deux et le locus d'ADNr semble se retrouver au centre des chromosomes séparant les petits bras des grands bras.

Enfin, nous avons montré que l'analyse des cartes de contacts était très utile pour identifier le type et la position des réarrangements (translocation, duplication ou délétion) qui se sont produits à la suite de l'activation du système de SCRaMbLE. Lors de l'analyse des différents chromosomes synthétiques, les cartes de contacts des différentes souches nous ont également permis de mettre en évidence certains événements non-désirés (translocations ou duplications) qui avaient pu se produire entre les chromosomes natifs.

## RESEARCH ARTICLE SUMMARY

## SYNTHETIC BIOLOGY

## 3D organization of synthetic and scrambled chromosomes

Guillaume Mercy,\* Julien Mozziconacci,\* Vittore F. Scolari, Kun Yang, Guanghou Zhao, Agnès Thierry, Yisha Luo, Leslie A. Mitchell, Michael Shen, Yue Shen, Roy Walker, Weimin Zhang, Yi Wu, Ze-xiong Xie, Zhouqing Luo, Yizhi Cai, Junbiao Dai, Huanming Yang, Ying-Jin Yuan, Jef D. Boeke, Joel S. Bader, Héloïse Muller,† Romain Koszul†

**INTRODUCTION:** The overall organization of budding yeast chromosomes is driven and regulated by four factors: (i) the tethering and clustering of centromeres at the spindle pole body; (ii) the loose tethering of telomeres at the nuclear envelope, where they form small, dynamic clusters; (iii) a single nucleolus in which the ribosomal DNA (rDNA) cluster is sequestered from other chromosomes; and (iv) chromosomal arm lengths. Hi-C, a genomic derivative of the chromosome conformation capture approach, quantifies the proximity of all DNA segments present in the nuclei of a cell population, unveiling the average multiscale organization of chromosomes in the nuclear space. We exploited Hi-C to investigate the trajectories of synthetic chromosomes within the *Saccharomyces cerevisiae* nucleus and compare them with their native counterparts.

**RATIONALE:** The Sc2.0 genome design specifies strong conservation of gene content and arrangement with respect to the native chromosomal sequence. However, synthetic chromosomes incorporate thousands of designer changes, notably the removal of transfer RNA genes and repeated sequences such as transposons and subtelomeric repeats to enhance stability. They also carry loxP sites, allowing for inducible genome SCRaMbLE (synthetic chromosome rearrangement and modification by *loxP*-mediated evolution) aimed at accelerating genomic plasticity. Whether these changes affect chromosome organization, DNA metabolism, and fitness is a critical question for completion of the Sc2.0 project. To address these questions, we used Hi-C to characterize the organization of synthetic chromosomes.

**RESULTS:** Comparison of synthetic chromosomes with native counterparts revealed no substantial changes, showing that the redesigned sequences, and especially the removal of repeated sequences, had little or no effect on average chromosome trajectories. Sc2.0 synthetic chromosomes have Hi-C contact maps with much smoother contact patterns than those of native chromosomes, especially in subtelomeric regions. This improved

## ON OUR WEBSITE

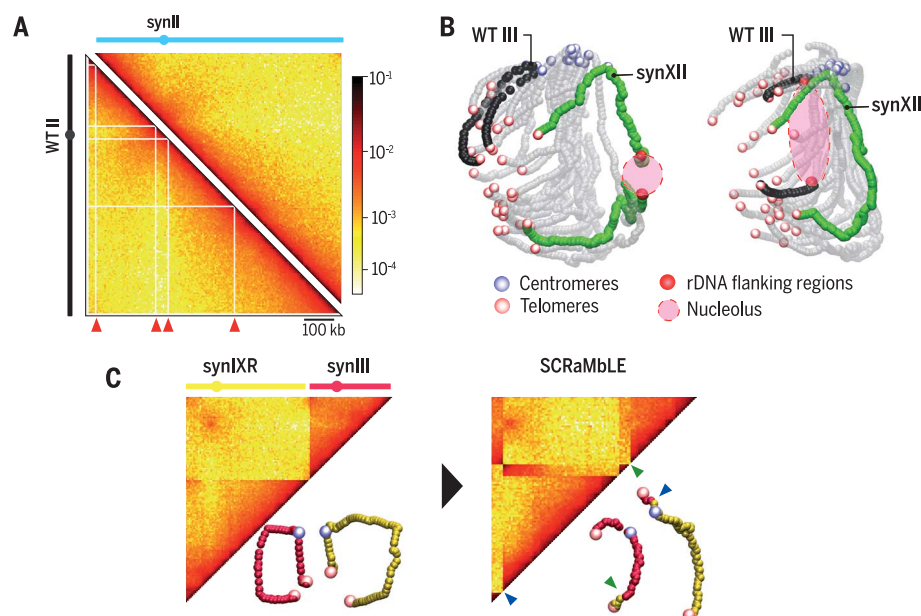
Read the full article at <http://dx.doi.org/10.1126/science.aaf4597>

“mappability” results directly from the removal of repeated elements all along the length of the synthetic chromosomes. These observations highlight a concep-

tual advance enabled by bottom-up chromosome synthesis, which allows refinement of experimental systems to make complex questions easier to address. Despite the overall similarity, differences were observed in two instances. First, deletion of the *HML* and *HMR* silent mating-type cassettes on chromosome III led to a loss of their specific interaction. Second, repositioning the large array of rDNA repeats nearer to the centromere cluster forced substantial genome-wide conformational changes—for instance, inserting the array in the middle of the small right arm of chromosome III split the arm into two noninteracting regions. The nucleolus structure was then trapped in the middle between small and large chromosome arms, imposing a physical barrier between them.

In addition to describing the Sc2.0 chromosome organization, we also used Hi-C to identify chromosomal rearrangements resulting from SCRaMbLE experiments. Inducible recombination between the hundreds of loxP sites introduced into Sc2.0 chromosomes enables combinatorial rearrangements of the genome structure. Hi-C contact maps of two SCRaMbLE strains carrying *synIII* and *synIXR* chromosomes revealed a variety of cis events, including simple deletions, inversions, and duplications, as well as translocations, the latter event representing a class of trans SCRaMbLE rearrangements not previously observed.

**CONCLUSION:** This large data set is a resource that will be exploited in future studies exploring the power of the SCRaMbLE system. By investigating the trajectories of Sc2.0 chromosomes in the nuclear space, this work paves the way for future studies addressing the influence of genome-wide engineering approaches on essential features of living systems. ■



**Synthetic chromosome organization.** (A) Hi-C contact maps of *synII* and native (wild-type, WT) chromosome II. Red arrowheads point to filtered bins (white vectors) that are only present in the native chromosome map. kb, kilobases. (B) Three-dimensional (3D) representations of Hi-C maps of strains carrying rDNA either on *synXII* or native chromosome III. (C) Contact maps and 3D representations of *synIXR* (yellow) and *synIII* (pink) before (left) and after (right) SCRaMbLE. Translocation breakpoints are indicated by green and blue arrowheads.

The list of author affiliations is available in the full article online.

\*These authors contributed equally to this work.

†Corresponding author. Email: [heloise.muller@curie.fr](mailto:heloise.muller@curie.fr) (H.M.); [romain.koszul@pasteur.fr](mailto:romain.koszul@pasteur.fr) (R.K.)

Cite this article as G. Mercy et al., *Science* **355**, eaaf4597 (2017). DOI: [10.1126/science.aaf4597](https://doi.org/10.1126/science.aaf4597)

## RESEARCH ARTICLE

## SYNTHETIC BIOLOGY

## 3D organization of synthetic and scrambled chromosomes

Guillaume Mercy,<sup>1,2,3\*</sup> Julien Mozziconacci,<sup>4\*</sup> Vittore F. Scolari,<sup>1,2</sup> Kun Yang,<sup>5</sup> Guanghou Zhao,<sup>6</sup> Agnès Thierry,<sup>1,2</sup> Yisha Luo,<sup>7</sup> Leslie A. Mitchell,<sup>8</sup> Michael Shen,<sup>8</sup> Yue Shen,<sup>9,10,7</sup> Roy Walker,<sup>7</sup> Weimin Zhang,<sup>6</sup> Yi Wu,<sup>11</sup> Ze-xiong Xie,<sup>11</sup> Zhouqing Luo,<sup>6</sup> Yizhi Cai,<sup>7</sup> Junbiao Dai,<sup>6</sup> Huanming Yang,<sup>12,9</sup> Ying-Jin Yuan,<sup>11</sup> Jef D. Boeke,<sup>8</sup> Joel S. Bader,<sup>5</sup> Héloïse Muller,<sup>1,2,†</sup> Romain Koszul<sup>1,2,†</sup>

Although the design of the synthetic yeast genome Sc2.0 is highly conservative with respect to gene content, the deletion of several classes of repeated sequences and the introduction of thousands of designer changes may affect genome organization and potentially alter cellular functions. We report here the Hi-C-determined three-dimensional (3D) conformations of Sc2.0 chromosomes. The absence of repeats leads to a smoother contact pattern and more precisely tractable chromosome conformations, and the large-scale genomic organization is globally unaffected by the presence of synthetic chromosome(s). Two exceptions are *synIII*, which lacks the silent mating-type cassettes, and *synXII*, specifically when the ribosomal DNA is moved to another chromosome. We also exploit the contact maps to detect rearrangements induced in SCRaMbLE (synthetic chromosome rearrangement and modification by *loxP*-mediated evolution) strains.

**G**enes in the genome of the budding yeast *Saccharomyces cerevisiae* are nonrandomly distributed along its 16 chromosomes. For instance, centromere- and telomere-proximal regions are enriched in families of co-regulated genes (1–4). Chromosome three-dimensional (3D) organization in this species has been studied through imaging and genetic studies (5–8), revealing a characteristic Rabl organization (9). Four main factors drive the overall organization: (i) 16 centromeres tethered and

clustered at the spindle pole body (SPB, the yeast microtubule organizing center); (ii) telomeres forming smaller groups tethered at the nuclear envelope; (iii) a single nucleolus across the nucleus from the SPB, where the ribosomal DNA (rDNA) is sequestered from other chromosomes; and (iv) chromosomal arm lengths. In addition to these factors, several sequences have been shown or suggested to influence the organization. For instance, transfer RNA (tRNA) genes have been proposed to influence chromosome folding through repositioning in the vicinity of either the SPB or the nucleolus (10, 11). Telomeric repeats may also play a role in telomere clustering and anchoring to the nuclear envelope (8, 12). Sc2.0 genome design specifies strong conservation of gene content and arrangement with respect to the native yeast genome (13). However, synthetic chromosomes encode thousands of designer changes intended to improve genetic stability and increase genetic flexibility (13). For example, tRNA genes were deleted during the design of Sc2.0 chromosomes, to be relocated to a separate “neochromosome” (13), and subtelomeric regions were substantially altered as well, with large repeated sequences corresponding either to Y or gene families deleted. Further, *loxP* sites encoded by Sc2.0 chromosomes enable inducible evolution by SCRaMbLE (synthetic chromosome rearrangement and modification by *loxP*-mediated evolution) (14, 15) to generate combinatorial genomic diversity through rearrangements (16). Several synthetic chromosomes are now built (14, 17–22); it is thus possible to experimentally address whether or not Sc2.0 modifications affect the overall chromosome organization in strains carrying synthetic chromosomes.

Genomic chromosome conformation capture approaches [e.g., Hi-C (23, 24)], which provide access to the average 3D chromosome organization from a population of cells, have highlighted the Rabl organization of the *S. cerevisiae* and other fungal genomes (Fig. 1A) (25, 26). In a Hi-C experiment, ligation frequencies between DNA restriction fragments are quantified through deep sequencing, reflecting contacts within the nuclear space and presumably their average respective positions in populations of cells. These frequencies are usually represented as 2D heat maps (fig. S1) and can be processed and visualized as 3D representations to facilitate their interpretation (Fig. 1A) (27). To investigate Sc2.0 chromosome organization, Hi-C contact maps (bin size, 5 kb) and 3D representations (see the Materials and methods section for interpretation of such structures) of *synII*, *synIII*, *synV*, *synVI*, *synIXR*, *synX*, and *synXII* were generated from daughter cells synchronized through elutriation (Fig. 1, B to D; figs. S2 to S10; tables S1 to S2; and movies S1 to S10) (28).

## Trajectories of synthetic chromosomes in the nuclear space

To determine whether individual or combinations of Sc2.0 chromosomes influence the overall genome organization, the contact map of the subset of invariant (native) chromosomes was generated for each strain. Euclidean distances between these maps were then computed, which revealed that the presence of one or more Sc2.0 chromosomes did not dramatically affect the overall structure of the remaining genome (fig. S11). The 3D representations also allowed for the qualitative comparison of native and synthetic sequences, showing that the average trajectories of individual Sc2.0 chromosomes did not appear to be substantively altered compared with those of their native counterparts, with synthetic chromosomes neighboring the same chromosomes as the native ones (Fig. 1 and figs. S2 to S10 for side-by-side comparisons). To quantitatively compare the internal folding of Sc2.0 chromosomes with that of their native counterparts, we computed the contact frequency decay as a function of the increasing genomic distance for all pairs of loci. No substantial differences between synthetic and native chromatin were detected (Fig. 1E), suggesting that the redesigned sequence has little or no effect on the internal folding of the chromosome beyond a ~5 to 10% shortening through the removal of repeats.

## Sc2.0 design improves mappability

A comparison of Sc2.0 maps with those of their native counterparts nevertheless revealed much smoother contact patterns in the designer chromosomes, as shown by the quasi-absence of white lines in synthetic contact maps. This stems from the intentional deletion of repeated elements, which leads to improved “mappability,” relative to the native sequence, along the length of the synthetic chromosome (Fig. 1, E and F, and figs. S2 to S10). For instance, for chromosome II, eight bins (5% of the chromosome) appeared insufficiently

<sup>1</sup>Spatial Regulation of Genomes, Department of Genomes and Genetics, Institut Pasteur, Paris 75015, France.

<sup>2</sup>UMR3525, Centre National de la Recherche Scientifique (CNRS), Paris 75015, France. <sup>3</sup>Sorbonne Universités,

Université Pierre et Marie Curie (Université Paris 6), Paris 75005, France. <sup>4</sup>Laboratoire de Physique Théorique de la Matière Condensée, CNRS UMR7600, Université Pierre et Marie Curie (Université Paris 6), Sorbonne Universités, Paris, France. <sup>5</sup>Department of Biomedical Engineering and High-Throughput Biology Center, Whiting School of Engineering,

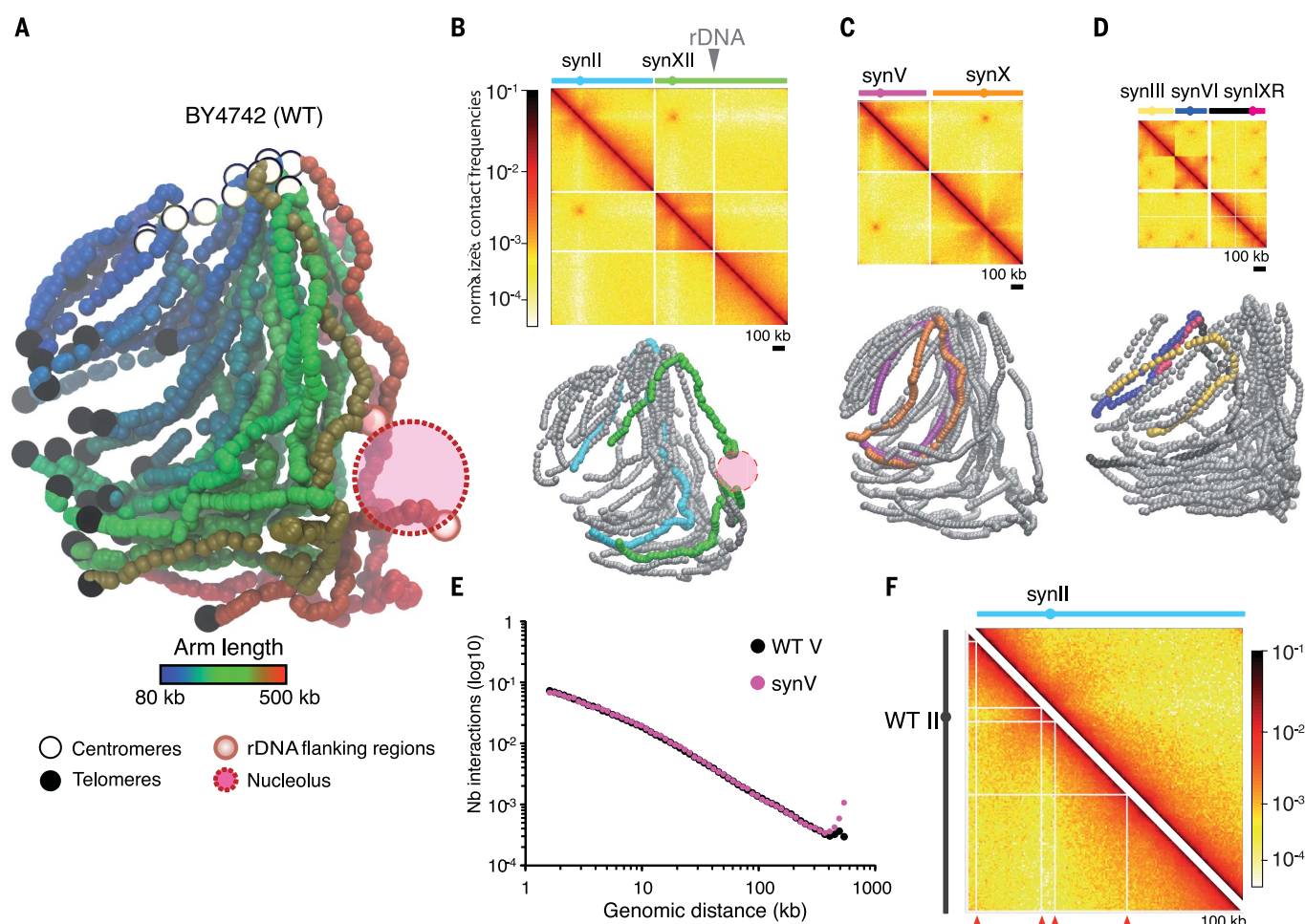
Johns Hopkins University, Baltimore, MD 21218, USA. <sup>6</sup>Key Laboratory for Industrial Biocatalysis (Ministry of Education), Key Laboratory of Bioinformatics (Ministry of Education), Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China. <sup>7</sup>School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK. <sup>8</sup>Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, Langone Medical Center, New York University, New York, NY 10016, USA. <sup>9</sup>BGI-Shenzhen, Shenzhen 518083, China. <sup>10</sup>BGI-Qingdao, Qingdao 266555, China. <sup>11</sup>Key Laboratory of Systems Bioengineering (Ministry of Education), SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China. <sup>12</sup>James D. Watson Institute of Genome Sciences, Hangzhou 310058, China.

\*These authors contributed equally to this work. †Present address: Institut Curie, UMR3664 Dynamique du Noyau, Paris, France.

†Corresponding author. Email: heloise.muller@curie.fr (H.M.); romain.koszul@pasteur.fr (R.K.)

†Corresponding author. Email: heloise.muller@curie.fr (H.M.); romain.koszul@pasteur.fr (R.K.)





**Fig. 1. 3D genome organization of native and synthetic chromosomes.**

(A) 3D representation of the native (wild-type, WT) yeast chromosomes, inferred from the Hi-C contact map displayed in fig. S1 (27). Each bead represents a 5-kb chromosome segment. Centromeres, telomeres, and rDNA-flanking regions are indicated with white, black, and red beads, respectively. Each chromosomal arm has been colored according to its length. (B to D) Contact maps and corresponding 3D representations of synthetic chromosomes in three different strains: JDY512 (synII and synXII) (B), yZX573 (synV and synX) (C), and yLM896 (synIII, synVI, and synIXR) (D). The top panels are normalized contact maps of the seven synthetic chromosomes (bin size, 5 kb). Normalized contact

frequencies (27) are indicated in a log<sub>10</sub> scale ranging from white (few contacts) to dark red (many contacts). Filtered bins are set to zero (white vectors). The solid gray triangle points to the position of the rDNA cluster. The bottom panels show whole-genome 3D representations of the contact maps displayed above, with synthetic and native chromosomes represented with colored and gray beads, respectively. (E) Distribution of the number of contacts as a function of the genomic distance for synV (pink) and native chromosome V (black). (F) Side-by-side comparison of synII (strain YSO31) and native chromosome II (strain BY4742) normalized contact maps. Nonmappable, repeated regions are highlighted with red arrowheads on the native map.

covered in the native (wild-type, WT) counterpart, whereas the synthetic counterpart (synII) had no bins that were insufficiently covered (figs. S2 and S12). In addition, the normalized contact frequencies between synVI subtelomeres appeared to be much higher than in the native counterpart (fig. S13). This variation results directly from the improved mappability of subtelomeric regions after the removal of subtelomeric repeats. On the other hand, subtelomeric contacts are underestimated in native chromosomes because of these repetitive sequences. These observations highlight the conceptual advance enabled by bottom-up chromosome design and synthesis to investigate fundamental questions; scientists can now design experimental systems at unprecedented levels of refinement, making complex questions easier to address experimentally. For

instance, investigating the elusive regulatory influences of repeated sequences in mammalian genomes could benefit from similar techniques.

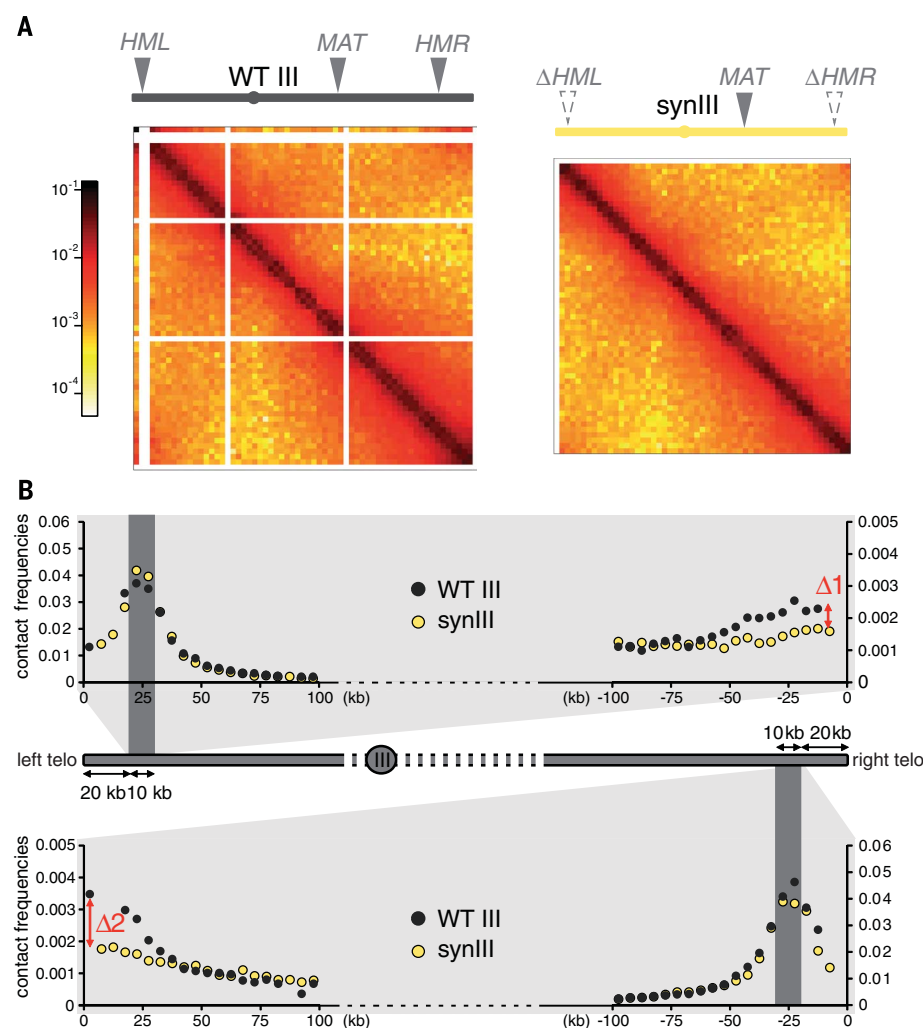
### SynIII conformation is modified by the design

Specific contacts between the duplicated *HML* and *HMR* silent mating-type loci that bridge the two arms of chromosome III have been identified (23, 29). The deletion of these cassettes led to a loss of contacts on synIII (Fig. 2, A and B). Moreover, chromosome III shows a mating-type-specific conformation that depends on the integrity of a recombination enhancer sequence (30). Despite the presence of the recombination enhancer sequence in synIII, the mating-type-specific conformation was lost (fig. S14). We speculate that the deletion of the silent loci in synIII underlies

this result, but additional experiments are required to discriminate this possibility from the potential influence of other changes in the sequences. synIII still exhibited enriched contacts between its subtelomeres, a feature characteristic of the small metacentric chromosomes I, III, and VI (12, 31).

### Repositioning the rDNA locus affects the global 3D structure of the genome

The rDNA locus on the right arm of chromosome XII consists of a cluster (100 to 200) of ~9-kb rDNA units, each containing two ribosomal precursor genes. This cluster assembles in the nucleolus, a discrete, crescent-shaped nuclear compartment that occupies ~one-third of the total nuclear space and is positioned opposite to the SPB. Two versions of the synXII chromosome



**Fig. 2. Loss of subtelomeric contacts in synIII.** (A) Normalized contact maps (bin size, 5 kb) of native (WT) chromosome III (left; strain BY4742) and synIII (right; strain yLM896). The color scale is as in Fig. 1. Mating-type sequences are indicated along the top x axis with solid gray triangles (MAT, mating-type locus; HML and HMR, left and right silent mating cassettes). The positions of deleted loci are indicated with gray dashed triangles. (B) Quantitative analysis of subtelomeric contacts in native chromosome III (black dots) and synIII (yellow dots) by means of a bait chromosome capture approach. Contacts of 10-kb subtelomeric regions positioned at 20 kb from the left (top) or right (bottom) telomeres (dark gray areas) are shown. The y axis shows normalized contact frequencies; the x axis shows the distance (in kilobases) from the left and right telomeres. Each point represents the mean contact frequency for the bait region, computed for each chromosome from three independent experiments (on strains BY4742, YS031, and JDY512 for native chromosome III and yLM896, yLM539, and JDY452 for synIII).  $\Delta 1$  and  $\Delta 2$  mark the normalized contacts discrepancies between both ends of synIII relative to native chromosome III.

were designed, with and without the rDNA cluster. When not located on synXII, the rDNA was positioned either on the multicopy  $2\mu$  plasmid (one unit per plasmid) or reintegrated as an array within the right arm of the small chromosome III (22). Chromosome organization was investigated in these strains. Because of their repetitive nature, rDNA unit reads cannot be uniquely mapped, and the rDNA cluster on the 3D reconstruction thus appears as an empty space flanked by the closest regions that can be mapped unambiguously (Figs. 1A and 3, A to C). In strains without the rDNA locus on synXII (JDY446, JDY448,

and JDY449), the continuity of the two regions flanking the original insertion site was clearly visible on the contact map and 3D representations of the right arm (Fig. 3, A to C, and figs. S5 and S6). The presence of  $2\mu$  plasmids carrying a rDNA unit (strain JDY446) did not result in a large reorganization of the genome (fig. S5), suggesting that the plasmids do not form large discrete heterochromatic structures, in agreement with the studies reporting the formation of multiple small clusters (22, 32). On the other hand, the insertion of the rDNA array in the middle of the right arm of chromosome III (strains JDY448 and JDY449) resulted

in substantial reorganization of the genome (Fig. 3C and fig. S6). The right arm of chromosome III appeared to be split into two noninteracting regions by the rDNA cluster, an arrangement similar to the native chromosome XII structure (25). Relocating the rDNA to a short chromosomal arm imposed new constraints on the genome: The nucleolus became trapped between short and long chromosome arms, forming a physical barrier between them, as shown by reduced contacts between these arms (Fig. 3C). In addition, contacts between the rDNA and subtelomeric regions were increased overall (Fig. 3D) as a result of the greater proximity of the nucleolus to the entire set of chromosomal extremities. Also, the III<sub>R</sub> subtelomere in this strain contacted long arm subtelomeres (blue arrow on Fig. 3C). No growth rate defects were detected in these strains, suggesting that the position of the rDNA locus has little effect on fitness in the laboratory growth conditions tested in our study. However, given the observed constraints, it would be informative to run long-term cultures to see whether the rDNA locus spontaneously relocalizes, and if so, where.

### Hi-C analysis of SCRaMbLE chromosomes

All synthetic chromosomes carry the SCRaMbLE system, a chromosome-wide expansion of the Cre/lox site-specific recombination assay used to measure local DNA concentrations in genomes (5, 33). The Sc2.0 design includes the integration of hundreds of loxP sites all along chromosomes, at the end of every nonessential gene and at specific landmarks. Upon activation of the Cre recombinase, two loxP sites will recombine if they encounter each other in the nuclear space, leading to a structural variant. The hundreds of sites therefore hold a large combinatory potential, leading to rapid and complex structural changes of synthetic chromosomes. Previous analysis of synIXR SCRaMbLE strains has revealed a large variety of cis events, ranging from simple duplications, inversions, and deletions to highly reorganized structures (14, 16). We took the opportunity offered by strains carrying two Sc2.0 chromosomes to further explore the power of the SCRaMbLE recombination assay in trans (14, 15). The Cre recombinase was induced for 48 hours in strain yLM539 carrying synIII and synIXR (Fig. 4A and fig. S15). Independent clones were sampled during the time course experiment, and two clones, HMSY029 (time  $T_2 = 2$  hours) and HMSY030 ( $T_8 = 8$  hours), were retained for in-depth analysis. Genome-wide Hi-C contact data have recently been shown to provide a convenient way to identify and solve chromosomal rearrangements (fig. S16) (34–37). We therefore applied Hi-C to investigate the consequences of genome structural alterations in these two isolates. The contact maps identify gross chromosomal rearrangements in both SCRaMbLE strains, with each strain exhibiting deletion and translocation events, the latter representing a class of trans SCRaMbLE events not previously observed (Fig. 4B). The resolution of the contact map (~2 to 5 kb) allowed identification of relatively large-scale rearrangements, but loxP sites are often



more closely spaced. To address this, a refined analysis of the reads coverage was performed (table S3 and fig. S17), unveiling multiple short-scale modifications of loxPsym-flanked segments (summarized in Fig. 4C, details in fig. S18). These results highlight the potential of the SCRaMbLE design to generate gross chromosomal rearrangements, with more rearrangements being achieved after longer induction times. Incidentally, Hi-C analysis of Sc2.0 strains identified several chromosomal rearrangements that spontaneously arose during chromosome assemblies, such as aneuploidy, duplication, or translocation events outside of the synthetic chromosome of interest (table S1 and fig. S19) (17). Hi-C is therefore a convenient way to validate the assembly and subsequent processing of strains carrying synthetic chromosomes.

## Outlook

This large data set is a resource that will be exploited in future studies exploring the power of the SCRaMbLE system. By showing that the spatial organization of the Sc2.0 genome is not substantially altered, this work paves the way for future studies addressing the influence of genome-wide engineering approaches on essential features of living systems.

## Materials and methods

### Media

Yeast were grown either in YPD [1% (w/v) Bacto peptone (Difco), 1% (w/v) Bacto yeast extract (Difco), 2% (w/v) glucose, and 2% (w/v) Bactoagar] or in selective SC media [0.67% Yeast Nitrogen Base without amino acids (Difco), supplemented with an amino acid mix minus those whose prototrophy is selected for, 2% (w/v) glucose and 2% (w/v) Bactoagar].

### Strains

All strains used in the study are described in table S1.

### Culture conditions

Cells were inoculated and grown overnight in 10 ml YPD. Overnight cultures were then diluted and grown to exponential phase in 500 ml YPD for ~15 hours at 30°C. For SCRaMbLE and mating-type specific organization experiments, Hi-C libraries were generated from asynchronous cultures (see table S2). For all other Sc2.0 strains, Hi-C libraries were generated from daughter cells recovered using elutriation and restarted for 30 min in YPD at 30°C (38).

### Generation of Hi-C libraries

Hi-C libraries were generated with a protocol adapted from (23) with introduction of a biotinylation step (24). Briefly, aliquots of  $1-3 \times 10^9$  cells were cross-linked for 30 min with fresh formaldehyde (3% final concentration) and quenched with glycine for 15 min. Pelleted cells were dissolved in 10 ml sorbitol 1 M and incubated for 30 min with DTT 5 mM and Zymolyase 100T ( $C_{\text{Final}} = 1 \text{ mg/ml}$ ). Spheroplasts were washed with 5 ml sorbitol 1M, then with 5 ml 1X restric-

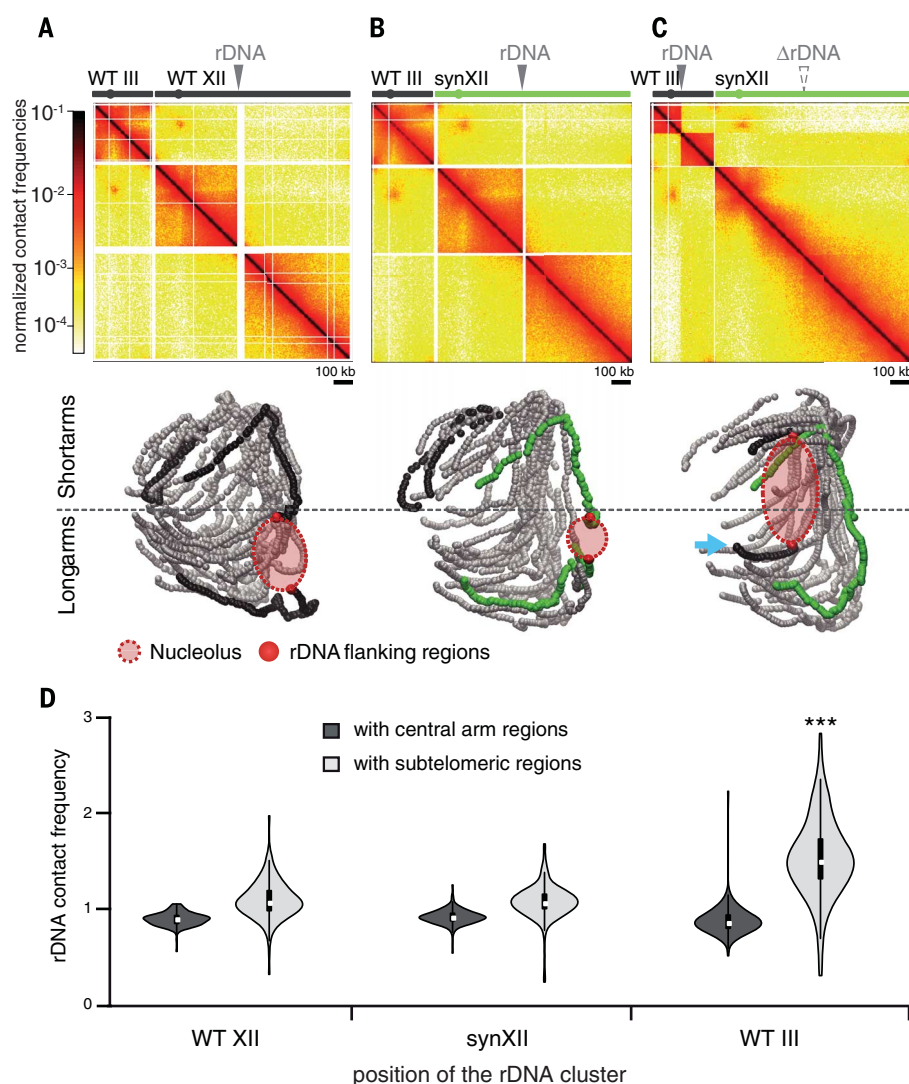
tion buffer (NEB), and suspended in 3.5 ml 1X restriction buffer. Cells were split into aliquots ( $V = 500 \mu\text{l}$ ) and incubated in SDS (3%) for 20 min at 65°C. Cross-linked DNA was digested at 37°C overnight with 150 units of DpnII restriction enzyme (NEB). The digestion mix was subsequently centrifuged 20 min at 18,000 g and the supernatant discarded. Pellets were suspended in cold water. DNA ends were repaired in the presence of 14-dCTP biotin (Invitrogen), and cross-linked complexes incubated for 4 hours at 16°C in presence of 250 U of T4 DNA ligase (Thermo Scientific, 12.5 ml final volume). DNA purification was achieved through an overnight incubation at 65°C with 250  $\mu\text{g/ml}$  proteinase K in 6.2 mM EDTA followed by a precipitation step and RNase treatment.

The resulting Hi-C libraries were sheared and processed into Illumina libraries according to manufacturer instructions (Paired-End DNA sample Prep Kit – Illumina – PE-930-1001) using custom-made versions of the Illumina PE adapters (28). Fragments between 400 and 800 bp were purified, amplified, and paired-end (PE) sequenced on an Illumina platform (HiSeq2000 or NextSeq500; table S2).

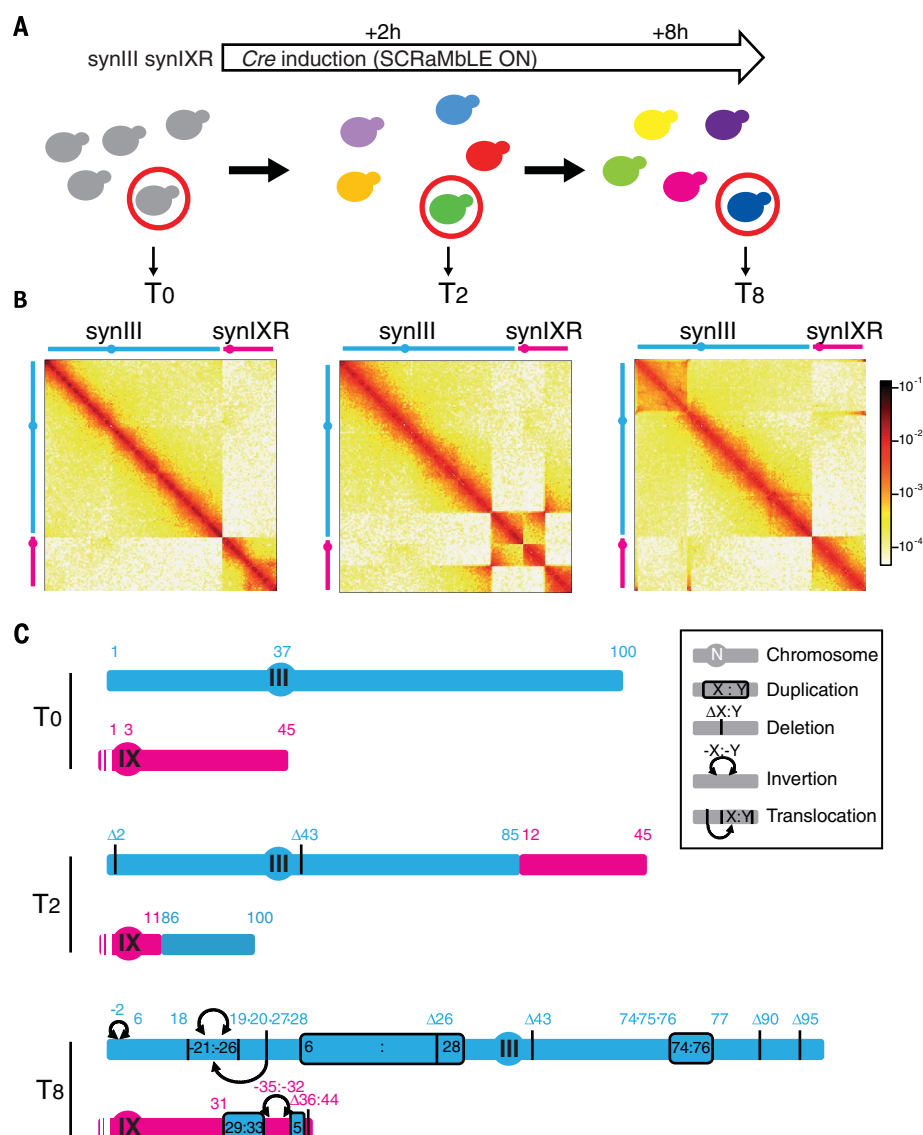
Raw sequences are accessible in the SRA database under accession number SRP070421.

### Generation and normalization of contact maps

Sequencing pair-end data was processed as follows. For each library, PCR duplicates were collapsed



**Fig. 3. Repositioning of rDNA affects the overall genomic organization.** (A to C) The top panels are normalized contact maps of strains BY4742 (A), JDY465 (B), and JDY449 (C) carrying the rDNA cluster at different positions. Solid gray triangles point to the position of the rDNA cluster in each strain. The position of the deleted rDNA cluster is indicated with a gray dashed triangle. The bottom panels show the 3D representations of the corresponding contact maps. The nucleolus and the rDNA cluster-flanking regions are represented on each structure. Chromosome coloring is the same as in Fig. 1. The blue arrow points to the chromosome III right arm that is displaced after rDNA insertion. (D) Violin plot of the contact frequencies between the rDNA cluster and either subtelomeric (light gray) or intra-arm (dark gray) chromosomal regions. \*\*\*P < 0.001.



**Fig. 4. SCRaMbLE induction of synthetic chromosomes results in multiple types of genome rearrangements.** (A) Illustration of the time course induction of the SCRaMbLE system in strain yLM539 carrying two synthetic chromosomes, synIII and synIXR. (B) Normalized contact maps (bin size, 2 kb) of the parental strain yLM539 (at  $T_0$ ; carrying synIII and synIXR) and two SCRaMbLE clones isolated after 2 hours ( $T_2$ ; HMSY029) and 8 hours ( $T_8$ ; HMSY030) of Cre induction. All Hi-C reads are mapped against the reference genome of the parental strain yLM539. (C) Schematic representations of the duplications, deletions, inversions, and translocations identified from these maps and the coverage analysis. DNA segments between two loxP sites are numbered from left to right (in blue for synIII and pink for synIX). The schematic representation shows the rearranged synIII and synIX by using numbering of the segments from the parental strain. Details are shown in fig. S18.

using the 6 Ns present on each of the custom-made adapters and trimmed. Reads were then aligned using Bowtie 2 in its most sensitive mode against the *S. cerevisiae* reference genome adapted with the synthetics sequences (39). An iterative alignment procedure was used: for each read the length of the mapped sequence increases gradually from 20 bp until mapping becomes unambiguous (mapping quality > 30). Read pairs were aligned independently and assigned to a restriction fragment (RF). Religation and other unwanted events were filtered out, taking into account the

orientation of the reads, as described (28, 40). Contact matrices were built for each strain by binning the aligned reads into units of single RF. Adjacent restriction fragments were then pooled into fixed size bins of either 2 or 5 kb. Bins exhibiting important contact frequencies variations (< or > two standard deviation) were filtered, with the corresponding vectors either set to zero (white band on contact maps) or removed (see fig. S12). Binned contact maps were then normalized using the sequential component normalization (SCN) described in (40).

### 3D representation of contact maps

The 3D representations of the contact maps were generated using ShRec3D (27) on the normalized contact maps filtered for low signal bins. These 3D structures are average representations from populations of billions of cells, and therefore do not represent the exact structure found in an individual cell. It must also be underlined that they are not polymer models and cannot be interpreted as such. They have to be interpreted as representations of DNA contact frequencies over a population of cells. For instance, telomeres loosely cluster together on these 3D representations. In a single nucleus, telomeres would rather form smaller groups scattered all around the nuclear membrane. Since in different cells these group gather different partners, they are regrouped together in the average structure that reflects the population average of contacts. All 3D structures presented here were rendered using VMD (41).

### Similarity between contact maps

To assess the differences of the different data sets, we proceeded as follows. First, normalized contact maps of native chromosomes I, IV, VII, VIII, XIII, XIV, XV, and XVI were binned at 50 kb and quantile normalized. We then computed the Euclidean distance between all pairs of maps and performed a principal component analysis on the resulting distance matrix.

### Genomic distance plot

Pairs of reads mapping in cis (i.e., intrachromosomal) positions along the genome were partitioned by chromosome. Reads oriented towards different directions or separated by less than 1.5 kb were discarded. For each chromosome, read pairs were log-binned according to the genomic distance  $s$  separating them (in kb)

$$\text{bin} = \lfloor \log_{10}(s) \rfloor$$

The genomic distance plot is the weighted histogram computed from the sum of read pairs for each bin, weighted with the expected number of pairs under the uniform null hypothesis. To compare synthetic and native chromosome both distributions were normalized by their mean computed over the 70% of the curve at lower distances.

### SCRaMbLE assay

Strain yLM539 (synIII, synIXR) was transformed with the plasmid pSCwII CRE-EBD\_HIS3 and grown in SC-HIS media (14). Cre expression was induced with 1  $\mu\text{M}$  estradiol. Samples of induced and non-induced cultures were analyzed over 48 hours (0, 1, 2, 4, 5, 6, 8, 24, and 48 hours). For each time point, serial dilutions were spotted on YPD and SC-HIS agar plates (fig. S15). Two clones, HMSY029 ( $T_2 = 2$  hours) and HMSY030 ( $T_8 = 8$  hours), were retained for sequencing and Hi-C analysis.

### Genomic analysis of SCRaMbLE strains

Illumina paired-end short-reads were trimmed to remove adapter sequences. Reads shorter than 100 bp, and/or with unknown bases, and/or exhibiting 1 or more bases with a Phred-score under 7, were filtered and discarded from the analysis.



Remaining reads were then mapped to the reference sequence using Bowtie2. The read coverage over the entire genome was computed, with a specific focus on the synthetic regions. The copy number of each segment in between loxPsym sites was assessed to detect deletions, duplications, and higher amplifications. To account for possible systematic experimental biases in the sequencing depth resulting from library preparation and mapping, we used an iterative algorithm to refine the copy number estimation (16).

We then focused on the unmapped reads carrying a loxPsym site to characterize SCRaMBLE rearrangements. The two segments flanking a loxPsym site define a junction. We trisected unmapped reads into a loxPsym site and its two flanking extremities, then mapped each of the latter to the reference genome using EMBOSS water (42) to identify novel junction. The average sequencing depths of novel junctions for the 2-hour and 8-hour Cre-induced strains were 16.3 and 14.3, respectively. We combined novel junctions with a depth higher than 5 and the copy number of segments to characterize the SCRaMBLE induced rearrangements (table S3).

That no off-target rearrangement occurred elsewhere in the genome was confirmed by analyzing unmapped reads without loxPsym site with EMBOSS water to detect potential recombination events.

### Analysis of rDNA contacts

Genome-wide rDNA contact frequencies were investigated in Sc2.0 strains carrying the rDNA either at its natural position on chromosome XII (data sets from strains YS031, yXZX538, yLM539, yLM896, HMSY012, yXZX573), on chromosome synXII (data sets from strains JDY465, JDY512, JDY452), on chromosome III (data sets from strains JDY448, JDY449), or on the 2 $\mu$  plasmid (data set from strain JDY446). For each data set, pair-end reads were aligned on the reference genome of the corresponding strain lacking rDNA sequences and on an extra contig corresponding to a single rDNA unit (~9 kb). Contact maps of trans interactions were generated (bin size, 5 kb) and normalized (40). To compare the contacts between the rDNA contig with the rest of the genome in different data sets, the values in the corresponding vector were divided by their median to alleviate differences of coverage and/or number of rDNA units. Chromosomes III and XII, which carry the rDNA cluster in three out of four data sets, were removed from the analysis to allow comparison of the rest of the genome. For each data set, the contacts made by the rDNA sequence with either the 28 subtelomeric regions (30 kb at the extremities of the 14 chromosomes investigated) or a 30-kb region positioned 30 kb from the centromere on each arm (central region) were plotted using violin plot function from the “vioplot” R library.

### Distribution of read coverage for synthetic and native chromosomes

The removal of repeated genetic elements in synthetic chromosomes relative to native counterparts

leads to smother contact maps. Indeed, Hi-C reads cannot be aligned against repeats without ambiguities, resulting in vectors with a low if not null number of contacts. Those vectors appear as empty rows and columns in the contact maps and are a source of noise and spurious contacts in the matrices after normalization. For this reason, these regions are filtered according to a threshold (see the section on Generation and normalization of contact maps). To quantify the improvement in the visibility of synthetic chromosomes all along their length, we compared the histogram of their coverage between the synthetic and the native data set (fig. S12). The histograms count the number of contacts made by each vector within the region of interest with the entire genome (histogram bin size, 500 reads). The filtering threshold (median – 2SD) was computed over the entire genome (reported on the fig. S12 histograms as dotted lines).

### REFERENCES AND NOTES

- P. M. Sharp, A. T. Lloyd, Regional base composition variation along yeast chromosome III: Evolution of chromosome primary structure. *Nucleic Acids Res.* **21**, 179–183 (1993). doi: [10.1093/nar/21.2.179](https://doi.org/10.1093/nar/21.2.179); pmid: [8441625](https://pubmed.ncbi.nlm.nih.gov/8441625/)
- K. R. Bradnam, C. Seoighe, P. M. Sharp, K. H. Wolfe, G+C content variation along and among *Saccharomyces cerevisiae* chromosomes. *Mol. Biol. Evol.* **16**, 666–675 (1999). doi: [10.1093/oxfordjournals.molbev.a026149](https://doi.org/10.1093/oxfordjournals.molbev.a026149); pmid: [10335660](https://pubmed.ncbi.nlm.nih.gov/10335660/)
- B. A. Cohen, R. D. Mitra, J. D. Hughes, G. M. Church, A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* **26**, 183–186 (2000). doi: [10.1038/79896](https://doi.org/10.1038/79896); pmid: [11017073](https://pubmed.ncbi.nlm.nih.gov/11017073/)
- E. J. Louis, A. V. Vershinin, Chromosome ends: Different sequences may provide conserved functions. *BioEssays* **27**, 685–697 (2005).
- S. M. Burgess, N. Kleckner, Collisions between yeast chromosomal loci in vivo are governed by three layers of organization. *Genes Dev.* **13**, 1871–1883 (1999). doi: [10.1101/gad.13.14.1871](https://doi.org/10.1101/gad.13.14.1871); pmid: [10421638](https://pubmed.ncbi.nlm.nih.gov/10421638/)
- V. Guacci, E. Hogan, D. Koshland, Centromere position in budding yeast: Evidence for anaphase A. *Mol. Biol. Cell* **8**, 957–972 (1997). doi: [10.1091/mbc.8.6.957](https://doi.org/10.1091/mbc.8.6.957); pmid: [9201708](https://pubmed.ncbi.nlm.nih.gov/9201708/)
- P. Therizols, T. Duong, B. Dujon, C. Zimmer, E. Fabre, Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2025–2030 (2010). doi: [10.1073/pnas.0914187107](https://doi.org/10.1073/pnas.0914187107); pmid: [20080699](https://pubmed.ncbi.nlm.nih.gov/20080699/)
- A. Taddei, S. M. Gasser, Structure and function in the budding yeast nucleus. *Genetics* **192**, 107–129 (2012). doi: [10.1534/genetics.112.140608](https://doi.org/10.1534/genetics.112.140608); pmid: [22964839](https://pubmed.ncbi.nlm.nih.gov/22964839/)
- C. Rabl, On cell division. *Morphol. Jahrb.* **10**, 214–330 (1985).
- M. Thompson, R. A. Haeusler, P. D. Good, D. R. Engelke, Nucleolar clustering of dispersed tRNA genes. *Science* **302**, 1399–1401 (2003). doi: [10.1126/science.1089814](https://doi.org/10.1126/science.1089814); pmid: [14631041](https://pubmed.ncbi.nlm.nih.gov/14631041/)
- R. A. Haeusler, D. R. Engelke, Genome organization in three dimensions: Thinking outside the line. *Cell Cycle* **3**, 273–275 (2004). doi: [10.4161/cc.3.3.732](https://doi.org/10.4161/cc.3.3.732); pmid: [14726665](https://pubmed.ncbi.nlm.nih.gov/14726665/)
- H. Schober et al., Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Res.* **18**, 261–271 (2008). doi: [10.1101/gr.6687808](https://doi.org/10.1101/gr.6687808); pmid: [18096749](https://pubmed.ncbi.nlm.nih.gov/18096749/)
- S. M. Richardson et al., Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
- J. S. Dymond et al., Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011). doi: [10.1038/nature10403](https://doi.org/10.1038/nature10403); pmid: [21918511](https://pubmed.ncbi.nlm.nih.gov/21918511/)
- J. Dymond, J. Boeke, The *Saccharomyces cerevisiae* SCRaMBLE system and genome minimization. *Bioeng. Bugs* **3**, 168–171 (2012). pmid: [22572789](https://pubmed.ncbi.nlm.nih.gov/22572789/)
- Y. Shen et al., SCRaMBLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res.* **26**, 36–49 (2016). doi: [10.1101/gr.193433.115](https://doi.org/10.1101/gr.193433.115); pmid: [26566658](https://pubmed.ncbi.nlm.nih.gov/26566658/)
- Y. Shen et al., Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355**, eaaf4791 (2017).
- N. Annaluru et al., Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–58 (2014). doi: [10.1126/science.1249252](https://doi.org/10.1126/science.1249252); pmid: [24674868](https://pubmed.ncbi.nlm.nih.gov/24674868/)
- Z.-X. Xie et al., “Perfect” designer chromosome V and behavior of a ring derivative. *Science* **355**, eaaf4704 (2017).
- L. A. Mitchell et al., Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355**, eaaf4831 (2017).
- Y. Wu et al., Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355**, eaaf4706 (2017).
- W. Zhang et al., Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355**, eaaf3981 (2017).
- J. Dekker, K. Rippe, M. Dekker, N. Kleckner, Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002). doi: [10.1126/science.1067799](https://doi.org/10.1126/science.1067799); pmid: [11847345](https://pubmed.ncbi.nlm.nih.gov/11847345/)
- E. Lieberman-Aiden et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009). doi: [10.1126/science.1181369](https://doi.org/10.1126/science.1181369); pmid: [19815776](https://pubmed.ncbi.nlm.nih.gov/19815776/)
- Z. Duan et al., A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010). doi: [10.1038/nature08973](https://doi.org/10.1038/nature08973); pmid: [20436457](https://pubmed.ncbi.nlm.nih.gov/20436457/)
- H. Marie-Nelly et al., Filling annotation gaps in yeast genomes using genome-wide contact maps. *Bioinformatics* **30**, 2105–2113 (2014). doi: [10.1093/bioinformatics/btu162](https://doi.org/10.1093/bioinformatics/btu162); pmid: [24711652](https://pubmed.ncbi.nlm.nih.gov/24711652/)
- A. Lesne, J. Riposo, P. Roger, A. Cournac, J. Mozziconacci, 3D genome reconstruction from chromosomal contacts. *Nat. Methods* **11**, 1141–1143 (2014). doi: [10.1038/nmeth.3104](https://doi.org/10.1038/nmeth.3104); pmid: [25240436](https://pubmed.ncbi.nlm.nih.gov/25240436/)
- A. Cournac, M. Marbouty, J. Mozziconacci, R. Koszul, Generation and analysis of chromosomal contact maps of yeast species. *Methods Mol. Biol.* **1361**, 227–245 (2016). doi: [10.1007/978-1-4939-3079-1\\_13](https://doi.org/10.1007/978-1-4939-3079-1_13); pmid: [26483025](https://pubmed.ncbi.nlm.nih.gov/26483025/)
- A. Miele, K. Bystrycky, J. Dekker, Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. *PLoS Genet.* **5**, e1000478 (2009). doi: [10.1371/journal.pgen.1000478](https://doi.org/10.1371/journal.pgen.1000478); pmid: [19424429](https://pubmed.ncbi.nlm.nih.gov/19424429/)
- J.-M. Belton et al., The conformation of yeast chromosome III is mating type dependent and controlled by the recombination enhancer. *Cell Rep.* **13**, 1855–1867 (2015). doi: [10.1016/j.celrep.2015.10.063](https://doi.org/10.1016/j.celrep.2015.10.063); pmid: [26655901](https://pubmed.ncbi.nlm.nih.gov/26655901/)
- K. Bystrycky, T. Laroche, G. van Houwe, M. Blaszczyk, S. M. Gasser, Chromosome looping in yeast. *J. Cell Biol.* **168**, 375–387 (2005). doi: [10.1083/jcb.200409091](https://doi.org/10.1083/jcb.200409091); pmid: [15684028](https://pubmed.ncbi.nlm.nih.gov/15684028/)
- M. Oakes et al., Mutational analysis of the structure and localization of the nucleolus in the yeast *Saccharomyces cerevisiae*. *J. Cell Biol.* **143**, 23–34 (1998). doi: [10.1083/jcb.143.1.23](https://doi.org/10.1083/jcb.143.1.23); pmid: [9763418](https://pubmed.ncbi.nlm.nih.gov/9763418/)
- E. R. Hildebrandt, N. R. Cozzarelli, Comparison of recombination in vitro and in *E. coli* cells: Measure of the effective concentration of DNA in vivo. *Cell* **81**, 331–340 (1995). doi: [10.1016/0092-8674\(95\)90386-0](https://doi.org/10.1016/0092-8674(95)90386-0); pmid: [7736586](https://pubmed.ncbi.nlm.nih.gov/7736586/)
- H. Marie-Nelly et al., High-quality genome (re)assembly using chromosomal contact data. *Nat. Commun.* **5**, 5695 (2014). doi: [10.1038/ncomms6695](https://doi.org/10.1038/ncomms6695); pmid: [25517223](https://pubmed.ncbi.nlm.nih.gov/25517223/)
- N. Kaplan, J. Dekker, High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat. Biotechnol.* **31**, 1143–1147 (2013). doi: [10.1038/nbt.2768](https://doi.org/10.1038/nbt.2768); pmid: [24270850](https://pubmed.ncbi.nlm.nih.gov/24270850/)
- J.-F. Flot, H. Marie-Nelly, R. Koszul, Contact genomics: Scaffolding and phasing (meta)genomes using chromosome 3D physical signatures. *FEBS Lett.* **589**, 2966–2974 (2015). doi: [10.1016/j.febslet.2015.04.034](https://doi.org/10.1016/j.febslet.2015.04.034); pmid: [25935414](https://pubmed.ncbi.nlm.nih.gov/25935414/)
- J. N. Burton et al., Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013). doi: [10.1038/nbt.2727](https://doi.org/10.1038/nbt.2727); pmid: [24185095](https://pubmed.ncbi.nlm.nih.gov/24185095/)
- M. Guidi et al., Spatial reorganization of telomeres in long-lived quiescent cells. *Genome Biol.* **16**, 206 (2015). doi: [10.1186/s13059-015-0766-2](https://doi.org/10.1186/s13059-015-0766-2); pmid: [26399229](https://pubmed.ncbi.nlm.nih.gov/26399229/)

39. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923); pmid: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
40. A. Cournac, H. Marie-Nelly, M. Marbouty, R. Koszul, J. Mozziconacci, Normalization of a chromosomal contact map. *BMC Genomics* **13**, 436 (2012). doi: [10.1186/1471-2164-13-436](https://doi.org/10.1186/1471-2164-13-436); pmid: [22935139](https://pubmed.ncbi.nlm.nih.gov/22935139/)
41. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996). doi: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5); pmid: [8744570](https://pubmed.ncbi.nlm.nih.gov/8744570/)
42. P. Rice, I. Longden, A. Bleasby, EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000). doi: [10.1016/S0168-9525\(00\)00204-2](https://doi.org/10.1016/S0168-9525(00)00204-2); pmid: [10827456](https://pubmed.ncbi.nlm.nih.gov/10827456/)

# ACKNOWLEDGMENTS

We thank A. Cournac, M. Marbouty, and L. Lazar-Stefanita for fruitful discussions and advice. This research was supported by funding to R.K. from the European Research Council (ERC) under

the 7th Framework Program (FP7/2007-2013, ERC grant agreement 260822), Agence Nationale pour la Recherche (ANR) (MeioRec ANR-13-BSV6-0012-02), and ERASynBio and ANR (IESY ANR-14-SYNB-0001-03). H.M. and V.F.S. are partly supported by Pasteur-Roux Fellowships. J.D. was funded by the National Science Foundation of China (31471254), Tsinghua University Initiative Scientific Research Program (2011Z02296), Ph.D. Programs Foundation of Ministry of Education of China (20110002120055), and the Chinese Ministry of Science and Technology (2012CB725201). Y.-J.Y. was supported by the Natural Science Foundation of China (21621004 and 21390203). Y.C. was funded by a Chancellor's Fellowship from the University of Edinburgh, a startup fund from the Scottish Universities Life Sciences Alliance, and grants from the U.K. Biotechnology and Biological Sciences Research Council (BB/M005690/1, BB/M025640/1, and BB/M00029X/1). Y.S. was supported by a research grant from the Shenzhen Engineering Laboratory for Clinical Molecular Diagnostic Promotion [JZF no. (2016)884]. This work was supported in part by funding from the U.S. National Science Foundation (grants

MCB-0718846 and MCB-1026068 to J.D.B. and J.S.B. and MCB-0546446 and MCB-1445545 to J.S.B.). J.D.B. and J.S.B. are founders and directors of Neochromosome. J.D.B. serves as a scientific advisor to Recombinetics and Sample6. These arrangements are reviewed and managed by the committees on conflict of interest at the New York University Langone Medical Center (J.D.B.) and Johns Hopkins University (J.S.B.). FASTQ files of the reads have been deposited in the Sequence Read Archive database under accession number SRP070421.

# SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/355/6329/eaaf4597/suppl/DC1](http://www.sciencemag.org/content/355/6329/eaaf4597/suppl/DC1)

Figs. S1 to S19

Tables S1 to S3

References (43–44)

Movies S1 to S10

16 February 2016; accepted 1 February 2017  
10.1126/science.aaf4597

## Résultats supplémentaires

### 4.4.1 Introduction

En plus des premières analyses faites sur certains des chromosomes du projet Sc2.0 et présentées dans l'article précédent, nous avons étudié un intermédiaire du chromosome regroupant les ADNt et plusieurs combinaisons de chromosomes fusionnés (voir section 4.1.3 de ce manuscrits). Le chromosome codant les ARNt est toujours en cours de synthèse dans le laboratoire de Patrick Cai (Université de Manchester, anciennement Université d'Édimbourg), nous n'avons donc pas pu analyser la version finale, mais ce dernier nous a envoyé une version intermédiaire codant 190 ARNt sur les 274 possibles. Comme les autres chromosomes du projet Sc2.0, ce nouveau chromosome est construit dans la souche BY4741 de *S. cerevisiae* avant d'être combiné aux autres chromosomes. Les chromosomes fusionnés, quant à eux, ont été obtenus par Jingshuan Luo, une étudiante de l'équipe du Pr. Jef Boeke, qui a effectué des tests de fusion de chromosomes pour préparer la construction du chromosome *synI*. Après s'être aperçue que *S. cerevisiae* supportait très bien la fusion de deux chromosomes, elle a continué à fusionner ces chromosomes les uns avec les autres pour essayer de regrouper tout son génome en un seul et même chromosome<sup>2</sup>. Elle a ainsi obtenu plusieurs souches de *S. cerevisiae* avec un nombre réduit de chromosomes : une souche à 14 chromosomes (JL277) ; une souche à 3 chromosomes (JL410) ; et récemment une souche à 2 chromosomes (JL498) qui doit encore être analysée (Tableau 4.1). Toutes ces souches nous ont été transmises pour que nous puissions analyser l'organisation de ces nouveaux génomes.

**TABLE 4.1** Combinaison de chromosomes fusionnés

| Souche | Nb. de chr. | Combinaisons des chr. fusionnés         | Centromères |
|--------|-------------|-----------------------------------------|-------------|
| JL277  | 14 chr.     | Chr. IX-III-I                           | CEN3        |
|        |             | Chr. XVI-IX-III-I-VI-II                 | CEN3        |
| JL410  | 3 chr.      | Chr. XI-VII-XIII-XIV                    | CEN7        |
|        |             | Chr. V-VIII-X-XV-VI-XII                 | CEN15       |
| JL498  | 2 Chr.      | Chr. XVI-IX-III-I-VI-II                 | CEN3        |
|        |             | Chr. XIV-XIII-VII-XI-V-VIII-X-XV-IV-XII | CEN15       |

Dans cette partie, je vais vous présenter et discuter les différents résultats que nous avons obtenus suite à l'analyse de ces nouvelles souches.

### 4.4.2 L'organisation du nouveau chromosome regroupant les ADNt

Dans le cadre du projet Sc2.0, il est prévu que les 274 copies ADNt de *S. cerevisiae* soient regroupées sur un nouveau chromosome. Bien que la synthèse de ce chromosome ne soit pas

2. Lors de la fusion de deux chromosomes un seul des deux centromères est conservé, et au niveau de leur jonction, les séquences des cores-X, les éléments Y et les STR des régions subtélomériques et les télomères sont éliminés.

encore terminée, nous avons pu analyser une version intermédiaire, circulaire, codant 190 ARNt. En appliquant la méthode de Hi-C sur une population asynchrone de cette souche (en phase de croissance exponentielle), nous avons généré une carte de contacts et une structure 3D moyenne de son génome (figure 4.6). Le zoom fait sur ce chromosome, nous permet de visualiser son signal circulaire, visible grâce à la présence d'interactions dans les coins supérieur droit et inférieur gauche de la carte de contacts (marqueurs noirs). Comme on pouvait s'y attendre, le centromère présent sur ce chromosome est regroupé avec ceux des autres chromosomes (marqueurs bleus). Sur la structure 3D, il semble que ce nouveau chromosome (représenté en orange) soit isolé du reste des autres chromosomes, au centre de la structure, mais il est très probable que ce soit le résultat d'un artefact dû au signal circulaire de ce chromosome. En effet si l'on regarde la proportion de contact *cis* et *trans* de ce chromosome, elle est équivalente à celle des autres chromosomes de la souche (tableau 4.2). S'il avait été réellement isolé des autres chromosomes, la proportion de contacts *cis* aurait été beaucoup plus élevée. Nous avons également cherché à voir si le regroupement de toutes ces copies d'ADNt permettait de rapprocher ce chromosome du locus d'ADNr, comme certaines études ont pu le suggérer (Thompson et al., 2003; Wang et al., 2005), mais aucun effet n'a pu être identifié.

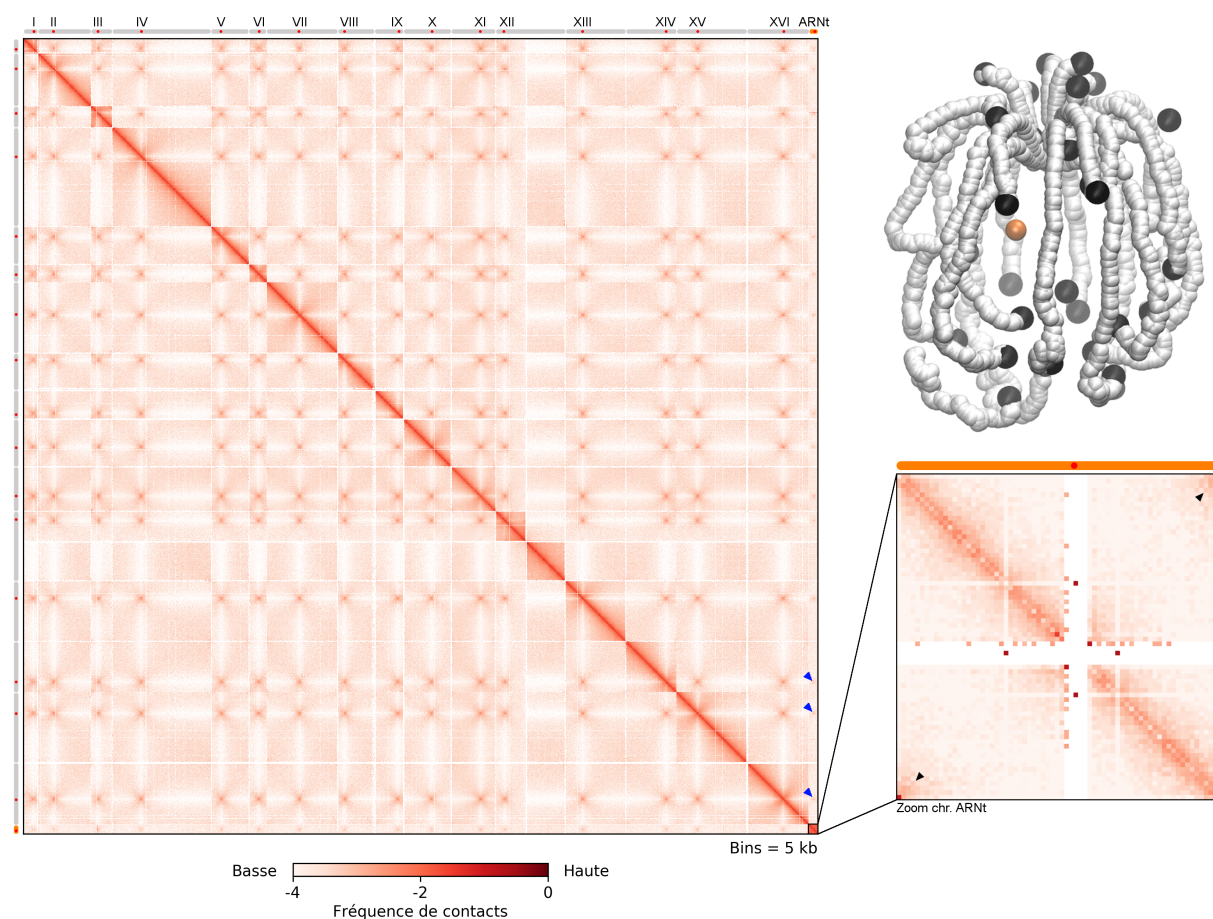
**TABLE 4.2** Proportion de contacts entre le chromosome d'ADNt et le reste du génome

|                | Nb. contacts globaux | Nb. contacts <i>cis</i> | Nb. contacts <i>trans</i> |
|----------------|----------------------|-------------------------|---------------------------|
| Génome complet | 25'920'705           | 10'999'333 - (42.4%)    | 14'921'372 - (57.6%)      |
| Chr. ARNt      | 592'469              | 244'163 - (41.2%)       | 348'306 - (58.8%)         |

Ces observations préliminaires, obtenues à partir d'une version intermédiaire de ce chromosome, devront être confirmées une fois qu'il sera entièrement synthétisé. D'après le design de ce nouveau chromosome, grâce au système de télomérisation (section 4.1.3 ; Mitchell and Boeke, 2014), deux versions pourront être utilisées : une version circulaire et une version linéarisée. Il sera intéressant de voir, en fonction du design choisi (circulaire ou linéaire), si l'organisation de ce chromosome est différente. Enfin la version de ce chromosome que nous avons analysée était maintenue dans la souche BY4741 de *S. cerevisiae*, or toutes les copies d'ADNt sont déjà présentes dans le génome natif. Dans cette configuration, ce nouveau chromosome n'est pas encore essentiel pour la cellule, ce qui peut également influencer son organisation. Il n'est donc pas impossible que son organisation soit différente une fois qu'il sera exprimé dans la souche Sc2.0 finale.

#### 4.4.3 Regroupement des cassettes HML et HMR indépendamment des télomères

Nous nous sommes ensuite intéressé à la souche ne contenant plus que 14 chromosomes (JL277), dans laquelle les chromosomes I, III et IX ont été fusionnés en un seul, en conservant la séquence CEN3 comme centromère (tableau 4.1). Pour analyser l'organisation des chromosomes de cette



**FIGURE 4.6 Organisation du nouveau chromosome d'ARNt :** La matrice de contacts de la souche contenant le nouveau chromosome regroupant les copies d'ADNt est représentée pour des bins de 5kb. Les contacts ont été normalisés. Le nouveau chromosome est situé en bas à droite dans la matrice, le regroupement de son centromère avec ceux des autres chromosomes est indiquée par les flèches bleues. Sur le grossissement, le signal circulaire du nouveau chromosome est indiqué par les flèches noires. Dans la structure 3D, les télomères sont représentés par les billes noires et le nouveau chromosome est représenté en orange.

souche nous avons effectué une expérience de Hi-C sur une population de cellules synchronisées en G1. D'après la carte de contacts et la structure 3D obtenues pour cette population de cellules, on remarque que l'organisation de pseudo-Rabl des chromosomes de *S. cerevisiae* est conservée, avec le regroupement des centromères d'une part, et celui des télomères d'autre part (figure 4.7). Plus localement, lorsque l'on s'intéresse à la trajectoire du chromosome fusionné (chr. IX-III-I), en *cis*, bien que les régions télomériques et subtélomérique aient été supprimées, les deux *loci* HML et HMR, aux extrémités du chromosome III, interagissent ensembles (flèche bleue, figure 4.7). Ce repliement du chromosome III avait déjà été décrit dans la souche sauvage de *S. cerevisiae* (Dekker et al., 2002 ; Belton et al., 2015 ; figure 4.7B) et nous avons nous même constaté que ce repliement était dépendant des cassettes HML et HMR (Mercy et al., 2017). En plus de leur interaction entre eux, ces deux *loci* interagissent également avec les télomères du chromosome (flèches noires, figure 4.7). Du fait de ces interactions, le chromosome prend alors une forme de « clef de sol », avec les cassettes HML et HMR aux extrémités du chromosome III qui se rapprochent l'une de l'autre (en vert dans la représentation 3D, figure 4.7C) et les télomères qui se replient également vers ces deux *loci*. En *trans*, on peut voir sur la carte de contacts que ces deux *loci* interagissent avec les télomères des autres chromosomes (flèches vertes, figure 4.7). Sur la représentation 3D les *loci* HML et HML sont localisés au niveau des foyers télomériques.

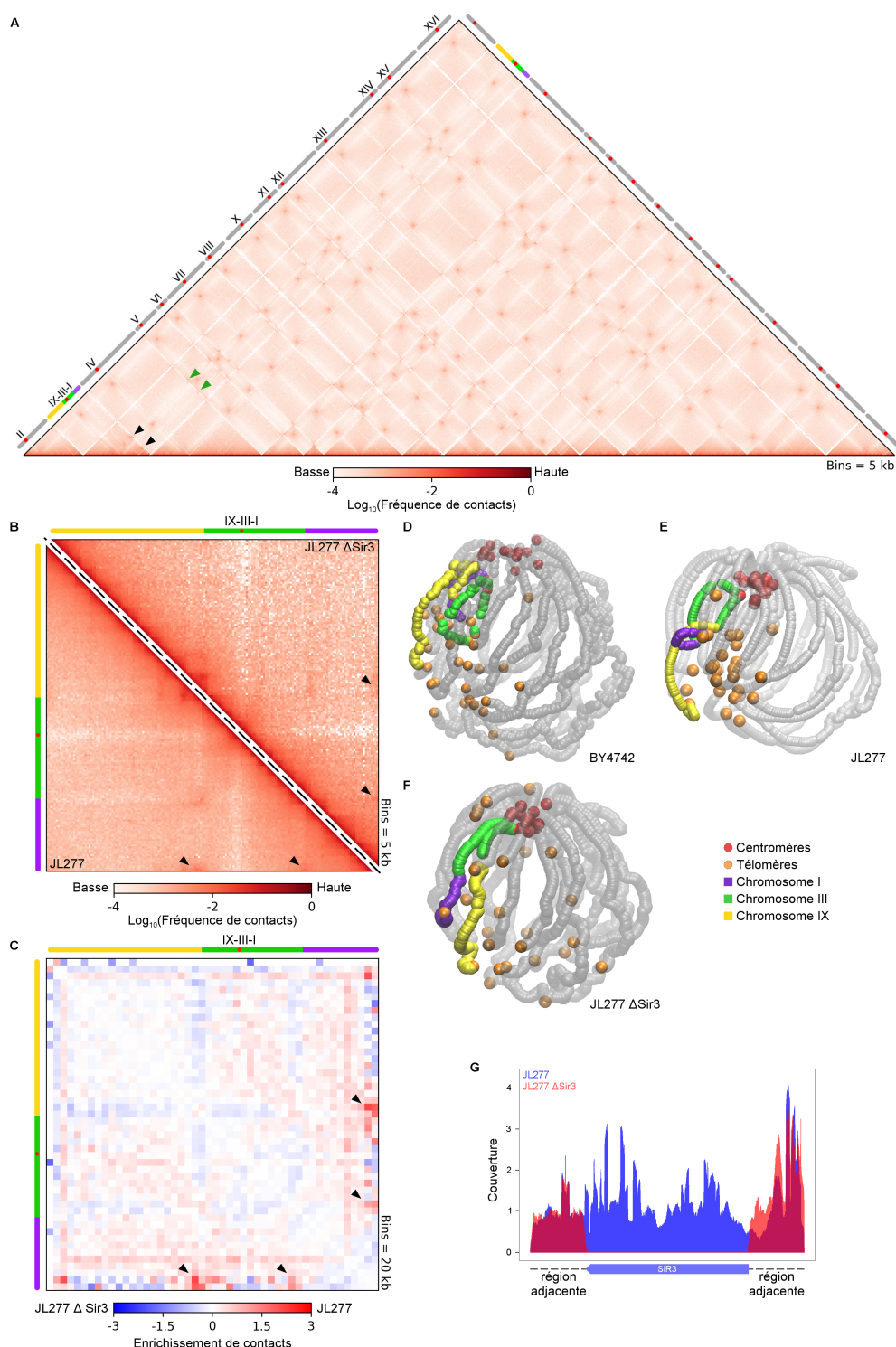
Ce regroupement des *loci* HML et HMR au niveau des foyers télomériques est probablement la conséquence de leur inhibition par les protéines du complexe SIR. Pour tester cette hypothèse, nous avons analysé l'organisation de ce chromosome IX-III-I après avoir supprimé le gène *SIR3* de la souche<sup>3</sup>. Dans cette condition (souche JL277  $\Delta$ Sir3), on peut voir que les interactions entre HML, HMR et les télomères du chromosome sont diminuées (figure 4.7). Cette diminution de fréquence de contacts est plus claire lorsque que l'on effectue le ratio entre les carte de contacts de ces deux souches (JL277 en rouge et JL277  $\Delta$ Sir3 en bleu, figure 4.7C). D'après ce résultat, et en accord avec notre hypothèse, la suppression de la protéine Sir3 entraîne une perte de la formation du regroupement des cassettes HML et HMR au niveau des foyers télomériques. Bien que ces *loci* ne soient plus à proximité des télomères dans le chromosome IX-III-I, la cellule continue à les rapprocher des régions subtélomériques, via l'intermédiaire des protéines du complexe SIR, probablement dans le but d'inhiber leur transcription.

#### 4.4.4 Organisation de la souche à trois chromosomes

En continuant à fusionner ces chromosomes, l'équipe du Pr. Jef Boeke a obtenu une souche de *S. cerevisiae* de trois chromosomes. La taille moyenne des chromosomes de cette nouvelle souche est ainsi passée de 750 kpb à 4'000 kpb, et, 13 centromères et 26 télomères ont été supprimés par rapport à la souche native de 16 chromosomes. Comme dans les cas précédents, afin de déterminer si de telles modifications ont changé l'organisation du génome de cette levure,

3. Les lectures de séquençage peuvent également être utiliser pour contrôler que la séquence du gène avait bien été supprimée (figure 4.7).





**FIGURE 4.7 L'organisation des cassettes HML et HMR dans le chromosome fusionné :** (A) Matrice de contacts normalisées (Bins = 5 kpb) de la souche JL277 synchronisée en G1. La position des chromosomes est indiquée sur le côté de la matrice. Les flèches noires indique le rapprochement des cassettes HML et HMR l'une de l'autre. Les flèches vertes indiquent leur rapprochement avec les télomères des autres chromosomes. (B) Grossissement du chr. IX-III-I pour la souche JL277 (moitié inférieure) et la souche JL277  $\Delta$ Sir3 (moitié supérieure). (C) Ratio des deux matrices précédentes. En bleu, les contacts sont plus fort dans la souche JL277  $\Delta$ Sir3, en rouge ils sont plus fort souche JL277. Sur les structures 3D des souches BY4742 (D), JL277 (E) et JL277  $\Delta$ Sir3 (F), les centromères sont représentés en rouge, les télomères en orange, le chr. I en violet, le chr. III en vert et le chr. IX en jaune. (G) Couverture du gène Sir3 dans la souche JL277 (bleu) et JL277  $\Delta$ Sir3 (rouge).

nous avons effectué une analyse de Hi-C sur une population de cellules asynchrones (phase de croissance exponentielle), et sur une population de cellules synchronisées en G1 par élutriation. Comme pour le génome natif de *S. cerevisiae*, le génome de cette souche suit une organisation de pseudo-Rabl. On y retrouve le regroupement des centromères ainsi que celui des télomères, et le nucléole qui semble s'isoler, représenté en bleu clair à l'arrière de la structure 3D (Figure 4.8B). Les télomères, qui ne sont plus que six, semblent être beaucoup plus proches que ce qu'ils peuvent être dans la souche sauvage. Il serait intéressant de les visualiser en microscopie à fluorescence pour vérifier s'ils ne forment plus qu'un seul foyer.

En G1, on peut remarquer quelques éléments supplémentaires dans l'organisation de ce génome. Tout comme on a pu l'observer dans la souche JL277, les cassettes HML et HMR, qui sont aux extrémités de ce qui était le chromosome III, se rapprochent l'une de l'autre (marqueur bleu), et se regroupent également avec les régions subtélomériques des autres chromosomes (marqueur noir). On peut également voir sur la matrice et sur la structure 3D que la séquence du chromosome I se replie et son extrémité droite interagit avec la cassette HML (marqueur vert). Tout comme nous avons pu le voir dans le cas précédent, ce rapprochement des cassettes HML et HMR avec les régions subtélomériques des autres chromosomes se fait probablement dans le but d'empêcher leur transcription. La disparition de ce phénomène dans les données obtenues à partir de la population asynchrone est probablement le résultat de la réplication des chromosomes au cours de la phase S.

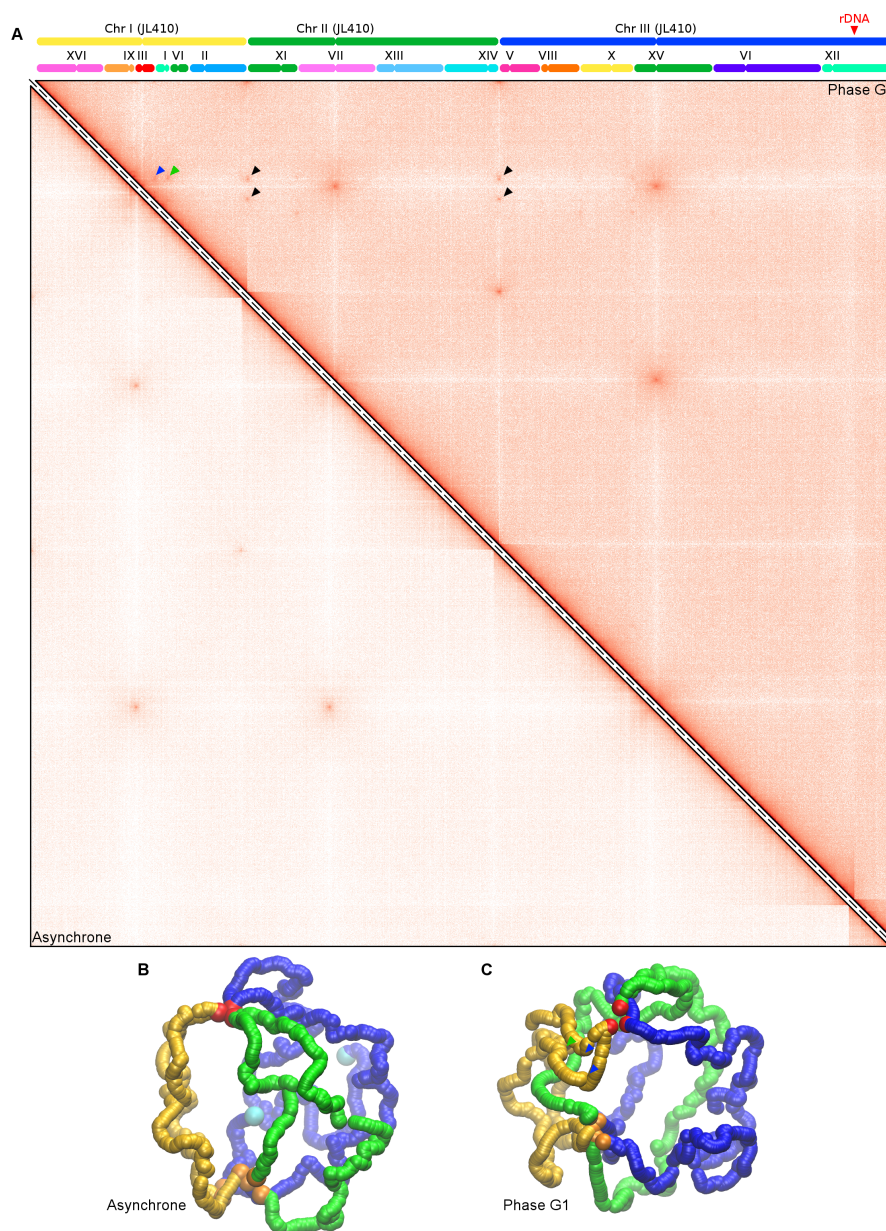
À partir de nos données de Hi-C, nous pouvons évaluer la condensation d'un chromosome en calculant la proportion de contacts entre deux régions de ce chromosome en fonction de la distance qui les sépare (loi de distance génomique,  $P(s)$ ). En utilisant les séquences de référence des trois chromosomes natifs dont les centromères ont été conservés dans la souche aux chromosomes fusionnés (chr. III, chr. VII et chr. XV), nous avons pu comparer la condensation de ces nouveaux chromosomes par rapport à leur homologue natif, pour les différents bras de ces chromosomes. Nous avons ainsi pu constater que la condensation de ces chromosomes n'était pas modifiée par rapport à la celle des chromosomes de la souche sauvage (figure 4.9).

En plus de ces observations, grâce aux données obtenues en G1 et en phase exponentielle, nous avons pu déterminer le profil de réplication de ces chromosomes fusionnés et le comparer à celui des chromosomes sauvages.

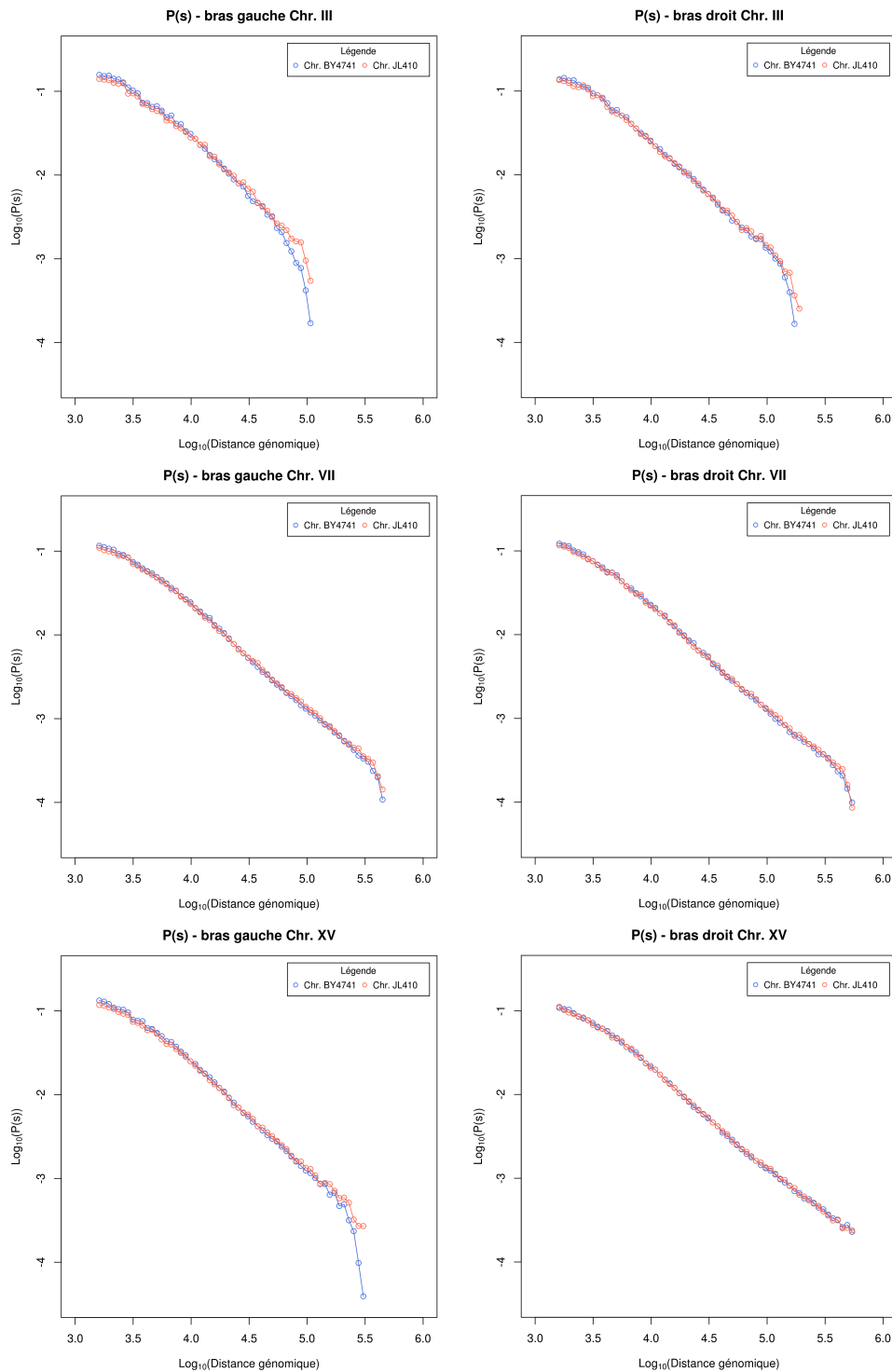
#### 4.4.5 Analyse de la réplication de la souche à trois chromosomes

Chez *S. cerevisiae*, la réplication est initiée de manière stochastique au niveau des séquences de réplication autonomes (ARS ; Brewer et Fangman, 1987), certaines d'entre elles commencent à se répliquer plus tôt que d'autres. La distribution de ces ARS est très inégale, avec un enrichissement des ARS initiant la réplication au niveau des régions centromérique, contrairement aux régions

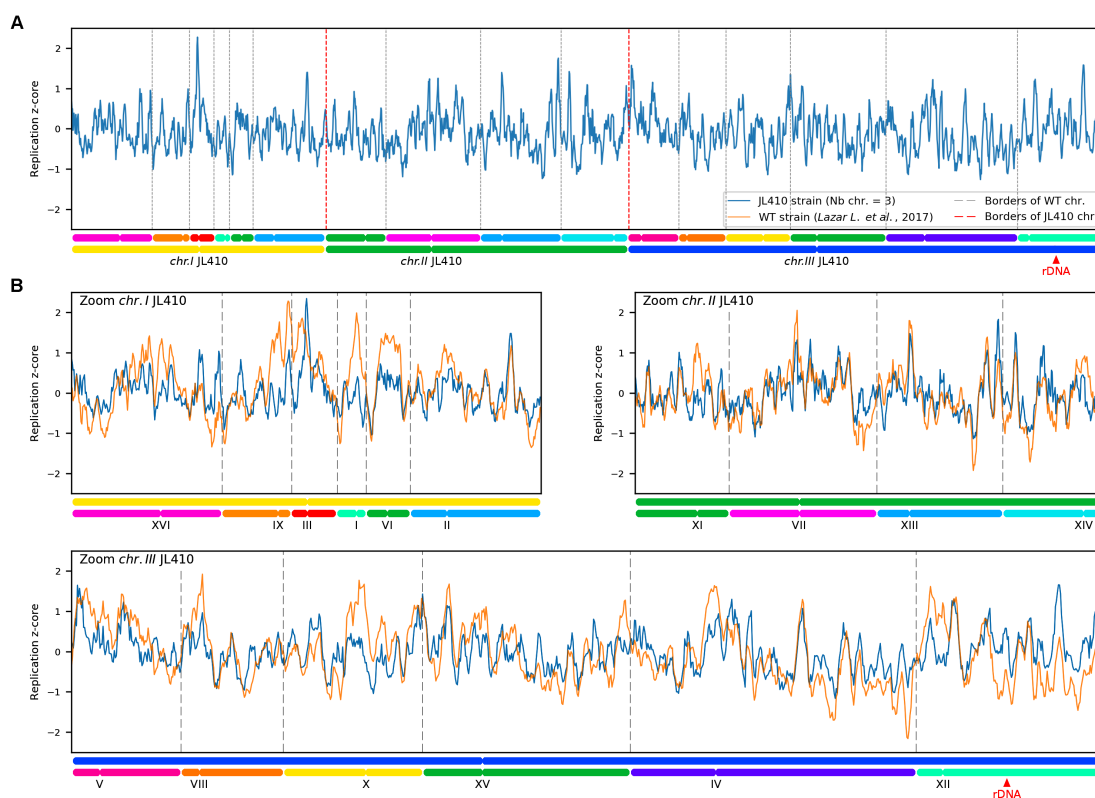




**FIGURE 4.8 Organisation de la souche à trois chromosomes :** (A) Matrice de contacts normalisées (Bins = 5 kpb) de la souche JL410 synchronisée en G1. La position des chromosomes est indiquée sur le côté de la matrice. La flèche bleue indique le rapprochement des cassettes HML et HMR l'une de l'autre. La flèche verte indique le rapprochement de la cassette HML et de l'extrémité du chr. I. Les flèches noires indiquent le rapprochement des cassettes HML et HMR avec les télomères des autres chromosomes. (B) Structure 3D moyenne des génomes de la population de la souche JL410 asynchrone. (C) Structure 3D moyenne des génomes de la population de la souche JL410 synchronisée en G1.



**FIGURE 4.9** Comparaison de la loi de distance génomique entre les chromosomes natifs et fusionnés : La loi de distance génomique des chromosomes natifs est représentée en bleu, celle des chromosomes fusionnés en rouge.



**FIGURE 4.10 Profil de réplication des chromosomes fusionnés :** (A) Profil de réplication des chromosomes de la souche JL410. (B) Superposition des profils de réplication des chromosomes de la souche JL410 (bleu) et BY4742 (orange).

subtélomériques qui en sont appauvries (Raghuraman et al, 2001 ; McCune et al, 2008 ; Lazar-Stefanita et al, 2017). En calculant le z-score du ratio de la couverture des données de Hi-C obtenues entre la phase G1 et exponentielle<sup>4</sup>, nous pouvons déduire le profil de réplication des chromosomes fusionnés (Lazar-Stefanita et al, 2017 ; voir aussi la méthode). Lorsqu'une région est répliquée en première, son z-score augmente, inversement si elle est répliquée plus tardivement, il diminue.

D'après le profil de réplication de la souche JL410 (figure 4.10A), les régions centromériques des chromosomes fusionnés sont répliquées en premières, mais d'autres régions de ces chromosomes sont également répliquées simultanément, elles correspondent principalement aux régions centromériques des chromosomes natifs. Bien que les centromères aient été retirés, les ARS qui se répliquaient très tôt dans les chromosomes natifs sont toujours présentes et continuent à être répliquées précocement. Bien que ces régions centromériques natives fassent partie des régions répliquées en premier, elles semblent cependant un peu retardées par rapport au profil fusionné. Par ailleurs, certaines des régions télomériques natives, qui étaient répliquées tardivement, sont maintenant répliquées plus tôt dans les chromosomes fusionnés. Globalement, le profil de la réplication des chromosomes fusionnés est plus homogène que le profil des chromosomes natifs. Les différences entre les régions répliquées tôt ou tardivement y sont moins importantes que dans les chromosomes natifs. La nouvelle distribution des ARS et la grande taille de ces chromosomes sont peut-être les conséquences de cette homogénéisation de la réplication. Au lieu d'être concentrée au niveau des centromères, la réplication de ces grands chromosomes débute au niveau de plusieurs régions de ces chromosomes. Ces résultats sont intéressants car ils permettent de distinguer un potentiel effet « localisation » dans l'activation des origines de réplifications, qu'elles soient précoces ou tardives. Cependant, des répliquas expérimentaux sont nécessaires pour renforcer ces conclusions préliminaires.

#### 4.4.6 Conclusion

L'étude de la version intermédiaire du nouveau chromosome codant les ARNt de *S. cerevisiae* ne nous a pas apporté énormément d'informations supplémentaires sur l'organisation des ADNt dans le génome de la levure. Il nous apportera probablement plus d'informations une fois sa synthèse terminée et quand il sera maintenu dans une souche contenant plusieurs chromosomes synthétiques. Les souches contenant les chromosomes fusionnés se sont, quant à elles, révélées très informatives. Bien que le nombre de centromères et de télomères soit considérablement réduit et que la taille de ces chromosomes soit amplement agrandie, le génome de levure conserve son organisation de pseudo-Rabl et les différentes structures fonctionnelles, telles que le recrutement des cassettes HML et HMR au niveau des régions subtélomériques ou l'isolement du nucléole, sont conservées.

---

4. Le z-score est une méthode permettant de standardiser un signal en fonction de sa moyenne et de sa déviation standard.

# **Design d'un chromosome synthétique pour améliorer la qualité des données de 3C**

## **Redesigning chromosomes for optimized Hi-C assay provides insights on loop formation and homologs pairing during meiosis**

Muller Héloïse<sup>1,2\*</sup>, Scolari F. Vittore<sup>1,2\*</sup>, Mercy Guillaume<sup>1,2</sup>, Agier Nicolas<sup>3,4</sup>, Aurèle Piazza<sup>1,2</sup>, Lazar-Stefanita Luciana<sup>1,2</sup>, Descorps-Declere Stephane<sup>5</sup>, Espeli Olivier<sup>6</sup>, Llorente Bertrand<sup>7</sup>, Fischer Gilles<sup>3,4</sup>, Mozziconacci Julien<sup>8§</sup>, and Koszul Romain<sup>1,2§</sup>

<sup>1</sup> Institut Pasteur, Department Genomes and Genetics, Groupe Régulation Spatiale des Génomes, 75015 Paris, France

<sup>2</sup> CNRS, UMR 3525, 75015 Paris, France

<sup>3</sup> Sorbonne Universités, UPMC Univ. Paris 06, Institut de Biologie Paris-Seine UMR 7238, Biologie Computationnelle et Quantitative, F-75005, Paris, France

<sup>4</sup> CNRS, Institut de Biologie Paris-Seine UMR7238, Biologie Computationnelle et Quantitative, F-75005, Paris, France

<sup>5</sup> Institut Pasteur, Center of Bioinformatics, Biostatistics and Integrative Biology (C3BI), Paris, F-75015, France

<sup>6</sup> Centre Interdisciplinaire de Recherche en Biologie, Collège de France, UMR-CNRS 7241, INSERM U1050, Paris, France.

<sup>7</sup> Cancer Research Center of Marseille, CNRS UMR7258, Inserm U1068, Institut Paoli-Calmettes, Aix-Marseille Université UM105, 13273 Marseille, France

<sup>8</sup> Sorbonne Universités, Theoretical Physics for condensed matter lab, CNRS UMR 7600, UPMC University Paris 06, 75005 Paris, France

Corresponding authors:

Tel: +33 1 40 61 33 25; E-mail: [romain.koszul@pasteur.fr](mailto:romain.koszul@pasteur.fr)

Tel: +33 1 44 27 45 40; E-mail: [mozziconacci@lptmc.jussieu.fr](mailto:mozziconacci@lptmc.jussieu.fr)

Keywords: Rec8, cohesin, synthetic yeast,

## Abstract

In all chromosome conformation capture based experiments the accuracy with which contacts are detected varies considerably because of the uneven distribution of restriction sites along genomes. In addition, repeated sequences as well as homologous, large identical regions remain invisible to the assay because of the ambiguities they introduce during the alignment of the sequencing reads along the genome. As a result, the investigation of homologs during meiosis prophase through 3C studies has been limited. Here, we redesigned and reassembled in yeast a 145kb region with regularly spaced restriction sites for various enzymes. Thanks to this Syn-3C design, we enhanced the signal to noise ratio and improved the visibility of the entire region. We also improved our understanding of Hi-C data and definition of resolution. The redesigned sequence is now distinguishable from its native homologous counterpart in an isogenic diploid strain. As a proof of principle, we track the establishment of homolog pairing during meiotic prophase in a synchronized population. This provides new insights on the individualization and pairing of homologs, as well as on their internal restructuration into arrays of loops during meiosis prophase. Overall, we show the interest of redesigned genomic regions to explore complex biological questions otherwise difficult to address.

## Introduction

Genomic derivatives of the capture of chromosome conformation assay (3C, Hi-C, Capture-C)(Lieberman-Aiden *et al*, 2009; Dekker *et al*, 2002; Hughes *et al*, 2014) are widely applied to decipher the average intra- and inter-chromosomal organization of eukaryotes and prokaryotes (Sexton *et al*, 2012; Le *et al*, 2013; Dekker *et al*, 2013; Marbouty *et al*, 2014a). Formaldehyde cross-linking followed by segmentation of the genome by a restriction enzyme (RE) are the first steps of the experimental protocol. The basic unit of “C” experiments therefore consists of restriction fragments (RFs) that are subsequently religated and captured to identify long range contacts. The best resolution that can be obtained is directly imposed by the positions of the RE sites along the genome. Both 6-cutter and 4-cutter REs have been used (Marie-Nelly *et al*, 2014; Sexton *et al*, 2012; Rao *et al*, 2014; Le *et al*, 2013), the latter with the expectation that the resolution increases with the number of sites. However, this approach

suffers from two major caveats. First, restriction sites (RSs) are not regularly spaced along genomes. The distribution of RFs lengths follows a geometric distribution, with important variations along the genome that depend on the local GC content and the specific sequence recognized by the RE. Given that the likelihood for a RF to be crosslinked by formaldehyde during the first step in the procedure depends on its length (Cournac *et al*, 2012), the probability to detect a given fragment in any 3C experiment will in turn be strongly affected by this parameter (**Fig 1A**). Normalization procedures have been developed in order to correct the signal (Cournac *et al*, 2012; Imakaev *et al*, 2012) but these methods involve filtering out fragments with unusually low or high signal and aggregating the contact data over several consecutive fragments in longer bins of fixed genomic length, at the expense of actual resolution (Lajoie *et al*, 2015). Overall, the definition of Hi-C resolution has remained empiric, because of the lack of a control sequence where RF biases would be alleviated. The second limitation reflects the fact that repetitive sequences cannot be tracked because the sequencing reads obtained from a Hi-C experiment cannot be mapped unambiguously along the genome, alleviating the possibility to track homologous chromosomes in isogenic backgrounds.

One consequence of these limitations has been the absence of in-depth studies of meiotic prophase through Hi-C approach. Meiosis is the cell division where a diploid cell gives rise to four haploid gametes through two rounds of chromosome segregation with no replication in-between them. The prophase of the first division, where the homologous paternal and maternal chromosomes segregate, comprises a series of regulated events involving the recognition and pairing of homologs all along their length. Homologs, to become physically connected, must meet each other's in space, which implies a dynamic reorganization of the overall genome and disentanglement between paired DNA molecules (Zickler & Kleckner, 2016). During the meiotic program shared by budding yeast and mammals (i.e. the succession of events mediated by evolutionary conserved molecular complexes), pairing occurs early on during the leptotene stage. This process can be accompanied and/or facilitated by dynamic movements of chromosomes, and lead to telomere clustering at the zygotene stage (bouquet stage; (Zickler & Kleckner, 2016) or other forms of movements mediated by chromosome ends directed by cytoskeletal components through direct association across the nuclear envelop (Koszul &



Kleckner, 2009). These events have mostly been described using imaging techniques, or, when it comes to the analysis of the underlying molecular events, through site-specific assays. Dekker and coworkers pioneered the analysis of meiosis in the original chromosome conformation capture study (Dekker *et al*, 2002). Using restriction polymorphism to distinguish the maternal and paternal versions of a locus along chromosome III, they notably showed that 3C was able to capture homolog pairing, as well as centromere declustering. However, the higher-order organization surrounding the recombining locus remained unexplored. Similarly, the influence on homolog pairing of the vigorous movements mediated by chromosome ends remain unexplored using Hi-C (Koszul & Kleckner, 2009).

In order to investigate the behavior of two homologous chromosomes sharing the same nuclear space, we designed and assembled a dedicated “synthetic” genomic region (Koszul, 2016), also aimed at increasing the resolution of 3C-based experiments. As a proof of concept of this strategy, we describe here a redesigned ~150kb region (called synIV-3C) of *Saccharomyces cerevisiae* yeast chromosome 4. We then investigated its behavior during the first stages of meiotic prophase.

## **Results and Discussion**

### **Design and assembly of the Syn-3C region**

Designer chromosome synIV-3C closely resembles the native chromosome with respect to genetic elements (see Material and Methods), but was “designed” to yield high resolution and high visibility in 3C experiments by providing nearly equally spaced restriction sites. The RSs of four different enzymes were removed from the native sequence with point mutations and subsequently reintroduced within the sequence at regularly spaced positions (400bp, 1,500bp, 2,000bp and 6,000bp for DpnII, XbaI, HindIII and NdeI, respectively; **Fig 1B** and **Fig S2**; **Table 1**). As shown on **Fig 1C**, the DpnII and HindIII RFs sizes in the redesigned synIV-3C region are normally distributed when compared to the skewed, native genome-wide distributions. Besides providing a way to increase the resolution of the 3C experiment, the design can also be used to focus on specific contacts, for instance between promoters and terminators (**Fig S2**). When

possible, coding sequences were targeted preferentially and modified using synonymous mutations (**Fig S1**). We identified a 150kb window on chromosome 4 for which the uniformity of RFs lengths was maximized while the number of potentially deleterious base changes was minimized (the final choice for the region can also take into account sequence annotation and be guided by specific interests of the end-user). From this design, DNA building blocks were purchased and assembled in yeast BY (S288C) and SK1 background strains as described (Annaluru *et al*, 2014; Muller *et al*, 2012) (Material and Methods). Sequencing confirmed that 144kb within the targeted region were replaced by the redesigned sequence and that 100% of the mutations were introduced at the correct positions corresponding to a total of ~2% divergence with the reference genomes (3,229 bp out of 150,000). Analysis of the growth profile did not reveal significant negative effects of the modifications introduced in the SynIV-3C region compared to the isogenic parental strains (**Fig S3**).

### ***Cis*-contact pattern of the SynIV-3C region**

To assess for the quality improvement of Hi-C data in the Syn-3C region, Hi-C experiments were performed in parallel on BY strain carrying the synIV-3C redesigned chromosome as well as on the native parental strain using DpnII and HindIII (Material and Methods). The raw DpnII contact map of chromosome 4 exhibited a remarkably “smooth” pattern within the redesigned region compared to the native flanking regions (**Fig 1D**). The read coverage over the region also exhibits a dramatic and compelling change, with a more homogeneous and regular distribution in the synthetic regions for both enzymes compared to a heterogeneous distribution in the native sequence (**Fig. 2A, B**). Interestingly, careful examination of this distribution indicates that besides its own length, the capture frequency of a given fragment is also influenced by the length of its neighbors, resulting in a bias. To quantify the improvement in the SynIV-3C region we compared the signal with the signal over the same region obtained in the WT strain using the same number of aligned read pairs and identical bins of various sizes (**Fig. 2C, D**). At the smallest resolution tested (600bp for DpnII and 2,400bp for HindIII) the WT contact map exhibited numerous blind regions with no detectable contacts (empty bins), in sharp contrast with its synthetic counterpart (**Fig. 2C, D**). When fragments were aggregated in bins of increasing sizes

(hence, resulting in a loss of resolution) these blind regions gradually disappear, although the heterogeneity of the data remains consistently higher in the WT compare to synIV-3C strain, as shown by the increased span of the color-scales of the WT maps.

In order to further quantify this heterogeneity, we computed the cumulative distributions of the number of contacts between bins separated by a given genomic distance  $s$  (bp) in the synIV-3C region and in its native counterpart for DpnII and HindIII (**Figs 2C and 2D**, respectively). The redesigned region systematically exhibited more homogeneous contacts counts and narrower distributions than the WT region, both at short ( $s = 2 \times$  bins sizes; **Figs 2C and D** middle panels) and longer distances (Material and Methods and **Fig. S5, S6**). Some of the bins in the native region remain almost invisible to the assay as a result of the heterogeneity in RF distribution (blue squares on **Figs 2C and D** middle panels). We computed the coefficient of variation CV (i.e. standard deviation /mean) of these distributions for multiple values of  $s$ . We use this value as an indication of the signal to noise ratio (**Figs 2C and D** right panels). Interestingly, we found that even for large bins, the CV is significantly and consistently smaller in the synthetic region, again indicating improved resolution. These results also clearly illustrate the advantage of using a frequent cutter (DpnII vs. HindIII) restriction enzyme with respect to resolution since the distribution of contact counts between bins remains much more spread with HindIII than with DpnII, even for native sequences (**Fig 2B**).

### Statistical analysis of Hi-C contact data

The sequencing step of a Hi-C library corresponds to the random sampling of all ligation events generated during the experiment. The outcome of this draw, for a given genomic distance  $s \pm \Delta s$  between pairs of loci, is expected to follow a Poisson distribution if we suppose that each pair of loci has a strictly equal probability  $p$  to be drawn. For real data, this cannot be the case since other factors (such as for instance the restriction fragment size, or the average 3D folding of chromatin) can influence this probability. When the number of events becomes large enough though, the differences between the re-ligation probabilities of different loci will start to kick in and the distribution of contacts should switch to a Gaussian behavior, assuming that these probabilities follow a Gaussian distribution. The standard deviations ( $\sigma$ ) of those two

distributions (Poisson and Gaussian) scale either as the square root of the mean ( $\mu$ ) or as the mean, respectively. We took advantage of the fact that the overall contact number  $\mu$  decreases with increasing genomic distances ( $s$ ) to check whether these relations between  $\mu$  and  $\sigma$  holds for real data (**Fig. 3A**).

To start our analysis with the highest contact numbers we aggregated the results from 12 Hi-C experiments performed in G1 or early S phase (from Lazar-Stefanita *et al.*, 2017). In the following, we focused our analysis on the Syn3C region. For five different distances (**Fig. 3A**;  $1,400 \pm 350$  bp in red;  $4,900 \pm 350$  bp in green;  $9,100 \pm 350$  bp in purple;  $1,330 \pm 350$  bp in cyan and  $17,500 \pm 350$  bp in yellow), the five contact distributions computed from each restriction fragments pairs were re-scaled by their mean and superposed on the same plot (**Fig. 3A**, inset). The collapse of these re-scaled distributions clearly indicates a Gaussian behavior (i.e., the mean scales as the standard deviation). We next explored the relation between  $\mu$  and  $s$  over a wider range of values. We used data for increasing values of  $s$ , corresponding to lower  $\mu$ , as well as data aggregated over multiple experiments and data from capture experiments, corresponding to higher values of  $\mu$  (**Fig. 3B**). When plotting the values of  $s$  for different  $\mu$ , we found that both in the native and in the Syn-3C sequence context there is a crossover from the Poisson to the Gaussian distribution (indicated by the red and blue lines, respectively) and that the standard deviation for the Syn3C experiment is lower than in the WT counterpart for all the values of  $\mu$ , as expected from the analysis done in **Fig. 2**. Interestingly, the two transition points can tell us about the importance of the bias of having uneven restriction fragments compared to other biases and/or biologically relevant variations. In the WT case, as soon as each pair of fragments receive one read, the distribution of counts switches to Gaussian, indicating that each pair has already been sampled unevenly. In the case of the Syn3C construct, where this bias is absent, one needs to aggregate 10 reads per fragment pairs to start to see variations among fragment pair re-ligation frequencies and switch to the Gaussian behavior. This highlights the strong effect of fragments length biases and justifies the use of large bins which will encompass many fragments as well as normalization procedures in all Hi-C experiments. For any genomic distance  $s$ , the value of  $\mu$  can arbitrarily be increased by increasing the bin size, or resolution.

The existence of the transition between Poisson and Gaussian behaviors thus enables us

to propose a rigorous way to determine the resolution of a Hi-C experiment by choosing a bin size which exactly corresponds to the transition point. It is worth underlying that according to this analysis the resolution of a Hi-C experiment can only be defined for a given genomic distance  $s$ .

### **Analysis of genome organization during meiosis prophase**

The SynIV-3C chromosome was designed and assembled with the aim to investigate the interplay between homologous chromosomes during a variety of DNA related metabolic processes, including the mitotic and meiotic cell cycle. A diploid SK1 strain carrying the SynIV-3C region on one homolog, its native counterpart on the other homolog, but isogenic for the rest of the genome, was processed into a synchronized meiotic culture (Material and Methods). The synchrony of meiotic progression was assessed by monitoring meiotic replication by FACS analysis and the two meiotic divisions by DAPI staining. Cells that have passed through anaphase I or anaphase II contain two or four DAPI-stained bodies, respectively. After 6h in sporulation medium, ~40% of the cells display two or more DAPI bodies. By 8h, 70% of the cells are passed anaphase II, showing that most cells have completed meiosis synchronously (**Fig. 4B**) (Hunter & Kleckner, 2001; Koszul *et al*, 2008). Hi-C contact maps were generated for cells sampled at 0, 3 and 4h from the synchronized culture, corresponding to WT mitotic, early zygotene and early pachytene cells (when a fraction of the population has passed anaphase I), respectively (**Fig. 4C**). The differences between mitotic and early pachytene (4h) cells were determined by computing the log-ratio between the contact maps (bin: 5 kb; **Fig. 4D**; Material and Methods). Under this representation, the color scale of the map reflects the variations in contact frequency for each bin between two different contact maps. Individual 2D maps can also be represented as 3D structures to facilitate their interpretation (**Fig. 4E**; Lesne *et al*, 2014; Lazar-Stefanita *et al*, 2017). These different representations illustrate and recapitulate known key features of meiotic prophase.

### *Declustering of centromeres*

A loss of inter-centromeric contacts was readily apparent during zygotene and pachytene

compared to the pre-replication stage (green arrowheads; **Fig. 4C, D, F**). No significant enrichment in contacts between telomeres was observed (purple arrowheads; **Fig. 4C, D, F**). This result reflects the rapid declustering of centromeres that accompanies entry into meiotic prophase also observed through microscopy and 3C (Dekker *et al*, 2002; Trelles-Sticken *et al*, 1999). On the other hand, the absence of inter-telomeric contacts doesn't immediately support the transient and small increase in telomere clustering described at this stage and corresponding to the bouquet stage ((Zickler & Kleckner, 1999; Trelles-Sticken *et al*, 1999). One possibility is that the subtelomeric sequences of the SK1 strain remain incomplete or incorrectly assembled and therefore escape monitoring using Hi-C. Alternatively, the transient increase in clustering is very rapid, or occurs a bit earlier, and was missed during this experiment. More experiments are therefore necessary to fully assess whether or not budding yeast meiotic program really present a "bouquet" stage.

In addition to the loss of discrete centromeric contacts, the ratio map also revealed the abolition of the "polymer brush" effect, which insulates centromeres and their flanking chromosomal regions from the rest of the chromosome arms (green arrows on the magnification of chromosome XV, **Fig. 4D**). As a result, the centromeric regions become almost indistinguishable from the rest of chromosome arms, and chromosomes appear as relatively individualized, homogenous entities. Once declustered, centromeres regions are more likely to contact other portions of the genome, resulting in the increased *trans* contact signal made by these regions with other chromosomes and visible on the ratio map (**Fig. 4D**). The transformation of chromosomes into well-individualized entities could be strikingly recapitulated by the 3D representation of the 2D maps (**Fig. 4E**) (Lesne *et al*, 2014).

### *Chromosome folding*

The ratio contact maps also display a strong increase in intra-chromosomal contacts at early-pachytene compared to mitotic cells, as reflected by a red and large diagonal that clearly appears when magnifying individual chromosomes (inset on **Fig. 4D**). The folding can also be assessed by computing the contact probability  $p$  as a function of genomic distance of all chromosome arms (Naumova *et al*, 2013; Lieberman-Aiden *et al*, 2009; Lazar-Stefanita *et al*,

2017). The  $p(s)$  curves were computed for two premeiotic replicates (0h), three mitotic G1 replicates (Lazar-Stefanita et al., 2017), zygotene (3h) and early-pachytene (4h) chromosomes. The two later curves display sharp differences compared to premeiotic and G1 mitotic cells. First, contacts frequencies increase between loci positioned 20 to 50 kb apart, with a peak around 50 kb (**Fig 4G**). Then, contacts sharply decrease, suggesting a general stiffening of the polymer associated with the loss of distant loci to contact each other's. This pattern suggests that chromosomes fold during meiotic prophase into a structure that favor contacts under a certain distance, while disfavoring local and long range interactions, and points at the formation of chromatin loops.

### **Visualization of meiotic loops using Hi-C**

After or during meiotic replication, the two sister chromatids of each homolog becomes organized as arrays of ~20 kb loops tethered to a protein structural axis composed (**Fig. 3A**) (Zickler & Kleckner, 2016; Blat *et al*, 2002; Zickler & Kleckner, 1999). The meiosis-specific cohesion subunit Rec8 is one of the component present along these axis (Klein *et al*, 1999). Rec8 is distributed into discrete domains corresponding to meiotic loops bases, and contributes to the establishment of cohesion between sister chromatids. These loops provide a highly-organized chromosome architecture context for the formation and resolution of meiotic DNA double-strand breaks generated during leptotene (Padmore *et al*, 1991; Blat *et al*, 2002; Kim *et al*, 2010; Sommermeyer *et al*, 2013; Acquaviva *et al*, 2013). It was proposed that the expansion state of the chromatin within the loops modulates the degree of compaction and stiffness of the chromosomes at the different stages of prophase (Kleckner *et al*, 2004; see also Koszul *et al*, 2008). The impact of loop formation on chromosomes had never been addressed through Hi-C, and we took advantage of this pilot study to have a glimpse on these processes.

First, we investigated whether Rec8-mediated loops were visible on the Hi-C contact maps of zygotene and early-pachytene cells by displaying the Rec8 binding regions along chromosome axis. Both at 3 and 4h, the short range intra-chromosomal contacts displayed a heterogeneous signal, with triangular shapes domains reminiscent of the topological domains observed along the mitotic chromosomes of other species appearing along chromosomes (**Fig.**

**5A)** (Dixon *et al*, 2012; Sexton *et al*, 2012; Nora *et al*, 2012). The boundaries between these domains correlate with Rec8 enriched regions (Ito *et al*, 2014; Glynn *et al*, 2004), suggesting that this signal may correspond to chromatid loops bridged at their bases to the chromosomal axis.

To further test whether these domains are indeed meiotic loops, we compared the aggregated intrachromosomal contacts made by Rec8-binding sites to random positions in mitotic (0 h) and early-pachytene cells (4h). The ratio between the cumulated contact maps made by 100kb windows (5 kb bins) centered on Rec8 binding sites and random maps display no significant enrichment in mitotic conditions (**Fig. 5B**, left column). At 4h into meiotic prophase, the same analysis display a strikingly different pattern, with the Rec8 bound sites now clearly delimiting two distinct domains (**Fig. 5B**, right column). In addition, these sites now display enriched contacts with DNA regions positioned on average 20 to 25 kb upstream and downstream the chromosome. This contact pattern is typical of loops bridging distant loci in *cis* (Rao *et al*, 2014; Lazar-Stefanita *et al*, 2017), and altogether these results demonstrate that Rec8 is positioned at the basis of chromatin loops during meiotic prophase. Therefore, Hi-C detects loop formation during meiotic prophase. The difference between zygotene and early-pachytene stages was not readily apparent from this work, although changes in chromosomes stiffness have been reported and proposed at these stages (Koszul *et al*, 2008; Kleckner *et al*, 2004). One possibility is that at 4h the amount of cells with fully complete synaptonemal complex axis (pachytene cells) remains limited. Another option is that the signal within the loops is so strong that variations over long distances between zygotene and pachytene remain negligible when assessed by Hi-C. More work on synchronous cultures of wild type and mutant cells should answer these questions and provide new insights on these mechanisms.

### **Homolog-homolog contacts using the SynIV-3C and native regions**

To characterize more specifically the inter-homolog contacts between the synthetic and native chromosome IV regions an enrichment step for this region was performed using a Capture-C strategy (Hughes *et al*, 2014). This led to a ~20 fold increase in reads from the SynIV-3C region and its native counterpart that were used to generate the contact map at 0, 3 and 4 hours (**Fig. 5C** right panels; see Supplementary methods for controls). To monitor whether



changes in contacts between the two homolog regions may result from their similar distance to the centromere, the contact pattern of two heterologous 150kb regions positioned at equal distances from the centromeres on arms of approximately similar sizes (XVR and XIIIIR) were generated (**Fig. 5C**, left panels). These heterologous regions became more insulated from each other's as cell enter meiotic prophase, reflecting the individualization of chromosomes and loss of Rabl organization (see also **Fig. 4D**). On the other hand, the SynIV-3C and native homologous regions, which are clearly distinguishable from each other's thanks to the SNP introduced in the synthetic design, display a contact frequency that increases over time, from 1.4% of reads before replication to 6.2% in early pachytene.

The two homologs do not appear to be closely and/or strongly juxtaposed, as shown by the weak diagonal that appears over time in zygotene and early pachytene cells. This observation strongly suggests that the synaptonemal complex plays a role not only in the maintenance of homolog structure by bridging chromosomal axis, but also in insulated the two homologs from each other's. To overcome this barrier, meiotic DNA double-strand breaks are indeed relocalized to the SC, where inter-homolog contact is promoted and eventually recombination can proceed (Sommermeier *et al*, 2013; Acquaviva *et al*, 2013). Because the SynIV-3C and native regions differ on average by 2%, we wondered whether this polymorphism could affect their folding and/or meiotic recombination processes. The computation of the p(s) for each of the two homologous region at the three timepoints reveal a similar trend than the one observed for the whole genome (**Fig. 5D**), showing that the design does not affect chromatin folding. To verify whether this design impairs meiotic recombination, we performed a restriction assay taking advantage of the restriction polymorphism introduced between the two regions (**Fig. 5E**). Crossing over were detected at the meiotic DSB hotspot tested within the region, suggesting that meiotic recombination can proceed in this genetic environment (**Fig. 5E**). More analyses remain to be performed to verify that this is the case over the entire region. Eventually, the synthetic design could be adapted to alleviate this concern, for instance by removing polymorphisms within 2 kb of DSB hotspots of interests. Nevertheless, this experiment shows that the present approach allows to track recombination events concomitantly to the higher-order architecture of the chromosomes.

## Conclusion

The yeast genome presents a relatively homogeneous GC content and a few repeated sequences. The gain in resolution achieved by redesigning RS along the genome should therefore be even higher in organisms with more heterogeneous genomic content and will enable unbiased tracking of entire regions that are otherwise inaccessible to the experiment. One could envision, for instance, assembling the redesigned chromosome in yeast (Benders *et al*, 2010), before replacing its native counterpart in the organisms of interest (such as a bacteria, or eventually on mammalian cells). Other advantages of the approach include the modularity of the assembly step (**Supplementary Note 2**), that allows the introduction of building blocks carrying genetic elements of interest within the redesigned region. For instance, one could introduce highly expressed promoters in the middle of “gene desert” areas, to investigate the effect of gene expression on the local chromatin structure. One can also “shuffle” some of these building blocks, to look at the influence of specific DNA binding proteins on the contact networks. This specific 3C-friendly design is the first time, to our knowledge, where a large (>100kb) region of chromosome is specifically redesigned and assembled for the purpose of improving an assay so that we can now address more precisely and accurately specific questions related to the biology of the cell. It paves the way to more studies exploiting the power of synthetic biology to boost, refine, and maybe reshape traditional molecular biology approaches through orthogonal ones.

Here, we take advantage of the design to track for the first time using Hi-C the large structural changes that affect chromosomes during meiosis prophase. We showed that Hi-C clearly points at the formation of loops, whose bases overlap with binding sites of the meiotic-specific cohesion subunit Rec8. The redesigned region now also allows to distinguish, to a large extent, the two homologs. Preliminary results suggest that the loops along each homolog are not closely juxtaposed in space, and that the synaptonemal complex constitutes a barrier in-between. This observation further supports the mechanical necessity to generate bridges between homologs for recombination to proceed.

## Material and Methods

### *Strains*

The SynIV-3C region was assembled in the s288C and SK1 genetic backgrounds by transformation of BY4742 and ORT4602 (Sollier *et al*, 2004) to generate strains RSG\_Y181 and RSG\_Y189, respectively (see above). The SK1 diploid heterozygous for the SynIV-3C region (RSG\_Y190) was obtained by crossing RSG\_Y189 and ORT4601 (Sollier *et al*, 2004). Genotypes are provided in **Table 2**.

### *Design principles of Syn-3C chromosomes.*

We aimed at modifying the native sequence of a budding yeast chromosome according to our design principles while introducing as little modifications as possible. Because we were planning on re-assembling only a 150kb window within the genome, we scanned through the overall sequence using a scoring quality function to look for the candidate regions qualifying as the ideal target, i.e. where our principles would introduce a minimal number of mutations. The starting material was the *S. cerevisiae* SK1 strain genome sequence and annotations (Liti *et al*, 2009) and a list of 9 restriction enzymes (EcoRI, HindIII, NdeI, PstI, SacI, SacII, SalI, XbaI, XhoI and DpnII). RE were selected based on their low cost and restriction efficiency. A genome index file was then computed, that contained the following information for each base pair:

- Whether it consists of a “forbidden mutation” site, defined by us as follow: i) start and stop codons of known ORFs, ii) regulatory transcription pre-initiation complexes binding regions identified through ChIP-Seq exo, encompassing TATA-box binding sites (Rhee & Pugh, 2012), iii) the consensus sequence of Autonomous Replicating Sequences (ARS), i.e. the core sequence within *S. cerevisiae* replication origins (list of ARS obtained from oridb (Siow *et al*, 2012), iv) intron borders, v) centromeres, vi) tRNA.
- Whether the position belongs to a restriction site.
- If it belongs to an intergenic or coding region, and in the latter case, the codon it belongs to and its position.

Sliding windows of 150 kb moving with 10kb steps were then generated over the entire genome. In parallel, we defined the restriction pattern we wanted to generate:

- Regularly spaced intervals for 400, 1,500, 2,000 and 6,000 bp
- Gene promoter/terminator (substitutions within a coding sequence strongly preferred)

For each window, we computed all possible changes to apply to the genome so that all combinations of five out of the eight chosen 6-cutter enzymes were repositioned to generate all expected new restriction patterns. For each combination of 5 enzymes, all sites were first removed from the genome before being reintroduced at ideal positions. A margin of error in the positioning of the “ideal” position was tolerated (10% of the window size) to maximize the probability of introducing only synonymous mutations within the coding sequence. Once a RS was positioned, the position of the adjacent RS was adjusted based on the newly positioned site so that overall, the distribution of RFs remains as close as possible to the theoretical distribution. Overall, for each enzyme, a quality score was computed for each window based on the difference between the expected distribution of the site, and the real distribution. For each combination of enzyme, a global score corresponding to the sum of the individual scores of each enzyme was computed (see **Fig S1** for schema).

Overall, we selected the 10 “best” windows located at least at 150 kb from either a centromere or a telomere. The quality score was weighted by the presence of “forbidden positions” within the window, for instance when a start codon overlaps a restriction site to be deleted. Finally, a manual curation, aiming at fixing potential conflicts (such as 2 RSs overlapping the same bases, or accidental re-creation of a RS of one enzyme when processing a second one), followed, and was performed on the genome windows presenting the best quality scores.

We chose the final window based also on our research interests, i.e. containing at least two early replicating replication origins (Siow *et al*, 2012; Raghuraman *et al*, 2001), and several hotspots of meiotic DNA double-strand breaks (Pan *et al*, 2011). We also attempted to avoid too many retrotransposable elements or other DNA repeats. The final window was positioned on chromosome IV::700,000-850,000, with restriction patterns as follow: DpnII ↔ 400 bp window; XbaI ↔ 1,500 bp window; HindIII ↔ 2,000 bp window; NdeI ↔ 6,000 bp window; HhaI ↔ promoter/terminator (see summary on **Fig EV2**). 1037 mutations were present in the sequence,

the vast majority corresponding to the modifications necessary to reorganize DpnII RS (**Table 1**).

Overall, 1037 mutations were introduced, corresponding to 0.7% divergence.

Other modifications were introduced into the sequence. First, PCRTags similar to those used in the Sc2.0 design(Annaluru *et al*, 2014; Dymond *et al*, 2011) specific to either the native or synthetic sequence were also introduced within the window. Performing PCR using these primers allow testing for the presence and absence of the synthetic and native sequence, respectively. PCRTags were manually curated to adapt them to the restriction design, and overall 59 PCR tags out of 154 needed to be modified accordingly.

Overall, a total of 3,229 bp were modified (~2% of the 150,000 bp window). 743 codons were modified, but no change in the sequences of the corresponding proteins were introduced.

Although we took great care in the design of the sequence and algorithm, our ongoing experiments nevertheless suggest this design is perfectible and could be simplified. First, windows of 400 and 1,000 bp are probably sufficient to assay the structure at a reasonable resolution. Also, SNPs have to be introduced in repeats or low complexity DNA to facilitate mapping of the reads. Finally, a possibility to find the best frame for each interval/pattern is to start with the positions of RS overlapping forbidden sites as seed.

#### *Assembly of the redesigned chromosome.*

The redesigned sequence was split into 52 fragments of ~3,000 bp (i.e., blocks), with 200 bp overlaps between them. In addition sequences corresponding either to the auxotrophic marker genes *URA3* or *LEU2* were added to blocks 20, 37, 52 (*URA3*) and blocks 11, 28, 47 (*LEU2*), followed by 200 bp sequences of the WT neighboring chromosomal region. The replacement of the native sequence of strain BY4742 with the redesigned blocks was performed through a succession of six transformations, up to 11 blocks at a time (Muller *et al*, 2012).

After each transformation, independent colonies were sampled and PCRs performed at the PCR tags positions to identify the transformants that have replaced all of the native sequence with the redesigned one (**Fig. EV3**). Upon the last transformation, the selected transformant genome was sequenced and the region 707,556-852,114 (144,558 bp) was found to be replaced by the synthetic blocks.

### *Growth rate analysis*

Growth assays were performed to see if the transformants exhibited changes in fitness. Little to no growth defect could be identified when blocks 1 to 47 replaced the native sequence. The final transformation with blocks 48 till 52 led repeatedly to the recovery of transformants exhibiting a slow-growth, petite phenotype (Slonimski, 1949), reflecting a block in the aerobic respiratory chain pathway and a decrease in ATP. Since the region concerned by the 6<sup>th</sup> transformation only involved a few kb, we decided to keep the first 145 kb already successfully reassembled and discard this last step. However, crossing this petite strain with a WT strain gave diploids without growth defects. Sporulation of these diploids gave offsprings with growth rates also similar to WT, suggesting stable complementation of mitochondrial genomes.

### *Pre-growth and sporulation of yeast*

Pre-growth and sporulation of the RSG\_Y190 strain was carried out as described (Oh *et al*, 2009). Briefly, cells patched on YPG-Agar plates (1% Yeast Extract, 2% Peptone, 1.5% Agar, 2% Glycerol) from -80°C stocks were streaked on YPD plates (1% Yeast Extract, 2% Peptone, 1.5% Agar, 2% D-glucose, 0.004% Adenine). A single colony was used to inoculate 5 mL YPD liquid culture and grown at 30°C up to saturation. The saturated culture was used to inoculate 350 mL of a freshly made (less than 48 hrs) pre-sporulation media (SPS; 0.5% Yeast extract, 1% Peptone, 1% Potassium Acetate, 1% Ammonium Sulfate, 0.5% Potassium Hydrogen Pthalate, 0.17% Yeast Nitrogen Base lacking all amino-acids, two drops of anti-foaming agent) and grown with robust agitation (320 rpm) in 5L baffled flasks at 30°C. Cells were washed, transferred in 500 mL of sporulation media (SPM; 1% Potassium Acetate, 0.2X of Uracil, Arginine and Leucine, two drops of anti-foaming agent) and put back with robust agitation at 30°C.

### *RNA Isolation from Yeast for RNA Sequencing*

Three independent RNA-seq libraries were generated for BY4742, SK1 and Syn3C strains. For each library, a single colony was grown in a 2 mL YPD culture overnight at 30°C. The next morning, 10 mL cultures in YPD were started from 10<sup>6</sup> cells/mL until they reached 2.10<sup>7</sup> cell/mL. The cells were pelleted by spinning at 5000 rpm at 4°C for 5 min. The pellet was resuspended in

0.5 mL of Tris-HCl (10 mM, pH 7,5) and transferred to a microfuge tube. The cells were pelleted again by spinning briefly and discarding the supernatant. The cells were resuspended in 400  $\mu$ L RNA TES buffer (10 mM Tris-HCl, pH 7.5 10 mM EDTA 0.5% SDS). 400  $\mu$ L of acid phenol/chloroform was then added to the cells and vortexed for 1 min, and heated at 65°C for 30 minutes, briefly vortexing some time to time. The cells were placed on ice for 5 min and centrifuge at 13000 rpm at 4°C for 5 min. The aqueous layer (~400  $\mu$ L) was then transferred into another microfuge tube; an equal amount of phenol/chloroform acid was added a second time, mixed well and centrifuged at 13000 rpm at 4°C for 5 min. The RNA (~400  $\mu$ L) was precipitated by adding 40  $\mu$ L of sodium acetate (3 M) and 1,1 mL of absolute ethanol and incubating the tube at -80°C for a least 30 min. The RNA was pelleted by centrifuging at 13000 rpm at 4°C for 20 min. The RNA pellet was then washed with 500  $\mu$ L of 70% ethanol, air-dried and then resuspended in 50  $\mu$ L of water. 15  $\mu$ g were treated with 2U of DNase TURBO (Invitrogen) and cleaned up by phenol extraction and ethanol precipitation before being prepared for sequencing.

#### *RNA-Seq Analysis of synIII*

Single-end non-strand-specific RNA-seq of the Syn3C and BY4742 were performed using Illumina Nextseq and standard TruSeq preparations kits, after depletion of ribosomal RNA. Reads were mapped using Bowtie2 to the reference *S. cerevisiae* BY4742 and Syn3C genome. For each gene, reads were counted if mapping quality was lower than 30 and analyzed for differential expression using DESeq2, with standard parameters.

#### *Hi-C experiments*

*S. cerevisiae* G1 daughter cells of the redesigned strain were recovered from an exponentially growing population through an elutriation procedure (Marbouty *et al*, 2014). Hi-C libraries were generated as described (Lazar-Stefanita *et al*, 2017; Dekker *et al*, 2002). G1 daughter cells were cross-linked for 20 minutes with fresh formaldehyde (3% final concentration). To generate the libraries with different restriction enzymes, aliquots of  $3 \times 10^9$  cells were resuspended in 10 ml sorbitol 1M and incubated 30 minutes with DTT 5mM and Zymolyase 100T ( $C_{\text{Final}}=1$  mg/ml) to

digest the cell wall. Spheroplasts were then washed first with 5 ml of sorbitol 1M, then with 5 ml of 1X restriction buffer (depending on the restriction enzyme used). The spheroplasts were then resuspended either in 3.5 ml of the corresponding restriction buffer (NEB). For each aliquot/experiment, the cells were then split into three tubes (V=500 $\mu$ L) and incubated in SDS (3%) for 20 minutes at 65°C.

Crosslinked DNA was digested at 37°C overnight with 15 units of the appropriate restriction enzyme (NEB, DpnII, HindIII or NdeI). The digestion mix was then centrifuged for 20 minutes at 18000 g and the supernatant discarded. The pellets were then resuspended and pooled into 400  $\mu$ L of cold water. Depending on the sequence of the restriction site overhangs, the extremities of the fragments were repaired in the presence of either 14-dCTP biotin or 14-dATP biotin (Invitrogen). Biotinylated DNA molecules were then incubated 4 hours at 16°C in presence of 250 U of T4 DNA ligase (Thermo Scientific, 12.5 ml final volume). DNA purification was achieved through an overnight incubation at 65°C in presence of 250 $\mu$ g/ml proteinase K in 6.2mM EDTA followed by precipitation step in presence of RNase.

The resulting 3C libraries were sheared and processed into Illumina libraries using custom-made versions of the Illumina PE adapters (Paired-End DNA sample Prep Kit – Illumina – PE-930-1001). Fragments of sizes between 400 and 800 bp were purified using a PippinPrep apparatus (SAGE Science), PCR amplified, and paired-end (PE) sequenced on an Illumina platform (HiSeq2000; 2 x 75 bp).

The accession number for the data reported in this paper is [Database]: [xxxx] (under completion).

### *Processing of the reads and contact maps generations*

The raw data from each 3C experiment was processed as follow: first, PCR duplicates were collapsed using the 6 Ns present on each of the custom-made adapter and trimmed. Reads were then aligned using Bowtie 2 in its most sensitive mode against *S. cerevisiae* reference genome (native genome) or against the *S. cerevisiae* reference adapted for the Syn-3C region on chromosome 4 (SynIV-3C genome). An iterative alignment procedure was used: for each read the length of the sequence mapped increases gradually from 20 bp until the mapping became



unambiguous (mapping quality > 30). Paired reads were aligned independently and each mapped read was assigned to a restriction fragment. Religation events have been filtered out through the information about the orientation of the sequences as described in (Cournac *et al*, 2012). The distribution of the reads along the synthetic region and its native counterpart is represented in **Fig 2**.

Contact matrices were built for the wild type and the mutant by binning the aligned reads into units of single restriction fragments. DpnII and HindIII contact maps for the SynIV-3C region and its native counterpart were randomly resampled in order to present the same number of contacts. The raw contact maps were then subsequently binned into units (i.e. bins) of 600, 1,200, 2,400, 4,800 and 9,600 base pairs. Contact maps were generated using the *levelplot* function of the R *lattice* package. Matrices for the synthetic region were subsequently obtained by extracting the diagonal blocks for bins falling in the 719,756bp to 849,206bp interval. Outliers have been removed from the matrices if the number of the contacts surpassed by 20 times the top 5% threshold of the number of contacts between restriction fragment pairs.

### *Statistical analysis*

The CV is defined as the ratio between the standard deviation and the mean of the contact histograms at fixed distance  $s$ ; to take into account the finite-size effect, we discarded bins at the edge of the contact matrix in order to keep the statistics (number of bins) for different values of  $s$  constant, up to  $s < 15,000$  bp in DpnII and  $s < 70,000$  bp in HindIII datasets. To show that the improvement is specific to the new restriction pattern and is unlikely to be found spontaneously within the genome, we compared the SynIV-3C results with seven regions of similar size along chromosome 4 (460,856-590,306 bp; 590,306-719,756 bp; 849,206-978,656 bp; 978,656-1,108,106 bp; 1,108,106-1,237,556 bp; 1,237,556-1,367,006 bp; 1,367,006-1,496,456 bp). The quality improvement was assessed by computing the logarithm of the ratio of the CVs of the SynIV-3C and native region (**Fig EV5**).

### *Southern blot analysis of crossing over formation at the CCT6 hotspot*

The *CCT6* locus was chosen because it is the strongest Spo11-induced DSB hotspot in the synthetic region (Pan *et al*, 2011). Cells were harvested and DNA was extracted as described (Oh *et al*, 2009), except that no crosslinking step was performed. The DNA was digested with NdeI and XbaI (New England Biolabs), migrated on a 1% UltraPure Agarose (Invitrogen) 1X TAE for 15 hours at 70 V, and capillary transferred onto a Hybond-XL membrane (GE Healthcare) following the manufacturer instructions. Southern blot hybridization was performed at 65°C in Church buffer (1% BSA, 0.25 M Na<sub>2</sub>HPO<sub>4</sub> pH 7.3, 7% SDS, 1 mM EDTA) with a 1,104 bp-long radiolabeled probe corresponding to the rightmost region common to both the native and SynIV-3C restriction fragments (obtained from SK1 genomic DNA with primers 5'-TGGTGAAGAACTCAGGATTC-3' and 5'-CAGTTACAATGAAGTCCAGG-3') and radiolabeled phage lambda DNA (molecular ladder). Radiolabeling was performed with P<sup>32</sup>-αdCTP with the High Prime labeling kit (Roche) following the manufacturer instructions. The membrane was washed, exposed O/N, and the storage Phosphor Screen (GE Healthcare) was scanned on a Typhoon phosphorimager (Molecular Dynamics). The length of the native and SynIV-3C parental fragments are 5,135 bp and 6,157 bp, respectively. CO formation generates two recombinants of 4,453 bp and 6,839 bp. Quantifications were performed under ImageQuant 5.2 (Molecular Dynamics).

## **Acknowledgements**

We thank Jef Boeke for the Sc2.0 PCRTags sequences for chromosome 4 and for fruitful discussions and comments on the manuscript. We thank Elodie Pirayre and Ivan Moszer for contributing to the initial steps of the design of the algorithm, and Nancy Kleckner, Gianni Liti, Axel Cournac, Martial Marbouty, Luciana Lazar-Stefanita for discussions. Vittore Scolari and Héloïse Muller are recipients of Pasteur-Roux-Cantarini fellowships. This research was supported by funding to R.K. from the European Research Council under the 7th Framework Program (FP7/2007-2013, ERC grant agreement 260822), from ERASynBio and Agence Nationale pour la Recherche (IESY ANR-14-SYNB-0001-03), and to R.K. O.E. and B.L. from Agence Nationale pour la Recherche (MeioRec ANR-13-BSV6-0012).

## Author contributions

HM, SDD, BL, OE, NA, GF and RK designed the sequence. HM assembled the chromosome and performed the Hi-C experiments with GM. AP analyzed meiotic recombination. VS and JM analyzed the data, with contribution from LLS. RK conceived the study and wrote the manuscript, with contributions from JM, VS, HM, AP, and BL.

## Conflict of interest

The authors declare no competing financial interests.

## References

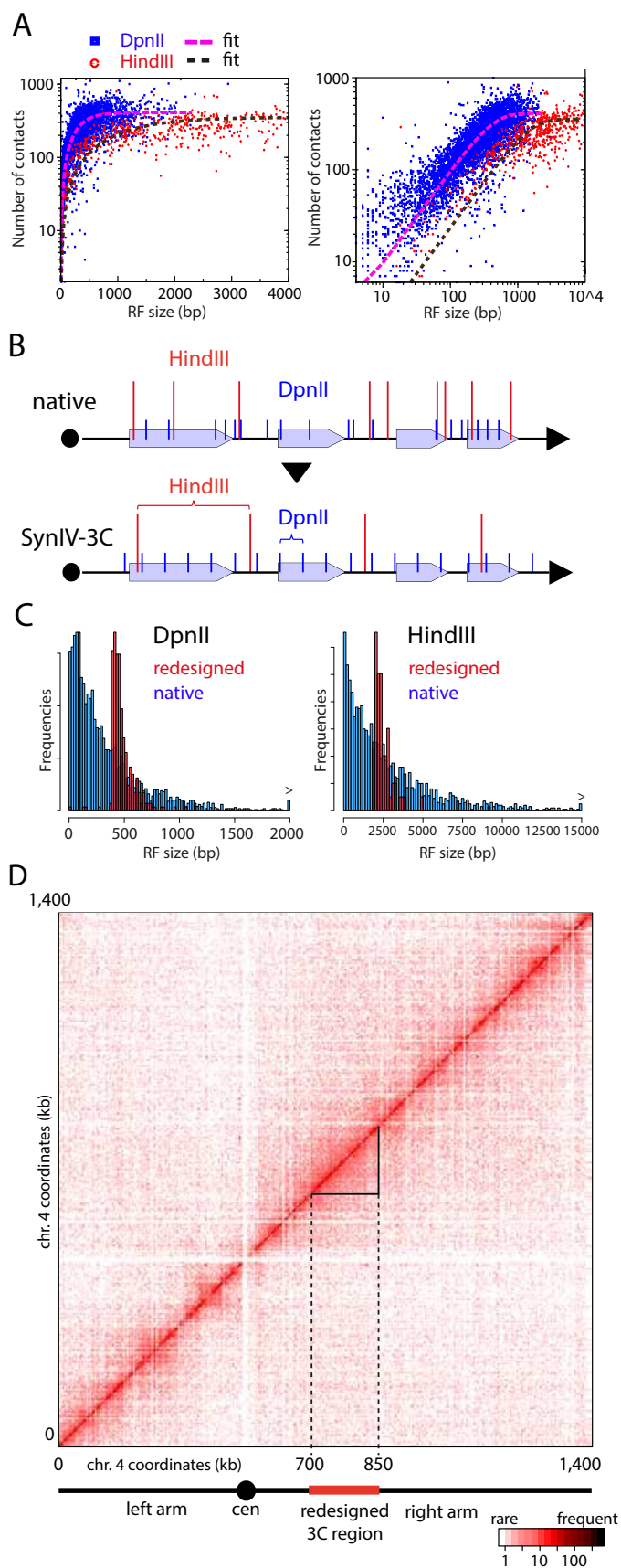
- Acquaviva L, Székvölgyi L, Dichtl B, Dichtl BS, de La Roche Saint André C, Nicolas A & Géli V (2013) The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science* **339**: 215–218
- Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, Cai Y, Zeller K, Agmon N, Han JS, Hadjithomas M, Tullman J, Caravelli K, Cirelli K, Guo Z, London V, et al (2014) Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science* **344**: 55–58
- Benders GA, Noskov VN, Denisova EA, Lartigue C, Gibson DG, Assad-Garcia N, Chuang R-Y, Carrera W, Moodie M, Algire MA, Phan Q, Alperovich N, Vashee S, Merryman C, Venter JC, Smith HO, Glass JI & Hutchison CA (2010) Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.* **38**: 2558–2569
- Blat Y, Protacio RU, Hunter N & Kleckner N (2002) Physical and Functional Interactions among Basic Chromosome Organizational Features Govern Early Steps of Meiotic Chiasma Formation. *Cell* **111**: 791–802
- Cournac A, Marie-Nelly H, Marbouty M, Koszul R & Mozziconacci J (2012) Normalization of a chromosomal contact map. *BMC Genomics* **13**: 436
- Dekker J, Rippe K, Dekker M & Kleckner N (2002) Capturing chromosome conformation. *Science* **295**: 1306–1311
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS & Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–380
- Dymond JS, Richardson SM, Coombes CE, Babatz T, Müller H, Annaluru N, Blake WJ, Schwerzmann JW, Dai J, Lindstrom DL, Boeke AC, Gottschling DE, Chandrasegaran S,

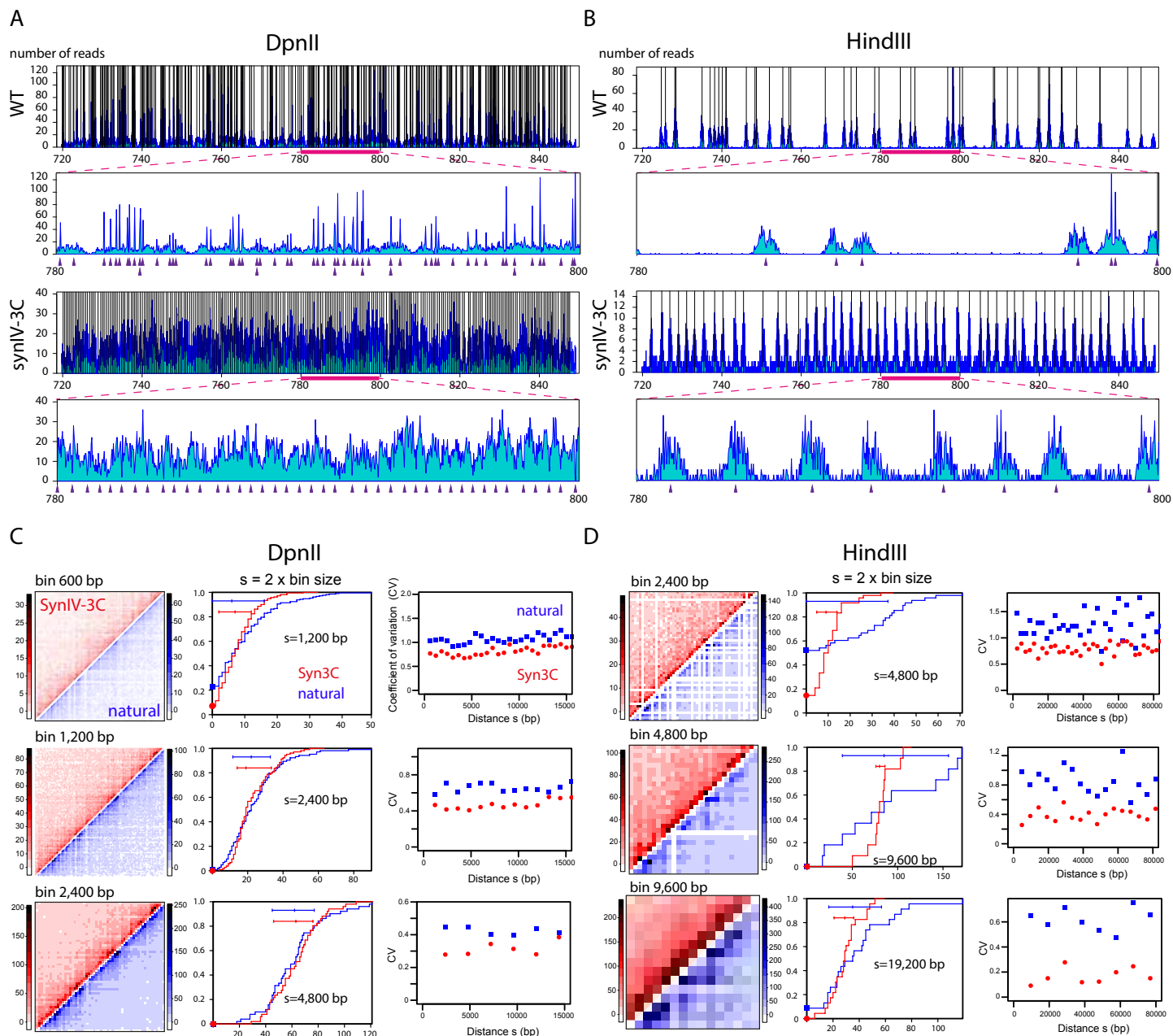
- Bader JS & Boeke JD (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**: 471–476
- Glynn EF, Megee PC, Yu H-G, Mistrot C, Unal E, Koshland DE, DeRisi JL & Gerton JL (2004) Genome-Wide Mapping of the Cohesin Complex in the Yeast *Saccharomyces cerevisiae*. *PLOS Biol.* **2**: e259
- Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, De Gobbi M, Taylor S, Gibbons R & Higgs DR (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**: 205–212
- Hunter N & Kleckner N (2001) The Single-End Invasion: An Asymmetric Intermediate at the Double-Strand Break to Double-Holliday Junction Transition of Meiotic Recombination. *Cell* **106**: 59–70
- Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J & Mirny LA (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**: 999–1003
- Ito M, Kugou K, Fawcett JA, Mura S, Ikeda S, Innan H & Ohta K (2014) Meiotic recombination cold spots in chromosomal cohesion sites. *Genes Cells Devoted Mol. Cell. Mech.* **19**: 359–373
- Kim KP, Weiner BM, Zhang L, Jordan A, Dekker J & Kleckner N (2010) Sister cohesion and structural axis components mediate homolog bias of meiotic recombination. *Cell* **143**: 924–937
- Kleckner N, Zickler D, Jones GH, Dekker J, Padmore R, Henle J & Hutchinson J (2004) A mechanical basis for chromosome function. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 12592–12597
- Klein F, Mahr P, Galova M, Buonomo SBC, Michaelis C, Nairz K & Nasmyth K (1999) A Central Role for Cohesins in Sister Chromatid Cohesion, Formation of Axial Elements, and Recombination during Yeast Meiosis. *Cell* **98**: 91–103
- Koszul R (2016) Beyond the bounds of evolution: Synthetic chromosomes... How and what for? *C. R. Biol.* Available at: <http://dx.doi.org/10.1016/j.crv.2016.05.007>
- Koszul R, Kim KP, Prentiss M, Kleckner N & Kameoka S (2008) Meiotic Chromosomes Move by Linkage to Dynamic Actin Cables with Transduction of Force through the Nuclear Envelope. *Cell* **133**: 1188–1201
- Koszul R & Kleckner N (2009) Dynamic chromosome movements during meiosis: a way to eliminate unwanted connections? *Trends Cell Biol.* **19**: 716–724
- Lajoie BR, Dekker J & Kaplan N (2015) The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods San Diego Calif* **72**: 65–75

- Lazar-Stefanita L, Scolari VF, Mercy G, Muller H, Guérin TM, Thierry A, Mozziconacci J & Koszul R (2017) Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J.*: e201797342
- Le TBK, Imakaev MV, Mirny LA & Laub MT (2013) High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**: 731–734
- Lesne A, Riposo J, Roger P, Cournac A & Mozziconacci J (2014) 3D genome reconstruction from chromosomal contacts. *Nat. Methods* **11**: 1141–1143
- Lieberman-Aiden E, Berkum NL van, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES & Dekker J (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**: 289–293
- Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O’Kelly MJT, van Oudenaarden A, Barton DBH, Bailes E, Nguyen AN, Jones M, Quail MA, et al (2009) Population genomics of domestic and wild yeasts. *Nature* **458**: 337–341
- Marbouty M, Ermont C, Dujon B, Richard G-F & Koszul R (2014) Purification of G1 daughter cells from different *Saccharomycetes* species through an optimized centrifugal elutriation procedure. *Yeast* **31**: 159–166
- Marie-Nelly H, Marbouty M, Cournac A, Liti G, Fischer G, Zimmer C & Koszul R (2014) Filling annotation gaps in yeast genomes using genome-wide contact maps. *Bioinforma. Oxf. Engl.* **30**: 2105–2113
- Mercy G, Mozziconacci J, Scolari VF, Yang K, Zhao G, Thierry A, Luo Y, Mitchell LA, Shen M, Shen Y, Walker R, Zhang W, Wu Y, Xie Z, Luo Z, Cai Y, Dai J, Yang H, Yuan Y-J, Boeke JD, et al (2017) 3D organization of synthetic and scrambled chromosomes. *Science* **355**: eaaf4597
- Muller H, Annaluru N, Schwerzmann JW, Richardson SM, Dymond JS, Cooper EM, Bader JS, Boeke JD & Chandrasegaran S (2012) Assembling large DNA segments in yeast. *Methods Mol. Biol. Clifton NJ* **852**: 133–150
- Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA & Dekker J (2013) Organization of the mitotic chromosome. *Science* **342**: 948–953
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J & Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381–385
- Oh SD, Jessop L, Lao JP, Allers T, Lichten M & Hunter N (2009) Stabilization and Electrophoretic Analysis of Meiotic Recombination Intermediates in *S. pombe*

Type='Italic'>*Saccharomyces cerevisiae*</Emphasis>. In *Meiosis* pp 209–234. Humana Press Available at: [https://link.springer.com/protocol/10.1007/978-1-59745-527-5\\_14](https://link.springer.com/protocol/10.1007/978-1-59745-527-5_14) [Accessed December 14, 2017]

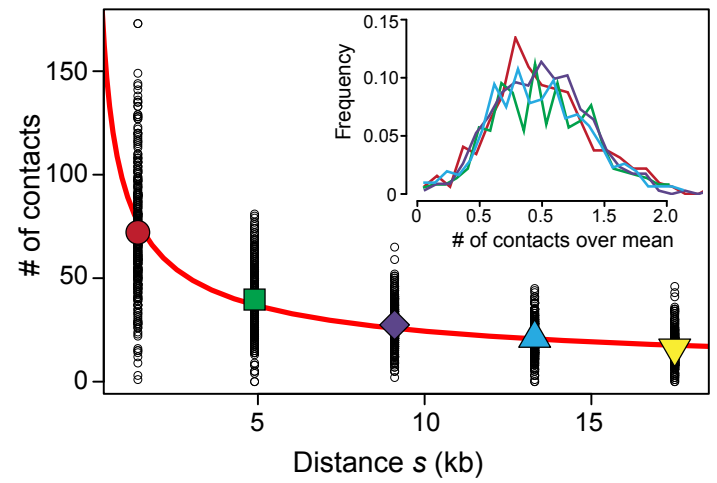
- Padmore R, Cao L & Kleckner N (1991) Temporal comparison of recombination and synaptonemal complex formation during meiosis in *S. cerevisiae*. *Cell* **66**: 1239–1256
- Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, Hochwagen A & Keeney S (2011) A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**: 719–731
- Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ & Fangman WL (2001) Replication dynamics of the yeast genome. *Science* **294**: 115–121
- Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES & Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–1680
- Rhee HS & Pugh BF (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**: 295–301
- Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A & Cavalli G (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**: 458–472
- Siow CC, Nieduszynska SR, Müller CA & Nieduszynski CA (2012) OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* **40**: D682–D686
- Sollier J, Lin W, Soustelle C, Suhre K, Nicolas A, Géli V & Saint-André C de LR (2004) Set1 is required for meiotic S-phase onset, double-strand break formation and middle gene expression. *EMBO J.* **23**: 1957–1967
- Sommermeier V, Béneut C, Chaplais E, Serrentino ME & Borde V (2013) Spp1, a Member of the Set1 Complex, Promotes Meiotic DSB Formation in Promoters by Tethering Histone H3K4 Methylation Sites to Chromosome Axes. *Mol. Cell* **49**: 43–54
- Trelles-Sticken E, Loidl J & Scherthan H (1999) Bouquet formation in budding yeast: initiation of recombination is not required for meiotic telomere clustering. *J. Cell Sci.* **112**: 651–658
- Zickler D & Kleckner N (1999) Meiotic chromosomes: integrating structure and function. *Annu. Rev. Genet.* **33**: 603–754
- Zickler D & Kleckner N (2016) A Few of Our Favorite Things: Pairing, the Bouquet, Crossover Interference and Evolution of Meiosis. *Semin. Cell Dev. Biol.* **54**: 135–148



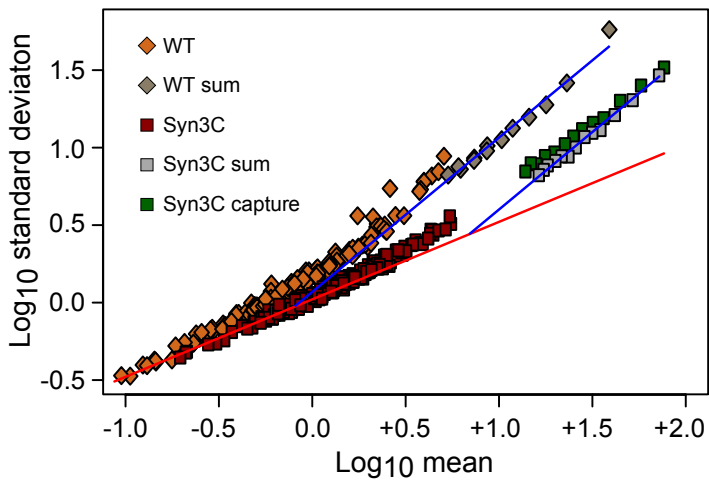


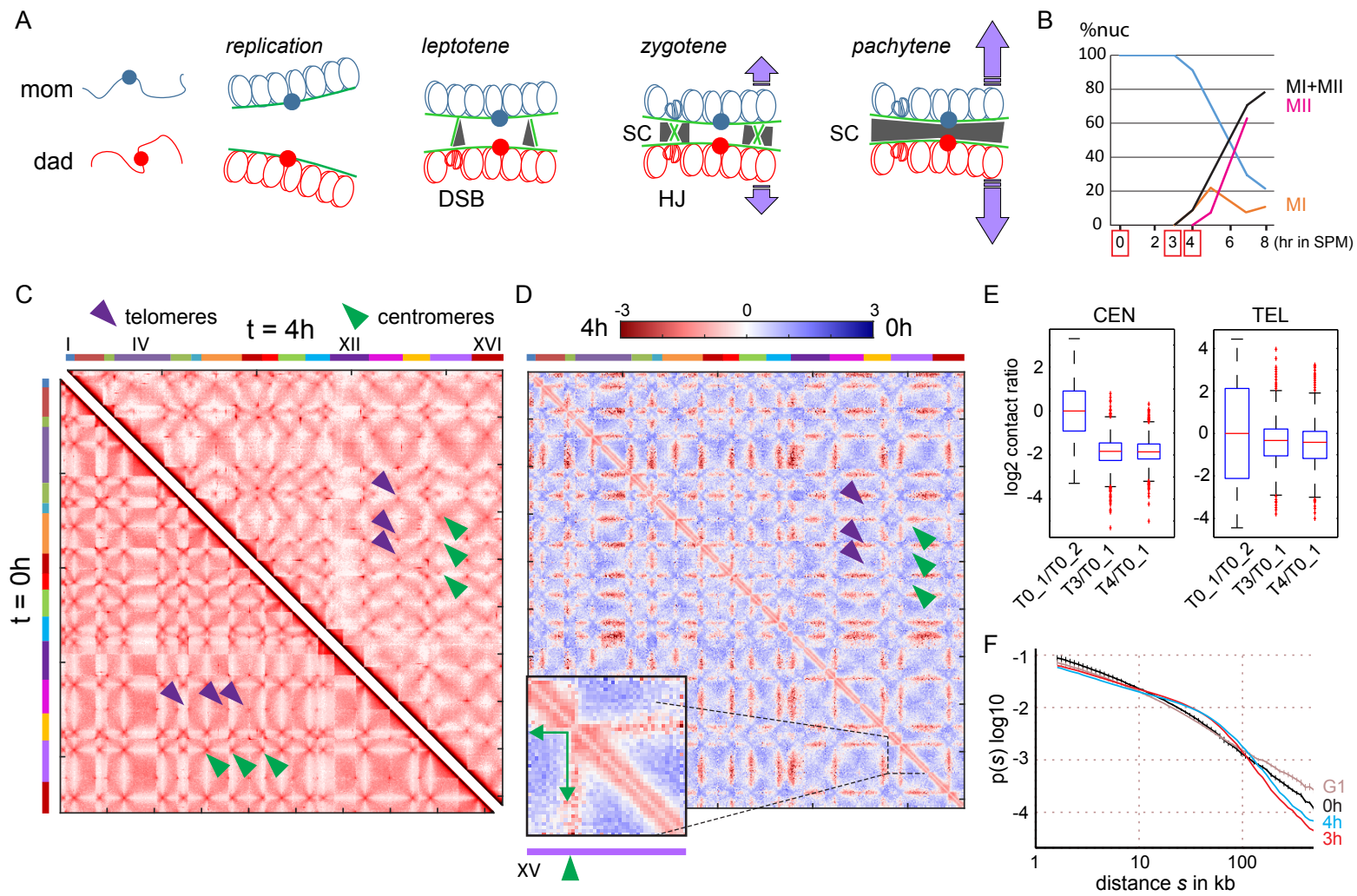


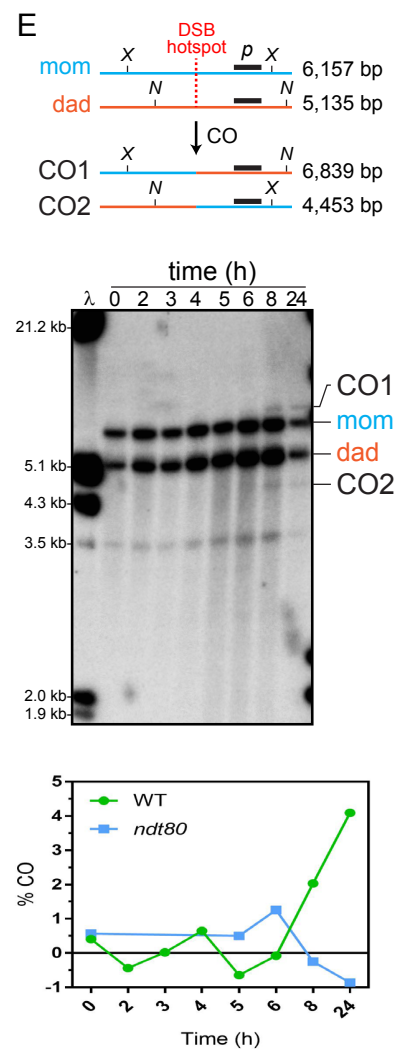
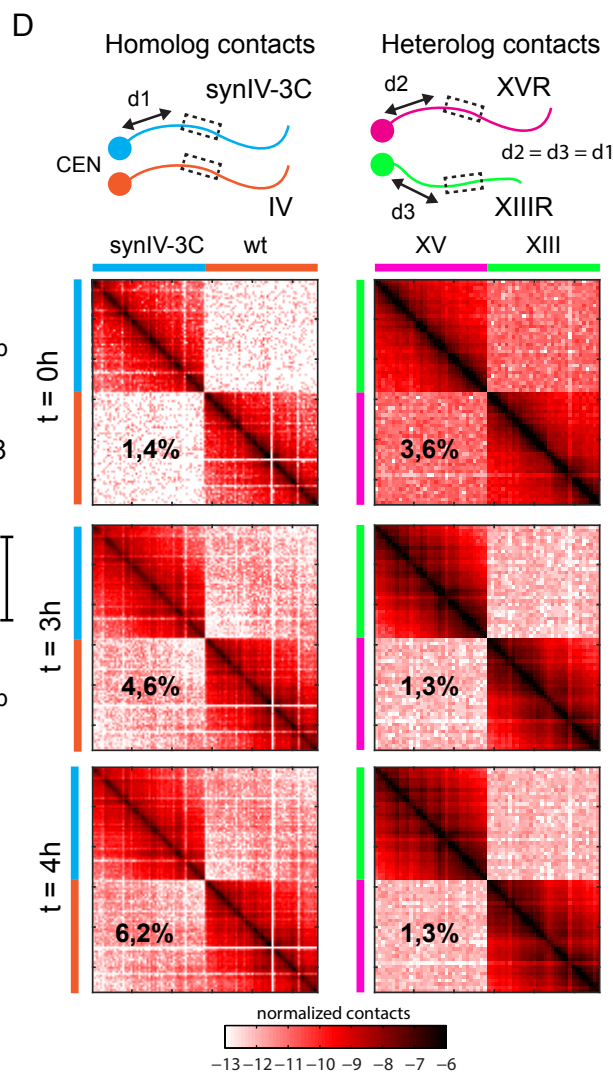
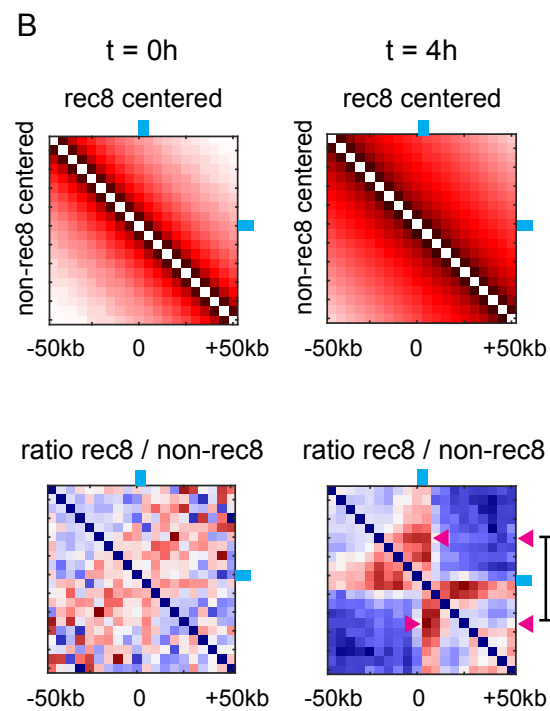
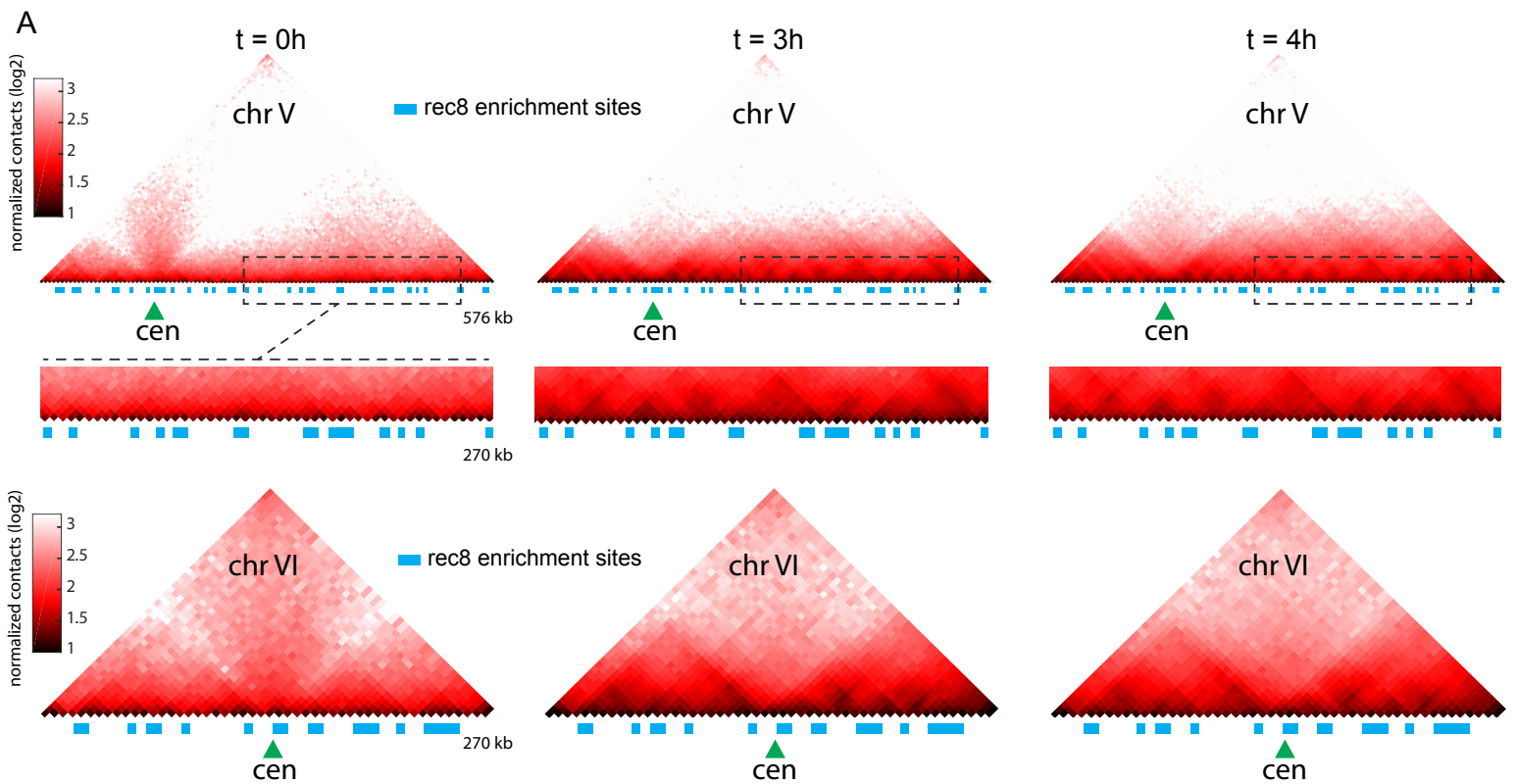
A



B







## Figure legends

### Figure 1 - synIV-3C design and assembly.

- A Number of contacts made by RFs as a function of their size (HindIII (red) or DpnII (blue) in the native sequence. Left panel: log-lin scale. Right panel: log-log scale.).
- B Illustration of the design principles of the synIV-3C sequence for the DpnII and HindIII RSs. Black arrow: chromosome. Grey rectangles: genetic elements. Blue and red vertical lines represent the RSs positions for the enzymes DpnII and HindIII, respectively. Top panel: restriction pattern of a (hypothetical) native sequence. Bottom panel: restriction pattern after synIV-3C design, with the RSs defining regularly spaced intervals.
- C Distribution of the DpnII (left) and HindIII (right) RFs sizes in both the native and synIV-3C 150kb redesigned sequence (red and blue, respectively).
- D Raw DpnII contact map of the Hi-C experiment performed on G1 daughter cells synchronized through elutriation (Marbouty *et al*, 2014). Dashed lines: borders of the redesigned region. Plain black lines: borders of the contact map analyzed in **Fig. 2**.

**Figure 2** – Sequence reads coverage from Hi-C experiments performed with DpnII (A) and HindIII (B) restriction enzymes in synIV-3C and native strains, and mapped against the synthetic region SynIV-3C and its natural WT counterpart, respectively. For each region, the magnification of a 20 kb window (pink line) is shown. Vertical lines indicates the position of restriction sites in the top panel and purple triangles their position in the magnified 20kb region. Note that the scale of the

y-axis illustrates the heterogeneity of the coverage, with some positions in the DpnII map being overrepresented with respect to others.

C, D Analysis of the contact counts along the synIV-3C region for DpnII (C) and HindIII (D). First column: synIV-3C (in red) and chromosome 4 native counterpart (blue) Hi-C contact maps. For each experiment, three different fixed bin sizes were analyzed (600 bp, 1200 bp and 2400 bp for DpnII, 2400 bp, 4800 bp and 9600 bp for HindIII). Middle panels: cumulative distribution of the number of contacts between bins located at a genomic distance  $s$  from each other's ( $s = 2 \times \text{bin size}$ ). Right panels: distribution of the coefficient of variation (CV) as a function of  $s$ .

**Figure 3. Relation between standard deviation and mean for each pair of fragments located at similar genomic distances  $s \pm \Delta s$**

A The mean number of contact of restriction fragments (big symbols) at a fixed range of distances ( $s \pm 350\text{bp}$ , see text) decreases as a power law (red line) with exponent -0.6. Small circles represent the contact number of individual pairs of restriction fragments. Inset: Distributions of contact numbers re-scaled by their means

B The relation between the variance and the mean of these distributions undergoes a transition from Poisson to Gaussian. WT and Syn3C data were obtained by aggregating data-sets respectively from (Mercy *et al*, 2017) and (Lazar-Stefanita *et al*, 2017). Capture-C data are from this study. The red line corresponds to the theoretical Poisson behavior, whereas the blue line corresponds to the theoretical Gaussian behavior fitting the CV from data.

**Figure 4. Individualization of chromosomes during prophase**

A Schematic representation of the structural changes affecting homologs during meiotic prophase. Sister chromatids organize as arrays of loops along each homolog axis (green lines) after replication. At leptotene, DNA double-strand breaks (DSBs) occur. During zygotene, homologs come together partially and the synaptonemal complex (SC, grey) originates at DSBs and centromeres. At pachytene, homologs are fully synapsed and undergo vigorous motion (purple arrows).

B Meiotic progression, as measured by completion of MI and MII divisions.

C Contacts maps of synchronized populations of cells after 0 (bottom left) and 4 (top right)

hours in sporulation medium. The 16 yeast chromosomes are displayed atop the maps. Purple arrowheads: inter-telomere contacts. Green arrowheads: inter-centromeric contacts.

D Log-ratio of the two contact maps from (C). The blue to red color scale reflects the enrichment in contacts in one population with respect to the other ( $\log_2$ ). The inset on the bottom left displays the magnification of chromosome XV. The green arrows point at the contacts made at  $t=4h$  by the centromere with the left and right arms.

E Boxplots representing the variation in number of normalized contacts for centromeres (left) and telomeres (right) between 0, 3 and 4h in sporulation media. Variations between  $t=0h$  replicates, 3h and 0h, and 4h and 0h, were computed. The relative Wilcoxon test supports a decrease in centromere contacts at  $t=3$  and 4h compared to  $t=0h$ .

F Average intra-chromosomal contact frequency  $p$  between two loci with respect to their genomic distance  $s$  along the chromosome ( $\log\text{-}\log$  scale;  $p(s)$ ) during mitotic G1 (brown curve; three replicates from Lazar-Stefanita et al., 2017), meiotic  $t=0h$  (black curve; two replicates), 3h (red curve) and 4h (blue curve).

### Figure 5. Organization of homologs into chromatin loops

A Magnification of chromosomes V and VI contact maps during meiotic time course at  $t=0$ , 3 and 4 hours (bin size: 5kb). The blue rectangles point at bins enriched in Rec8 protein. Green triangles: centromere position.

B Top panels: aggregated intrachromosomal contacts made by 100kb windows (5 kb bins) centered on Rec8 enriched bins (top right) or randomly chosen (bottom left). Middle panels: ratio between the cumulated contact maps of the above panel. Blue color shows a depletion of contacts in the random maps, whereas the red signal points at an enrichment in contacts in the maps centered on Rec8 enriched bins. Purple arrowheads point at looping signal between the center of the window (corresponding to the Rec8 enriched bin) and flanking regions. Bottom panel: schematic representation of the disposition of the chromatin at the corresponding timepoints.

C Normalized frequency of contacts as a function of genomic distance within the SynIV-3C and its WT counterpart for  $t=0$ , 3, and 4h.

D Contacts between the SynIV-3C region and its native homolog (left panels) and between two windows of similar sizes positions at an equal distance ( $d$ ) to the centromere but belonging to two other chromosomes (right panels). The percentages on each column reflects different measurements. For the inter-homolog panels, they represent the number of pairs of reads

bridging homologs (*trans* contacts) divided by the total number of pairs aligning within the regions. On the other hand, for inter-heterologs, the percentage represents the amount of pairs of reads bridging the two distinct regions, divided by the total number of pairs aligning within the distinct regions. This illustrates the gradual chromosomal individualization.

E Crossing over between the SynIV-3C region and the native counterpart. Top: schematic representation of how the redesigned restriction pattern allows characterization of CO events using a probe (black line *p*) at a DSB hotspot (Material and Methods). Middle: meiotic recombination of cells progressing into meiosis and analyzed using restriction pattern similar to (Hunter & Kleckner, 2001). Parental homologs, “Dad” and “Mom”, and COs are distinguished on Southern blots via restriction site polymorphism. Bottom panel: CO as percent of total hybridizing signal with time after transfer to sporulation medium.

## Tables

**Table 1.** Mutations necessary to remove and generate new sites along chromosome 4 700,000::850,000 window.

|              | deletion | new sites |
|--------------|----------|-----------|
| HindIII      | 58       | 61        |
| NdeI         | 34       | 23        |
| XbaI         | 25       | 76        |
| DpnII        | 442      | 310       |
| <b>Total</b> | 559      | 470       |

**Table 2: Genotype of *Saccharomyces cerevisiae* strains used in this study.**

| Name     | Genotype                                                                                                                | Background | Source                |
|----------|-------------------------------------------------------------------------------------------------------------------------|------------|-----------------------|
| RSG_Y181 | <i>MATa, ura3Δ0, leu2Δ0, his3Δ1, lys2Δ0, SynIV-3C-URA3</i>                                                              | S288C      | This study            |
| RSG_Y189 | <i>MATα, ura3, lys2, ho::LYS2, leu2-R, arg4-nsp,bgl, SynIV-3C-URA3</i>                                                  | SK1        | This study            |
| RSG_Y190 | <i>MATa/MATα ura3/ura3, lys2/lys2, ho::LYS2/ho::LYS2 leu2-K/leu2-R, arg4-nsp,bgl/arg4-nsp,bgl, SynIV-3C-URA3/native</i> | SK1        | This study            |
| ORT4601  | <i>MATα, ura3, lys2, ho::LYS2, leu2-K, arg4-nsp,bgl</i>                                                                 | SK1        | (Sollier et al, 2004) |

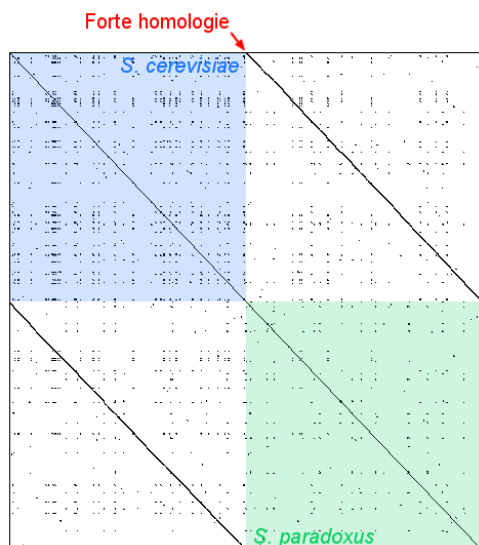


# Étude de l'interaction des chromosomes homologue d'une levure hybride

## 4.5 Introduction

Le génome de la levure est très organisé au sein de son noyau, aussi bien dans l'espace que dans le temps. Cette organisation est d'ailleurs très importante pour les processus biologiques tels que la transcription, la réparation de l'ADN et la recombinaison (Taddei et al., 2010). La principale caractéristique du génome de *S. cerevisiae* est son organisation de pseudo-Rabl (Jin et al., 1998) : au cours de l'interphase, les centromères sont tous regroupés à un pôle du noyau, attachés au *spindle pole body* ; les chromosomes se replient au niveau de ces centromères, permettant à leurs bras de s'étendre vers la périphérie nucléaire où les télomères s'associent en petits foyers (Therizols et al., 2010) ; les répétitions d'ADNr forment quant à elles le nucléole qui se localise au pôle opposé au SPB (Yang et al., 1989), divisant ainsi le chromosome en deux domaines séparés. Cette organisation est très bien conservée pendant la durée du cycle cellulaire (Jin et al., 1998) et même au cours de la quiescence, bien que le regroupement des télomères soit plus important et celui des centromères un peu plus faible (Guidi et al., 2015). Le développement de la méthode Hi-C, en couplant la méthode de capture de conformation de chromosome au séquençage de seconde génération, a permis de confirmer ces observations faites en microscopie (Duan et al., 2010 ; Guidi et al., 2015). Jusqu'à présent, la méthode de Hi-C n'a pu être utilisée que sur des populations de levures haploïdes. Si elle est effectuée sur une cellule diploïde, les chromosomes homologues ne pourront pas être distingués l'un de l'autre lors de l'analyse des lectures de séquençage.

Chez la levure diploïde, les chromosomes homologues peuvent être appariés non seulement lors de la méiose, mais aussi lors de la croissance mitotique (Burgess et al., 1999, Burgess et al., 1999, Weiner et Kleckner, 1994), comme cela existe chez la Drosophile (Metz, 1916). Cependant, l'ampleur de l'appariement des homologues mitotiques a été débattue en raison des divergences entre les études (Barzel et Kupiec, 2008). Une explication de ces discordances réside dans la présence d'artefacts potentiels dans les méthodes de microscopie utilisées pour détecter l'appariement. En FISH, la perte de signal peut conduire à une fausse inférence de colocalisation (Lorenz et al., 2003). Il a également été suggéré que le marquage de *locus* génomiques avec des réseaux répétitifs de GFP pour l'imagerie des cellules vivantes peut directement provoquer



**FIGURE 4.11** Matrice de similarité des génomes de *S. cerevisiae* et *S. paradoxus* : Matrice d'homologie générée avec Dotmatcher (EMBOSS package) pour des bins de 50 kpb.

l'appariement via la dimérisation de la GFP (Mirkin et al., 2014). De plus, l'orientation de pseudo-Rabl peut, à elle seule, créer l'apparence d'un appariement d'homologues si elle n'est pas contrôlée en juxtaposant des *loci* chromosomiques à la même distance des centromères, y compris des *locus* homologues (Lorenz et al., 2003). Les écarts dans l'ampleur de l'appariement entre les études pourraient être attribués à la variation de la force d'appariement à travers le génome ; cependant, l'appariement des homologues mitotiques n'a été examiné que dans quelques *loci*.

Dans notre précédent manuscrit sur le chromosome syn3C, nous avons démontré que la redistribution des sites de restriction générait suffisamment de SNPs entre ce chromosome synthétique et son homologue natif pour nous permettre d'étudier leur organisation grâce à la méthode de Hi-C. Cependant, ce design ne concerne qu'une région de 150 kpb du chromosome IV, un autre modèle est donc nécessaire pour nous permettre d'étudier l'interaction des chromosomes homologues à l'échelle du génome. Pour cela, nous avons utilisé une souche de levure hybride contenant une copie du génome de *S. cerevisiae* et de *S. paradoxus*, qui sont deux espèces de levures très similaires. Leurs génomes sont tous les deux composés de 16 chromosomes linéaires, représentant environ 12 millions de paires de bases chacun. Avec une similarité d'approximativement 90%, elles sont les parents connus les plus proches du genre des *Saccharomyces* (Kellis et al., 2003). Le taux d'hétérozygotie qui existe entre les génomes de *S. cerevisiae* et de *S. paradoxus*, ainsi que la machinerie de réparation des mésappariements, empêchent les chromosomes homologues de s'apparier au début de la méiose, entraînant une baisse de la fertilité de cet hybride (Hunter et al., 1996). Bien que cet hybride ne puisse pas être utilisé pour étudier l'organisation des chromosomes au cours de la méiose, grâce aux 10 % de différence entre les deux génomes, il représente un modèle de choix pour l'étude de l'interaction des chromosomes homologues au cours du cycle cellulaire mitotique.

Au travers de notre analyse de l'organisation du génome de cet hybride, nous nous sommes aperçus que certains artefacts pouvaient apparaître du fait de la haute similarité des deux génomes. La correction de ces artefacts est délicate car il nous faut trouver un équilibre optimal entre l'élimination des données ambiguës et la qualité des données restantes. En utilisant l'hybride de *S. cerevisiae* et *S. paradoxus* comme modèle de levure diploïde, nous proposons, dans un premier temps, une méthode pour réduire cet artefact, nous permettant ensuite d'analyser l'organisation des chromosomes homologues dans les conditions standard de culture en laboratoire.

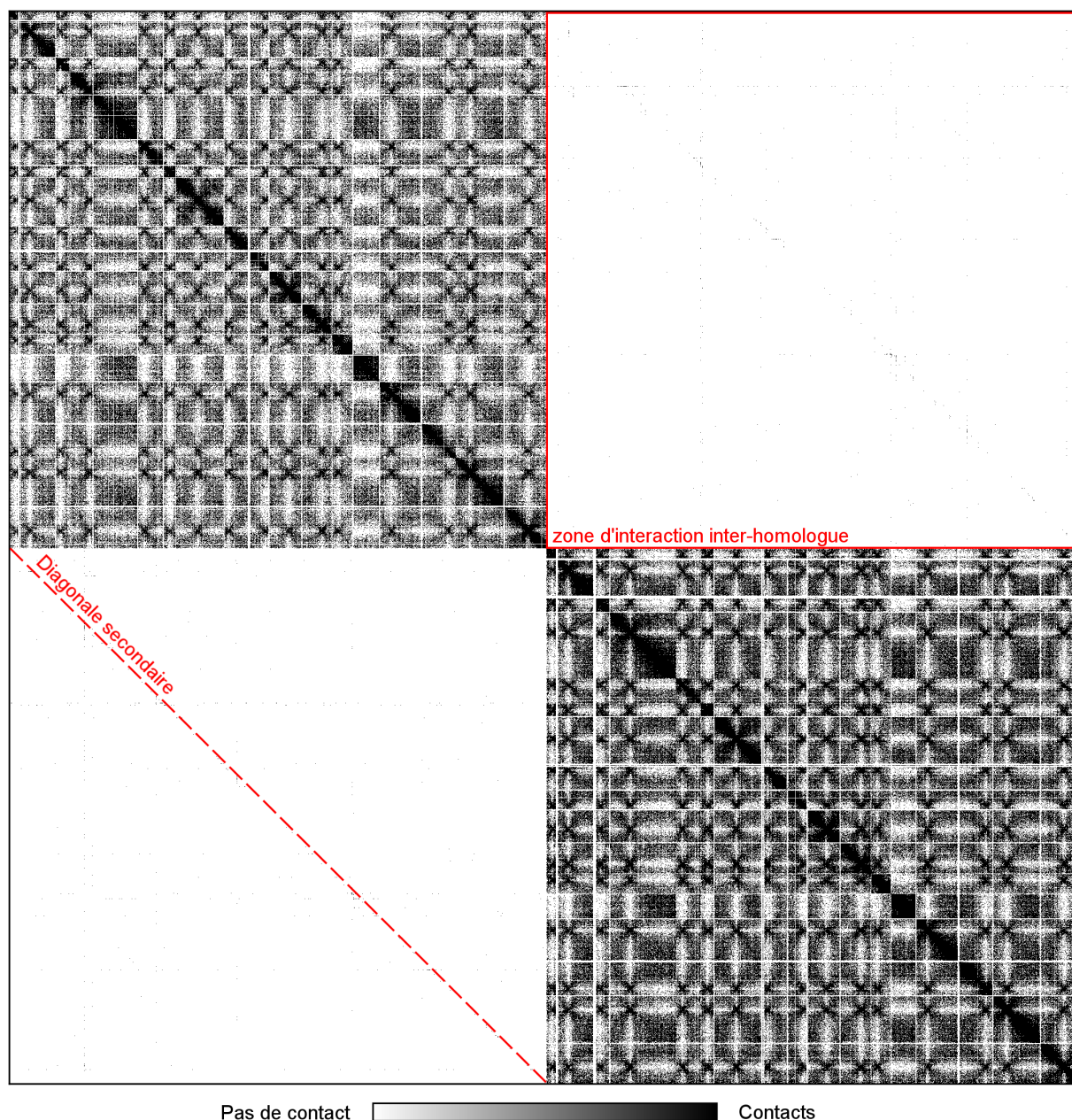
## 4.6 Résultats et discussions

### 4.6.1 Alignement des données haploïdes sur le génome hybride

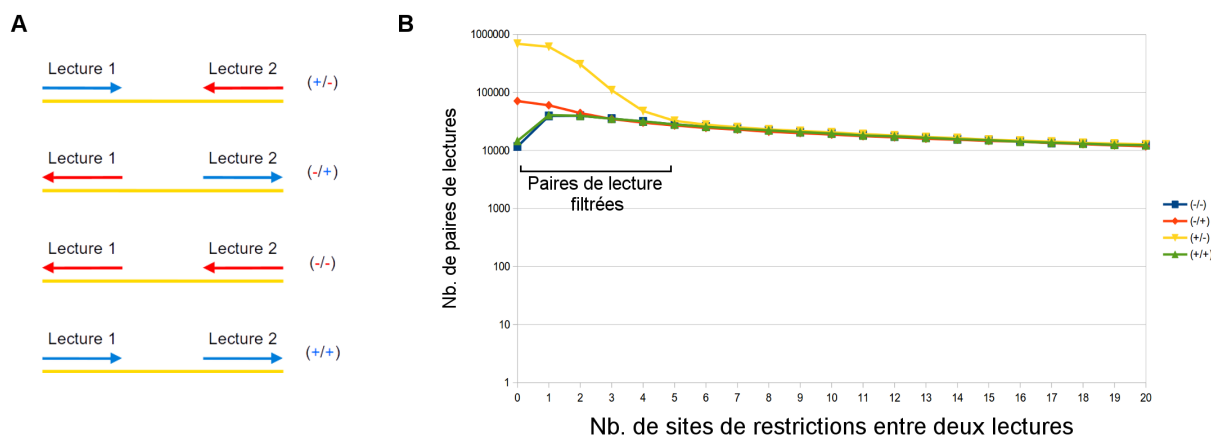
Les méthodes de génomique ne permettent pas d'analyser un génome diploïde car il est impossible de distinguer les lectures de séquençage appartenant à un chromosome ou à son homologue, lors de l'étape d'alignement. Chez l'hybride de *S. cerevisiae* et *S. paradoxus*, les séquences des chromosomes homologues partagent 90% d'identité (figure 4.11), dans cette configuration, il est alors possible de distinguer les deux chromosomes homologues. Pour vérifier que notre méthode d'alignement nous permette d'attribuer les lectures de séquençage au bon génome, nous avons dans un premier temps effectué notre expérience de Hi-C sur des populations des souches haploïdes de *S. cerevisiae* et *S. paradoxus* synchronisées en G1 par élutriation. Une fois les données de séquençage obtenues, les lectures ont été mélangées puis alignées sur le génome de référence complet de l'hybride (figure 4.12). Bien que la majorité des lectures de séquençage aient été convenablement attribuées à l'organisme correspondant, une petite partie d'entre elles ont été alignées sur le mauvais génome, laissant apparaître une très faible diagonale secondaire dans la zone censée correspondre aux interactions entre homologues. Notre alignement étant déjà très strict, ces erreurs semblent compliquées à éviter lors de cette étape. Nous avons alors développé une méthode permettant d'éliminer, en partie, ces artefacts d'alignement lors de la construction de la matrice.

### 4.6.2 Filtrage de la matrice pour corriger les artefacts

Tout comme les deux contrôles haploïdes, la population de l'hybride a été synchronisée en G1 par élutriation avant que nous analysions l'organisation de son génome en Hi-C. Après alignement des données de séquençage, en suivant la méthode utilisée pour aligner les données des haploïdes, la carte de contacts présente elle aussi une diagonale secondaire au niveau des interactions inter-homologue (figure 4.14). Comme nous avons pu l'observer sur les cartes de contacts des populations haploïdes, il s'agit probablement, en partie, d'artefacts issus d'erreurs d'alignement. Il est très probable que lors de l'étape d'alignement, qui est réalisée à partir de paires de lectures, l'une des deux lectures ait été attribuée au mauvais génome. Il est très probable que



**FIGURE 4.12** Matrice de contrôle de l'alignement du génome de l'hybride : Alignement des deux jeux de donnée de *S. cerevisiae* et *S. paradoxus* haploïde sur le génome de l'hybride. Le code couleur est binaire, blanc quand il n'y a pas de contacts, noir s'il y en a.



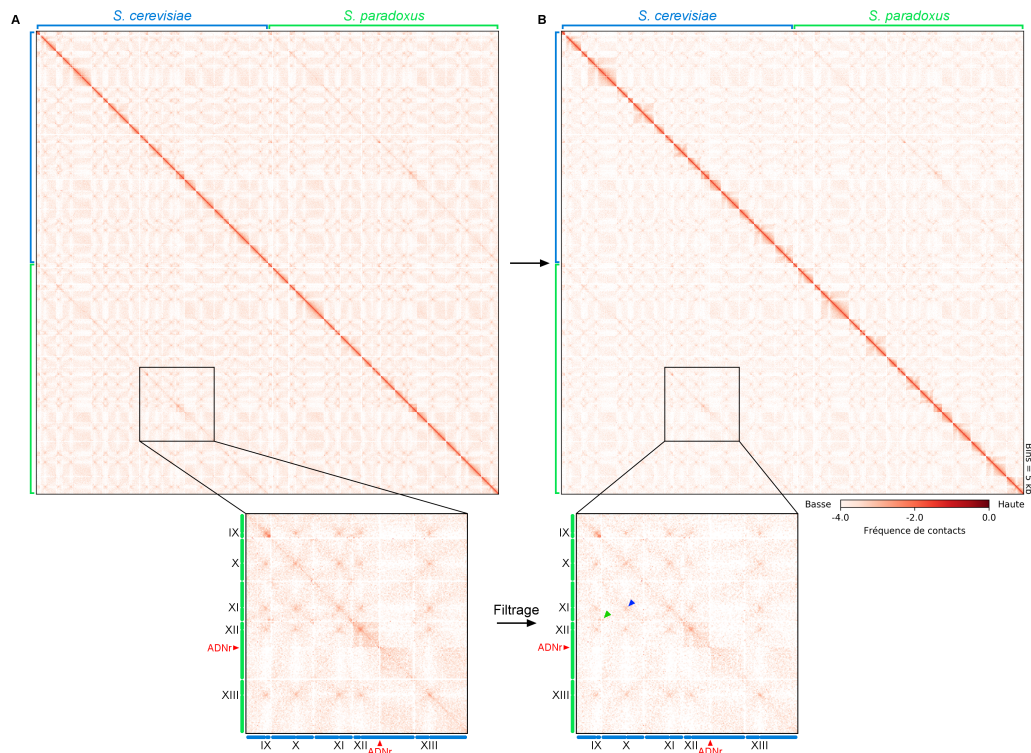
**FIGURE 4.13 Orientation des paires de lectures :** (A) Il y a quatre combinaisons d'orientation possible pour une paire de lectures : (+/-), (-/+), (+/+) et (-/-). (B) La proportion de ces différentes orientations n'étant pas égale à moins de 6 fragments de restriction, ces données sont filtrées.

cette diagonale secondaire soit aussi forte car la majorité de ces lectures mal attribuées aurait dû être filtrées, ce qui n'a pas pu être réalisé du fait de ces erreurs d'alignement.

Les lectures de séquençage peuvent être alignées selon deux orientations, (+) ou (-), en fonction qu'elles correspondent respectivement au brin sens ou anti-sens de l'ADN. Étant donné, que les fragments d'ADN d'un jeu de données de Hi-C ont été coupés puis réparés, quatre combinaisons d'orientation sont possibles pour une paire de lectures : (+/-), (-/+), (+/+) et (-/-). Ces quatre combinaisons devrait être retrouvées en proportions égales dans nos données, or pour les paires de lectures qui s'alignent dans des régions voisines, à une distance d'environ 6 fragments de restrictions, les paires de (+/-) et (-/+) sont surreprésentées par rapport aux paires (+/+) et (-/-) (figure 4.13). Elles correspondent en général à des fragments d'ADN qui n'ont pas été coupés ou qui se sont réparés avec leur voisin initial, ces fragments étant présents en quantité importante et n'apportant pas d'informations d'organisation structurale, ils sont éliminés de notre analyse (Cournac et al., 2012). Nous supposons que ce sont ces fragments qui posent problèmes dans notre analyse de la levure hybride. En effet, si l'une des lectures s'est alignée sur un chromosome de *S. cerevisiae* et que l'autre est alignée sur son homologue de *S. paradoxus*, alors même qu'elles auraient dû être alignées sur le même chromosome à une distance inférieure ou égale aux 6 fragments de restrictions, ces dernières ne pourront pas être filtrées lors de la construction de la matrice de contacts.

Pour réduire cet artefact, lors de la construction de la matrice des interactions *trans*, nous proposons d'éliminer toutes les paires de lectures ayant été alignées en orientation (+/-) et (-/+), et de ne garder que les paires de lectures alignées en (+/+) et (-/-). Les matrices avant et après filtrage sont représentées respectivement à gauche et à droite de la figure 4.14. Grâce à l'agrandissement, avant filtrage, on peut voir la fine diagonale d'artefacts au niveau des interactions inter-homologues qui était présente dans les contrôles haploïdes. En comparant plus précisément le nombre de





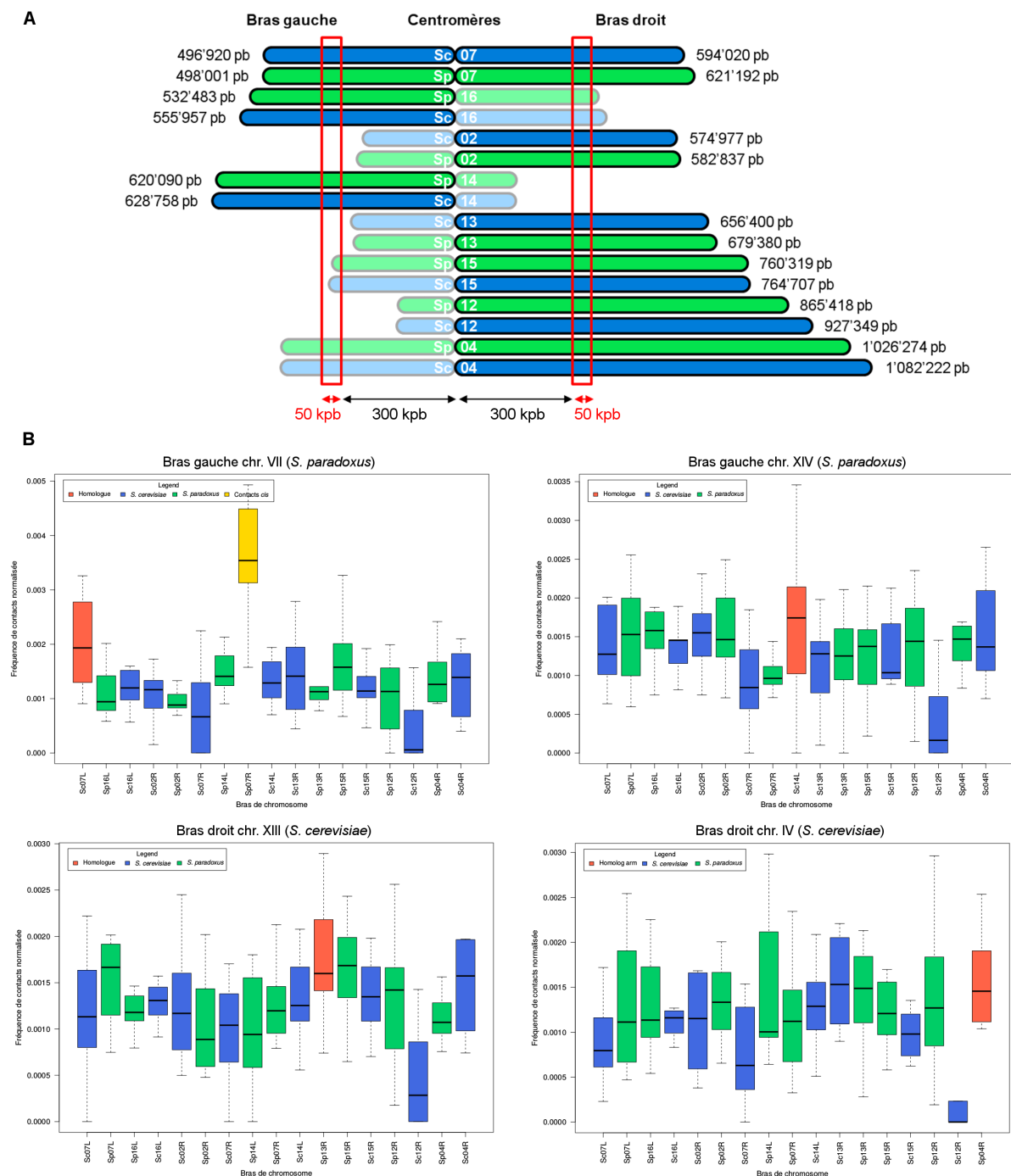
**FIGURE 4.14 Filtrage de la matrice de contacts :** Matrices de contacts normalisées de l'hybride en phase G1 (bins = 5 kpb), avant filtrage des interaction *trans* (à gauche) et après filtrage (à droite). La flèche verte indique le regroupement des télomères et la flèche bleue celui des centromères.

contacts, avant et après cette procédure de filtrage, les interactions *trans* sont globalement réduites de 50%<sup>5</sup>, tandis que, plus localement, au niveau des zones d'interactions inter-homologues, la réduction du nombre de contacts est légèrement plus importante, de l'ordre de 52%. En se rapprochant de cette diagonale, la proportion de lectures éliminées augmente, confirmant ainsi notre hypothèse selon laquelle une partie des contacts de cette diagonale secondaire sont des artefacts. Grâce à cette correction apportée par notre méthode de filtrage, nous pouvons maintenant analyser l'organisation des chromosomes de ce modèle de levure diploïde.

### 4.6.3 L'organisation du génome de l'hybride et les contacts inter-homologues au cours de la phase G1

Tout comme pour les levures haploïdes, sur la carte de contacts de la levure hybride, synchronisée en G1, nous observons le regroupement des centromères et celui des télomères (flèches bleue et verte, respectivement, figure 4.14). On remarque également une interaction légèrement plus importante entre les deux chromosomes XII, portant chacun un *locus* d'ADNr. Cette interaction est le résultat de l'isolation des deux *loci* d'ADNr en un pôle du noyau, comme cela a déjà pu être montré en FISH (Lorenz et al., 2002) et très récemment en Hi-C (Kim et al., 2017). Les interactions

5. Résultat attendu considérant qu'au-delà de 6 fragments de restrictions, les paires de lecture orientées (+/-), (-/+), (+/+) et (-/-) sont représentées en proportion égale (figure 4.13B)



**FIGURE 4.15 Effet de Rab1 ou appariement d'homologue :** (A) Les interactions dans une fenêtre de 50 kpb à une distance de 300 kpb du centromère, pour tous les bras de chromosome dont la taille est supérieure à 480 kpb, ont été extraites et représentées sous forme de boîtes à moustaches, pour les bras droits des chromosomes XIII et IV de *S. cerevisiae*, et les bras gauches des chromosomes VII et XIV de *S. paradoxus*. (B). Les interactions entre bras de chromosomes homologues sont représentées en orange, celles impliquant un bras de chromosome de *S. cerevisiae* en bleu, celles impliquant un bras de chromosome de *S. paradoxus* en vert, et pour le bras gauche du chromosome VII de *S. paradoxus* son interaction *cis* est représentée en jaune.

entre bras de chromosome de petite taille, notamment les bras droits des chromosomes IX et ceux des chromosomes I, sont très marquées par rapport aux autres interactions *trans*. Il est possible que cette observation soit le résultat du regroupement préférentiel des bras de chromosome en fonction de leur taille, résultant en l'isolation de ces petits bras de chromosome. Cette hypothèse pourrait éventuellement être vérifiée en déplaçant l'un des deux centromères des chromosomes IX, de manière à allonger le bras droit de ce chromosome, si cette interaction est fonctionnelle (conséquence d'un mécanisme moléculaire), elle devrait être conservée.

Pour déterminer si les interactions entre homologue sont spécifiques ou bien s'il s'agit d'un effet Rabl, nous avons calculé la fréquence de contacts entre une région de 50 kpb à une distance de 300 kpb du centromère, pour les bras droits des chromosomes XIII et IV, et les bras gauches des chromosomes VII et XIV, avec tous les bras de chromosome dont la taille est supérieure à 480 kpb (Figure 4.15A). Les régions d'intérêt ont été choisies de manière à ce qu'elles soient suffisamment éloignées des centromères et des télomères pour essayer de limiter l'effet dû à leur rapprochement. Les résultats sont représentés sous forme de « boîtes à moustache », où les interactions entre la région d'intérêt et son homologue sont représentées en orange, en bleu pour les interactions avec des chromosome de *S. cerevisiae* et en vert pour celles avec ceux de *S. paradoxus* (boxplots, Figure 4.15B). Les régions des bras droits des chromosomes XIII et IV, ainsi que celle du bras gauche du chromosome XIV ne semblent pas présenter d'affinités particulières pour leur homologue respectif, par rapport à leurs interactions avec les autres chromosomes. En ce qui concerne la région du bras gauche du chromosome VII de *S. paradoxus*, sa fréquence de contacts avec son homologue, le bras gauche du chromosome VII de *S. cerevisiae*, est légèrement supérieure à celles avec les autres chromosomes (Figure 4.15B, panel du haut à gauche). Cette région étant relativement proche du télomère ( $\approx 100$  kpb), il n'est pas impossible que cette augmentation soit un effet du rapprochement des régions télomériques de ces deux chromosomes.

## 4.7 Conclusion

De manière générale, ces résultats ne suggèrent pas l'existence d'un appariement spécifique des homologues entre eux au cours de la phase G1, mais plutôt un effet de l'organisation de pseudo-Rabl qui rapprocherait les bras de chromosomes de tailles similaires entre eux. Ce qui est en accord avec les résultats de Kim et al., qui ont analysé un hybride de *S. cerevisiae* et *Saccharomyces uvarum* et qui avancent que l'appariement des chromosomes homologues se ferait plutôt lorsque les cellules entrent en phase stationnaire (Kim et al., 2017). Cependant, il existe tout de même une spécificité entre les bras droit des chromosomes IX et les chromosomes XII qui portent chacun un *locus* d'ADNr.



# **L'organisation 3D des chromosomes de *S. cerevisiae* au cours du cycle cellulaire**



# Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle

Luciana Lazar-Stefanita<sup>1,2,3,4</sup>, Vittore F Scolari<sup>1,2,3</sup>, Guillaume Mercy<sup>1,2,3,4</sup>, Héloïse Muller<sup>1,2,3</sup>, Thomas M Guérin<sup>5</sup>, Agnès Thierry<sup>1,2,3</sup>, Julien Mozziconacci<sup>6,7,\*</sup> & Romain Koszul<sup>1,2,3,\*\*</sup>

## Abstract

Duplication and segregation of chromosomes involves dynamic reorganization of their internal structure by conserved architectural proteins, including the structural maintenance of chromosomes (SMC) complexes cohesin and condensin. Despite active investigation of the roles of these factors, a genome-wide view of dynamic chromosome architecture at both small and large scale during cell division is still missing. Here, we report the first comprehensive 4D analysis of the higher-order organization of the *Saccharomyces cerevisiae* genome throughout the cell cycle and investigate the roles of SMC complexes in controlling structural transitions. During replication, cohesion establishment promotes numerous long-range intra-chromosomal contacts and correlates with the individualization of chromosomes, which culminates at metaphase. In anaphase, mitotic chromosomes are abruptly reorganized depending on mechanical forces exerted by the mitotic spindle. Formation of a condensin-dependent loop bridging the centromere cluster with the rDNA loci suggests that condensin-mediated forces may also directly facilitate segregation. This work therefore comprehensively recapitulates cell cycle-dependent chromosome dynamics in a unicellular eukaryote, but also unveils new features of chromosome structural reorganization during highly conserved stages of cell division.

**Keywords** chromosome segregation; Hi-C; loop extrusion; replication profile; SMC

**Subject Categories** Cell Cycle; Chromatin, Epigenetics, Genomics & Functional Genomics; DNA Replication, Repair & Recombination

**DOI** 10.15252/embj.201797342 | Received 21 May 2017 | Revised 29 June 2017 | Accepted 4 July 2017

## Introduction

The chromosomes of prokaryotes and eukaryotes display multiple levels of hierarchical organization, whose dynamic changes influence or regulate metabolic processes including gene expression and DNA replication and repair (Taddei & Gasser, 2012; Wang *et al*, 2013; Dekker & Mirny, 2016). The improper coordination of chromosome condensation and segregation during the cell cycle can lead to important structural abnormalities and result in cell death or diseases such as cancer (Valton & Dekker, 2016). In recent years, major advances in imaging and chromosome conformation capture approaches (Dekker *et al*, 2002; Lieberman-Aiden *et al*, 2009; 3C, Hi-C) have complemented earlier work by describing at an unprecedented resolution the multiple hierarchical layers of genome organization. A variety of remarkable 3D chromosomal structures have been described in a number of species, including in unicellular organisms such as bacteria and yeasts.

The genome of budding yeast *Saccharomyces cerevisiae* presents a Rabl organization driven by (i) centromeres clustering at the spindle pole body (SPB, *S. cerevisiae* microtubule organizing center), (ii) telomeres tethering to the nuclear envelope, (iii) the nucleolus where the rDNA is sequestered opposite to the SPB, and (iv) chromosome arm length (Burgess & Kleckner, 1999; Taddei & Gasser, 2012). Hi-C experiments have confirmed this Rabl organization, but the existence of sub-megabase structures within yeast chromosomes similar to mammalian topological associated domains or their bacterial equivalent is still controversial (Duan *et al*, 2010; Hsieh *et al*, 2015; Eser *et al*, 2017). Importantly, genomic analysis of chromosome 3D architectures has usually been done using asynchronous populations, in which cells are found in various stages of the cell cycle. However, the initiation and progression of replication, followed by the segregation of the sister chromatids (SCs) into daughter cells, is expected to modify the genome higher-order

<sup>1</sup> Institut Pasteur, Department Genomes and Genetics, Unité Régulation Spatiale des Génomes, Paris, France

<sup>2</sup> CNRS, UMR 3525, Paris, France

<sup>3</sup> Institut Pasteur, CNRS, Center of Bioinformatics, Biostatistics and Integrative Biology (C3BI), USR 3756, Paris, France

<sup>4</sup> Sorbonne Universités, UPMC Université Paris 6, Complexité du Vivant, Paris, France

<sup>5</sup> Laboratoire Télomères et Réparation du Chromosome, CEA, INSERM, UMR 967, IRCM, Université Paris-Saclay, Fontenay-aux-Roses, France

<sup>6</sup> Sorbonne Universités, Theoretical Physics for Condensed Matter Lab, UPMC Université Paris 06, Paris, France

<sup>7</sup> CNRS, UMR 7600, Paris, France

\*Corresponding author. Tel: +33 1 44 27 45 40; E-mail: mozziconacci@lptmc.jussieu.fr

\*\*Corresponding author. Tel: +33 1 40 61 33 25; E-mail: romain.koszul@pasteur.fr

organization. Recent studies have unveiled cell-cycle stage-specific genome-wide topological variations in bacteria, yeast, fly, and mammals (Naumova *et al*, 2013; Guidi *et al*, 2015; Marbouty *et al*, 2015; Hug *et al*, 2017). As expected, in all species the largest reorganization transition is associated with SC condensation, a fundamental process occurring concomitantly to their individualization, and facilitating their proper segregation.

Pioneer studies on yeasts proved essential to study these processes. Mutations in cell-division cycle (*cdc*; Hartwell *et al*, 1973) genes can block the cell cycle progression, enabling the study of global and/or local chromosome reorganization at specific cycle phases (Hartwell *et al*, 1973; Guacci *et al*, 1994; Sullivan *et al*, 2004; Renshaw *et al*, 2010; Rock & Amon, 2011). The evolutionary conserved structural maintenance of chromosomes (SMC) proteins bind to chromosomes and modify their structure in spatially and temporarily regulated manner during the cell cycle (Aragon *et al*, 2013; Uhlmann, 2016). Cohesins, such as Scc1, promote SC cohesion during DNA replication (Blat & Kleckner, 1999; Glynn *et al*, 2004) and get cleaved at the metaphase-to-anaphase transition (Uhlmann *et al*, 1999). At the same time, condensins such as Smc2 are loaded onto SCs to facilitate their segregation (Renshaw *et al*, 2010; Stephens *et al*, 2011; Hirano, 2012). In fission yeast, the binding of SMCs modifies the level of chromosome compaction defined as the ratio between long (> 10 kb)- and short-range (< 10 kb) contacts at specific loci (Mizuguchi *et al*, 2014; Kim *et al*, 2016).

While Hi-C studies on mammalian and *drosophila* cells have confirmed this compaction change and provided important insights on the organization of mitotic chromosomes' internal structure (Naumova *et al*, 2013; Hug *et al*, 2017), no comprehensive analysis of the 4D dynamics of the chromosomes during an entire eukaryotic cell cycle has been achieved. To explore new chromosomal structural features over the cell cycle progression, we analyzed the internal folding and overall organization of *S. cerevisiae* genome over 15 synchronized time points and the role of cohesin and condensin using Hi-C (Dekker *et al*, 2002; Lieberman-Aiden *et al*, 2009). This analysis provides a broad overview and in-depth insight on SMC-dependent structural transitions resulting in chromosome individualization and segregation, including a potential role for a condensin-dependent loop in contributing to the segregation of the rDNA cluster.

## Results

### Comparison of chromosome contact maps of synchronized cells

Hi-C libraries were generated from cell cultures synchronized in G1 with elutriation (Marbouty *et al*, 2014) and/or arrested at different stages of the cell cycle through thermosensitive (ts) *cdc* mutations (Fig 1A; Hartwell *et al*, 1973). After sequencing, the corresponding normalized genome-wide contact maps were computed (bin: 5 kb; Fig 1C and D, left panels; Fig EV1; Materials and Methods; Cournac *et al*, 2012).

These 2D maps were translated into 3D representations to visualize the main folding features (Lesne *et al*, 2014; e.g., centromeres and telomeres clustering in G1; Figs 1B and EV1). These 3D structures are average representations of the contact frequencies quantified over a population of cells and therefore do not represent the

exact structure found in individual cells. For instance, on these 3D representations all the telomeres loosely cluster together. In a single nucleus, telomeres rather form small groups scattered all around the nuclear membrane (Taddei & Gasser, 2012). Since in different cells the composition of these clusters differs, all telomeres end up being regrouped together in the average 3D structure that reflects the population average of contacts. In addition, they are not polymer models and cannot be interpreted as such. Nevertheless, these representations conveniently highlight important structural features not readily apparent in the 2D maps (Mercy *et al*, 2017).

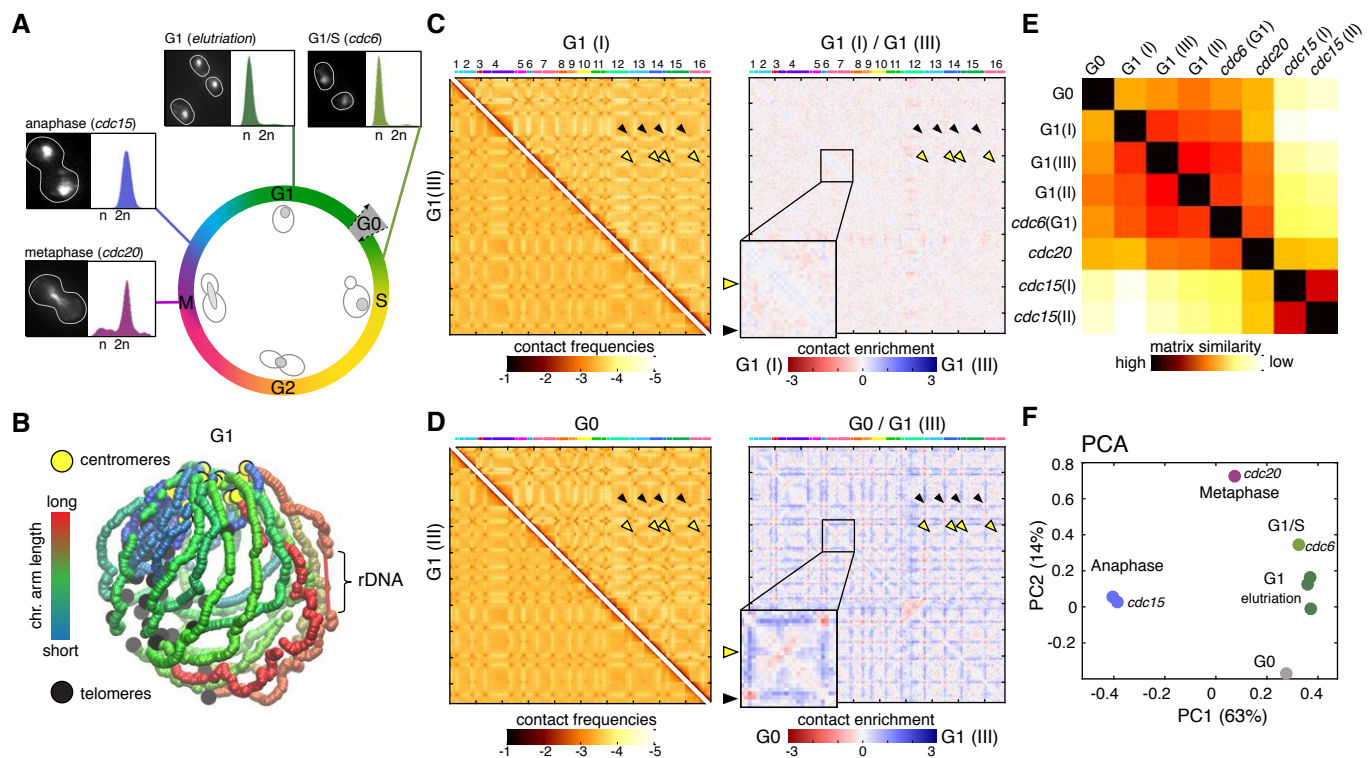
The differences between two conditions were determined by computing the log-ratio between the maps (bin: 5 kb; Fig 1C; Materials and Methods). The color scale reflects the variations in contact frequency for each bin between two different contact maps. The ratio of contact maps generated from two independent G1 cell populations (experimental replicates) displays a relatively homogenous white (i.e., null) signal, corresponding to little differences between them (Fig 1C, right panel). These minor variations between the maps result in occasional faint colored areas and reflect experimental noise (Appendix and Materials and Methods). On the other hand, the ratio between exponentially growing G1 and quiescent G0 cells contact maps (Fig 1D, right panel) shows a strong difference in inter-telomere contact frequencies, reflecting the formation of the telomeres hyper-cluster characteristic of the G0 metabolic state (Guidi *et al*, 2015; Fig 1D, black arrowheads).

Multiple maps can also be compared altogether by computing their pairwise distance matrix, showing that the genome organization of cells in anaphase (*cdc15*) differs the most compared to other time points (Fig 1E; Materials and Methods). The overall similarities/differences between datasets can then be summarized using principal component analysis (PCA; Fig 1F). This 2D representation shows that the experimental duplicates (such as G1, or anaphase *cdc15*) clustered together, while the distance increases progressively between G1 (obtained with either elutriation or *cdc6* ts mutant), metaphase (*cdc20*), and the distant anaphase (*cdc15*) datasets.

Altogether, these comparisons highlight major changes in chromosome higher-order architecture taking place in cells progressing throughout the cell cycle into metaphase and anaphase.

### Cohesin-mediated compaction during S phase

To decipher the chromosome structural changes that take place during replication, synchronized G1 cells were released into S phase and Hi-C maps generated for six time points sampled from two independent kinetics (Figs 2A and EV2; Materials and Methods). The PCA reveals a progressive structural evolution from G1 to late S/G2 phase (Fig 2B). The dependency of the contact probability *P* on genomic distance reflects the chromosome compaction state (Lieberman-Aiden *et al*, 2009; Naumova *et al*, 2013; Mizuguchi *et al*, 2014). The *P*(*s*) shows a gradual and consistent enrichment in long-range intra-chromosomal contacts (> 20 kb) with respect to short-range (< 10 kb) during replication (Fig 2C). This compaction change is absent when replication is impaired, for instance, in the absence of the replication-checkpoint regulator *cdc6* (Piatti *et al*, 1995), even though cells enter mitosis and engage into segregation of non-replicated chromosomes (Fig 2D, left panel). The progressive increase in long-range contacts stops with the completion of S



**Figure 1. Comparison of genome structures recovered from five synchronized stages over the cell cycle.**

- A Overview of the different synchronization time points with corresponding FACS profiles and representative images of DAPI-stained cells.
- B 3D average representation of the Hi-C contact map of a yeast G1 population. The color code reflects chromosomal arm lengths, and centromeres, telomeres, and rDNA are highlighted.
- C, D Comparison of contact maps. The 16 yeast chromosomes are displayed atop the maps. Black arrowheads: inter-telomere contacts. Yellow arrowheads: inter-centromeric contacts. Left panels: Hi-C maps obtained from two G1 cell populations synchronized independently (C) and from G1 and G0 populations (D). Brown to yellow color scales reflect high to low contact frequencies, respectively (log10). Right panels: log-ratio between each pair of maps. Insets display magnifications of chr4. Blue to red color scales reflect the enrichment in contacts in one population with respect to the other (log2).
- E Pairwise Euclidean distances between contact maps of populations of G0, G1 either synchronized with elutriation or blocked using a *cdc6* mutant, metaphase (*cdc20* mutant), and anaphase (*cdc15* mutant) cells. Color code: contact map similarity.
- F Principal component analysis (PCA) of the distance matrix in (E).

phase, when it reaches the level observed in cells arrested at the G2/metaphase transition (G/M) with the microtubule-depolymerizing drug nocodazole (Jacobs *et al*, 1988; Fig 2D, middle panel). The crossing of the  $P(s)$  slopes from the early to late replication time points occurs around 10–20 kb (Fig 2C, highlighted in gray), a window within the range of the spacing reported between cohesin binding sites (~11 kb on average; Glynn *et al*, 2004), suggesting that this change in compaction could be due to cohesin activity. In agreement with the key role of cohesin in sister-chromatid folding during replication, *Scc1* depletion using an auxin-inducible degron *scc1-aid* strain prevents the enrichment in long-range contacts in S/G2 (Fig 2D, right panel). This result supports the hypothesis that distant regions enriched in cohesin are tethered together, resulting in chromatin loops (Guillou *et al*, 2010).

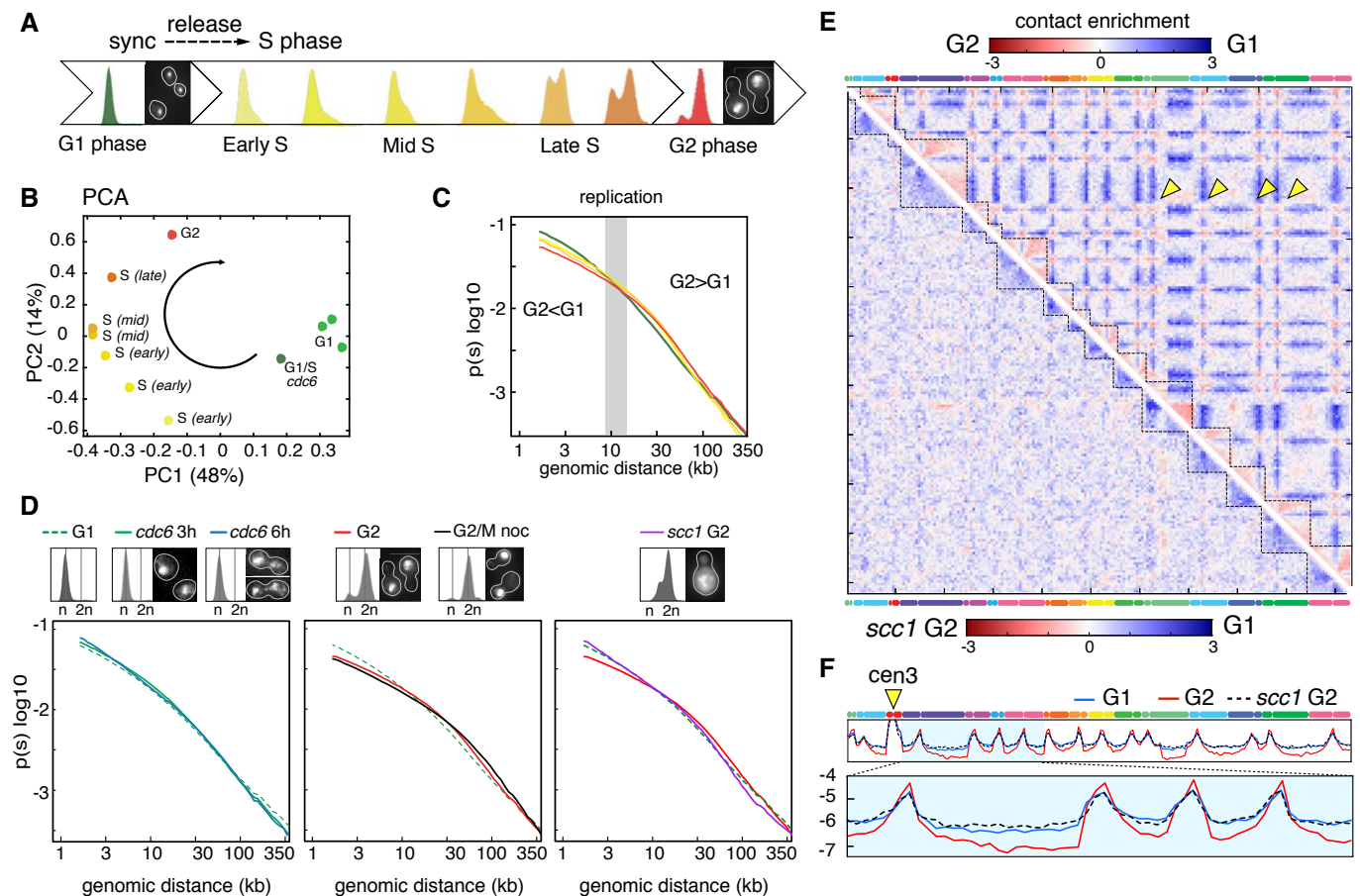
#### Chromosome compaction is concomitant with chromosome individualization

The *Scc1*-dependent compaction occurs concomitantly with a gradual individualization of the SC pairs throughout replication, as shown by the overall increase in the ratio between intra- and

inter-chromosomal contacts from  $63 \pm 10\%$  in G1 (six time points) to  $73 \pm 4\%$  in S/G2 (four time points) and illustrated by the ratio between G1 and G2 maps (Fig 2E, top right ratio). In sharp contrast to this overall decrease in inter-chromosomal contacts, the centromeres of different chromosome tend to strongly cluster in G2. In the absence of the cohesin *Scc1*, intra-chromosomal contacts in G2 cells decrease to levels similar to or even below G1 (Fig 2E, bottom left ratio), while the major binding sites for cohesin (i.e., centromeres; Glynn *et al*, 2004) also exhibit a reduced level of contacts (Fig 2F; Appendix Fig S1). These results suggest that cohesins affect the genome organization through the gradual compaction of SC, the clustering of centromeres, and chromosome individualization. Although yeast chromosomes are shorter than mammalian chromosomes, they similarly change their internal conformation and individualize themselves prior to entering metaphase.

#### Spatial resolution of the replication timing program

In budding yeast, replication initiates at discrete autonomously replicating sequences (ARSs; Brewer & Fangman, 1987). ARSs display

**Figure 2. Dynamic reorganization of chromosomes during replication.**

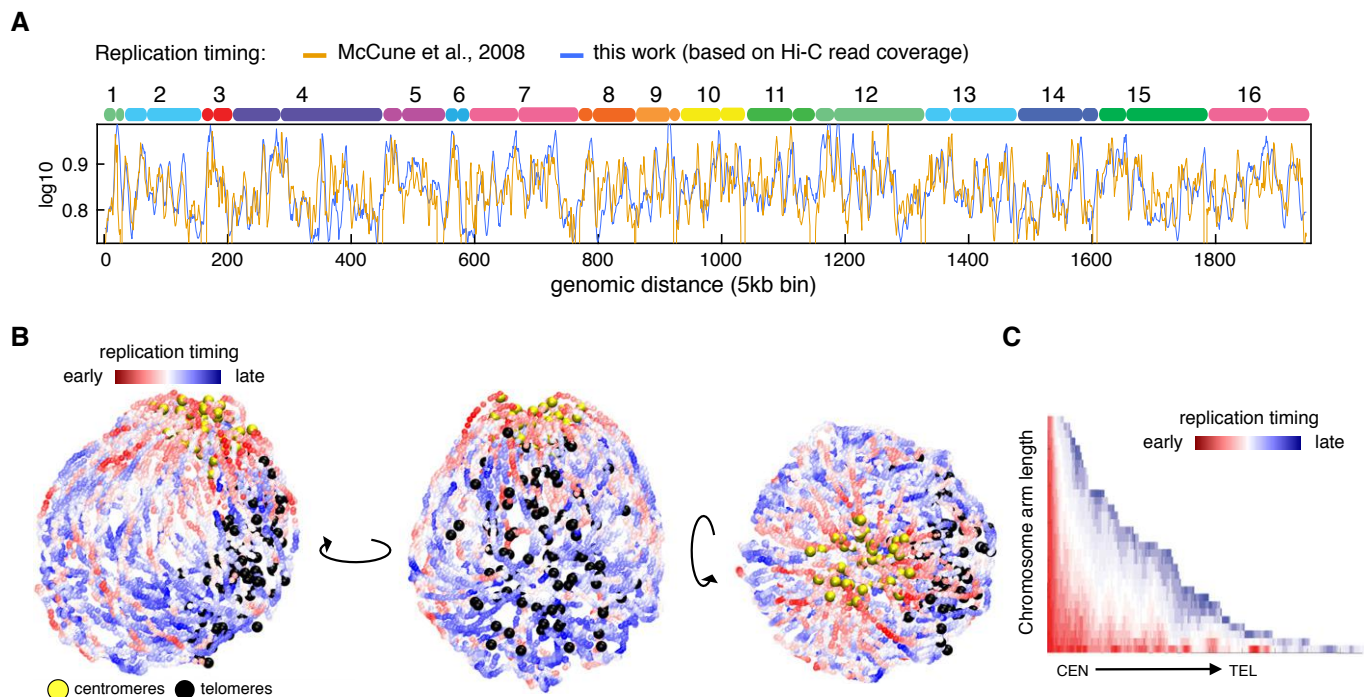
**A** FACS profiles and representative DAPI-stained cells of G1 synchronized cells released in S phase.  
**B** PCA of the distance matrix between the contact maps of the population displayed in (A).  
**C**  $P(s)$ , that is, average intra-chromosomal contact frequency  $P$  between two loci with respect to their genomic distance  $s$  along the chromosome (log-log scale) during replication (color code identical to FACS profiles and PCA).  
**D** Left panel:  $P(s)$  of replication-defective cells (*cdc6* thermosensitive mutant). G1 elutriated cells were released for 3 h and 6 h in non-permissive conditions. The corresponding FACS profiles show no S-phase progression. Middle panel:  $P(s)$  of cells that completed replication. G1 elutriated cells were released in S phase in the absence or presence of nocodazole (G2/M noc). Right panel:  $P(s)$  of cohesin-depleted (*scc1* G2) and nocodazole-arrested cells.  
**E** Log-ratio of contact maps between G2 and G1 cells (top right) and *scc1* G2 and G1 cells (bottom left). Blue to red color scales reflect the enrichment in contacts in one population with respect to the other (log2). Yellow arrowheads: inter-centromere contacts.  
**F** Normalized contact frequencies between chr3 centromere (cen3; yellow arrowhead) and the rest of the genome for G1, G2, and *scc1* G2.

partially stochastic activation, with only a subset of origins activated early during S phase. The distribution of early origins is uneven, with an enrichment in pericentromeric regions, and a depletion in subtelomeric regions. The genome-wide pattern of ARS activation timing defines a population-average replication timing program (Raghuraman *et al*, 2001). To investigate the link between genome organization and replication timing, the read coverage of the Hi-C libraries was used to compute the replication timing profile of the cell population for each of the time point, and follow their progression through S phase. The average profile correlates well with previously published pattern (Raghuraman *et al*, 2001; McCune *et al*, 2008; Fig 3A; Materials and Methods). To visualize the progression of replication on the higher-order architecture of the genome, we colored the 3D structures recovered from three early replication time points according to their replication progression status. The

superimposition of the three structures recapitulates intuitive properties of yeast replication program, with a “replication wave” propagating from the centromeric regions enriched in early origins, through chromosomal arms, and toward the late replicating subtelomeric regions (Fig 3B and C; red and blue signal, respectively).

We also asked whether our data support the proposed co-localization of adjacent early replication origins (Kitamura *et al*, 2006; Knott *et al*, 2012; Saner *et al*, 2013). We found a statistically significant enrichment in contacts between these positions and their surrounding regions, but whether it results from an active co-localization or from their positioning in the pericentromeric regions co-localized due to the Rab1 organization remains unclear (not shown). More analyses are required to solve this question and integrate the different observations.



**Figure 3. 3D replication profile.**

**A** Comparison of replication profiles of the synchronized populations used for the analysis displayed in Fig 2G. The read coverage of raw Hi-C libraries reflects the replication progression throughout S phase, plotted along the 16 chromosomes of the yeast genome (top axis; blue curve). The replication timing obtained in this study is highly similar to the one from McCune *et al* (2008) (yellow curve).

**B** Superposition of three 3D representations of chromosomes in early replication (I, II, III). The color scale indicates the replication timing. Centromeres and telomeres are highlighted. Different views of the structure are presented.

**C** Pattern of the replication profile for each of the chromosomal arms. The color code reflects the timing of replication.

### Global structural changes during mitotic transitions

After replication, cells progress into mitosis (M phase). During metaphase, microtubules originating from opposite SPBs attach to the kinetochores of the two SCs (London & Biggins, 2014). The anaphase-promoting complex (APC) co-activator Cdc20 is essential for the proper activation of separase, resulting in the cleavage of cohesin and SC segregation in anaphase (Uhlmann *et al*, 1999; Visintin *et al*, 1997; 20). In the absence of Cdc20, cohesins are not cleaved and cells remain blocked in metaphase. Another key player in mitosis progression is the Cdc15 kinase which promotes mitotic exit at the end of anaphase by activating cytokinesis (Rock & Amon, 2011). In the absence of Cdc15, cells are therefore blocked into late anaphase. The higher-order changes in the organization of chromosomes that take place during metaphase and anaphase were investigated using populations of cells synchronized with conditional mutants of *cdc20* and *cdc15*, respectively. Contact maps of *cdc20*-, *cdc15*-, and *cdc15*-arrested cells released into permissive conditions were generated to characterize chromosome reorganization throughout M phase (Figs 4A and EV3; Materials and Methods). PCA shows that the major structural change occurs during mitotic exit and that cells released from the *cdc15* arrest display after 60 min a G1-like genome structure, reflecting the fact that the entire cell cycle is now covered by our analysis (Fig 4B). The *P(s)* reveals a strong increase in short-range contacts (< 10–20 kb) from G2 to anaphase,

exceeding G1 levels which are only restored after anaphase completion (Fig 4C, left panel). This increase in short-range contacts and the accompanying drop in long-range contacts suggest the formation of an elongated, stretched structure. Upon spindle destabilization using the microtubule-depolymerizing drug nocodazole in *cdc15*-arrested cells (*cdc15* noc), the two segregated chromosomal masses get closer as shown by imaging of DAPI-stained cells (Fig 4C, inset; Fig EV4; Materials and Methods), in agreement with former reports (Jacobs *et al*, 1988). In these cells, the stretched chromosomal structure disappears as shown by a *P(s)* that now overlaps the G2 curve (Fig 4C, right panel). Besides the change in *P(s)*, the global contact pattern of *cdc15*-arrested cells remains unaltered following nocodazole treatment (Fig 4D, upper right ratio). Altogether, these results show that microtubule-dependent segregation forces contribute to the stretching the chromosomes in anaphase, possibly in combination with additional constraints resisting this force such as the cohesion of SC arm extremities (see Discussion).

### Nocodazole affects chromosome 12 conformation

Nocodazole is commonly used to synchronize cells at the G2/M transition. We took advantage of having contact maps of *cdc20*-arrested cells in metaphase to compare them with those obtained from nocodazole-arrested cells (Fig EV4; Materials and Methods). The ratio map appeared globally similar, although we noticed in the

presence of nocodazole a small drop in inter-chromosomal contacts (Fig 4D, bottom left ratio). Chromosome 12 (chr12) also presents a peculiar signal at the level of the rDNA cluster (Fig 4E, left panel), with an enrichment in contacts between the two flanking regions of the rDNA cluster in G2/M nocodazole-treated cells compared to

*cdc20*-arrested cells (Fig 4E, right panel). These results indicate that the G2/M nocodazole arrest is associated with a destabilization of the chr12 structure at the level of the rDNA locus. The intra-chromosomal contact increase within chr12 is also accompanied by a global decrease in inter-chromosomal contacts in the presence of

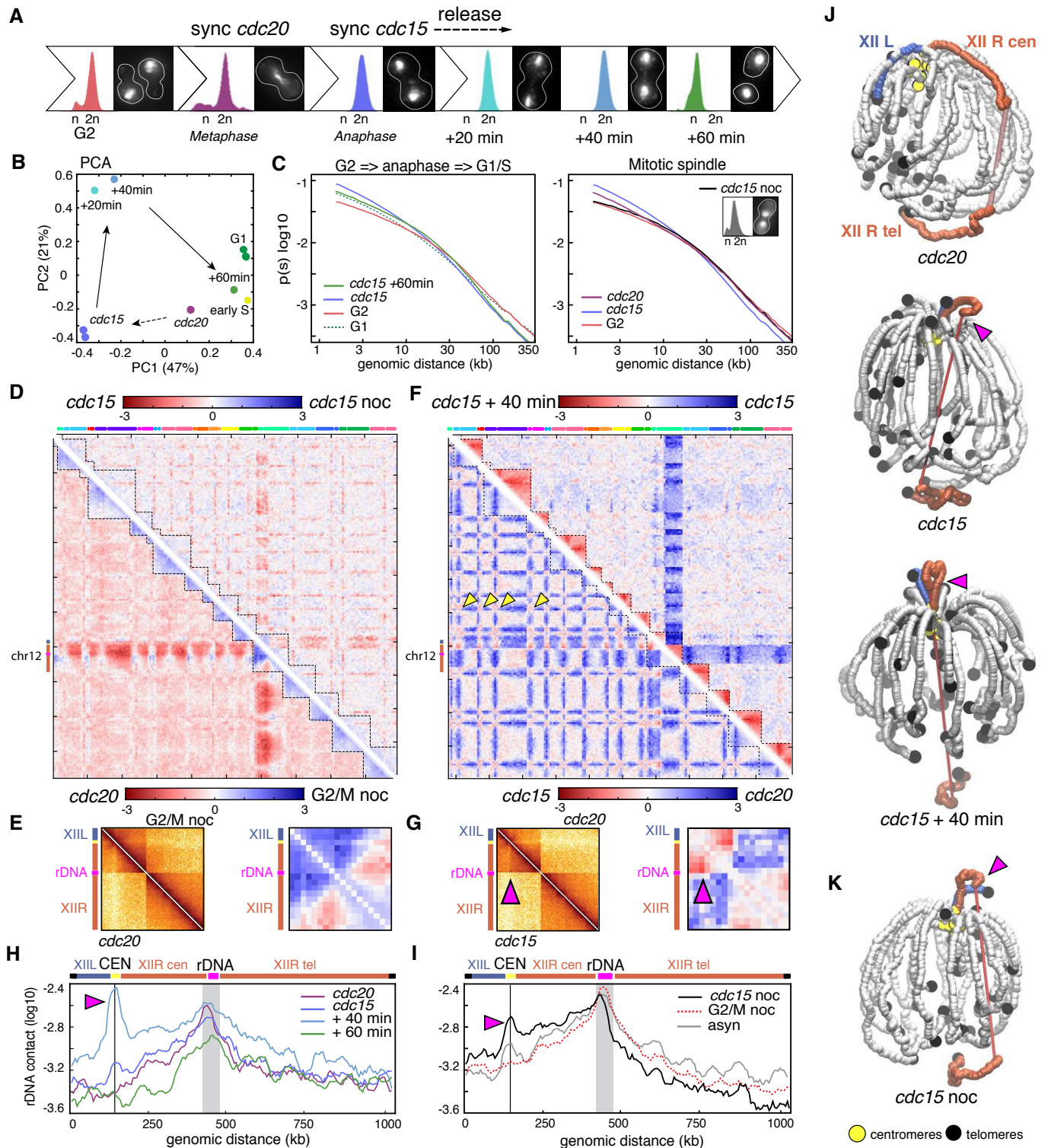


Figure 4.

nocodazole (Fig 4D, bottom left ratio). Remarkably, chr12 organization was not affected when *cdc15*-arrested cells were treated with nocodazole (Fig 4D, top right ratio). Altogether, these observations point to a role for the microtubule array in maintaining the organization of the nucleolus inside the nucleus, before its segregation in anaphase. In summary, while chromosome structures are overall similar in cell synchronized in G2 by nocodazole or in a *cdc20* ts mutant, nocodazole-arrested cells present a slightly different nucleolus structure (and, by extension, chr12). One interpretation could be that the condensation of the rDNA is not yet completed in G2/M nocodazole arrest and that as a result, rDNA flanking regions are freer to contact each other's.

### Chromosome 12 looping during anaphase

The comparison of *cdc15* and *cdc20* maps shows an increase in centromere clustering in anaphase, leading to the formation of a prominent polymer brush structure (Daoud & Cotton, 1982; Fig 4F, bottom left ratio, yellow arrowheads). Such increase is in agreement with the role of condensin in forming a "spring" of chromatin at pericentromeric regions at the metaphase-to-anaphase transition (Stephens et al, 2011). Surprisingly, a peculiar loop pattern appears on chr12 in *cdc15*-arrested cells, bridging the centromere and the centromere-proximal left flanking region of the rDNA cluster (see pink arrowheads in Fig 4F, G and H). Upon release from the *cdc15* arrest, the telomere-proximal right flanking region of the rDNA cluster becomes strongly isolated from the rest of the genome (Fig 4F, upper right ratio; *cdc15*+40 min), while the contacts of the centromere-rDNA loop intensify (Fig 4H; *cdc15*+40 min). After completion of mitosis and re-entry in interphase (*cdc15*+60 min), the loop disappears (Fig 4H). Interestingly, this loop can be seen in asynchronous populations while it is only present in anaphase (Fig 4I). 3D representations illustrate the dramatic reorganization of chr12 and the formation of the loop bridging centromeric region and the rDNA (Fig 4J, pink arrowheads). Microtubules are not required to maintain this loop in anaphase, since it remains present in *cdc15*-arrested cells treated with nocodazole (Fig 4D, upper right ratio; Fig 4I and K), suggesting that the left flanking region of the rDNA is physically bound through an unknown mechanism to the

centromeric regions. These results complement imaging studies showing that the rDNA exhibits a dense, line-like shape that extends throughout the nucleus at anaphase (2.1 SD, 0.2  $\mu$ m; Sullivan et al, 2004).

### Condensin promotes dramatic reorganization of chromosomes during anaphase

The proper condensation and segregation of the rDNA cluster requires the nucleolar release of the Cdc14 phosphatase. Cdc14 mediates a shutdown of rDNA transcription, facilitating the loading of the Smc2 condensin and hence the condensation of the cluster (Yoshida et al, 2002; D'Amours et al, 2004; Sullivan et al, 2004, 14; Machin et al, 2006; Clemente-Blanco et al, 2009). In addition, topoisomerase II (Top2), which decatenates the intertwining structures that appear between SCs during replication, is also required for rDNA segregation to proceed (Sullivan et al, 2004; D'Ambrosio et al, 2008; Baxter et al, 2011; Leonard et al, 2015). We investigated the influence of those factors on the 3D structure of the rDNA locus during anaphase (Figs 5A and EV5; Materials and Methods).

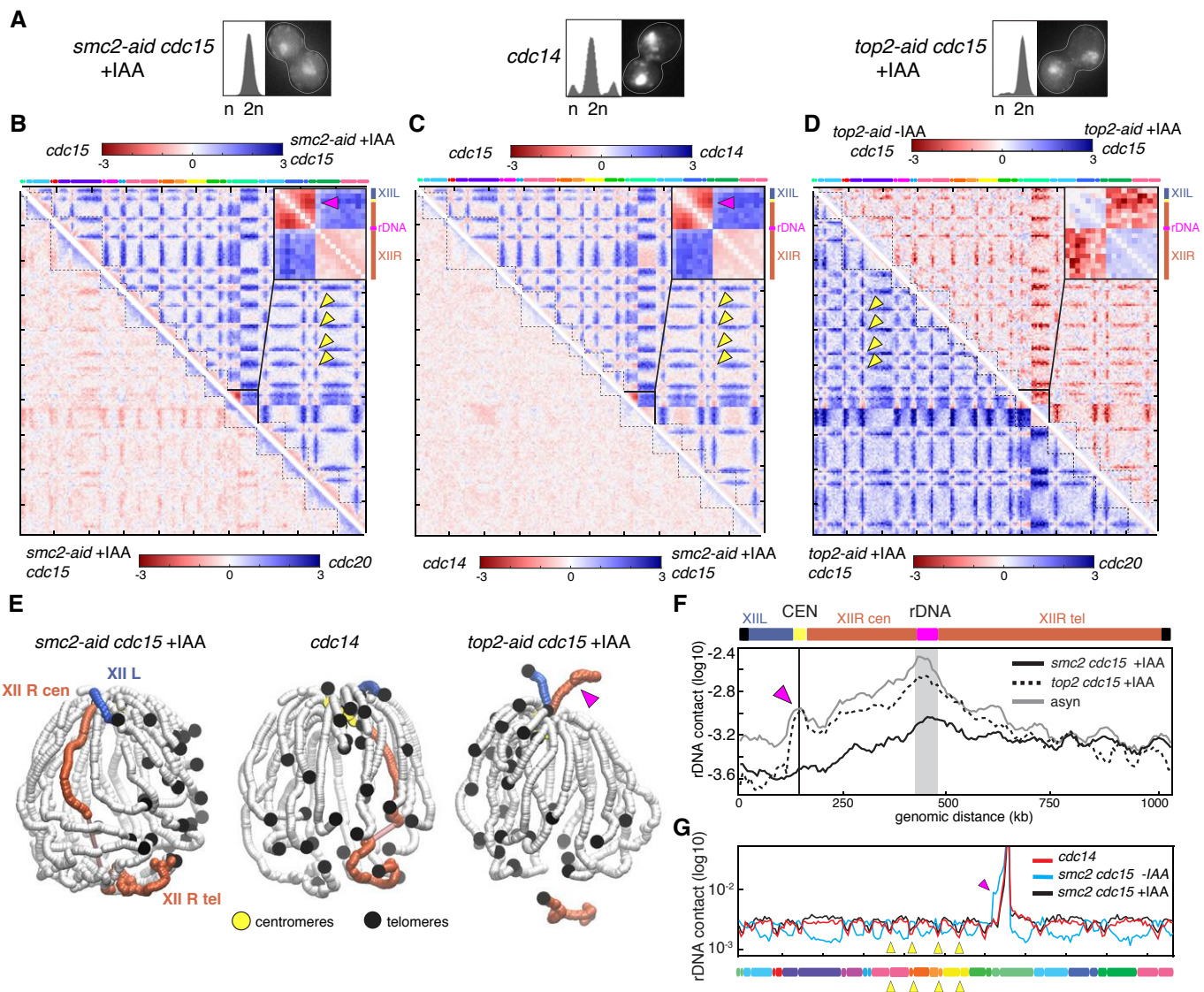
First, Smc2 depletion in *smc2-aid cdc15*-arrested strain affects anaphase genome organization by (i) reducing centromere clustering and (ii) suppressing the formation of the rDNA loop, with a resulting contact map highly similar to the *cdc20* map (Fig 5B, bottom left ratio). Therefore, condensins are responsible for the observed increase in inter-centromere contacts at anaphase compared to metaphase (Fig 4F, bottom left ratio), while they are also required for the formation of the loop bridging the centromere of chromosome 12 with the rDNA cluster (two loci enriched in condensin deposition). The *smc2 cdc15* and *cdc14* maps are strikingly similar (Fig 5C, bottom left ratio). The 3D representations of *smc2 cdc15* and *cdc14* cells (Fig 5E) and the rDNA contact plots with the rest of chr12 (Fig 5F) illustrate the loss of the rDNA loop in the absence of Smc2 and/or Cdc14. In addition to this effect, both mutants also display the same decrease in centromere clustering compared to *cdc15* cells (Fig 5C, upper right ratio; Fig 5G), pointing at their functional relationship on the same pathway.

The organization of the genome was also compared in *cdc15*-arrested cells in the presence or absence of Top2, *top2-aid*

**Figure 4. Dynamic reorganization of chromosomes during mitosis.**

- A FACS profiles and representative DAPI-stained cells of synchronized and/or released populations, from G2 until re-entry in G1/S.
- B PCA of the distance matrix between the contact maps of the populations described in (A).
- C Left panel:  $P(s)$  of cells in G1, G2, and anaphase (*cdc15*) and released from a *cdc15* arrest (*cdc15*+60 min). Right panel:  $P(s)$  of G2, *cdc20*-, and *cdc15*-arrested cells in the absence or presence of nocodazole (*cdc15* noc).
- D Log-ratio of contact maps. Bottom left: ratio between cells arrested in metaphase (*cdc20*) or at the G2/M transition with nocodazole (G2/M noc). Top right: ratio of cells blocked in anaphase and treated or not with nocodazole (*cdc15* noc and *cdc15*, respectively). Blue to red color scales reflect the enrichment in contacts in one population with respect to the other (log2).
- E Left: chr12-normalized contact maps of cells arrested at the G2/M transition and *cdc20*-arrested cells. Right: magnification of the log-ratio map from (D, bottom left).
- F Log-ratio of contact maps. Bottom left: log-ratio between *cdc20*- and *cdc15*-arrested cells. Top right: log-ratio of *cdc15*-arrested and *cdc15*-released (40 min) cells. Blue to red color scales reflect the enrichment in contacts in one population with respect to the other (log2). Yellow arrowheads: inter-centromere contacts.
- G Left: chr12-normalized contact maps in *cdc20*- and *cdc15*-arrested cells. Right: magnification of the log-ratio map from (F, bottom left). Pink arrowheads point at the right arm anaphase loop.
- H, I Distributions of intra-chromosomal contacts made by a 20-kb cen-proximal rDNA flanking region (highlighted in gray) with the rest of chr12 in *cdc20*-, *cdc15*-, *cdc15*-released (+40 min, +60 min), nocodazole-treated (G2/M noc, *cdc15* noc), and asynchronous (asyn) cells. Schematic representations of chr12 are displayed atop the graphs. Gray lines indicate centromere position. Pink arrowheads point at the right arm anaphase loop.
- J 3D representations of the contact maps from *cdc20*- and *cdc15*-arrested and *cdc20*- and *cdc15*-released (+40 min) cells. The right (XIIR) and left (XIIIL) arms of chr12 are highlighted in red and blue, respectively. Pink arrowheads point at the right arm anaphase loop. Centromeres and telomeres are highlighted.
- K 3D representation of the contact map from *cdc15* noc cells. Pink arrowhead points at the right arm anaphase loop.





**Figure 5. The anaphase rDNA loop is condensin-dependent.**

- A** FACS profiles and representative DAPI-stained cells of cells blocked in anaphase, in the absence of condensin (*smc2-aid cdc15* +IAA and *cdc14*) or topoisomerase 2 (*top2-aid cdc15* +IAA).
- B–D** Log-ratio of contact maps. Yellow arrowheads: inter-centromere contacts. The pink arrowheads point at the right arm anaphase loop on chr12. Insets display magnification of the chr12 ratio map. (B) Ratio map between (bottom left) *cdc20* and *smc2-aid cdc15* +IAA cells and between (top right) cells blocked in anaphase with our without condensin depletion (*cdc15* and *smc2-aid cdc15* +IAA). (C) Ratio map between (bottom left) *cdc14* and *smc2-aid cdc15* +IAA cells and between (top right) *cdc14* and *cdc15* cells. (D) Ratio map between (bottom left) *top2-aid cdc15* +IAA and *cdc20* cells and (top right) *top2-aid cdc15* -IAA and *top2-aid cdc15* +IAA cells.
- E** 3D representations of the contact maps from *smc2-aid cdc15* +IAA-, *cdc14*-, and *top2-aid cdc15* +IAA-arrested cells. The right (XIIR) and left (XII L) arms of chr12 are highlighted in red and blue, respectively. Pink arrowhead points at the right arm anaphase loop.
- F** Distribution of intra-chromosomal contacts of a cen-proximal rDNA flanking region (highlighted in gray) with the rest of chr12 in *smc2 cdc15*, *top2 cdc15*, and asynchronous (asyn) cells. Pink arrowhead points at the right arm anaphase loop.
- G** Normalized contact frequencies between the left rDNA flanking region (50 kb) and the rest of the genome in *cdc15 smc2-aid* (-IAA) and *cdc15 smc2-aid* (+IAA) cells. Yellow arrowheads point at a subset of centromeric positions. Pink arrowhead points at the right arm anaphase loop.

*cdc15*-arrested (Fig 5D, upper right ratio; Fig 5E and F; Materials and Methods). Top2-depleted cells display a strong decrease in contacts between the telomere–proximal region of chr12R and the rest of the genome (including chr12L). The signal is consistent with the essential role played by Top2 in rDNA segregation, showing that

the non-segregated regions are isolated from the segregated chromosomal sets. The comparison between *top2 cdc15* and *cdc20* cells reveals an enrichment in contacts at centromeres and the persistence of the centromere–rDNA loop in the Top2 mutant (Fig 5D, bottom left ratio). These results indicate that the formation of these

condensin-dependent structures in anaphase is independent from the decatenation and/or the segregation of the rDNA cluster.

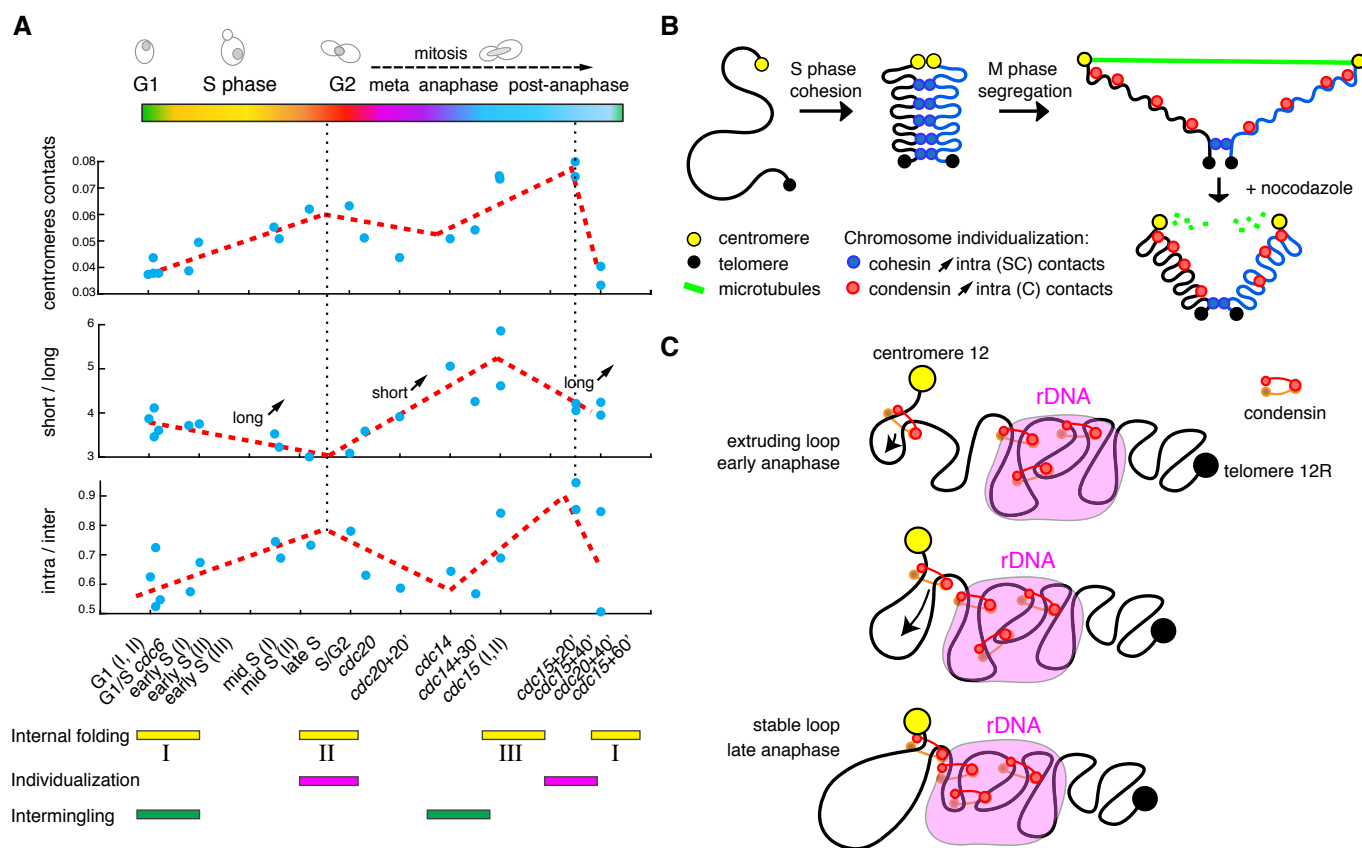
## Discussion

This study consists of an experimental and analysis framework to systematically investigate and compare chromosome folding and organization at different stages of the cell cycle. We applied Hi-C to populations of cells synchronized at different points of the cycle, generating genome-wide, 5-kb-resolution contact maps which unveil their average 3D genome organization. The global influence of cohesin, condensin, and topoisomerase 2 has been investigated in the corresponding mutants, as well as the effects of the microtubule-depolymerizing drug nocodazole. Comparative approaches between contact maps provided a global view of the structural transitions between the different stages of the cycle, some expected, such as

chromosome compaction during replication, and others that had not been described before, such as topological structures involving the rDNA cluster.

An overview of chromosome structural changes during the cell cycle can be summarized from centromere contacts, intra-/inter-chromosomal contact ratio, and short-/long-range contact ratio computed for each of the time points (Fig 6A).

Centromere clustering gradually increases during the cell cycle, through the establishment of sister-chromatid cohesion during replication, and through condensin-dependent clustering during anaphase (Fig 6A, upper panel). A potential consequence of this increased clustering in anaphase could be the generation of a stronger polymer brush, that is, the mechanical phenomenon that leads to the self-organization of a polymer tethered to a surface into stretched, non-intermingling structure (de Gennes, 1987). Interestingly, the strengthening of the polymer brush organization could consequently contribute to chromosome individualization during



**Figure 6. 4D reorganization of the yeast genome.**

- A** Dynamics of centromere contacts (top panel), Short-/long-range contact ratio (middle panel) and intra-/inter-chromosomal contact ratio (bottom panel) for each of the 20 time points (blue dots; see bottom x-axis) during the cell cycle. The three folding states (I, II, and III; Fig EV4) identified in the analysis are indicated under the panels, as well as interpretation with respect to individualization status.
- B** Illustration of the three chromatin folding states characteristic of each of the cell cycle phases. Establishment of sister-chromatid (SC) cohesion during S phase increases intra-SC long-range contacts and leads to the individualization of the replicated chromosomes. Then during M phase, the two sisters are segregated and each chromatid (C) individualized thanks to the action combination of cohesin cleavage, condensin loading and spindle elongation. The chromosomes display a stretched internal structure, which relaxed upon destabilization of the spindle with nocodazole.
- C** Model of loop extrusion generating the condensin-dependent loop formation between the centromere and the rDNA cluster, two regions enriched in condensin deposition. A loop formed in between the centromere and the rDNA cluster may extend until it reaches these two discrete positions, and stall because of mechanic impediment blocking further extrusion.

anaphase. The intra-/inter-chromosomal contact variations reflect the successive phases of chromosome individualization and intermingling, with individualization taking place during replication (cohesin-dependent) and during anaphase (spindle-dependent; Fig 6A, bottom panel). The intra-/inter-chromosomal contact ratio correlates strongly with centromere clustering ( $c = 0.72$ ,  $p = 10^{-4}$ ), with both ratios peaking during anaphase exit.

Short-/long-range contact ratio recapitulates the three different internal folding (I, II, and III) states of chromosomes (G1, G2, and anaphase; Fig 6B, middle panel). These three states can be determined based on a quantitative analysis of the significance of changes between  $P(s)$  curves obtained using several replicates in different phases of the cycle (Fig 7). During replication, cohesins mediate the compaction of chromosomes from state I to II. The chromosomes are then stretched by the mitotic apparatus during anaphase (state III) before returning to state I in G1. The mechanical constraint imposed by the anaphase spindle appears responsible for the state III stretching, as a nocodazole treatment results in relaxation of chromosomes, which switch back to state II. Imaging of the two sets of segregated chromosome during nocodazole treatment supports this spring relaxation effect, with the two masses being brought back together upon the depolymerization of microtubules. The nature of the mechanical constraints remains unknown, but it is tempting, in light of our observation of chr12 behavior (below), to propose a role for condensins in actively promoting this movement. In this scenario, condensins could favor the segregation of sister chromatids by pulling the chromosomes toward the centromere cluster. As a result, the loss of microtubule and tethering to the SPB may lead cohesin to actively pull back the segregated region together. We anticipate that whether condensins play an active role

in the segregation of chromosomes in addition to the pulling force imposed by the microtubule spindle will be thoroughly investigated in the years to come.

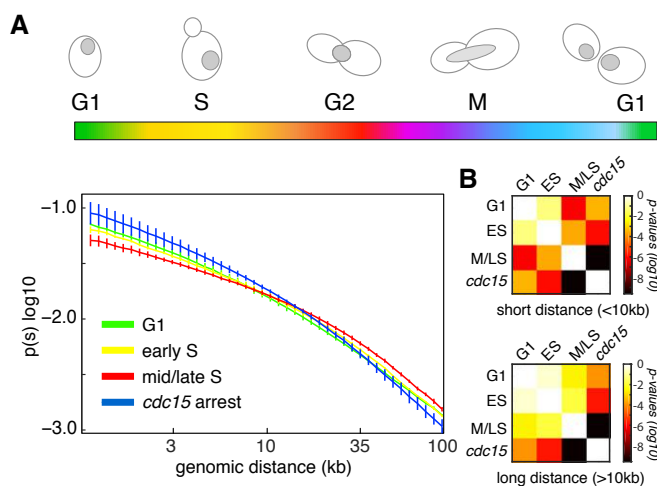
In addition, we also show that the two main regions of condensin deposition, *that is*, the centromeres and the rDNA locus, are bridged during anaphase through a condensin-dependent mechanism resulting in a loop-like structure on the right arm of chromosome 12. Whether this structure is systematically found in all cells, or only in a subset of the population, remains to be determined through single-cell imaging approaches such as FISH analysis. Although the precise mechanisms of formation remain unknown as well as its functional importance, we show that the setting up of the loop depends on condensin. Several mechanisms can be envisioned for the generation of this loop. One possibility is that starting from regions with a high condensin density, an active mechanism such as DNA extrusion through the action of condensins would pull the centromere and the rDNA cluster together (Fig 6C). Condensin depletion (leading to disruption of the loop) is associated with segregation defects. Overall, this structure therefore appears to *de facto* play a role in the segregation of the rDNA cluster, potentially through the application of a force that would drag the rDNA region to the centromere cluster before the completion of anaphase. A consequence of this model, would be that a similar loop extrusion mechanism could facilitate the segregation of other chromosomes as well. In this case, one or more loops could actively facilitate the segregation of large regions of chromosomes toward the tethered centromeres, down the telomeric regions. Chromosome 12, in this scenario, would appear as an exception with the large rDNA cluster generating a physical barrier in the middle of the right arm that is not present in other chromosomes. More experiments are nevertheless needed to investigate this proposed role. Yeast chromosome 12 could therefore prove a convenient model to study the action of loop extrusion mechanism (Alipour & Marko, 2012).

The importance of the rDNA loop remains to be further characterized as well as its similarity with loops found in other eukaryotic species. Overall, our exhaustive dataset opens new avenues for the comprehensive analysis of the 3D chromosome choreography during replication and segregation and brings to light new perspectives regarding these fundamental processes.

## Materials and Methods

### Media and culture conditions

All strains were grown in rich medium (YPD: 1% bacto peptone (Difco), 1% bacto yeast extract (Difco), and 2% glucose), except for YKL051 (*MET3-HA-CDC20*) that was grown in synthetic complete medium deprived of methionine (SC: 0.67% yeast nitrogen base without amino acids (Difco), supplemented with a mix of amino acids, uracil and adenine, and 2% glucose). Cells were grown at either 30°C or 23–25°C (the later temperature corresponding to the permissive temperature of the conditional thermosensitive mutations *cdc6-1*, *cdc14-3*, and *cdc15-2*; see below for details). Dataset corresponding to the quiescent state (G0) comes from already published data by Guidi *et al* (2015) and was obtained by carbon source exhaustion. All strains are described in Table EV1.



**Figure 7. Variation in  $P(s)$  for different phase of the cell cycle.**

- A  $P(s)$  for four different time points along the cycle. Each curve represents the average between three replicates with error bars corresponding to the standard deviation.
- B To assess the statistical significance of the differences between short (resp. long)-range contacts between these three time points, we computed a  $P$ -value for each pairwise comparison between two time points using Wilcoxon signed-rank test. The two distributions to compared were built by aggregating all the data points below (resp. above) 10 kb for the three replicates for each time point.

### Elutriation (recovery of G1 cells)

To recover G1 daughter cells, the exponentially growing cultures were elutriated—a physical method of synchronization, used to separate cells according to their density and sedimentation velocity (see Appendix Supplementary Methods; Marbouty *et al*, 2014). The G1 daughter cells recovered through elutriation were suspended in fresh YPD at 30°C for 30 min, so they could recover from the elutriation procedure (i.e., stay in PBS). To minimize the potential variability introduced by the age heterogeneity of the bulk population, G1 daughter cells were used as starting point for all cell cycle synchrony and in combination with genetic and chemical synchronization methods (see below).

### Release into S phase

G1 elutriated cells were released into S phase to analyze genome conformation during this stage.  $2 \times 10^9$  G1 cells—originating from the same elutriated fraction to minimize heterogeneity in replication initiation—were inoculated into 150 ml YPD at 25°C (to slow down replication fork progression). Upon release, the synchronized cultures were sampled every 5 min and the cells analyzed through FACS, revealing an approximate lag of 130 min before replication restart. Therefore, aliquots were cross-linked and processed into Hi-C libraries at 135, 140, 145, 150, 155, 160, and 165 min. The progression of each fraction throughout the S phase (from G1 to G2) was monitored with flow cytometry.

### Synchronization through thermosensitive mutations

Synchronizations using thermosensitive (ts) *cdc* strains (Hartwell *et al*, 1973) were all performed starting from elutriated G1 daughter cells growing in non-permissive temperature conditions designed to arrest the progression of the cycle at specific phases. See Appendix Supplementary Methods for details of synchronization procedures of strains YKL052 (*cdc14-3*), YKL053 (*cdc15-2*), and YKL054 (*cdc6-1*).

### Synchronization through chemical compounds

Chemical synchronization was also performed on elutriated G1 daughter cells.

Synchronization at the G2/M transition was achieved by restarting G1 cells (strain YKL050) in YPD at 30°C for 1 h, followed by the addition of nocodazole (Calbiochem; 15 µg/ml) and incubation for another 2 h at 30°C. Cells arrested in G2/M with nocodazole were either processed into Hi-C libraries, or washed and inoculated in fresh YPD medium at 30°C. The washing of nocodazole allowed G2/M synchronized cells to proceed into M phase (cells sampled after 20, 45, 60 and 90 min were processed into Hi-C libraries).

To investigate the constraints imposed by the spindle during anaphase, elutriated YKL053 cells were elutriated and the recovered G1 daughter cells processed and blocked into anaphase using the *cdc15-2* thermosensitive mutation. A sample of the population was then incubated with nocodazole (15 µg/ml) for 20 min. A sample was released at permissive temperature in the presence of nocodazole for 20 min. Finally, a sample was released at the permissive

temperature for 20 min before being incubated with nocodazole for 20 min.

For synchronization in metaphase, a system allowing induced depletion of *cdc20* was used (MET3-HA-CDC20; strain YKL051). Elutriated G1 daughter cells were restarted in YPD complemented with 50 µg/ml methionine for 5 h at 30°C. Cells arrested in metaphase were split into different aliquots. One sample was immediately processed into a Hi-C library, while two others were washed, suspended in SC medium without methionine, and processed into Hi-C after 20 and 40 min.

To investigate the influence of SMC on chromosome organization, strains carrying auxin-inducible degron (*aid*) versions of Scc1 (strain YKL055) and Smc2 (YKL056) proteins were processed into Hi-C libraries. The degradation of these proteins is induced when auxin (IAA) is added to the medium at a final concentration of 2 mM. Both asynchronous populations of strains YKL055 and YKL056 were elutriated in the absence of IAA. G1 daughter cells were incubated in YPD supplemented with IAA at 30°C. A sample of the YKL055 population (*scc1-aid*) was processed into a Hi-C library in late S/G2 (see Release into S phase). For the YKL056 population (*smc2-aid*), the cells were arrested in late anaphase using the *cdc15-2* mutation also present in the genome of this strain, before being processed into a Hi-C library.

To study the influence of topoisomerase II-mediated decatenation on chromosome organization, we used a strain (YKL057) in which *TOP2* gene is tagged by *aid* (*top2-aid*) and that also carries the *cdc15-2* mutation. An asynchronous exponentially growing culture of YKL057 cells was split into two fractions incubated for 3 h at the non-permissive temperature of 37°C in either the presence or absence of IAA (20 mM). The synchrony of each time point was monitored with flow cytometry and microscopy, and the cells were processed by Hi-C.

### Flow cytometry

About  $5 \times 10^6$  cells were fixed in ethanol 70% and stored at 4°C overnight. Cells were then pelleted, washed, and incubated in sodium citrate 50 mM (pH 7.4) complemented with RNase A (10 mg/ml; Roche) for 2 h at 37°C. Next, Sytox green (2 µM in sodium citrate 50 mM; ThermoFisher) was added and cells incubated for 1 h at 4°C. Flow cytometry was performed on a MACS-Quant Analyzer (Miltenyi Biotec), and data were analyzed using FlowJo X 10.0.7 software (Tree Star).

### Microscopy

Fractions of cells fixed in ethanol 70% and stored at 4°C overnight were pelleted and washed three times for 5 min in 1× PBS. Cells were permeabilized by immersion in 0.2% Triton X-100 (Biosolve) for 5 min. To remove the Triton, cells were pelleted and washed three times in 1× PBS. The liquid was aspirated and cells were suspended in DAPI labeling solution (2 µg/ml in 1× PBS) for 10 min at room temperature. Before imaging acquisition, the labeling solution was aspirated and the cells were washed three times for 5 min in 1× PBS. Cells were imaged at 350 nm excitation wavelength with Nikon fluorescence microscope (Camera Andor Neo sCMOS, software Andor IQ2 2.7.1, LED Lumencor Spectra X).



## Hi-C libraries

Hi-C libraries were generated using the four-cutter enzyme DpnII through a protocol adapted from Belton *et al* (2012). The protocol is detailed in Appendix Supplementary Methods. The resulting libraries were used as template for the Illumina amplification by PE-PCR primers and paired-end-sequenced on the NextSeq 500 or HiSeq 2000 Illumina platform ( $2 \times 75$  or  $2 \times 150$  bp kits; see Table EV2 for details).

## Generation and normalization of contact maps

Raw Hi-C data were processed as follows. PCR duplicates were removed using the 6 Ns present on each of the custom-made adapter and the 2 trimmed Ns. Paired-end reads were mapped independently using Bowtie 2.1.0 (mode: `-very-sensitive -rdg 500,3 -rfg 500,3`) against the *S. cerevisiae* reference genome (S288C). An iterative alignment, with an increasing truncation length of 20 bp, was used to maximize the yield of valid Hi-C reads (mapping quality > 30). Only uniquely mapped reads were retained. On the basis of their DpnII restriction fragment assignment and orientation, reads were classified as either valid Hi-C products or unwanted events to be filtered out (i.e., loops and non-digested fragments; for details, see Cournac *et al*, 2012, 2016). To generate contact matrices for all time points along the cycle, filtered Hi-C reads were binned into units of single restriction fragments, and then, successive fragments were assigned to fixed size bins of either 5 or 50 kb. Bins that exhibited a high contact frequency variance ( $< 1.5$  Standard Deviation or  $1.5\text{--}2$  SD. from the mean) were filtered out for all maps to allow pairwise comparison of the data. On average, around 15 million of valid reads were used to build each contact map. To remove potential biases resulting from the uneven distribution of restriction sites and variation in GC content and mappability, the contact maps were normalized using the sequential component normalization (SCN) procedure (Cournac *et al*, 2012).

## Similarity between contact maps

To assess the similarity between normalized matrices, these were binned at 50 kb and quantile-normalized (Hicks & Irizarry, 2015). We then measured their similarity by computing the Euclidean distance between them. In order to visualize similarities between sets of matrices, we did a principal component analysis (PCA) of the pairwise distance matrix between samples.

## Contact probability within increasing genomic distance

Polymers display a decrease in contact probability,  $P(s)$ , as a function of the genomic distance,  $s$ . The degree of decay of  $P(s)$  was often interpreted as informative of the polymer state. To compute the intra-chromosomal  $P(s)$  plots, pair of reads aligned in intra-chromosomal positions were partitioned by chromosome arms. Reads oriented toward different directions or separated by  $< 1.5$  kb were discarded to filter for self-circularizing events. For each chromosome, read pairs were log-binned in function of their genomic distance  $s$  (in kb), according to the following formula:

$$\text{bin} = \lceil \log_{1.1}(s) \rceil$$

The  $P(s)$  plot is the histogram computed on the sum of read pairs for each bin. This sum is weighted by the bin size  $1.1^{(1+\text{bin})}$  (because of the log-binning), as well as the difference between the length of the chromosome and the genomic distance  $s$ . The difference acts as a proxy for the number of possible events.

## 4C-like interaction plots

To obtain the 4C-like intra- and inter-chromosomal contact profiles for rDNA and centromeres, adjacent bins were indexed on the respective chromosomes. The resulting indexed and filtered matrices at either 5- or 50-kb bin were normalized using SCN (see Generation and normalization of contact maps). The profiles for the selected bins were plotted and compared using Matlab (no smoothing was applied).

## Computation of the replication profile from Hi-C data

The replication profile was computed from the raw 5-kb-binned contact maps. Firstly, G1 replicates were averaged and the sum of contact over each 5-kb bin was computed. The same computation was repeated for datasets obtained from cells released into S phase. To obtain the replication timing, we computed the ratio of these two signals and smoothed this ratio using a running-average window of six bins.

## 3D representation of contact maps

The 3D representations of the contact maps were generated using ShRec3D (Lesne *et al*, 2014) on the normalized contact maps, filtered for low-signal bins. First, the algorithm computes the distance matrix from the contact map, by assuming that distances are inversely proportional to the normalized contact counts. A shortest path algorithm is then used to insure that the distance matrix satisfies the triangular inequality. Finally, we used Sammon mapping to recover the optimal 3D coordinates from the distance matrix (Morlot *et al*, 2016). All the 3D structures presented here were rendered using VMD (Humphrey *et al*, 1996). Besides the cautiousness regarding the interpretation of 3D structure we mention in the main text, we also underline that the 3D structures are not used to compare datasets: All computational analyses are performed using the contact map data.

## Comparison of centromeres, intra-/inter-, and short-/long-range contacts between datasets

To compare contacts between centromeric regions, the sum of normalized inter-chromosomal contacts between 100-kb regions centered on centromeres was computed and divided by the total number of normalized inter-chromosomal contacts between all chromosomes. To compare short- versus long-range contacts, a ratio of intra contacts was computed as follows. The number of intra contacts involving fragments positioned  $< 30$  kb apart was divided by the number of intra contacts involving fragments positioned more than 30 kb apart, for all chromosomes. For intra-versus inter-chromosomal contacts, the total number of normalized

intra-chromosomal contacts was divided by the sum of normalized inter-chromosomal contacts.

### Quantification of variability between replicates

To assess for the contribution of experimental variability to the variations in contacts between different conditions, we proceeded as follows. Density histograms displaying the distribution of the log2 contact ratio of all elements of Hi-C matrices (50-kb bins) between pairs of biological and experimental replicates (3×G1, 2×G2, 3×M) were computed and compared to similar histograms computed from pairs of Hi-C matrix obtained in different experimental conditions (see Appendix Fig S2).

An estimation of the replicate variability at the centromeres was obtained by plotting the boxplots representing the distribution of the log2 contact ratios between pairs of biological and experimental replicates only of the bins encompassing the centromeres (50-kb bins; see mask; Appendix Figs S3 and S4). The same computation was performed on pairs of matrices obtained in different conditions to estimate the statistical significance of the variations. All replicates were taken into account. *P*-values were obtained by the Wilcoxon signed-rank test.

**Expanded View** for this article is available online.

### Acknowledgements

We thank Martial Marbouty and Axel Cournac for contributions to the early stage of this project. Sample description and raw contact maps are accessible on the GEO database through the following accession number: GSE90902. Raw sequences are accessible on SRA database through the following accession number: SRP094582. We are also grateful to Armelle Lengronne, Stephane Marcand, Philippe Pasero, Etienne Schwob, Emmanuelle Fabre, Nancy Kleckner, and Angela Taddei for sharing strains and for discussions. Vittore Scolari and Heloise Muller were partly supported by Pasteur-Roux-Cantarini postdoctoral fellowships. This research was supported by funding to R.K. from the European Research Council under the 7th Framework Program (FP7/2007-2013, ERC grant agreement 260822), from Agence Nationale pour la Recherche (MeioRec ANR-13-BSV6-0012-02), and from ERASynBio and Agence Nationale pour la Recherche (IESY ANR-14-SYNB-0001-03).

### Author contributions

LL-S and RK designed research. LL-S performed the experiments, with contributions from GM, AT, and HM. TMG generated the smc2-aid mutant. VFS and JM analyzed the data, with contributions from LL-S, JM, and RK interpreted the data and wrote the manuscript.

### Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Alipour E, Marko JF (2012) Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res* 40: 11202–11212
- Aragon L, Martinez-Perez E, Merkenschlager M (2013) Condensin, cohesin and the control of chromatin states. *Curr Opin Genet Dev* 23: 204–211
- Baxter J, Sen N, Martínez VL, Carandini MEMD, Schwartzman JB, Diffley JFX, Aragón L (2011) Positive supercoiling of mitotic DNA drives decatenation by topoisomerase II in eukaryotes. *Science* 331: 1328–1332
- Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58: 268–276
- Blat Y, Kleckner N (1999) Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* 98: 249–259
- Brewer BJ, Fangman WL (1987) The localization of replication origins on ARS plasmids in *Saccharomyces cerevisiae*. *Cell* 51: 463–471
- Burgess SM, Kleckner N (1999) Collisions between yeast chromosomal loci *in vivo* are governed by three layers of organization. *Genes Dev* 13: 1871–1883
- Clemente-Blanco A, Mayán-Santos M, Schneider DA, Machín F, Jarmuz A, Tschochner H, Aragón L (2009) Cdc14 inhibits transcription by RNA polymerase I during anaphase. *Nature* 458: 219–222
- Cournac A, Marie-Nelly H, Marbouty M, Koszul R, Mozziconacci J (2012) Normalization of a chromosomal contact map. *BMC Genom* 13: 436
- Cournac A, Marbouty M, Mozziconacci J, Koszul R (2016) Generation and analysis of chromosomal contact maps of yeast species. In *Yeast functional genomics*, Devaux F (ed) pp 227–245. New York: Springer Available at: [https://doi.org/10.1007/978-1-4939-3079-1\\_13](https://doi.org/10.1007/978-1-4939-3079-1_13) [Accessed April 25, 2017]
- D'Ambrosio C, Kelly G, Shirahige K, Uhlmann F (2008) Condensin-dependent rDNA decatenation introduces a temporal pattern to chromosome segregation. *Curr Biol* 18: 1084–1089
- D'Amours D, Stegmeier F, Amon A (2004) Cdc14 and condensin control the dissolution of cohesin-independent chromosome linkages at repeated DNA. *Cell* 117: 455–469
- Daoud M, Cotton JP (1982) Star shaped polymers: a model for the conformation and its concentration dependence. *J Phys* 43: 531–538
- Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295: 1306–1311
- Dekker J, Mirny L (2016) The 3D genome as moderator of chromosomal communication. *Cell* 164: 1110–1121
- Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, Noble WS (2010) A three-dimensional model of the yeast genome. *Nature* 465: 363–367
- Eser U, Chandler-Brown D, Ay F, Straight AF, Duan Z, Noble WS, Skotheim JM (2017) Form and function of topologically associating genomic domains in budding yeast. *Proc Natl Acad Sci USA* 114: E3061–E3070
- de Gennes PG (1987) Polymers at an interface; a simplified view. *Adv Colloid Interface Sci* 27: 189–209
- Glynn EF, Megee PC, Yu H-G, Mistrot C, Unal E, Koshland DE, DeRisi JL, Gerton JL (2004) Genome-wide mapping of the cohesin complex in the yeast *Saccharomyces cerevisiae*. *PLoS Biol* 2: e259
- Guacci V, Hogan E, Koshland D (1994) Chromosome condensation and sister chromatid pairing in budding yeast. *J Cell Biol* 125: 517–530
- Guidi M, Ruault M, Marbouty M, Loïdiche I, Cournac A, Billadeau C, Hoher A, Mozziconacci J, Koszul R, Taddei A (2015) Spatial reorganization of telomeres in long-lived quiescent cells. *Genome Biol* 16: 206
- Guillou E, Ibarra A, Coulon V, Casado-Vela J, Rico D, Casal I, Schwob E, Losada A, Méndez J (2010) Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev* 24: 2812–2822
- Hartwell LH, Mortimer RK, Culotti J, Culotti M (1973) Genetic control of the cell division cycle in yeast: V. Genetic analysis of cdc mutants. *Genetics* 74: 267–286

- Hicks SC, Irizarry RA (2015) Quantro: a data-driven approach to guide the choice of an appropriate normalization method. *Genome Biol* 16: 117
- Hirano T (2012) Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev* 26: 1659–1678
- Hsieh T-HS, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ (2015) Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* 162: 108–119
- Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM (2017) Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* 169: 216–228.e19
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33–38, 27–28
- Jacobs CW, Adams AE, Szanislo PJ, Pringle JR (1988) Functions of microtubules in the *Saccharomyces cerevisiae* cell cycle. *J Cell Biol* 107: 1409–1426
- Kim K-D, Tanizawa H, Iwasaki O, Noma K (2016) Transcription factors mediate condensin recruitment and global chromosomal organization in fission yeast. *Nat Genet* 48: 1242–1252
- Kitamura E, Blow JJ, Tanaka TU (2006) Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell* 125: 1297–1308
- Knott SRV, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, Tavaré S, Aparicio OM (2012) Forkhead transcription factors establish origin timing and long-range clustering in *Saccharomyces cerevisiae*. *Cell* 148: 99–111
- Leonard J, Sen N, Torres R, Sutani T, Jarmuz A, Shirahige K, Aragón L (2015) Condensin relocation from centromeres to chromosome arms promotes Top2 recruitment during anaphase. *Cell Rep* 13: 2336–2344
- Lesne A, Riposo J, Roger P, Cournac A, Mozziconacci J (2014) 3D genome reconstruction from chromosomal contacts. *Nat Methods* 11: 1141–1143
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326: 289–293
- London N, Biggins S (2014) Signalling dynamics in the spindle checkpoint response. *Nat Rev Mol Cell Biol* 15: 736–748
- Machín F, Torres-Rosell J, Piccoli GD, Carballo JA, Cha RS, Jarmuz A, Aragón L (2006) Transcription of ribosomal genes can cause nondisjunction. *J Cell Biol* 173: 893–903
- Marbouty M, Ermont C, Dujon B, Richard G-F, Koszul R (2014) Purification of G1 daughter cells from different *Saccharomycetes* species through an optimized centrifugal elutriation procedure. *Yeast Chichester Engl* 31: 159–166
- Marbouty M, Le Gall A, Cattoni DI, Cournac A, Koh A, Fiche J-B, Mozziconacci J, Murray H, Koszul R, Nollmann M (2015) Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by Hi-C and super-resolution imaging. *Mol Cell* 59: 588–602
- McCune HJ, Danielson LS, Alvino GM, Collingwood D, Delrow JJ, Fangman WL, Brewer BJ, Raghuraman MK (2008) The temporal program of chromosome replication: genomewide replication in *clb5{Delta}* *Saccharomyces cerevisiae*. *Genetics* 180: 1833–1847
- Mercy G, Mozziconacci J, Scolari VF, Yang K, Zhao G, Thierry A, Luo Y, Mitchell LA, Shen M, Shen Y, Walker R, Zhang W, Wu Y, Xie ZX, Luo Z, Cai Y, Dai J, Yang H, Yuan YJ, Boeke JD et al (2017) 3D organization of synthetic and scrambled chromosomes. *Science* 355: eaaf4597
- Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, FitzGerald P, Dekker J, Mirny L, Barrowman J, Grewal SIS (2014) Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516: 432–435
- Morlot J-B, Mozziconacci J, Lesne A (2016) Network concepts for analyzing 3D genome structure from chromosomal contact maps. *EPJ Nonlinear Biomed Phys* 4: 2
- Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, Dekker J (2013) Organization of the mitotic chromosome. *Science* 342: 948–953
- Piatti S, Lengauer C, Nasmyth K (1995) Cdc6 is an unstable protein whose *de novo* synthesis in G1 is important for the onset of S phase and for preventing a “reductional” anaphase in the budding yeast *Saccharomyces cerevisiae*. *EMBO J* 14: 3788–3799
- Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. *Science* 294: 115–121
- Renshaw MJ, Ward JJ, Kanemaki M, Natsume K, Nédélec FJ, Tanaka TU (2010) Condensins promote chromosome recoiling during early anaphase to complete sister chromatid separation. *Dev Cell* 19: 232–244
- Rock JM, Amon A (2011) Cdc15 integrates Tem1 GTPase-mediated spatial signals with Polo kinase-mediated temporal cues to activate mitotic exit. *Genes Dev* 25: 1943–1954
- Saner N, Karschau J, Natsume T, Gierliński M, Retkute R, Hawkins M, Nieduszynski CA, Blow JJ, de Moura APS, Tanaka TU (2013) Stochastic association of neighboring replicons creates replication factories in budding yeast. *J Cell Biol* 202: 1001–1012
- Stephens AD, Haase J, Vicci L, Taylor RM, Bloom K (2011) Cohesin, condensin, and the intramolecular centromere loop together generate the mitotic chromatin spring. *J Cell Biol* 193: 1167–1180
- Sullivan M, Higuchi T, Katis VL, Uhlmann F (2004) Cdc14 phosphatase induces rDNA condensation and resolves cohesin-independent cohesion during budding yeast anaphase. *Cell* 117: 471–482
- Taddei A, Gasser SM (2012) Structure and function in the budding yeast nucleus. *Genetics* 192: 107–129
- Uhlmann F, Lottspeich F, Nasmyth K (1999) Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature* 400: 37–42
- Uhlmann F (2016) SMC complexes: from DNA to chromosomes. *Nat Rev Mol Cell Biol* 17: 399–412
- Valton A-L, Dekker J (2016) TAD disruption as oncogenic driver. *Curr Opin Genet Dev* 36: 34–40
- Visintin R, Prinz S, Amon A (1997) CDC20 and CDH1: a family of substrate-specific activators of APC-dependent proteolysis. *Science* 278: 460–463
- Wang X, Montero Llopis P, Rudner DZ (2013) Organization and segregation of bacterial chromosomes. *Nat Rev Genet* 14: 191–203
- Yoshida S, Asakawa K, Toh-e A (2002) Mitotic exit network controls the localization of Cdc14 to the spindle pole body in *Saccharomyces cerevisiae*. *Curr Biol* 12: 944–950



**License:** This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.







**Figure EV1. Contact maps and 3D genome representations of the five cell cycle synchronization states.**

- A, B Contact maps generated from synchronized cell populations described in this study, with each vector (or bin) corresponding to 5 kb. x- and y-axis represent the 16 chromosomes of the yeast genome, displayed atop the maps. Brown to yellow color scales reflect high to low contact frequencies, respectively ( $\log_{10}$ ). Magnification panels in (B) show variations of the contact frequencies between synchronized populations. Yellow and pink arrowheads point at centromeres and rDNA positions, respectively.
- C 3D average representations of the Hi-C contact maps of synchronized cell populations of panel (A). The color code represents the chromosomal arm length, and centromeres, telomeres, and rDNA flanking regions are highlighted.

**Figure EV2. Contact maps and 3D genome representations during replication.**

- A, B Contact maps recovered from cell populations undergoing replication after G1 release. For each contact map, the FACS profile is displayed. x- and y-axis represent the 16 chromosomes of the yeast genome. The same color code as in Fig. EV1. Magnification panels in (B) highlight changes of the contact frequencies during S-phase progression. Yellow and pink arrowheads point at centromeres and rDNA positions, respectively.
- C 3D average representations of the Hi-C contact maps of synchronized cell populations of panel (A). The color code represents the chromosomal arm length, and centromeres, telomeres, and rDNA flanking regions are highlighted.



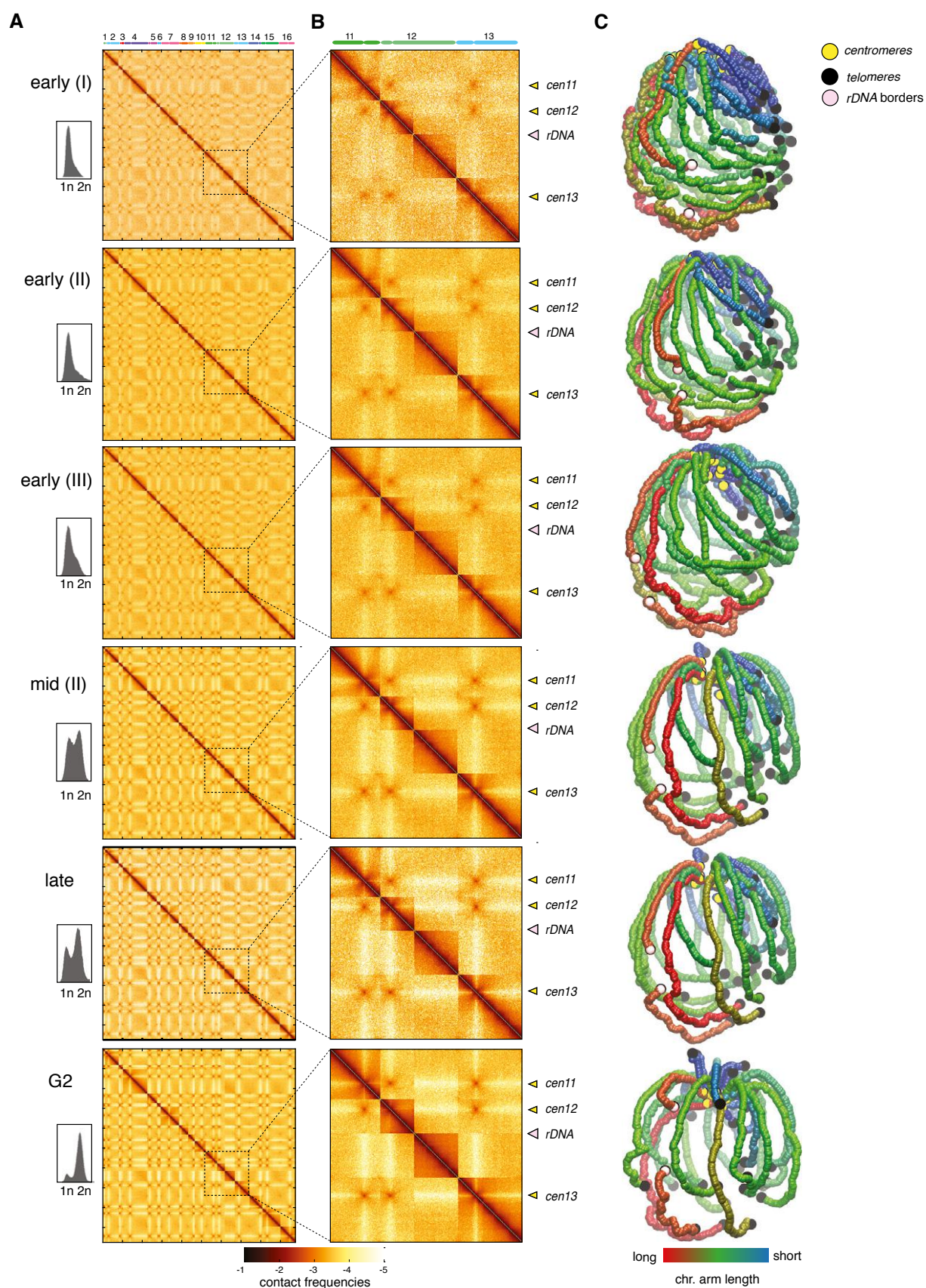


Figure EV2.

**Figure EV3. Contact maps and 3D genome representations during M phase.**

- A, B Contact maps of cell populations synchronized in metaphase (*cdc20*) and anaphase (*cdc15*) and released into mitosis from *cdc15* block (+20 min, +40 min, and +60 min). The corresponding FACS profiles and representative DAPI-stained cells are displayed on the left on the maps. x- and y-axis represent the 16 chromosomes of the yeast genome. The same color code as in Fig EV1. Magnification panels in (B) display variations of the contact frequencies during mitotic progression. Yellow and pink arrowheads point at centromeres and rDNA positions, respectively.
- C 3D average representations of the Hi-C contact maps of panel (A). The color code represents the chromosomal arm length, and centromeres, telomeres, and rDNA flanking regions are highlighted.



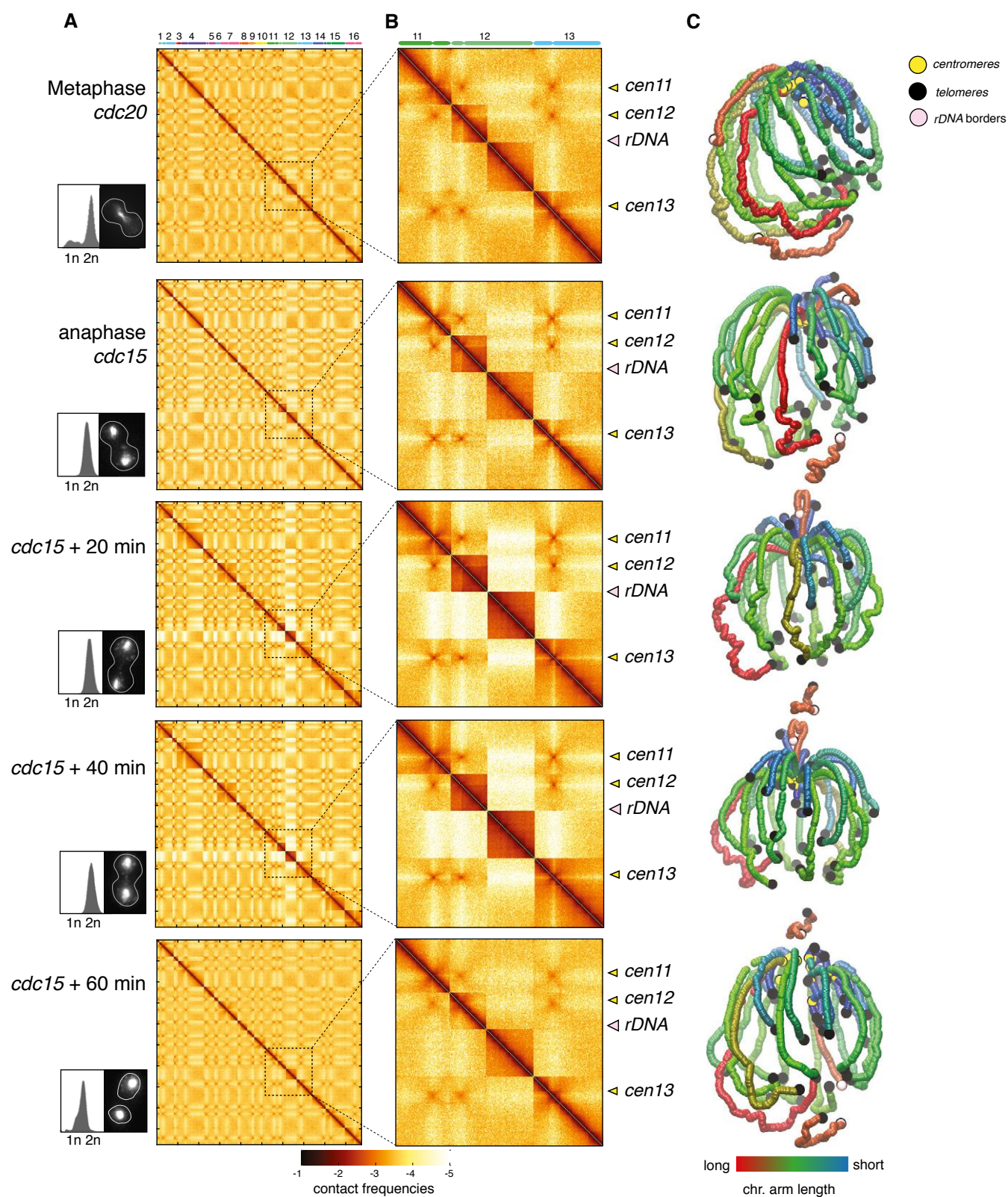
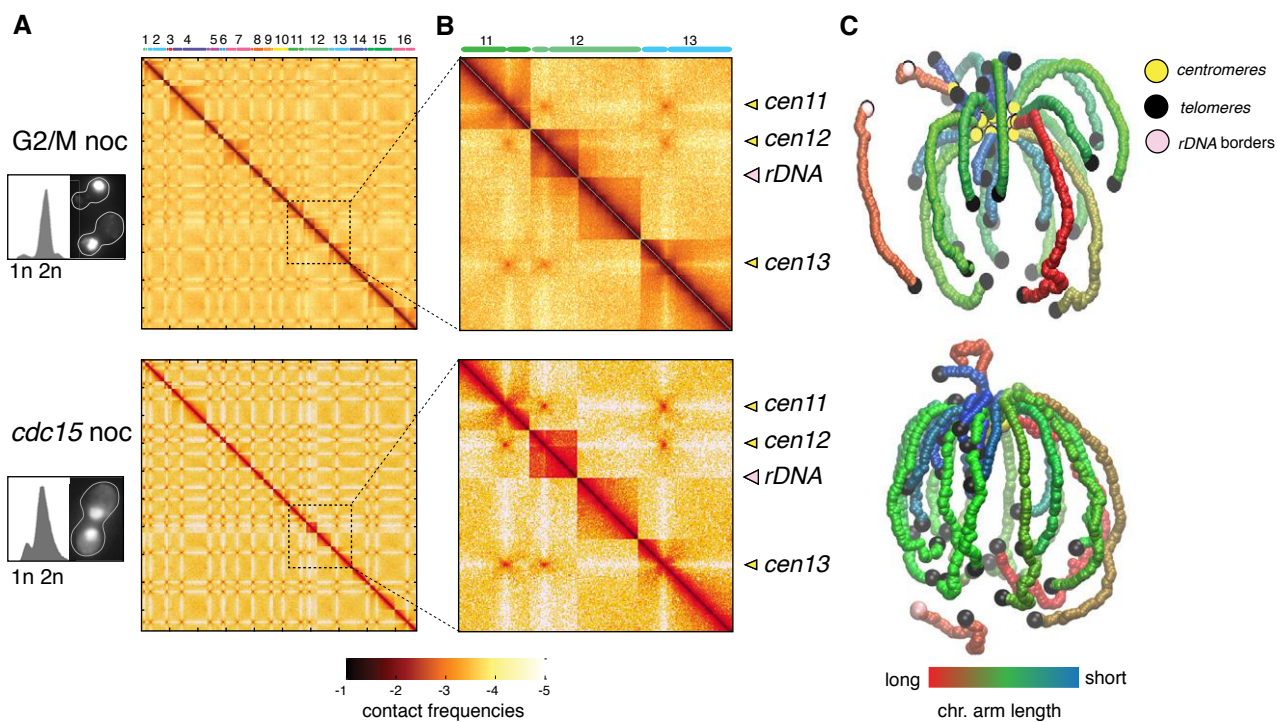


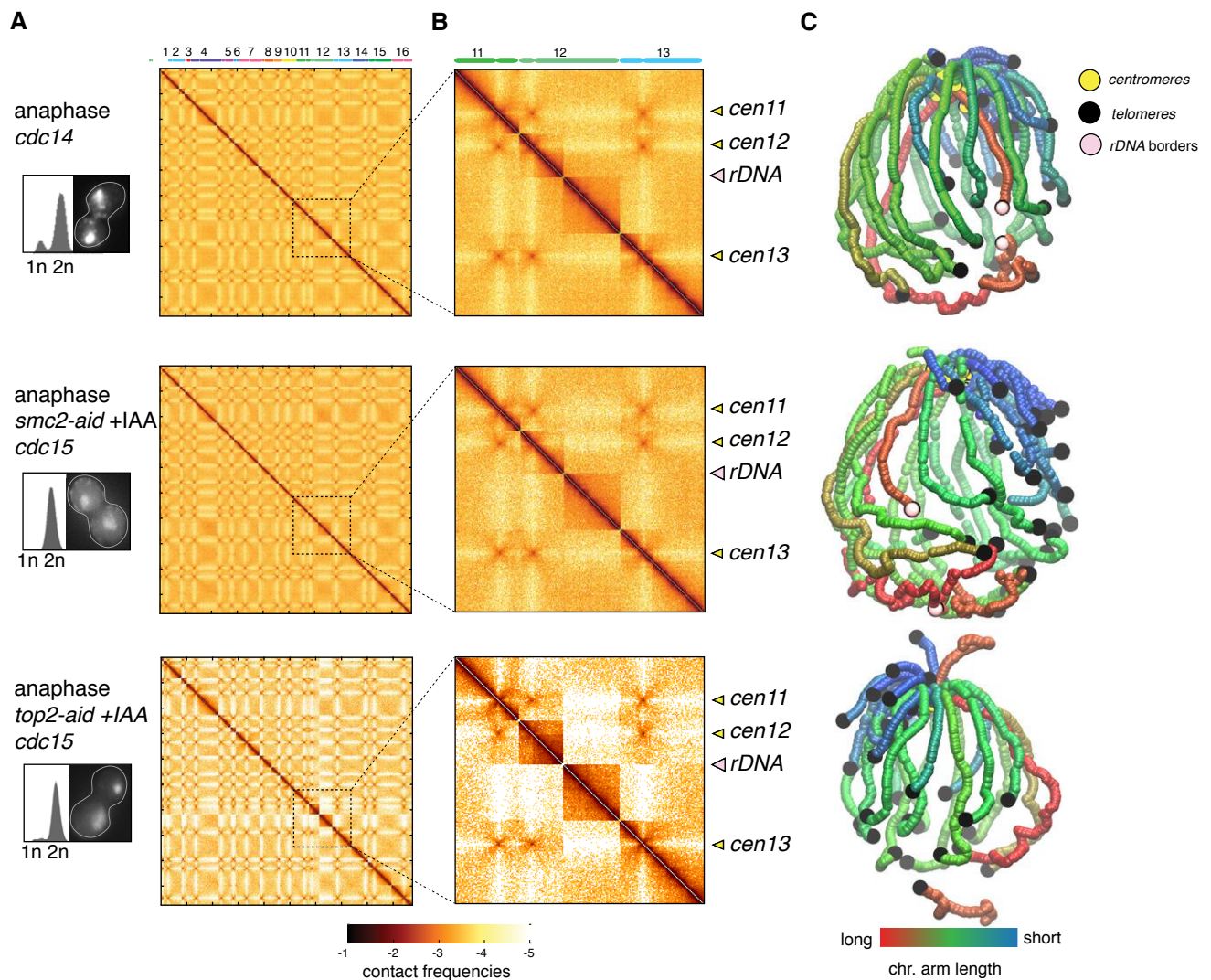
Figure EV3.



**Figure EV4. Nocodazole affects chromosome 12 conformation.**

- A, B Contact maps of G1 synchronized cell populations released either in the presence of nocodazole (G2/M noc) or at *cdc15* non-permissive temperature followed by a nocodazole treatment (*cdc15* noc). The corresponding FACS profiles and representative DAPI-stained cells are displayed on the left on the maps. x- and y-axis represent the 16 chromosomes of the yeast genome. The same color code as in Fig EV1. Magnification panels in (B) display variations of the contact frequencies. Yellow and pink arrowheads point at centromeres and rDNA positions, respectively.
- C 3D average representations of the Hi-C contact maps of panel (A). The color code represents the chromosomal arm length, and centromeres, telomeres, and rDNA flanking regions are highlighted.





**Figure EV5. Condensin and decatenation influence on chromatin structure during mitosis.**

A, B Contact maps of cell populations arrested in anaphase, either defective with condensation (*cdc14* or *smc2-aid cdc15 +IAA*) or depleted of topoisomerase 2 (*top2-aid cdc15 +IAA*). The corresponding FACS profiles and representative DAPI-stained cells are displayed on the left on the maps. x- and y-axis represent the 16 chromosomes of the yeast genome. The same color code as in Fig EV1. Magnification panels in (B) display variations of the contact frequencies. Yellow and pink arrowheads point at centromeres and *rDNA* positions, respectively.

C 3D average representations of the Hi-C contact maps of panel (A). The color code represents the chromosomal arm length, and centromeres, telomeres, and *rDNA* flanking regions are highlighted.

**Table EV1. List of strains used in this study.**

\*: strain YKL050 carries a ~130 kb region where some restriction sites have been reordered to generate fixed size restriction fragments to investigate the resolution of the Hi-C assay, a work that will be described elsewhere.

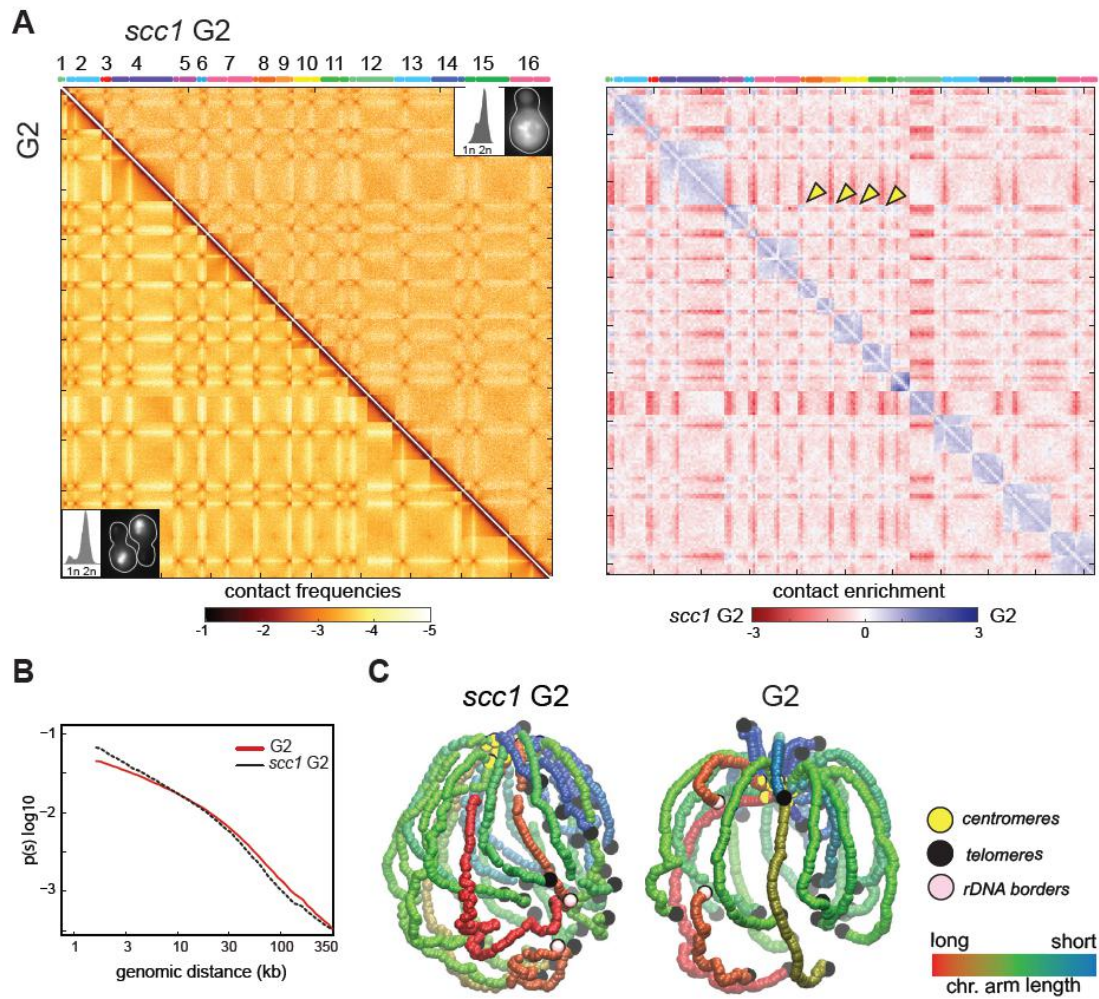
| Strain  | Genetic background | Genotype                                                                                                                                               | Used to study:                                            | Reference                                   |
|---------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|---------------------------------------------|
| BY4741  | S288C              | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>                                                                                                               | asynchronous, G0 and G1                                   | Mortimer and Johnston, 1986                 |
| YKL050* | BY4742             | <i>MATa his3Δ1 lys2Δ0 ura3Δ0 IV(715448-845757)::synIV(715448-845757 LEU2)</i>                                                                          | asynchronous, G1, S and G2 phase and nocodazole treatment |                                             |
| YKL051  | W303               | <i>MATa his3Δ leu2Δ ura3Δ0 can1Δ ade2Δ cdc20::MET-HA3-CDC20::TRP1</i>                                                                                  | M phase (metaphase)                                       | Uhlmann et al., 2000                        |
| YKL052  | W303               | <i>MATa his3Δ leu2Δ ura3Δ0 trp1Δ can1Δ ade2Δ cdc14-3(ts)</i>                                                                                           | M phase (early anaphase)                                  | Visinti et al., 1998                        |
| YKL053  | W303               | <i>MATa his3Δ leu2Δ ura3Δ trp1Δ can1Δ ade2Δ cdc15-2(ts)</i>                                                                                            | M phase (late anaphase) and mitosis exit                  | Surana et al., 1993                         |
| YKL054  | W303               | <i>MATa his3Δ ura3Δ0 trp1Δ can1Δ ade2Δ DBF4-Myc18::LEU2 cdc6-1(ts)</i>                                                                                 | G1/S and M phase                                          | Piatti et al., 1995                         |
| YKL055  | W303               | <i>MATa (rad5?) pADH1-OsTIR1-myc9::TRP1 GPD-TK(x7)::URA3 scc1-AID::KANr</i>                                                                            | G2 phase                                                  | Courtesy of Philippe Pasero (strain PP1792) |
| YKL056  | W303               | <i>MATa cdc15-2(ts) smc2-AID-9myc::NAT OsTIR1-9myc::URA3 SPO20<sup>51-91</sup>-TTYeGFP::TRP1 HTA1-mCherry::HPH KANr-pGAL1-CEN6-pGAL1-skHIS3 leu2-2</i> | M phase (late anaphase)                                   | Courtesy of Stephane Marcand                |
| YKL057  | W303               | <i>MATa cdc15-2(ts) top2-AID::HPH TIR1::URA3 KANr-pGAL1-CEN6-pGAL1-skHIS3</i>                                                                          | M phase (late anaphase)                                   | Courtesy of Stephane Marcand                |

## **Appendix**

### **Table of content**

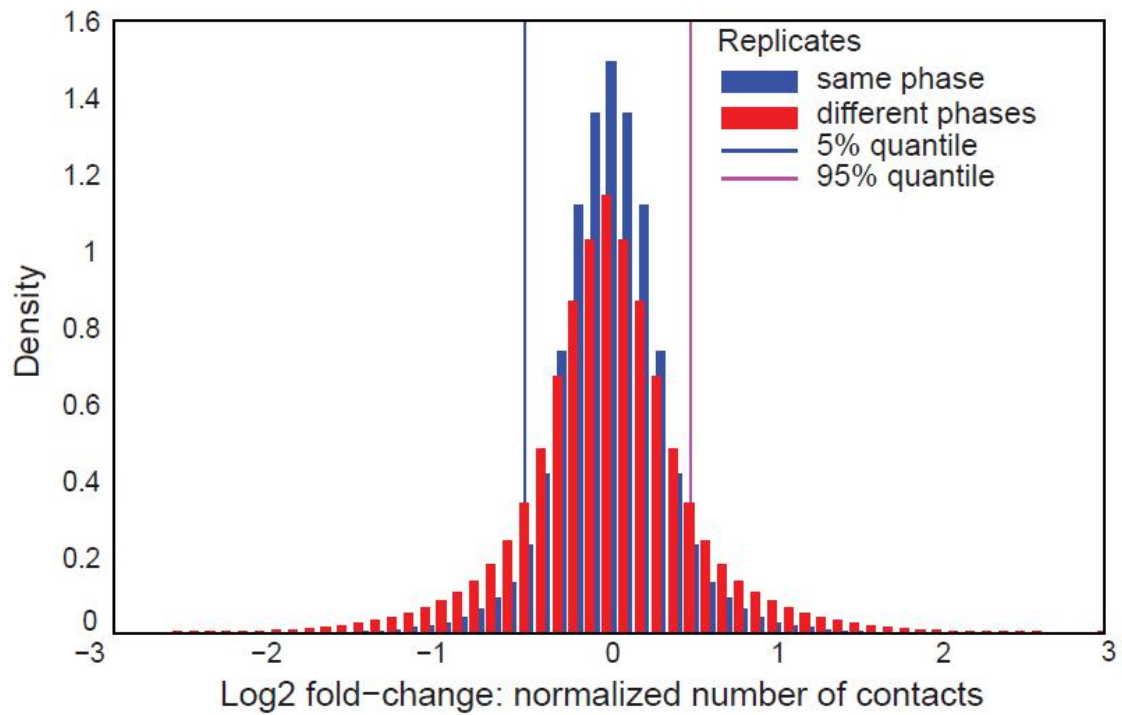
|                                |        |
|--------------------------------|--------|
| Appendix Figure S1             | Page 2 |
| Appendix Figure S2             | Page 3 |
| Appendix Figure S3             | Page 4 |
| Appendix Figure S4             | Page 5 |
| Appendix Supplementary Methods | Page 6 |





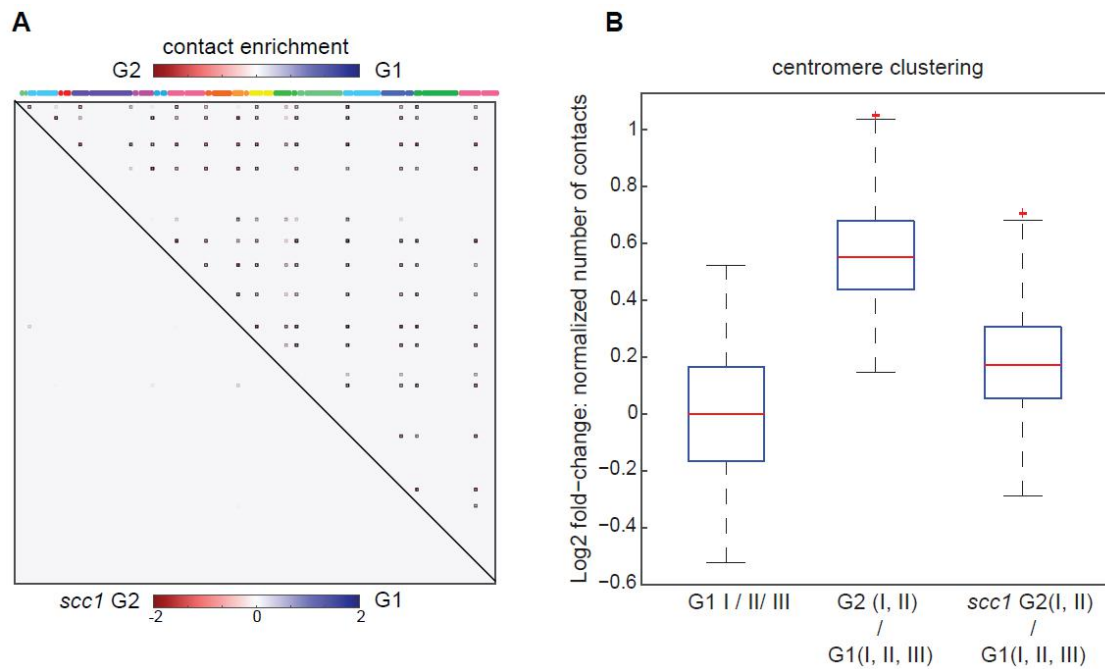
**Appendix Figure S1. Cohesin required for genome-wide chromatin organization during replication.**

(A) Left panel: Hi-C contact maps of cell populations G1 released and processed in G2 wild-type (G2, bottom left) and in cohesin depleted (*scc1* G2, upper right) cells. The corresponding FACS profiles as well as representative DAPI-stained cells are displayed in boxes on the bottom left and upper right corners, respectively. X and y axis represent the 16 chromosomes of the yeast genome. Same color code as in Figure EV1. Right panel: log-ratio of G2 and *scc1* G2 contact maps. Yellow arrowheads: inter-centromere contacts. The color code reflects the enrichment in contacts in one population with respect to the other. (B) Contact probability  $p(s)$  decay in G2 and *scc1* G2. (C) 3D average representations of the Hi-C contact maps of panel (A). Color code represents chromosomal arm length and centromeres, telomeres and rDNA flanking regions are highlighted.



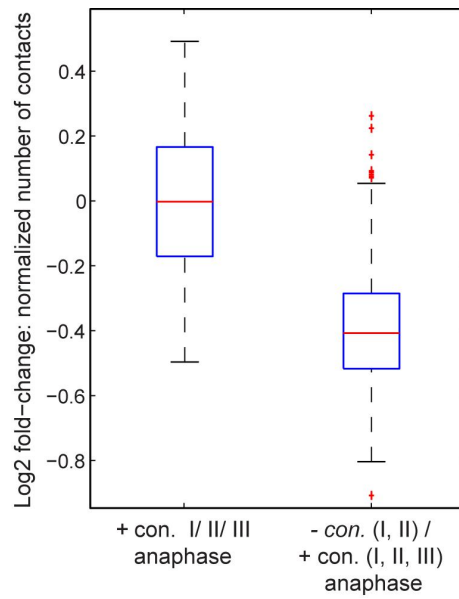
**Appendix Figure S2. Analysis of the variability of Hi-C data.**

Distribution of the fold-change in contacts made by bins in over different conditions. Blue: fold change in contacts in replicates. Red: fold change in contacts between all non-replicated conditions. This results backs the co-localization in the PC space of replicates in the PCA.



**Appendix Figure S3. Centromere clustering is cohesin-dependent.**

**A)** Positions within the contact map of bins whose specific variations in contacts in trans will be investigated between two conditions (here, the 16 bins corresponding to centromeres, between G1 and G2 [top right] and between *scc1* and G1 [bottom left]). **B)** Boxplots representing the variation in number of normalized contacts for a subset of bins between different contact maps (corresponding to different conditions). Left: variations between three G1 replicates. Middle: variations between G2 and G1 cells replicates. The relative Wilcoxon test provides a p-value  $< 10^{-10}$ , supporting an increase in contacts in G2 compared to G1. Right: variations between *scc1* arrested cells and G1 replicates.



**Appendix Figure S4. Condensins increase centromere clustering in anaphase.**

Left boxplot shows the centromere contact variability between anaphase replicates (*cdc15* I, II, III) in presence of condensins (+con.). Right boxplot shows the variation of contacts between centromeres between *cdc15* replicates and condensin defective strains arrested in anaphase (*cdc14* and *smc2 cdc15*; -con. I, II). The boxplots show that the centromere cluster in anaphase is condensin-dependent ( $P < 0.05$ ).

## Appendix Supplementary Methods

**Generation of Hi-C libraries.** Aliquots of  $1-3 \times 10^9$  cells in 150 ml YPD/synthetic medium were fixed in 3% formaldehyde (Sigma) for 20 min at room temperature and quenched with 25 ml glycine 2.5 M for 20 min at 4°C. Cross-linked cells were recovered through centrifugation, washed with YPD and a 150 mg pellet was stored at -80°C. The pellet was thawed on ice and incubated for 30 min in 10 ml of sorbitol 1M, DTT 5mM and Zymolyase 100T (C<sub>Final</sub>=1 mg/ml; Armsbio). Spheroplasts were washed once with 5 ml sorbitol 1M, once with 5 ml 1X RE buffer (DpnII NEB buffer) and finally suspended in 3.5 ml of 1X RE buffer. The spheroplasts were treated with 3% SDS for 20 min at 65°C and the lysate was digested overnight with DpnII (C<sub>Final</sub>=450 U/pellet; NEB) at 37°C. The digestion was centrifuged for 20 min at 18000 g, the supernatant discarded and the pellet suspended in 400  $\mu$ L of water. The 5' overhangs from DpnII digestion were filled in using dNTP 30  $\mu$ M (biotin-14-dCTP, dATP, dGTP and dTTP; Invitrogen), at 37°C for 45 min. The biotinylated DNA fragments were ligated by T4 DNA ligase (C<sub>Final</sub>=250 Weiss U/pellet; Thermo Scientific) for 4 h at 16°C. DNA purification was achieved through an overnight incubation at 65°C in presence of proteinase K 250  $\mu$ g/ml and EDTA 6.2 mM, followed by a phenol/chloroform extraction on the precipitated DNA and an RNase A DNase-free 500  $\mu$ g/ml treatment. The biotinylated but not ligated DNA fragments were removed by T4 DNA polymerase (C<sub>Final</sub>=5 U/pellet; NEB) treatment. Hi-C DNA libraries were 500 bp sheared, using CovarisS220 apparatus, and the biotin-labeled fragments were selectively captured by Dynabeads Myone Streptavidin C1 (Invitrogen). The resulting libraries were used as template for the Illumina amplification by PE-PCR primers and paired-end sequenced on the NextSeq500 or HiSeq 2000 Illumina platforms (2x75 or 2x150 bp kits; see [Table EV2](#)).

**Elutriation (recovery of G1 cells).** 800 ml overnight culture was centrifuged, washed in 1X PBS and pelleted cells were suspended in 1000 ml of fresh YPD for 2 h at 30°C. This additional growing step allowed cells in stationary phase to reenter exponential phase before being elutriated. For each elutriation experiment,  $1.2-1.8 \times 10^{11}$  cells were washed and suspended in 30 ml of 1X PBS and injected in the 40 ml elutriation chamber at an average flow rate ranging from 20 ml/min to 25 ml/min (MasterFlex L/S pump from Cole-Parmer), at 2,500 r.p.m. and 23°C. Cells were then left to equilibrate in 1X PBS for 45 min at a constant

flow and rotational speed. To start collecting the first fractions containing the small G1 cells, a periodic 2 ml/min increment of the flux was applied between each fraction. The resulting 600 ml fractions were centrifuged and approximately  $2.5 \times 10^9$  G1 cells/fraction were recovered. Before fixating the G1 state, cells were suspended in fresh YPD at 30°C for 30 min, so they could recover from their stay in PBS during the elutriation. To minimize the potential variability introduced by the age heterogeneity of the bulk population, G1 daughter cells were used as starting point for all cell cycle synchrony and in combination with genetic and chemical synchronization methods (see below).

**Synchronization through thermosensitive mutations.** Synchronizations using thermosensitive (ts) *cdc* strains (Hartwell *et al*, 1973) were all performed starting from elutriated G1 daughter cells growing in non-permissive temperature conditions designed to arrest the progression of the cycle at specific phases.

The G1/S checkpoint (*cdc6-1* mutation; YKL054 strain) was activated by growing cells overnight at 25°C, restarted in fresh YPD media and elutriated while in exponentially growing stage, still at 25°C. The elutriated G1 cells were incubated in fresh YPD at the non-permissive temperature of 37°C for 3 h. To study non-replicated mitotic chromosomes, the *cdc6-1* arrested cells were maintained in non-permissive growing conditions for an extended period of 6 h. During this time period, G1 cells bypass the G1/S checkpoint and proceed into M phase without having replicated their chromosomes.

A similar protocol was applied to *cdc14-3* (YKL052) and *cdc15-2* (YKL053) G1 cells exposed to non-permissive temperature of 30°C and 37°C after elutriation, respectively. *Cdc15* and *cdc14* arrested cells blocked into anaphase were shifted at the permissive temperatures of 25°C and 23°C, with different time-points were sampled after release (YKL052: 30 min; YKL053: 20 min, 40 min and 60 min). The synchrony of each time point (in G1/S, anaphase and release) was monitored with flow cytometry and microscopy.

# **Perspectives**





Au travers de mon manuscrit, j'ai rapporté les travaux auxquels j'ai participé dans le cadre de mon doctorat, au cours de ces trois dernières années. Dans le cadre du premier projet auquel j'ai participé, nous avons étudié l'organisation des chromosomes synthétiques issus du projet Sc2.0, que nous avons comparé à leur homologue natif. Au travers d'un second projet, nous avons démontré que la technologie de synthèse d'ADN pouvait être utilisée dans le but d'améliorer l'analyse de Hi-C et palier certaines de ses limites. Dans un premier temps, la réorganisation des sites de restriction, essentiels au protocole de Hi-C, nous a permis d'augmenter la résolution des données générées. Puis, grâce aux différents SNPs entre la séquence synthétique et la séquence native, elle nous a permis de suivre la dynamique des chromosomes de levure au cours de la méiose. En utilisant un hybride de *S. cerevisiae* et *S. paradoxus* comme modèle, nous également avons pu nous faire une idée de l'organisation globale de tout le génome d'une levure diploïde. Enfin, j'ai également participé à un dernier projet, au cours duquel nous avons décrit l'organisation des chromosomes de *S. cerevisiae* au cours du cycle cellulaire, et les rôles que peuvent jouer certaines condensines et cohésines dans le maintien de l'architecture des chromosomes à ces différentes étapes.

Il était originellement prévu d'étudier les souches évoluées par SCRaMbLE de génomes « Sc2.0 » mais l'avancement du projet ne nous a pas permis d'arriver à cette étape. Comme nous l'avons présenté dans notre premier article (Mercy et al., 2017), les cartes de contacts obtenues suite à une expérience de Hi-C sont très utiles pour identifier les différents réarrangements ayant pu se produire au cours de la réaction de SCRaMbLE. Nous aurions ainsi pu utiliser ces données pour reconstruire les structures génomiques de ces nouveaux génomes, avant de pouvoir les étudier. En plus des réarrangements produits par les événements de recombinaisons entre sites LoxP, il aurait été très intéressant de voir si la structure des chromosomes synthétiques était elle aussi modifiée en fonctions des conditions de sélections. En effet notre analyse des différents chromosomes issus du projet Sc2.0 a nous a permis de mettre en lumière la conservation de la structure de pseudo-Rabl de *S. cerevisiae*, guidée principalement par l'ancrage des centromères au SPB et le regroupement des télomères en périphérie du noyau. Ainsi en induisant la réaction de SCRaMbLE dans un milieu contenant du galactose comme source de carbone, plutôt que du glucose, il pourrait être très intéressant de suivre l'évolution de la position des gènes, tels que GAL1 (Casolari et al., 2004), INO1 (Brickner and Walter, 2004) ou HXK1 (Taddei et al., 2006), connus pour être associés aux pores nucléaires dans ces conditions. À l'image du nouveau chromosome construit pour coder tous les ARNt de la souche synthétique, il serait intéressant de voir, si la levure, au cours de la réaction de SCRaMbLE, regroupe spontanément des gènes codant pour une même voie métabolique.

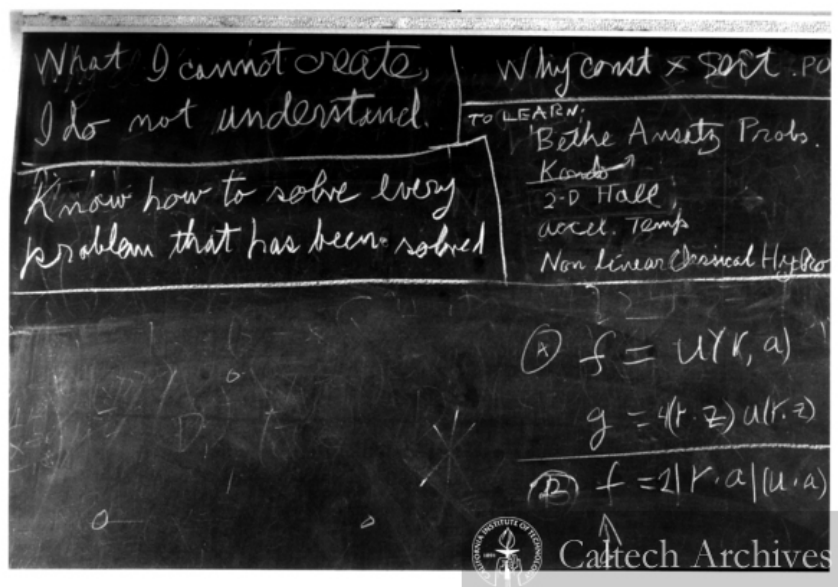
Avec la diminution du coût de la synthèse d'ADN, le projet Sc2.0 touchant bientôt à sa fin et le développement du projet de construction de génomes eucaryotes (Boeke et al., 2016), la génomique synthétique commence à entrer dans une nouvelle ère, alors que le nombre de projet de synthèse de génome augmente. Nos études sur l'organisation des chromosomes synthétiques de

levure et sur un design optimisé pour l'analyse de l'organisation des chromosomes peuvent dans ce cadre-là apporter des idées pour les designs futurs. Notre première étude met en avant l'importance de considérer, lors de l'étape de design du chromosome synthétique, l'impact qu'une modification au sein de ce chromosome pourrait avoir sur l'organisation tridimensionnelle générale du génome. Par exemple, lorsque les ADN<sub>r</sub> ont été déplacés au milieu du bras droit du chromosome III de *S. cerevisiae*, l'organisation du génome a été relativement modifiée. Bien que dans ce cas-là, cette modification ne semble pas poser de problème à la croissance des cellules, il est très probable que ce design, ne soit pas idéal. L'approche de synthèse de génome étant une approche « bottom-up », qui nécessite de passer par les étapes de design et de construction avant de pouvoir être testée. Pour chaque projet de synthèse les différents designs devront être testés et analysés pour pouvoir mesurer l'impact d'une modification sur l'organisation du génome, comme par exemple le regroupement des ADN<sub>t</sub> sur un nouveau chromosome dans le projet Sc2.0.

À partir des différents résultats que nous avons obtenus, si le projet de recherche et l'organisme modèle si prêtent, nous pouvons suggérer certaines modifications qui peuvent être considérées pour améliorer un design de chromosomes synthétiques : (1) la suppression des séquences répétées qui permet d'augmenter la visibilité des cartes de contacts ; (2) la réorganisation d'au moins un site de restriction qui permet d'augmenter la résolution de l'analyse de Hi-C ; et (3) s'il s'agit d'un génome diploïde, l'introduction de SNPs entre les chromosomes homologues permettant l'identification de ces derniers après séquençage.

L'organisation du génome de *S. cerevisiae* a été intensément étudiée au cours de ces trente dernières années. Certains mécanismes et certaines fonctions responsable de cette organisation nécessitent encore d'être approfondies pour être entièrement élucidées. C'est le cas par exemple de l'appariement des chromosomes homologues, qui a été observé par microscopie à l'échelle d'une cellule, mais qui ne l'a toujours pas été par les méthodes de génomique à l'échelle de la population. Nous avons commencé à démontrer, dans ce manuscrit, que l'utilisation de la génomique synthétique pouvait permettre de d'apporter une solution pour résoudre ce problème. Le modèle de l'hybride nous a ainsi qu'à Kim et al. (Kim et al., 2017) permis d'observer l'organisation des chromosomes lorsque les cellules sont en croissance végétative. Cependant, ce modèle reste imparfait, puisqu'on pourrait se demander si le taux d'hétérozygotie n'affecte pas l'appariement des chromosomes au cours du cycle cellulaire. La construction d'un génome synthétique, sur le modèle du Syn3C, présentant un taux d'hétérozygotie moins élevé pourrait nous permettre de lever cette interrogation concernant le modèle hybride.

La synthèse de génome nécessite un certain nombre d'étapes intermédiaires, qui très généralement, dans ce cas-là sont des souches intermédiaires. Ces « produits dérivés » de la synthèse de génome, peuvent s'avérer être des modèles d'études très intéressants. Par exemple, les souches de *S. cerevisiae* qui ont été utilisées pour modifier les génomes de différentes souches de *Mycoplasma*



**FIGURE 4.16 Tableau noir de Richard Feynman :** Ce dernier y avait inscrit sa maintenant célèbre citation : « *What I cannot create, I do not understand* » (« *Ce que je ne peux construire, je ne le comprends pas* »). L'image provient des archives de Caltech).

(Lartigue et al., 2009) pourraient être utilisées comme plateformes d'étude de protéines responsables de l'organisation des génomes bactériens. Les souches de *S. cerevisiae*, dont le génome est organisé sur deux ou trois chromosomes au lieu de seize initialement, qui sont disponibles au laboratoire, sont elles aussi des modèles très intéressants. En effet, ces chromosomes de grandes tailles pourraient être utilisés pour étudier la condensation des chromosomes de *S. cerevisiae* sur de plus large distance que ce qu'il est possible dans la souche native.

Tout comme le suggère Richard Feynman au travers de sa célèbre citation : « *What I cannot create, I do not understand* », la synthèse de chromosomes synthétiques devrait permettre d'expliquer et de confirmer un certain nombre de processus biologiques, dans les années à venir. Au cours de notre étude les chromosomes synthétiques de levure nous ont également permis de repousser certaines limites de la méthode de Hi-C, et d'améliorer notre compréhension de l'organisation fonctionnelle des chromosomes de levure.



## Références



# Bibliographie

- Anthony Admire, Lisa Shanks, Nicole Danzl, Mei Wang, Ulli Weier, William Stevens, Elizabeth Hunt, and Ted Weinert. Cycles of chromosome instability are associated with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. *Genes and Development*, 20(2) :159–173, 2006. ISSN 08909369. doi : 10.1101/gad.1392506.
- K. L. Agarwal, H. Büchi, M. H. Caruthers, N. Gupta, H. G. Khorana, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. Rajbhandary, J. H. Van De Sande, V. Sgaramella, H. Weber, and T. Yamada. Total Synthesis of the Gene for an Alanine Transfer Ribonucleic Acid from Yeast. *Nature*, 227 (5253) :27–34, jul 1970. ISSN 0028-0836. doi : 10.1038/227027a0.
- Benjamin Albert, Isabelle Léger-Silvestre, Christophe Normand, Martin K. Ostermaier, Pérez-Fernández, Kostya I. Panov, Joost C B M Zomerdijs, Patrick Schultz, and Olivier Gadal. RNA polymerase I-specific subunits promote polymerase clustering to enhance the rRNA gene transcription cycle. *Journal of Cell Biology*, 192(2) :277–293, 2011. ISSN 00219525. doi : 10.1083/jcb.201006040.
- E D Andrulis, a M Neiman, D C Zappulla, and R Sternglanz. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature*, 394(6693) :592–595, 1998. ISSN 0028-0836. doi : 10.1038/29100.
- Erik D Andrulis, David C Zappulla, Athar Ansari, Severine Perrod, Catherine V Laiosa, Marc R Gartenberg, and Rolf Sternglanz. Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. *Molecular and cellular biology*, 22(23) :8292–301, 2002. ISSN 0270-7306. doi : 10.1128/MCB.22.23.8292.
- Narayana Annaluru, Héloïse Muller, Leslie A. Mitchell, Sivaprakash Ramalingam, Giovanni Stracquadanio, Sarah M. Richardson, Jessica S. Dymond, Zheng Kuang, Lisa Z. Scheifele, Eric M. Cooper, Yizhi Cai, Karen Zeller, Neta Agmon, Jeffrey S. Han, Michalis Hadjithomas, Jennifer Tullman, Katrina Caravelli, Kimberly Cirelli, Zheyuan Guo, Viktoriya London, Apurva Yeluru, Sindurathy Murugan, Karthikeyan Kandavelou, Nicolas Agier, Gilles Fischer, Kun Yang, J. Andrew Martin, Murat Bilgel, Pavlo Bohutski, Kristin M. Boulier, Brian J. Capaldo, Joy Chang, Kristie Charoen, Woo Jin Choi, Peter Deng, James E. DiCarlo, Judy Doong, Jessilyn Dunn, Jason I. Feinberg, Christopher Fernandez, Charlotte E. Floria, David Gladowski, Pasha Hadidi, Isabel Ishizuka, Javaneh Jabbari, Calvin Y L Lau, Pablo A. Lee, Sean Li, Denise Lin, Matthias E. Linder, Jonathan Ling, Jaime Liu, Jonathan Liu, Mariya London, Henry Ma, Jessica Mao, Jessica E. McDade, Alexandra McMillan, Aaron M. Moore, Won Chan Oh, Yu Ouyang, Ruchi Patel, Marina Paul, Laura C. Paulsen, Judy Qiu, Alex Rhee, Matthew G. Rubashkin, Ina Y. Soh, Nathaniel E. Sotuyo, Venkatesh Srinivas, Allison Suarez, Andy Wong, Remus Wong, Wei Rose Xie, Yijie Xu, Allen T. Yu, Romain Koszul, Joel S. Bader, Jef D. Boeke, and Srinivasan Chandrasegaran. Total synthesis of a functional designer eukaryotic chromosome. *Science (New York, N.Y.)*, 344(6179) :55–8, 2014. ISSN 1095-9203. doi : 10.1126/science.1249252.
- Katsutoshi Ara, Katsuya Ozaki, Kouji Nakamura, Kunio Yamane, Junichi Sekiguchi, and Naotake Ogasawara. *Bacillus minimum* genome factory : effective utilization of microbial genome

- information. *Biotechnology and Applied Biochemistry*, 46(3) :169, 2007. ISSN 08854513. doi : 10.1042/BA20060111.
- O T Avery, C M Macleod, and M McCarty. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *The Journal of experimental medicine*, 79 (2) :137–58, feb 1944. ISSN 0022-1007. doi : 10.1084/jem.79.2.137.
- Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A. Datsenko, Masaru Tomita, Barry L. Wanner, and Hirotada Mori. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants : The Keio collection. *Molecular Systems Biology*, 2, 2006. ISSN 17444292. doi : 10.1038/msb4100050.
- Maria Barbi, Julien Mozziconacci, J.-M. Victor, Hua Wong, and Christophe Lavelle. On the topology of chromatin fibres. *Interface Focus*, 2(5) :546–554, oct 2012. ISSN 2042-8898. doi : 10.1098/rsfs.2011.0101.
- J. G.J. Bauman, J. Wiegant, P. Borst, and P. van Duijn. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Experimental Cell Research*, 128(2) :485–490, 1980. ISSN 00144827. doi : 10.1016/0014-4827(80)90087-7.
- Andrew S. Belmont and Aaron F. Straight. In vivo visualization of chromosomes using lac operator-repressor binding. *Trends in Cell Biology*, 8(3) :121–124, 1998. ISSN 09628924. doi : 10.1016/S0962-8924(97)01211-7.
- Jon Matthew Belton, Bryan R. Lajoie, Sylvain Audibert, Sylvain Cantaloube, Imen Lassadi, Isabelle Goiffon, Davide Baù, Marc A. Marti-Renom, Kerstin Bystricky, and Job Dekker. The Conformation of Yeast Chromosome III Is Mating Type Dependent and Controlled by the Recombination Enhancer. *Cell Reports*, 13(9) :1855–1867, 2015. ISSN 22111247. doi : 10.1016/j.celrep.2015.10.063.
- Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science (New York, N.Y.)*, 313(5793) : 1642–5, sep 2006. ISSN 1095-9203. doi : 10.1126/science.1127344.
- J D Boeke, C A Styles, and G R Fink. *Saccharomyces cerevisiae* SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements. *Molecular and cellular biology*, 6(11) :3575–81, 1986. ISSN 0270-7306. doi : 10.1128/MCB.6.11.3575.
- Jef D Boeke, George Church, Andrew Hessel, Nancy J Kelley, Adam Arkin, Yizhi Cai, Rob Carlson, Aravinda Chakravarti, Virginia W Cornish, Liam Holt, Farren J Isaacs, Todd Kuiken, Marc Lajoie, Tracy Lessor, Jeantine Lunshof, Matthew T Maurano, Leslie A Mitchell, Jasper Rine, Susan Rosser, Neville E Sanjana, Pamela A Silver, David Valle, Harris Wang, Jeffrey C Way, and Luhan Yang. GENOME ENGINEERING. The Genome Project-Write. *Science (New York, N.Y.)*, 353(6295) :126–7, jul 2016. ISSN 1095-9203. doi : 10.1126/science.aaf6850.
- Theodor Boveri. *Die Blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualität*. Engelmann, 1909.
- Alan P. Boyle, Sean Davis, Hennady P. Shulha, Paul Meltzer, Elliott H. Margulies, Zhiping Weng, Terrence S. Furey, and Gregory E. Crawford. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*, 132(2) :311–322, 2008. ISSN 00928674. doi : 10.1016/j.cell.2007.12.014.



- Donna Garvey Brickner, Ivelisse Cajigas, Yvonne Fondufe-Mittendorf, Sara Ahmed, Pei Chih Lee, Jonathan Widom, and Jason H. Brickner. H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biology*, 5(4) : 704–716, 2007. ISSN 15449173. doi : 10.1371/journal.pbio.0050081.
- Jason H. Brickner and Peter Walter. Gene recruitment of the activated INO1 locus to the nuclear membrane. *PLoS Biology*, 2(11), 2004. ISSN 15449173. doi : 10.1371/journal.pbio.0020342.
- Antoine Bridier-Nahmias and Pascale Lesage. Two large-scale analyses of Ty1 LTR-retrotransposon de novo insertion events indicate that Ty1 targets nucleosomal DNA near the H2A/H2B interface. *Mobile DNA*, 3(1) :3–6, 2012. ISSN 17598753. doi : 10.1186/1759-8753-3-22.
- Kristin Brogaard, Liqun Xi, Ji-ping Wang, and Jonathan Widom. A map of nucleosome positions in yeast at base-pair resolution. *Nature*, pages 2–7, jun 2012. ISSN 0028-0836. doi : 10.1038/nature11142.
- David M. Brown, Yujia A. Chan, Prashant J. Desai, Peter Grzesik, Lauren M. Oldfield, Sanjay Vashee, Jeffrey C. Way, Pamela A. Silver, and John I. Glass. Efficient size-independent chromosome delivery from yeast to cultured cell lines. *Nucleic Acids Research*, 45(7), 2017. ISSN 13624962. doi : 10.1093/nar/gkw1252.
- Mary Bryk, Mukti Banerjee, Michael Murphy, Karen E Knudsen, David J Garfinkel, and M Joan Curcio. Transcriptional silencing of Ty1 elements in the RDNI locus of yeast. *Genes & Development*, 11 :255–269, 1997. ISSN 08909369. doi : 10.1101/gad.11.2.255.
- E Bullitt, M P Rout, J V Kilmartin, and C W Akey. The yeast spindle pole body is assembled around a central crystal of Spc42p. *Cell*, 89(7) :1077–1086, 1997. ISSN 00928674. doi : 10.1016/S0092-8674(00)80295-0.
- Jennifer M. Bupp, Adriana E. Martin, Elizabeth S. Stensrud, and Sue L. Jaspersen. Telomere anchoring at the nuclear periphery requires the budding yeast Sad1-UNC-84 domain protein Mps3. *Journal of Cell Biology*, 179(5) :845–854, 2007. ISSN 00219525. doi : 10.1083/jcb.200706040.
- Sean M. Burgess and Nancy Kleckner. Collisions between yeast chromosomal loci in vivo are governed by three layers of organization. *Genes and Development*, 13(14) :1871–1883, 1999. ISSN 08909369. doi : 10.1101/gad.13.14.1871.
- B. Byers and L. Goetsch. Duplication of Spindle Plaques and Integration of the Yeast Cell Cycle. *Cold Spring Harbor Symposia on Quantitative Biology*, 38 :123–131, jan 1974. ISSN 0091-7451. doi : 10.1101/SQB.1974.038.01.016.
- Kerstin Bystricky, Patrick Heun, Lutz Gehlen, Jörg Langowski, and Susan M Gasser. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proceedings of the National Academy of Sciences of the United States of America*, 101(47) :16495–500, 2004. ISSN 0027-8424. doi : 10.1073/pnas.0402766101.
- Kerstin Bystricky, Thierry Laroche, Griet Van Houwe, Marek Blaszczyk, and Susan M. Gasser. Chromosome looping in yeast : Telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *Journal of Cell Biology*, 168(3) :375–387, 2005. ISSN 00219525. doi : 10.1083/jcb.200409091.
- Ghislain G. Cabal, Auguste Genovesio, Susana Rodriguez-Navarro, Christophe Zimmer, Olivier Gadal, Annick Lesne, Henri Buc, Frank Feuerbach-Fournier, Jean Christophe Olivo-Marin,

- Eduard C. Hurt, and Ulf Nehrbass. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, 441(7094) :770–773, 2006. ISSN 14764687. doi : 10.1038/nature04752.
- Maya Capelson, Yun Liang, Roberta Schulte, William Mair, Ulrich Wagner, and Martin W. Hetzer. Chromatin-Bound Nuclear Pore Components Regulate Gene Expression in Higher Eukaryotes. *Cell*, 140(3) :372–383, 2010. ISSN 00928674. doi : 10.1016/j.cell.2009.12.054.
- Lenny M. Carruthers, Jan Bednar, Christopher L. Woodcock, and Jeffrey C. Hansen. Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays : Mechanistic ramifications for higher-order chromatin folding. *Biochemistry*, 37(42) :14776–14787, 1998. ISSN 00062960. doi : 10.1021/bi981684e.
- Jason M. Casolari, Christopher R. Brown, Suzanne Komili, Jason West, Haley Hieronymus, and Pamela A. Silver. Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell*, 117(4) :427–439, 2004. ISSN 00928674. doi : 10.1016/S0092-8674(04)00448-9.
- T Caspersson, G Lomakka, and L Zech. The 24 fluorescence patterns of the human metaphase chromosomes - distinguishing characters and variability. *Hereditas*, 67(1) :89–102, 1972. ISSN 0018-0661.
- Jeronimo Cello. Chemical Synthesis of Poliovirus cDNA : Generation of Infectious Virus in the Absence of Natural Template. *Science*, 297(5583) :1016–1018, 2002. ISSN 00368075. doi : 10.1126/science.1072266.
- Janet N.Y. Chan, Betty P.K. Poon, Jayesh Salvi, Jonathan B. Olsen, Andrew Emili, and Karim Mekhail. Perinuclear Cohibin Complexes Maintain Replicative Life Span via Roles at Distinct Silent Chromatin Domains. *Developmental Cell*, 20(6) :867–879, jun 2011. ISSN 15345807. doi : 10.1016/j.devcel.2011.05.014.
- Leon Y Chan, Sriram Kosuri, and Drew Endy. Refactoring bacteriophage T7. *Molecular Systems Biology*, 1(1) :E1–E10, 2005. ISSN 1744-4292. doi : 10.1038/msb4100025.
- Iain M. Cheeseman, Scott Anderson, Miri Jwa, Erin M. Green, Jung seog Kang, John R. Yates, Clarence S.M. Chan, David G. Drubin, and Georjana Barnes. Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell*, 111(2) :163–172, 2002. ISSN 00928674. doi : 10.1016/S0092-8674(02)00973-X.
- C. Chen, H. H. Lim, J. Shi, S. Tamura, K. Maeshima, U. Surana, and L. Gan. Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo. *Molecular Biology of the Cell*, 27(21) :3357–3368, 2016. ISSN 1059-1524. doi : 10.1091/mbc.E16-07-0506.
- Tammy M.K. Cheng, Sebastian Heeger, Raphaël A.G. Chaleil, Nik Matthews, Aengus Stewart, Jon Wright, Carmay Lim, Paul A. Bates, and Frank Uhlmann. A simple biophysical model emulates budding yeast chromosome condensation. *eLife*, 2015(4) :1–22, 2015. ISSN 2050084X. doi : 10.7554/eLife.05565.
- Sau Wai Cheung, Peter V. Tishler, Leonard Atkins, Sisir K. Sengupta, Edward J. Modest, and Bernard G. Forget. Gene mapping by fluorescent in situ hybridization. *Cell Biology International Reports*, 1(3) :255–262, 1977. ISSN 03091651. doi : 10.1016/0309-1651(77)90050-9.
- Heidi J Chial and Mark Winey. Mechanisms of genetic instability revealed by analysis of yeast spindle pole body duplication. *Biology of the Cell*, 91(6) :439–450, jul 1999. ISSN 02484900. doi : 10.1111/j.1768-322X.1999.tb01099.x.

- Michelle Christian, Tomas Cermak, Erin L. Doyle, Clarice Schmidt, Feng Zhang, Aaron Hummel, Adam J. Bogdanove, and Daniel F. Voytas. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186(2) :756–761, 2010. ISSN 00166731. doi : 10.1534/genetics.110.120717.
- Cedric R. Clapier and Bradley R. Cairns. The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, 78(1) :273–304, 2009. ISSN 0066-4154. doi : 10.1146/annurev.biochem.77.062706.153223.
- Louise Clarke and John Carbon. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, 287(5782) :504–509, oct 1980. ISSN 0028-0836. doi : 10.1038/287504a0.
- G Cottarel, J H Shero, P Hieter, and J H Hegemann. A 125-base-pair CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 9(8) :3342–3349, aug 1989. ISSN 0270-7306. doi : 10.1128/MCB.9.8.3342.
- Axel Cournac, Hervé Marie-Nelly, Martial Marbouty, Romain Koszul, and Julien Mozziconacci. Normalization of a chromosomal contact map. *BMC Genomics*, 13(1) :436, 2012. ISSN 1471-2164. doi : 10.1186/1471-2164-13-436.
- Emily Crane, Qian Bian, Rachel Patton McCord, Bryan R. Lajoie, Bayly S. Wheeler, Edward J. Ralston, Satoru Uzawa, Job Dekker, and Barbara J. Meyer. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*, 523(7559) :240–244, 2015. ISSN 14764687. doi : 10.1038/nature14450.
- Gregory E Crawford, Sean Davis, Peter C Scacheri, Gabriel Renaud, Mohamad J Halawi, Michael R Erdos, Roland Green, Paul S Meltzer, Tyra G Wolfsberg, and Francis S Collins. DNase-chip : a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature Methods*, 3(7) :503–509, 2006. ISSN 1548-7091. doi : 10.1038/nmeth888.
- Roberto Crea and Adam Kraszewski. Chemical synthesis of genes for human insulin. *Proceedings of the National Academy of Sciences of the United States of America*, 75(12) :5765–5769, 1978.
- Thomas Cremer and Marion Cremer. Chromosome territories. *Cold Spring Harbor perspectives in biology*, 2(3) :a003889, mar 2010. ISSN 1943-0264. doi : 10.1101/cshperspect.a003889.
- F. Crick. Central dogma of molecular biology. *Nature*, 227(5258) :561–3, 1970. ISSN 0028-0836. doi : 10.1038/227561a0.
- Y. Cui and C. Bustamante. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proceedings of the National Academy of Sciences*, 97(1) :127–132, 2000. ISSN 0027-8424. doi : 10.1073/pnas.97.1.127.
- Sara Cuylen, Jutta Metz, and Christian H. Haering. Condensin structures chromosomal DNA through topological links. *Nature Structural and Molecular Biology*, 18(8) :894–901, 2011. ISSN 15459993. doi : 10.1038/nsmb.2087.
- Antoine Danchin and Gang Fang. Unknown unknowns : essential genes in quest for function. *Microbial Biotechnology*, 9(5) :530–540, 2016. ISSN 17517915. doi : 10.1111/1751-7915.12384.
- K A Datsenko and B L Wanner. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12) :6640–5, 2000. ISSN 0027-8424. doi : 10.1073/pnas.120163297.

- James O.J. Davies, A. Marieke Oudelaar, Douglas R. Higgs, and Jim R. Hughes. How best to identify chromosomal interactions : A comparison of approaches. *Nature Methods*, 14(2) : 125–134, 2017. ISSN 15487105. doi : 10.1038/nmeth.4146.
- Job Dekker. Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *Journal of Biological Chemistry*, 283(50) :34532–34540, 2008. ISSN 00219258. doi : 10.1074/jbc.M806479200.
- Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *science*, 295(5558) :1306–1311, 2002.
- G. Dieppois, N. Iglesias, and F. Stutz. Cotranscriptional Recruitment to the mRNA Export Receptor Mex67p Contributes to Nuclear Pore Anchoring of Activated Genes. *Molecular and Cellular Biology*, 26(21) :7858–7870, 2006. ISSN 0270-7306. doi : 10.1128/MCB.00870-06.
- P J Dillon and C A Rosen. A rapid method for the construction of synthetic genes using the polymerase chain reaction. *BioTechniques*, 9(3) :298, 300, sep 1990. ISSN 0736-6205.
- B. Dorigo. Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. *Science*, 306(5701) :1571–1573, 2004. ISSN 0036-8075. doi : 10.1126/science.1103124.
- J.F. Dorn, K. Jaqaman, D.R. Rines, G.S. Jelson, P.K. Sorger, and G. Danuser. Yeast Kinetochore Microtubule Dynamics Analyzed by High-Resolution Three-Dimensional Microscopy. *Biophysical Journal*, 89(4) :2835–2854, 2005. ISSN 00063495. doi : 10.1529/biophysj.104.058461.
- Jessica A. Downs, Effie Kosmidou, Alan Morgan, and Stephen P. Jackson. Suppression of homologous recombination by the *Saccharomyces cerevisiae* linker histone. *Molecular Cell*, 11(6) :1685–1692, 2003. ISSN 10972765. doi : 10.1016/S1097-2765(03)00197-7.
- Zhijun Duan, Mirela Andronescu, Kevin Schutz, Sean McIlwain, Yoo Jung Kim, Choli Lee, Jay Shendure, Stanley Fields, C. Anthony Blau, and William S. Noble. A three-dimensional model of the yeast genome. *Nature*, 465(7296) :363–367, 2010. ISSN 0028-0836. doi : 10.1038/nature08973.
- J Dubochet, M Adrian, P Schultz, and P Oudet. Cryo-electron microscopy of vitrified SV40 minichromosomes : the liquid drop model. *The EMBO journal*, 5(3) :519–28, 1986. ISSN 0261-4189.
- John J Dunn, F William Studier, and M Gottesman. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of molecular biology*, 166(4) :477–535, 1983.
- Jessica S. Dymond, Sarah M. Richardson, Candice E. Coombes, Timothy Babatz, Héloïse Muller, Narayana Annaluru, William J. Blake, Joy W. Schwerzmann, Junbiao Dai, Derek L. Lindstrom, Annabel C. Boeke, Daniel E. Gottschling, Srinivasan Chandrasegaran, Joel S. Bader, and Jef D. Boeke. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, 477(7365) :471–476, 2011. ISSN 00280836. doi : 10.1038/nature10403.
- Thoma F., Th Koller, and A Klug. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell. Biol.*, 83(November) :403–427, 1979. ISSN 0021-9525. doi : 10.1083/jcb.83.2.403.
- Peggy J. Farnham. Insights from genomic profiling of transcription factors. *Nature Reviews Genetics*, 10(9) :605–616, 2009. ISSN 14710056. doi : 10.1038/nrg2636.

- J. T. Finch and A. Klug. Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences*, 73(6) :1897–1901, 1976. ISSN 0027-8424. doi : 10.1073/pnas.73.6.1897.
- M Fitzgerald-Hayes, L Clarke, and J Carbon. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell*, 29(1) :235–244, 1982. ISSN 0092-8674. doi : 10.1016/0092-8674(82)90108-8.
- Rosalind E Franklin and Raymond G Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171 :740–741, 1953.
- I Freidkin and D J Katcoff. Specific distribution of the *Saccharomyces cerevisiae* linker histone homolog HHO1p in the chromatin. *Nucleic acids research*, 29(19) :4043–4051, 2001. ISSN 1362-4962. doi : 10.1093/nar/29.19.4043.
- C E Fritze, K Verschueren, R Strich, and R Easton Esposito. Direct evidence for SIR2 modulation of chromatin structure in yeast rDNA. *The EMBO journal*, 16(21) :6495–509, nov 1997. ISSN 0261-4189. doi : 10.1093/emboj/16.21.6495.
- Eden Fussner, Reagan W. Ching, and David P. Bazett-Jones. Living without 30nm chromatin fibers. *Trends in Biochemical Sciences*, 36(1) :1–6, 2011. ISSN 09680004. doi : 10.1016/j.tibs.2010.09.002.
- Thomas Gaj, Charles A. Gersbach, and Carlos F. Barbas. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, 31(7) :397–405, 2013. ISSN 01677799. doi : 10.1016/j.tibtech.2013.04.004.
- D J Galas and A Schmitz. DNase footprinting : a simple method for the detection of protein-DNA binding specificity. *Nucleic acids research*, 5(9) :3157–70, 1978. ISSN 0305-1048. doi : 10.1093/nar/5.9.3157.
- Joseph G Gall and Mary Lou Pardue. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proceedings of the National Academy of Sciences of the United States of America*, 63(2) :378–83, jun 1969. ISSN 0027-8424. doi : 10.1073/pnas.63.2.378.
- Marc R. Gartenberg, Frank R. Neumann, Thierry Laroche, Marek Blaszczyk, and Susan M. Gasser. Sir-mediated repression can occur independently of chromosomal and subnuclear contexts. *Cell*, 119(7) :955–967, 2004. ISSN 00928674. doi : 10.1016/j.cell.2004.11.008.
- D. G. Gibson, G. A. Benders, K. C. Axelrod, J. Zaveri, M. A. Algire, M. Moodie, M. G. Montague, J. C. Venter, H. O. Smith, and C. A. Hutchison. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proceedings of the National Academy of Sciences*, 105(51) :20404–20409, 2008. ISSN 0027-8424. doi : 10.1073/pnas.0811011106.
- Daniel G. Gibson. Programming biological operating systems : genome design, assembly and activation. *Nature methods*, 11(5) :521–6, may 2014. ISSN 1548-7105. doi : 10.1038/nmeth.2894.
- Daniel G Gibson, John I Glass, Carole Lartigue, Vladimir N Noskov, Ray-Yuan Chuang, Mikkel A Algire, Gwynedd A Benders, Michael G Montague, Li Ma, Monzia M Moodie, and Others. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 329(5987) :52–56, 2010.
- Yuko Giga-Hama, Hideki Tohda, Kaoru Takegawa, and Hiromichi Kumagai. Schizosaccharomyces pombe minimum genome factory. *Biotechnology and Applied Biochemistry*, 46(3) :147, 2007. ISSN 08854513. doi : 10.1042/BA20060106.

- Paul G Giresi, Jonghwan Kim, R. M. McDaniel, Vishwanath R Iyer, and Jason D Lieb. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research*, 17(6) :877–885, jun 2007. ISSN 1088-9051. doi : 10.1101/gr.5533506.
- D V Goeddel, D G Kleid, F Bolivar, H L Heyneker, D G Yansura, R Crea, T Hirose, A Kraszewski, K Itakura, and A D Riggs. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences of the United States of America*, 76(1) :106–110, 1979. ISSN 0027-8424. doi : 10.1073/pnas.76.1.106.
- A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 Genes. *Science*, 274(5287) :546–567, 1996. ISSN 0036-8075. doi : 10.1126/science.274.5287.546.
- M Gotta, T Laroche, A Formenton, L Maillet, H Scherthan, and S M Gasser. The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. *The Journal of cell biology*, 134(6) :1349–63, sep 1996. ISSN 0021-9525. doi : 10.1083/jcb.134.6.1349.
- Shoshanna Gottlieb and Rochelle Easton Esposito. A new role for a yeast transcriptional silencer gene, SIR2, in regulation of recombination in ribosomal DNA. *Cell*, 56(5) :771–776, mar 1989. ISSN 00928674. doi : 10.1016/0092-8674(89)90681-8.
- Daniel E. Gottschling, Oscar M. Aparicio, Barbara L. Billington, and Virginia A. Zakian. Position effect at *S. cerevisiae* telomeres : Reversible repression of Pol II transcription. *Cell*, 63(4) : 751–762, 1990. ISSN 00928674. doi : 10.1016/0092-8674(90)90141-Z.
- S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proceedings of the National Academy of Sciences*, 106(32) :13317–13322, 2009. ISSN 0027-8424. doi : 10.1073/pnas.0903280106.
- Sergei A. Grigoryev, Gavin Bascom, Jenna M. Buckwalter, Michael B. Schubert, Christopher L. Woodcock, and Tamar Schlick. Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. *Proceedings of the National Academy of Sciences*, 113(5) :1238–1243, 2016. ISSN 0027-8424. doi : 10.1073/pnas.1518280113.
- V. Guacci, E Hogan, and D Koshland. Chromosome condensation and sister chromatid pairing in budding yeast. *The Journal of cell biology*, 125(3) :517–30, may 1994. ISSN 0021-9525. doi : 10.1083/jcb.125.3.517.
- V Guacci, E Hogan, and D Koshland. Centromere position in budding yeast : evidence for anaphase A. *Molecular biology of the cell*, 8(6) :957–72, 1997. ISSN 1059-1524. doi : 10.1091/MBC.8.6.957.
- Micol Guidi, Myriam Ruault, Martial Marbouty, Isabelle Loïodice, Axel Cournac, Cyrille Billau-deau, Antoine Hoher, Julien Mozziconacci, Romain Koszul, and Angela Taddei. Spatial reorganization of telomeres in long-lived quiescent cells. *Genome Biology*, 16(1) :206, 2015. ISSN 1474-760X. doi : 10.1186/s13059-015-0766-2.
- James E. Haber. Mating-Type Genes and MAT Switching in *Saccharomyces cerevisiae*. *Genetics*, 191(1) :33–64, may 2012. ISSN 0016-6731. doi : 10.1534/genetics.111.134577.
- Rebecca A Haeusler, M. Pratt-Hyatt, Paul D Good, Theresa A Gipson, and David R Engelke. Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. *Genes & Development*, 22(16) :2204–2214, aug 2008. ISSN 0890-9369. doi : 10.1101/gad.1675908.

- Adrian Halme, Stacie Bumgarner, Cora Styles, and Gerald R. Fink. Genetic and Epigenetic Regulation of the FLO Gene Family Generates Cell-Surface Variation in Yeast. *Cell*, 116(3) : 405–415, feb 2004. ISSN 00928674. doi : 10.1016/S0092-8674(04)00118-7.
- Housheng Hansen He, Clifford A. Meyer, Mei Wei Chen, V. Craig Jordan, Myles Brown, and X. Shirley Liu. Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Research*, 22(6) :1015–1025, 2012. ISSN 10889051. doi : 10.1101/gr.133280.111.
- Andreas Hecht, Thierry Laroche, Sabine Strahl-Bolsinger, Susan M Gasser, and Michael Grunstein. Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins : A molecular model for the formation of heterochromatin in yeast. *Cell*, 80(4) :583–592, feb 1995. ISSN 00928674. doi : 10.1016/0092-8674(95)90512-X.
- Florence Hediger, Karine Dubrana, and Susan M. Gasser. Myosin-like proteins 1 and 2 are not required for silencing or telomere anchoring, but act in the Tel1 pathway of telomere length control. *Journal of Structural Biology*, 140(1-3) :79–91, 2002. ISSN 10478477. doi : 10.1016/S1047-8477(02)00533-6.
- Danièle Hernandez-Verdun, Pascal Roussel, and Jeannine Gébrane-Younès. Emerging concepts of nucleolar assembly. *Journal of cell science*, 115(Pt 11) :2265–2270, 2002. ISSN 0021-9533.
- H L Heyneker, J Shine, H M Goodman, H W Boyer, J Rosenberg, R E Dickerson, S A Narang, K Itakura, S.-Y. Lin, and A D Riggs. Synthetic lacoperator DNA is functional in vivo. *Nature*, 263(5580) :748–752, 1976. ISSN 0028-0836. doi : 10.1038/263748a0.
- A Hill and K Bloom. Genetic manipulation of centromere function. *Molecular and cellular biology*, 7(7) :2397–405, 1987. ISSN 0270-7306. doi : 10.1128/MCB.7.7.2397.Updated.
- N J Hillson, R D Rosengarten, and J D Keasling. j5 DNA assembly design automation software. *ACS Synth Biol*, 1 :14–21, 2012.
- R H Hoess, A Wierzbicki, and K Abremski. The role of the loxP spacer region in P1 site-specific recombination. *Nucleic acids research*, 14(5) :2287–300, 1986. ISSN 0305-1048. doi : 10.1093/nar/gkn942.
- R. W. Holley, G. A. Everett, J. T. Madison, and A. Zamir. Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid. *The Journal of Biological Chemistry*, 240(5) :2122–2128, 1965.
- P J Horn and C L Peterson. Chromatin higher order folding : wrapping up transcription. *Science*, 297(5588) :1824–1827, 2002. ISSN 1095-9203. doi : 10.1126/science.1074200.
- Amanda L. Hughes and Oliver J. Rando. Mechanisms Underlying Nucleosome Positioning In Vivo. *Annual Review of Biophysics*, 43(1) :41–63, 2014. ISSN 1936-122X. doi : 10.1146/annurev-biophys-051013-023114.
- C A Hutchison, Scott N. Peterson, Steven R. Gill, Robin T. Cline, Owen White, Calire M. Fraser, Hamilton O Smith, and J Craig Venter. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science (New York, N.Y.)*, 286(5447) :2165–9, dec 1999. ISSN 0036-8075. doi : 10.1126/science.286.5447.2165.
- Clyde A. Hutchison, Ray-Yuan Chuang, Vladimir N. Noskov, Nacyra Assad-Garcia, Thomas J. Deerinck, Mark H. Ellisman, John Gill, Krishna Kannan, Bogumil J. Karas, Li Ma, James F. Pelletier, Zhi-Qing Qi, R. Alexander Richter, Elizabeth A. Strychalski, Lijie Sun, Yo Suzuki, Billyana Tsvetanova, Kim S. Wise, Hamilton O. Smith, John I. Glass, Chuck Merryman,

- Daniel G. Gibson, and J. Craig Venter. Design and synthesis of a minimal bacterial genome. *Science (New York, N.Y.)*, 351(6280) :aad6253, mar 2016. ISSN 1095-9203. doi : 10.1126/science.aad6253.
- Maxim Imakaev, Geoffrey Fudenberg, Rachel Patton McCord, Natalia Naumova, Anton Goloborodko, Bryan R Lajoie, Job Dekker, and Leonid A Mirny. Iterative correction of hi-c data reveals hallmarks of chromosome organization. *Nature methods*, 9(10) :999–1003, 2012.
- Farren J Isaacs, Peter A Carr, Harris H Wang, Marc J Lajoie, Bram Sterling, Laurens Kraal, Andrew C Tolonen, Tara A Gianoulis, Daniel B Goodman, Nikos B Reppas, Christopher J Emig, Duhee Bang, Samuel J Hwang, Michael C Jewett, Joseph M Jacobson, and George M Church. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science (New York, N.Y.)*, 333(6040) :348–53, jul 2011. ISSN 1095-9203. doi : 10.1126/science.1205822.
- K. Itakura, N. Katagiri, S. A. Narang, C. P. Bahl, K. J. Mariani, and R. Wu. Chemical synthesis and sequence studies of deoxyribonucleotides which constitute the duplex sequence of the lactose operator of *Escherichia coli*. *Journal of Biological Chemistry*, 250(12) :4592–4600, 1975. ISSN 00219258.
- K Itakura, T Hirose, R Crea, A. Riggs, H. Heyneker, F Bolivar, and H. Boyer. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science*, 198 (4321) :1056–1063, dec 1977. ISSN 0036-8075. doi : 10.1126/science.412251.
- C W Jacobs, A E Adams, P J Szaniszló, and J R Pringle. Functions of microtubules in the *Saccharomyces cerevisiae* cell cycle. *The Journal of cell biology*, 107(4) :1409–26, oct 1988. ISSN 0021-9525. doi : 10.1083/jcb.107.4.1409.
- H. Ji, D. P. Moore, M. A. Blomberg, L. T. Braiterman, D. F. Voytas, G. Natsoulis, and J. D. Boeke. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell*, 73(5) :1007–1018, 1993. ISSN 00928674. doi : 10.1016/0092-8674(93)90278-X.
- Q W Jin, J Fuchs, and J Loidl. Centromere clustering is a major determinant of yeast interphase nuclear organization. *Journal of Cell Science*, 113 ( Pt 1 :1903–1912, 2000. ISSN 0021-9533.
- Quan Wen Jin, Edgar Trelles-Sticken, Harry Scherthan, and Josef Loidl. Yeast nuclei display prominent centromere clustering that is reduced in nondividing cells and in meiotic prophase. *Journal of Cell Biology*, 141(1) :21–29, 1998. ISSN 00219525. doi : 10.1083/jcb.141.1.21.
- Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096) :816–21, aug 2012. ISSN 1095-9203. doi : 10.1126/science.1225829.
- Ajit P. Joglekar, Kerry Bloom, and E. D. Salmon. In Vivo Protein Architecture of the Eukaryotic Kinetochore with Nanometer Scale Accuracy. *Current Biology*, 19(8) :694–699, 2009. ISSN 09609822. doi : 10.1016/j.cub.2009.02.056.
- H. A. John, M. L. Birnstiel, and K. W. Jones. RNA-DNA hybrids at the cytological level. *Nature*, 223(5206) :582–7, aug 1969. ISSN 0028-0836. doi : 10.1038/223582a0.
- R A Jorgensen, S J Rothstein, and W S Reznikoff. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. *Molecular & general genetics : MGG*, 177 (1) :65–72, 1979. ISSN 0026-8925.



- Noam Kaplan, Irene K. Moore, Yvonne Fondufe-Mittendorf, Andrea J. Gossett, Desiree Tillo, Yair Field, Emily M. LeProust, Timothy R. Hughes, Jason D. Lieb, Jonathan Widom, and Eran Segal. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236) : 362–366, mar 2009. ISSN 0028-0836. doi : 10.1038/nature07667.
- Jonathan R. Karr, Jayodita C. Sanghvi, Derek N. MacKlin, Miriam V. Gutschow, Jared M. Jacobs, Benjamin Bolival, Nacyra Assad-Garcia, John I. Glass, and Markus W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2) :389–401, 2012. ISSN 00928674. doi : 10.1016/j.cell.2012.05.044.
- M. A. Keene, V. Corces, K. Lowenhaupt, and S. C. Elgin. DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription. *Proceedings of the National Academy of Sciences*, 78(1) :143–146, 1981. ISSN 0027-8424. doi : 10.1073/pnas.78.1.143.
- H. G. Khorana, K. L. Agarwal, H. Büchi, M. H. Caruthers, N. K. Gupta, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. RajBhandary, J. H. van de Sande, V. Sgaramella, T. Tebao, H. Weber, and T. Yamada. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. *Journal of Molecular Biology*, 72(2) :209–217, 1972. doi : [https://doi.org/10.1016/0022-2836\(72\)90146-5](https://doi.org/10.1016/0022-2836(72)90146-5).
- Sepideh Khorasanizadeh. The nucleosome : from genomic organization to genomic regulation. *Cell*, 116(2) :259–72, jan 2004. ISSN 0092-8674. doi : 10.1016/S0092-8674(04)00044-3.
- Naomi Kitamura, Bert L. Semler, Paul G. Rothberg, Glenn R. Larsen, Cheryl J. Adler, Andrew J. Dorner, Emilio A. Emini, Ronnie Hanecak, James J. Lee, Sylvie Van Der Werf, Carl W. Anderson, and Eckard Wimmer. Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature*, 291(5816) :547–553, 1981. ISSN 00280836. doi : 10.1038/291547a0.
- Gregory D. Koblenz. The <i>De Novo</i> Synthesis of Horsepox Virus : Implications for Biosecurity and Recommendations for Preventing the Reemergence of Smallpox. *Health Security*, 15(5) :hs.2017.0061, 2017. ISSN 2326-5094. doi : 10.1089/hs.2017.0061.
- Sarah J. Kodumal and Daniel V. Santi. DNA ligation by selection. *BioTechniques*, 37(1) :34–40, 2004. ISSN 07366205.
- Peter König, Michael B. Braunfeld, John W. Sedat, and David A. Agard. The three-dimensional structure of in vitro reconstituted *Xenopus laevis* chromosomes by EM tomography. *Chromosoma*, 116(4) :349–372, 2007. ISSN 00095915. doi : 10.1007/s00412-007-0101-0.
- Philipp Korber and Wolfram Hürz. In vitro assembly of the characteristic chromatin organization at the yeast PHO5 promoter by a replication-independent extract system. *Journal of Biological Chemistry*, 279(33) :35113–35120, 2004. ISSN 00219258. doi : 10.1074/jbc.M405446200.
- R. D. Kornberg. Chromatin structure : a repeating unit of histones and DNA. *Science (New York, N.Y.)*, 184(4139) :868–71, may 1974. ISSN 0036-8075. doi : 10.1126/science.184.4139.868.
- Roger D. Kornberg and Yahli Lorch. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, 98(3) :285–294, 1999. ISSN 00928674. doi : 10.1016/S0092-8674(00)81958-3.
- Roger D. Kornberg and Lubert Stryer. Statistical distributions of nucleosomes : nonrandom locations by a stochastic mechanism. *Nucleic Acids Research*, 16(14) :6677–6690, 1988. ISSN 0305-1048. doi : 10.1093/nar/16.14.6677.

- Sriram Kosuri, Nikolai Eroshenko, Emily M LeProust, Michael Super, Jeffrey Way, Jin Billy Li, and George M Church. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nature biotechnology*, 28(12) :1295–1299, 2010.
- Nils Krietenstein, Megha Wal, Shinya Watanabe, Bongsoo Park, Craig L. Peterson, B. Franklin Pugh, and Philipp Korber. Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell*, 167(3) :709–721.e12, 2016. ISSN 10974172. doi : 10.1016/j.cell.2016.09.045.
- Maarten Kruithof, Fan Tso Chien, Andrew Routh, Colin Logie, Daniela Rhodes, and John Van Noort. Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nature Structural and Molecular Biology*, 16(5) :534–540, 2009. ISSN 15459993. doi : 10.1038/nsmb.1590.
- Katrin Küpper, Alexandra Kölbl, Dorothee Biener, Sandra Dittrich, Johann von Hase, Tobias Thormeyer, Heike Fiegler, Nigel P. Carter, Michael R. Speicher, Thomas Cremer, and Marion Cremer. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma*, 116(3) :285–306, 2007. ISSN 00095915. doi : 10.1007/s00412-007-0098-4.
- V Larionov, N Kouprina, J Graves, and M a Resnick. Highly selective isolation of human DNAs from rodent-human hybrid cells as circular yeast artificial chromosomes by transformation-associated recombination cloning. *Proceedings of the National Academy of Sciences of the United States of America*, 93(November) :13925–13930, 1996. ISSN 00278424. doi : 10.1073/pnas.93.24.13925.
- Thierry Laroche, Sophie G. Martin, Monica Gotta, Hazel C. Gorham, Fiona E. Pryde, Edward J. Louis, and Susan M. Gasser. Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Current Biology*, 8(11) :653–657, 1998. ISSN 09609822. doi : 10.1016/S0960-9822(98)70252-0.
- Carole Lartigue, John I Glass, Nina Alperovich, Rembert Pieper, Prashanth P Parmar, Clyde A Hutchison, Hamilton O Smith, and J Craig Venter. Genome transplantation in bacteria : changing one species to another. *Science*, 317(5838) :632–638, 2007.
- Carole Lartigue, Sanjay Vashee, Mikkel A Algire, Ray-yuan Chuang, Gwynedd A Benders, Li Ma, Vladimir N Noskov, Evgeniya A Denisova, Daniel G Gibson, Nacyra Assad-Garcia, Nina Alperovich, David W Thomas, Chuck Merryman, Clyde A Hutchison, Hamilton O Smith, J Craig Venter, and John I Glass. Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science (New York, N.Y.)*, 325(5948) :1693–6, sep 2009. ISSN 1095-9203. doi : 10.1126/science.1173759.
- Imen Lassadi and Kerstin Bystricky. Tracking of single and multiple genomic loci in living yeast cells. *Methods in molecular biology (Clifton, N.J.)*, 745 :499–522, 2011. ISSN 1940-6029. doi : 10.1007/978-1-61779-129-1\_29.
- Brigitte D. Lavoie, Eileen Hogan, and Douglas Koshland. In vivo dissection of the chromosome condensation machinery : Reversibility of condensation distinguishes contributions of condensin and cohesin. *Journal of Cell Biology*, 156(5) :805–815, 2002. ISSN 00219525. doi : 10.1083/jcb.200109056.
- Tung B K Le, Maxim V Imakaev, Leonid a Mirny, and Michael T Laub. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science (New York, N.Y.)*, 342(6159) : 731–4, 2013. ISSN 1095-9203. doi : 10.1126/science.1242059.
- William Lee, Desiree Tillo, Nicolas Bray, Randall H. Morse, Ronald W. Davis, Timothy R. Hughes, and Corey Nislow. A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics*, 39(10) :1235–1244, 2007. ISSN 10614036. doi : 10.1038/ng2117.

- Isabelle Léger-Silvestre, Stéphanie Trumtel, Jacqueline Noaillac-Depeyre, and Nicole Gas. Functional compartmentalization of the nucleus in the budding yeast *Saccharomyces cerevisiae*. *Chromosoma*, 108(2) :103–13, may 1999. ISSN 0009-5915. doi : 10.1007/s004120050357.
- Jerome Lejeune, Marion Gautier, and Rodman Turpin. Study of somatic chromosomes from 9 mongoloid children. *Comptes rendus de l'Academie des sciences*, 248 11 :1721–2, 1959.
- Francene J. Lemoine, Natasha P. Degtyareva, Kirill Lobachev, and Thomas D. Petes. Chromosomal translocations in yeast induced by low levels of DNA polymerase : A model for chromosome fragile sites. *Cell*, 120(5) :587–598, 2005. ISSN 00928674. doi : 10.1016/j.cell.2004.12.039.
- Annick Lesne, Julien Riposo, Paul Roger, Axel Cournac, and Julien Mozziconacci. 3D genome reconstruction from chromosomal contacts. *Nature Methods*, 11(11) :1141–1143, 2014. ISSN 15487105. doi : 10.1038/nmeth.3104.
- Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4) :291–305, 2012. ISSN 17401526. doi : 10.1038/nrmicro2737.
- E. Lieberman-Aiden, N. L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A Mirny, E. S. Lander, and J. Dekker. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326(5950) :289–293, oct 2009. ISSN 0036-8075. doi : 10.1126/science.1181369.
- Derek L. Lindstrom and Daniel E. Gottschling. The mother enrichment program : A genetic system for facile replicative life span analysis in *Saccharomyces cerevisiae*. *Genetics*, 183(2) : 413–422, 2009. ISSN 00166731. doi : 10.1534/genetics.109.106229.
- G. Ling, A. Sugathan, T. Mazor, E. Fraenkel, and D. J. Waxman. Unbiased, Genome-Wide In Vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-Specific Liver Gene Expression. *Molecular and Cellular Biology*, 30(23) :5531–5544, 2010. ISSN 0270-7306. doi : 10.1128/MCB.00601-10.
- Gianni Liti and Edward J. Louis. Yeast Evolution and Comparative Genomics. *Annual Review of Microbiology*, 59(1) :135–153, 2005. ISSN 0066-4227. doi : 10.1146/annurev.micro.59.030804.121400.
- Edward J. Louis. The chromosome ends of *Saccharomyces cerevisiae*. *Yeast*, 11(16) :1553–1573, 1995. ISSN 10970061. doi : 10.1002/yea.320111604.
- P.T Lowary and J Widom. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, 276(1) :19–42, 1998. ISSN 00222836. doi : 10.1006/jmbi.1997.1494.
- K Luger, a W Mäder, R K Richmond, D F Sargent, and T J Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648) :251–260, 1997a. ISSN 0028-0836. doi : 10.1038/38444.
- Karolin Luger, Thomas J Rechsteiner, Andrew J Flaus, Mary M.Y Waye, and Timothy J Richmond. Characterization of nucleosome core particles containing histone proteins made in bacteria 1 Edited by A. Klug. *Journal of Molecular Biology*, 272(3) :301–311, 1997b. ISSN 00222836. doi : 10.1006/jmbi.1997.1235.

- Karolin Luger, Mekonnen L. Dechassa, and David J. Tremethick. New insights into nucleosome and chromatin structure : An ordered state or a disordered affair ? *Nature Reviews Molecular Cell Biology*, 13(7) :436–447, 2012. ISSN 14710072. doi : 10.1038/nrm3382.
- Roopa Luthra, Shana C. Kerr, Michelle T. Harreman, Luciano H. Apponi, Milo B. Fasken, Suneela Ramineni, Shyam Chaurasia, Sandro R. Valentini, and Anita H. Corbett. Actively transcribed GAL genes can be physically linked to the nuclear pore by the SAGA chromatin modifying complex. *Journal of Biological Chemistry*, 282(5) :3042–3049, 2007. ISSN 00219258. doi : 10.1074/jbc.M608741200.
- K. Maeshima, S. Hihara, and H. Takata. New Insight into the Mitotic Chromosome Structure : Irregular Folding of Nucleosome Fibers Without 30-nm Chromatin Structure. *Cold Spring Harbor Symposia on Quantitative Biology*, 75(2009) :439–444, jan 2010a. ISSN 0091-7451. doi : 10.1101/sqb.2010.75.034.
- Kazuhiro Maeshima, Saera Hihara, and Mikhail Eltsov. Chromatin structure : Does the 30-nm fibre exist in vivo ? *Current Opinion in Cell Biology*, 22(3) :291–297, 2010b. ISSN 09550674. doi : 10.1016/j.ceb.2010.03.001.
- Kazuhiro Maeshima, Satoru Ide, Kayo Hibino, and Masaki Sasai. Liquid-like behavior of chromatin. *Current Opinion in Genetics and Development*, 37 :36–45, 2016. ISSN 18790380. doi : 10.1016/j.gde.2015.11.006.
- Laurent Maillet, Florence Gaden, Vanessa Brevet, Geneviève Fourel, Sophie G. Martin, Karine Dubrana, Susan M. Gasser, and Eric Gilson. Ku-deficient yeast strains exhibit alternative states of silencing competence. *EMBO Reports*, 2(3) :203–210, 2001. ISSN 1469221X. doi : 10.1093/embo-reports/kve044.
- W. Mandecki and T. J. Boiling. FokI method of gene synthesis. *Gene*, 68(1) :101–107, 1988. ISSN 03781119. doi : 10.1016/0378-1119(88)90603-8.
- Wlodek Mandecki, Mark A. Hayden, Mary Ann Shallcross, and Elizabeth Stotland. A totally synthetic plasmid for general cloning, gene expression and mutagenesis in *Escherichia coli*. *Gene*, 94(1) :103–107, sep 1990. ISSN 03781119. doi : 10.1016/0378-1119(90)90474-6.
- J Manning, N Hershey, T Broker, M Pellegrini, H Mitchell, and N Davidson. A New Method of in situ Hybridization. *Chromosoma*, 117(5121) :107–117, 1975.
- Martial Marbouty, Antoine Le Gall, Diego I. Cattoni, Axel Cournac, Alan Koh, Jean Bernard Fiche, Julien Mozziconacci, Heath Murray, Romain Koszul, and Marcelo Nollmann. Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Molecular Cell*, 59(4) :588–602, 2015. ISSN 10974164. doi : 10.1016/j.molcel.2015.07.020.
- Hervé Marie-Nelly, Martial Marbouty, Axel Cournac, Jean François Flot, Gianni Liti, Dante Poggi Parodi, Sylvie Syan, Nancy Guillén, Antoine Margeot, Christophe Zimmer, and Romain Koszul. High-quality genome (re)assembly using chromosomal contact data. *Nature Communications*, 5 :1–10, 2014. ISSN 20411723. doi : 10.1038/ncomms6695.
- M. P.F. Marsden and U. K. Laemmli. Metaphase chromosome structure : Evidence for a radial loop model. *Cell*, 17(4) :849–858, 1979. ISSN 00928674. doi : 10.1016/0092-8674(79)90325-8.
- Fabrizio Martino, Stephanie Kueng, Philip Robinson, Monika Tsai-Pflugfelder, Fred van Leeuwen, Mathias Ziegler, Fabien Cubizolles, Moira M. Cockell, Daniela Rhodes, and Susan M. Gasser. Reconstitution of Yeast Silent Chromatin : Multiple Contact Sites and O-AADPR Binding Load SIR Complexes onto Nucleosomes In Vitro. *Molecular Cell*, 33(3) :323–334, 2009. ISSN 10972765. doi : 10.1016/j.molcel.2009.01.009.

- N Mathias, S L Johnson, M Winey, a E Adams, L Goetsch, J R Pringle, B Byers, and M G Goehl. Cdc53p acts in concert with Cdc4p and Cdc34p to control the G1-to-S-phase transition and identifies a conserved family of proteins. *Molecular and cellular biology*, 16(12) :6634–6643, 1996. ISSN 0270-7306. doi : 10.1128/MCB.16.12.6634.
- a M Maxam and W Gilbert. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2) :560–564, feb 1977. ISSN 0027-8424. doi : 10.1073/pnas.74.2.560.
- Andrew D. McAinsh, Jessica D. Tytell, and Peter K. Sorger. Structure, Function, and Regulation of Budding Yeast Kinetochores. *Annual Review of Cell and Developmental Biology*, 19(1) : 519–539, 2003. ISSN 1081-0706. doi : 10.1146/annurev.cellbio.19.1.11301.155607.
- Karim Mekhail and Danesh Moazed. The nuclear envelope in genome organization, expression and stability. *Nature Reviews Molecular Cell Biology*, 11(5) :317–328, 2010. ISSN 1471-0072. doi : 10.1038/nrm2894.
- Karim Mekhail, Jan Seebacher, Steven P. Gygi, and Danesh Moazed. Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature*, 456(7222) :667–670, 2008. ISSN 00280836. doi : 10.1038/nature07460.
- Pamela B. Meluh, Peirong Yang, Lynn Glowczewski, Douglas Koshland, and M. Mitchell Smith. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. *Cell*, 94(5) :607–613, 1998. ISSN 00928674. doi : 10.1016/S0092-8674(00)81602-5.
- Christine Mézard, Denis Pompon, and Alain Nicolas. Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. *Cell*, 70(4) :659–670, 1992. ISSN 00928674. doi : 10.1016/0092-8674(92)90434-E.
- Tom Misteli. The concept of self-organization in cellular architecture. *The Journal of Cell Biology*, 155(2) :181–186, oct 2001. ISSN 0021-9525. doi : 10.1083/jcb.200108110.
- Leslie A. Mitchell and Jef D. Boeke. Circular permutation of a synthetic eukaryotic chromosome with the telomerase. *Proceedings of the National Academy of Sciences*, 111(48) :17003–17010, 2014. ISSN 0027-8424. doi : 10.1073/pnas.1414399111.
- Leslie A. Mitchell and Tom Ellis. Synthetic genome engineering gets infectious. *Proceedings of the National Academy of Sciences*, 114(42) :11006–11008, 2017. ISSN 0027-8424. doi : 10.1073/pnas.1715365114.
- Takeshi Mizuguchi, Geoffrey Fudenberg, Sameet Mehta, Jon Matthew Belton, Nitika Taneja, Hernan Diego Folco, Peter FitzGerald, Job Dekker, Leonid Mirny, Jemima Barrowman, and Shiv I.S. Grewal. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature*, 516(7531) :432–435, 2014. ISSN 14764687. doi : 10.1038/nature13833.
- P L Nagy, M L Cleary, P O Brown, and J D Lieb. Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc Natl Acad Sci U S A*, 100(11) :6364–6369, 2003. ISSN 0027-8424. doi : 10.1073/pnas.1131966100.
- Markus Noll and Roger D. Kornberg. Action of micrococcal nuclease on chromatin and the location of histone H1. *Journal of Molecular Biology*, 109(3) :393–404, 1977. ISSN 00222836. doi : 10.1016/S0022-2836(77)80019-3.
- M Oakes, Y Nogi, M W Clark, and M Nomura. Structural alterations of the nucleolus in mutants of *Saccharomyces cerevisiae* defective in RNA polymerase I. *Molecular and Cellular Biology*, 13(4) :2441–2455, apr 1993. ISSN 0270-7306. doi : 10.1128/MCB.13.4.2441.

- Lauren M. Oldfield, Peter Grzesik, Alexander A. Voorhies, Nina Alperovich, Derek MacMath, Claudia D. Najera, Diya Sabrina Chandra, Sanjana Prasad, Vladimir N. Noskov, Michael G. Montague, Robert M. Friedman, Prashant J. Desai, and Sanjay Vashee. Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proceedings of the National Academy of Sciences*, 114(42) :E8885–E8894, 2017. ISSN 0027-8424. doi : 10.1073/pnas.1700534114.
- A L Olins and Donald E Olins. Spheroid chromatin units (v bodies). *Science (New York, N.Y.)*, 183(4122) :330–2, jan 1974. ISSN 0036-8075. doi : 10.1126/science.183.4122.330.
- S G Oliver, Q. J. M. van der Aart, M L Agostoni-Carbone, M Aigle, L Alberghina, D Alexandraki, G Antoine, R Anwar, J. P. G. Ballesta, P Benit, G. Berben, E. Bergantino, N. Biteau, P. A. Bolle, M. Bolotin-Fukuhara, A. Brown, A. J. P. Brown, J. M. Buhler, C. Carcano, G. Carignani, H. Cederberg, R. Chanet, R. Contreras, M. Crouzet, B. Daignan-Fornier, E. Defoor, M. Delgado, J. Demolder, C. Doira, E. Dubois, B. Dujon, A. Dusterhoft, D. Erdmann, M. Esteban, F. Fabre, C. Fairhead, G. Faye, H. Feldmann, W. Fiers, M. C. Francingues-Gaillard, L. Franco, L. Frontali, H. Fukuhara, L. J. Fuller, P. Galland, M. E. Gent, D. Gigot, V. Gilliquet, N. Glansdorff, A. Goffeau, M. Grenson, P. Grisanti, L. A. Grivell, M. de Haan, M. Haasemann, D. Hatat, J. Hoenicka, J. Hegemann, C. J. Herbert, F. Hilger, S. Hohmann, C. P. Hollenberg, K. Huse, F. Iborra, K. J. Indje, K. Isono, C. Jacq, M. Jacquet, C. M. James, J. C. Jauniaux, Y. Jia, A. Jimenez, A. Kelly, U. Kleinhans, P Kreisl, G Lanfranchi, C Lewis, C. G. VanderLinden, G Lucchini, K Lutzenkirchen, M.J. Maat, L. Mallet, G. Mannhaupt, E. Martegani, A. Mathieu, C. T. C. Maurer, D. McConnell, R. A. McKee, F. Messenguy, H. W. Mewes, F. Molemans, M. A. Montague, M. Muzi Falconi, L. Navas, C. S. Newlon, D. Noone, C. Pallier, L. Panzeri, B. M. Pearson, J. Perea, P. Philippsen, A. Pierard, R. J. Planta, P. Plevani, B. Poetsch, F. Pohl, B. Purnelle, M. Ramezani Rad, S. W. Rasmussen, A. Raynal, M. Remacha, P. Richterich, A. B. Roberts, F. Rodriguez, E. Sanz, I. Schaaff-Gerstenschlager, B. Scherens, B. Schweitzer, Y. Shu, J. Skala, P. P. Slonimski, F. Sor, C. Soustelle, R. Spiegelberg, L. I. Stateva, H. Y. Steensma, S. Steiner, A. Thierry, G. Thireos, M. Tzermia, L. A. Urrestarazu, G. Valle, I. Vetter, Van, J. C. Vliet-Reedijk, M. Voet, G. Volckaert, P. Vreken, H. Wang, J. R. Warmington, D. von Wettstein, B. L. Wicksteed, C. Wilson, H. Wurst, G. Xu, A. Yoshikawa, F. K. Zimmermann, and J. G. Sgouros. The complete DNA sequence of yeast chromosome III. *Nature*, 357(6373) :38–46, 1992. ISSN 0028-0836. doi : 10.1038/357038a0.
- P Oudet, M Gross-Bellard, and P Chambon. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4(4) :281–300, apr 1975. ISSN 00928674. doi : 10.1016/0092-8674(75)90149-X.
- Julie Parenteau, Mathieu Durand, Steeve Véronneau, Andrée-Anne Lacombe, Geneviève Morin, Valérie Guérin, Bojana Cecez, Julien Gervais-Bird, Chu-Shin Koh, David Brunelle, Raymund J Wellinger, Benoit Chabot, and Sherif Abou Elela. Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Molecular biology of the cell*, 19(5) : 1932–41, may 2008. ISSN 1939-4586. doi : 10.1091/mbc.E07-12-1254.
- Julie Parenteau, Mathieu Durand, Genevieve Morin, Jules Gagnon, Jean Franois Lucier, Raymund J. Wellinger, Benoit Chabot, and Sherif Abou Elela. Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell*, 147(2) :320–331, 2011. ISSN 00928674. doi : 10.1016/j.cell.2011.08.044.
- Hee-Sung Park, Michael J Hohn, Takuya Umehara, Li-Tao Guo, Edith M Osborne, Jack Benner, Christopher J Noren, Jesse Rinehart, and Dieter Söll. Expanding the genetic code of escherichia coli with phosphoserine. *Science*, 333(6046) :1151–1154, 2011.

- Philippe Pasero and Monique Marilley. Size variation of rDNA clusters in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *MGG Molecular & General Genetics*, 236-236(2-3) :448–452, jan 1993. ISSN 0026-8925. doi : 10.1007/BF00277147.
- Riccardo Percudani, Angelo Pavesi, and Simone Ottonello. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae* 1 Edited by J. Karn. *Journal of Molecular Biology*, 268(2) :322–330, 1997. ISSN 00222836. doi : 10.1006/jmbi.1997.0942.
- György Pósfai, Guy Plunkett, Tamás Fehér, David Frisch, Günther M Keil, Kinga Umenhoffer, Vitaliy Kolisnychenko, Buffy Stahl, Shamik S Sharma, Monika De Arruda, et al. Emergent properties of reduced-genome *Escherichia coli*. *Science*, 312(5776) :1044–1046, 2006.
- T Powers and P Walter. Regulation of ribosome biogenesis by the rapamycin-sensitive TOR-signaling pathway in *Saccharomyces cerevisiae*. *Molecular biology of the cell*, 10(4) :987–1000, apr 1999. ISSN 1059-1524. doi : 10.1091/mbc.10.4.987.
- C Prodromou and L H Pearl. Recursive PCR : a novel technique for total gene synthesis. *Protein engineering*, 5(8) :827–829, dec 1992. ISSN 0269-2139 (Print).
- Fiona E. Pryde and Edward J. Louis. Limitations of silencing at native yeast telomeres. *EMBO Journal*, 18(9) :2538–2550, 1999. ISSN 02614189. doi : 10.1093/emboj/18.9.2538.
- C Rabl. Über zelltheilung. morphol. jahrb. 10, 214-330. *Rabl21410Morphol. Jahrb*, 1885.
- V R Racaniello and D Baltimore. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proceedings of the National Academy of Sciences of the United States of America*, 78(8) :4887–4891, 1981. ISSN 0027-8424. doi : 10.1073/pnas.78.8.4887.
- M. K. Raghuraman and Bonita J. Brewer. Molecular analysis of the replication program in unicellular model organisms. *Chromosome Research*, 18(1) :19–34, 2010. ISSN 09673849. doi : 10.1007/s10577-009-9099-x.
- Robert J D Reid, Ivana Sunjevaric, Warren P. Voth, Samantha Ciccone, Wendy Du, Aileen E. Olsen, David J. Stillman, and Rodney Rothstein. Chromosome-scale genetic mapping using a set of 16 conditionally stable *Saccharomyces cerevisiae* chromosomes. *Genetics*, 180(4) :1799–1808, 2008. ISSN 00166731. doi : 10.1534/genetics.108.087999.
- H. Renauld, O. M. Aparicio, P. D. Zierath, B. L. Billington, S. K. Chhablani, and D. E. Gottschling. Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes and Development*, 7(7 A) :1133–1145, 1993. ISSN 08909369. doi : 10.1101/gad.7.7a.1133.
- Sarah M Richardson, Leslie A Mitchell, Giovanni Stracquadanio, Kun Yang, Jessica S Dymond, James E DiCarlo, Dongwon Lee, Cheng Lai Victor Huang, Srinivasan Chandrasegaran, Yizhi Cai, Jef D Boeke, and Joel S Bader. Design of a synthetic yeast genome. *Science (New York, N.Y.)*, 355(6329) :1040–1044, 2017. ISSN 1095-9203. doi : 10.1126/science.aaf4557.
- C C Robinett, A Straight, G Li, C Wilhelm, G Sudlow, A Murray, and A S Belmont. In vivo localization of DNA sequences and visualization of large scale chromatin organization using lac-operator/repressor recognition. *J Cell Biol*, 135(6) :1685–1700, 1996.
- Philip J J Robinson, Woojin An, Andrew Routh, Fabrizio Martino, Lynda Chapman, Robert G. Roeder, and Daniela Rhodes. 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *Journal of molecular biology*, 381(4) :816–25, sep 2008. ISSN 1089-8638. doi : 10.1016/j.jmb.2008.04.050.

- Philip JJ Robinson and Daniela Rhodes. Structure of the '30nm' chromatin fibre : A key role for the linker histone. *Current Opinion in Structural Biology*, 16(3) :336–343, jun 2006. ISSN 0959440X. doi : 10.1016/j.sbi.2006.05.007.
- Milena Rondón-Lagos, Ludovica Verdun Di Cantogno, Caterina Marchiò, Nelson Rangel, Cesar Payan-Gomez, Patrizia Gugliotta, Cristina Botta, Gianni Bussolati, Sandra R. Ramírez-Clavijo, Barbara Pasini, and Anna Sapino. Differences and homologies of chromosomal alterations within and between breast cancer cell lines : A clustering analysis. *Molecular Cytogenetics*, 7(1) :1–14, 2014. ISSN 17558166. doi : 10.1186/1755-8166-7-8.
- David M. Roof, Pamela B. Meluh, and Mark D Rose. Kinesin-related proteins required for Assembly of the Mitotic Spindle. *Cell*, 118(1) :95–108, 1992. ISSN 0092-8674. doi : [http://dx.doi.org/10.1016/0092-8674\(92\)90169-D](http://dx.doi.org/10.1016/0092-8674(92)90169-D).
- Andrew Routh, Sara Sandin, and Daniela Rhodes. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proceedings of the National Academy of Sciences*, 105(26) :8872–8877, 2008. ISSN 0027-8424. doi : 10.1073/pnas.0802336105.
- Rajat Roy, Bettina Meier, Andrew D. McAinsh, Heidi M. Feldmann, and Stephen P. Jackson. Separation-of-function Mutants of Yeast Ku80 Reveal a Yku80p-Sir4p Interaction Involved in Telomeric Silencing. *Journal of Biological Chemistry*, 279(1) :86–94, 2004. ISSN 00219258. doi : 10.1074/jbc.M306841200.
- G T Rudkin and B D Stollar. High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. *Nature*, 265(5593) :472–3, feb 1977. ISSN 0028-0836.
- R. Saiki, D. Gelfand, S Stoffel, S. Scharf, R Higuchi, G. Horn, K. Mullis, and H. Erlich. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839) :487–491, 1988. ISSN 0036-8075. doi : 10.1126/science.2448875.
- R K Saiki, S Scharf, Fred Faloona, Kary B Mullis, G T Horn, H A Erlich, and N Arnheim. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230(4732) :1350–1354, 1985.
- F Sanger, G M Air, B G Barrell, N L Brown, a R Coulson, C a Fiddes, C a Hutchison, P M Slocombe, and M Smith. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 265(5596) :687–695, 1977. ISSN 0028-0836. doi : 5,386basepairs...
- Georgia Schäfer, Christopher R E McEvoy, and Hugh-G Patterton. The *Saccharomyces cerevisiae* linker histone Hho1p is essential for chromatin compaction in stationary phase and is displaced by transcription. *Proceedings Of The National Academy Of Sciences (Of The United States Of America)*, 105(39) :14838–14843, 2008. ISSN 1091-6490. doi : 10.1073/pnas.0806337105.
- Thomas Schalch, Sylwia Duda, David F. Sargent, and Timothy J. Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047) :138–141, 2005. ISSN 0028-0836. doi : 10.1038/nature03686.
- Margit Schardin, T. Cremer, H. D. Hager, and M. Lang. Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Human Genetics*, 71(4) :281–287, 1985. ISSN 03406717. doi : 10.1007/BF00388452.
- M. P. Scheffer, M. Eltsov, and A. S. Frangakis. Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proceedings of the National Academy of Sciences*, 108(41) :16992–16997, 2011. ISSN 0027-8424. doi : 10.1073/pnas.1108268108.



- Maria Schmid, Andreas Jaedicke, Tung G. Du, and Ralf Peter Jansen. Coordination of Endoplasmic Reticulum and mRNA Localization to the Yeast Bud. *Current Biology*, 16(15) :1538–1543, 2006. ISSN 09609822. doi : 10.1016/j.cub.2006.06.025.
- Dustin E. Schones, Kairong Cui, Suresh Cuddapah, Tae-Young Roh, Artem Barski, Zhibin Wang, Gang Wei, and Keji Zhao. Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132(5) :887–98, mar 2008. ISSN 1097-4172. doi : 10.1016/j.cell.2008.02.022.
- Edward A. Sekinger, Zarmik Moqtaderi, and Kevin Struhl. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Molecular Cell*, 18(6) :735–748, 2005. ISSN 10972765. doi : 10.1016/j.molcel.2005.05.003.
- Tom Sexton, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, 148(3) :458–472, 2012. ISSN 00928674. doi : 10.1016/j.cell.2012.01.010.
- Christian Seybold and Elmar Schiebel. Spindle pole bodies. *Current biology : CB*, 23(19) : R858–60, oct 2013. ISSN 1879-0445. doi : 10.1016/j.cub.2013.07.024.
- Yue Shen, Giovanni Stracquadanio, Yun Wang, Kun Yang, Leslie A. Mitchell, Yaxin Xue, Yizhi Cai, Tai Chen, Jessica S. Dymond, Kang Kang, Jianhui Gong, Xiaofan Zeng, Yongfen Zhang, Yingrui Li, Qiang Feng, Xun Xu, Jun Wang, Jian Wang, Huanming Yang, Jef D. Boeke, and Joel S. Bader. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Research*, 26(1) :36–49, 2016. ISSN 15495469. doi : 10.1101/gr.193433.115.
- Michael Shogren-Knaak. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science*, 311(5762) :844–847, feb 2006. ISSN 0036-8075. doi : 10.1126/science.1124000.
- David A. Sinclair and Leonard Guarente. Extrachromosomal rDNA circles - A cause of aging in yeast. *Cell*, 91(7) :1033–1042, 1997. ISSN 00928674. doi : 10.1016/S0092-8674(00)80493-6.
- Valentina Sirri, Silvio Urcuqui-Inchima, Pascal Roussel, and Danièle Hernandez-Verdun. Nucleolus : The fascinating nuclear body. *Histochemistry and Cell Biology*, 129(1) :13–31, 2008. ISSN 09486143. doi : 10.1007/s00418-007-0359-6.
- H. O. Smith, C. A. Hutchison, C. Pfannkoch, and J. C. Venter. Generating a synthetic genome by whole genome assembly : X174 bacteriophage from synthetic oligonucleotides. *Proceedings of the National Academy of Sciences*, 100(26) :15440–15445, 2003. ISSN 0027-8424. doi : 10.1073/pnas.2237126100.
- J S Smith and J D Boeke. An unusual form of transcriptional silencing in yeast ribosomal DNA. *Genes & development*, 11(2) :241–254, 1997. ISSN 0890-9369. doi : 10.1101/gad.11.2.241.
- Willem P C Stemmer, Andreas Cramer, Kim D. Ha, Thomas M. Brennan, and Herbert L. Heyneker. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, 164(1) :49–53, 1995. ISSN 03781119. doi : 10.1016/0378-1119(95)00511-4.
- A. T. Sumner, H. J. Evans, and R. A. Buckland. New technique for distinguishing between human chromosomes. *Nature New Biology*, 232(27) :31–32, 1971. ISSN 00900028. doi : 10.1038/newbio232031a0.
- Jack W. Szostak, Terry L. Orr-Weaver, Rodney J. Rothstein, and Franklin W. Stahl. The double-strand-break repair model for recombination. *Cell*, 33(1) :25–35, 1983. ISSN 00928674. doi : 10.1016/0092-8674(83)90331-8.

- Angela Taddei, Florence Hediger, Frank R. Neumaan, Christoph Bauer, and Susan M. Gasser. Separation of silencing from perinuclear anchoring functions in yeast Ku80, Sir4 and Esc1 proteins. *EMBO Journal*, 23(6) :1301–1312, 2004. ISSN 02614189. doi : 10.1038/sj.emboj.7600144.
- Angela Taddei, Griet Van Houwe, Shigeki Nagai, Ionas Erb, Erik van Nimwegen, and Susan M Gasser. The functional importance of telomere clustering : global changes in gene expression result from SIR factor dispersion. *Genome research*, 19(4) :611–25, apr 2009. ISSN 1088-9051. doi : 10.1101/gr.083881.108.
- Angela Taddei, Heiko Schober, and Susan M. Gasser. The budding yeast nucleus. *Cold Spring Harbor perspectives in biology*, 2(8) :a000612–a000612, aug 2010. ISSN 19430264. doi : 10.1101/cshperspect.a000612.
- Kosei Tanaka, Christopher S. Henry, Jenifer F. Zinner, Edmond Jolivet, Matthew P. Cohoon, Fangfang Xia, Vladimir Bidnenko, S. Dusko Ehrlich, Rick L. Stevens, and Philippe Noirot. Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Research*, 41(1) :687–699, 2013. ISSN 03051048. doi : 10.1093/nar/gks963.
- Rahul Thadani, Frank Uhlmann, and Sebastian Heeger. Condensin, Chromatin Crossbarring and Chromosome Condensation. *Current Biology*, 22(23) :R1012–R1021, dec 2012. ISSN 09609822. doi : 10.1016/j.cub.2012.10.023.
- Wai Hong Tham, J. B. Stuart Wyithe, Paul Ko Ferrigno, Pamela A. Silver, and Virginia A. Zakian. Localization of Yeast Telomeres to the Nuclear Periphery Is Separable from Transcriptional Repression and Telomere Stability Functions. *Molecular Cell*, 8(1) :189–199, 2001. ISSN 10972765. doi : 10.1016/S1097-2765(01)00287-8.
- P. Therizols, T. Duong, B. Dujon, C. Zimmer, and E. Fabre. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proceedings of the National Academy of Sciences*, 107(5) :2025–2030, 2010. ISSN 0027-8424. doi : 10.1073/pnas.0914187107.
- Martin Thompson, Rebecca A Haeusler, Paul D Good, and David R Engelke. Nucleolar clustering of dispersed tRNA genes. *Science (New York, N.Y.)*, 302(5649) :1399–401, nov 2003. ISSN 1095-9203. doi : 10.1126/science.1089814.
- JOE HIN TJIO and ALBERT LEVAN. THE CHROMOSOME NUMBER OF MAN. *Hereditas*, 42(1-2) :1–6, jul 1956. ISSN 00180661. doi : 10.1111/j.1601-5223.1956.tb03010.x.
- Jordi Torres-Rosell, Giacomo De Piccoli, Violeta Cordon-Preciado, Sarah Farmer, Adam Jarmuz, Felix Machin, Philippe Pasero, Michael Lisby, James E. Haber, and Luis Aragón. Anaphase onset before complete DNA replication with intact checkpoint responses. *Science (New York, N.Y.)*, 315(5817) :1411–5, mar 2007. ISSN 1095-9203. doi : 10.1126/science.1134025.
- David J. Tremethick. Higher-order structures of chromatin : the elusive 30 nm fiber. *Cell*, 128(4) : 651–4, feb 2007. ISSN 0092-8674. doi : 10.1016/j.cell.2007.02.008.
- S Trumtel, I Léger-Silvestre, P E Gleizes, F Teulières, and N Gas. Assembly and functional organization of the nucleolus : ultrastructural analysis of *Saccharomyces cerevisiae* mutants. *Molecular biology of the cell*, 11(6) :2175–89, jun 2000. ISSN 1059-1524.
- Alexander M. Tsankov, Dawn Anne Thompson, Amanda Socha, Aviv Regev, and Oliver J. Rando. The role of nucleosome positioning in the evolution of gene regulation. *PLoS biology*, 8(7) : e1000414, jul 2010. ISSN 1545-7885. doi : 10.1371/journal.pbio.1000414.

- Tudorita Tumber, Gail Sudlow, and Andrew S. Belmont. Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *The Journal of cell biology*, 145(7) : 1341–54, jun 1999. ISSN 0021-9525. doi : 10.1083/jcb.145.7.1341.
- Harm van Bakel, Kyle Tsui, Marinella Gebbia, Sanie Mnaimneh, Timothy R. Hughes, and Corey Nislow. A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. *PLoS genetics*, 9(5) :e1003479, may 2013. ISSN 1553-7404. doi : 10.1371/journal.pgen.1003479.
- Lilien N. Voong, Liqun Xi, Amy C. Sebeson, Bin Xiong, Ji-Ping Wang, and Xiaozhong Wang. Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell*, 167(6) :1555–1570.e15, dec 2016. ISSN 1097-4172. doi : 10.1016/j.cell.2016.10.049.
- Ty C. Voss, R. Louis Schiltz, Myong Hee Sung, Paul M. Yen, John A. Stamatoyannopoulos, Simon C. Biddie, Thomas A. Johnson, Tina B. Miranda, Sam John, and Gordon L. Hager. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell*, 146(4) :544–554, 2011. ISSN 00928674. doi : 10.1016/j.cell.2011.07.006.
- Harris H. Wang, Farren J. Isaacs, Peter A. Carr, Zachary Z. Sun, George Xu, Craig R. Forest, and George M. Church. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257) :894–898, 2009. ISSN 00280836. doi : 10.1038/nature08187.
- Li Wang, Rebecca A. Haeusler, Paul D. Good, Martin Thompson, Sapna Nagar, and David R. Engelke. Silencing near tRNA genes requires nucleolar localization. *The Journal of biological chemistry*, 280(10) :8637–9, mar 2005. ISSN 0021-9258. doi : 10.1074/jbc.C500017200.
- J D WATSON and F H CRICK. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171(4361) :964–7, may 1953. ISSN 0028-0836. doi : 10.1038/171964b0.
- Assaf Weiner, Amanda Hughes, Moran Yassour, Oliver J Rando, and Nir Friedman. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome research*, 20(1) :90–100, jan 2010. ISSN 1549-5469. doi : 10.1101/gr.098509.109.
- J Widom. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proceedings of the National Academy of Sciences of the United States of America*, 89(3) :1095–9, feb 1992. ISSN 0027-8424. doi : 10.1073/pnas.89.3.1095.
- M H F WILKINS, A R STOKES, and H R WILSON. Molecular structure of deoxypentose nucleic acids. *Nature*, 171(4356) :738–40, apr 1953. ISSN 0028-0836. doi : 10.1038/nature01396.
- S. P. Williams, B. D. Athey, L. J. Muglia, R. S. Schappe, A. H. Gough, and J. P. Langmore. Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophysical journal*, 49(1) :233–48, jan 1986. ISSN 0006-3495. doi : 10.1016/S0006-3495(86)83637-2.
- Jonathan Wong, Yuko Nakajima, Stefan Westermann, Ching Shang, Jung-Seog Kang, Crystal Goodner, Pantea Houshmand, Stanley Fields, Clarence S M Chan, David Drubin, Georjana Barnes, and Tony Hazbun. A protein interaction map of the mitotic spindle. *Molecular biology of the cell*, 18(10) :3800–9, oct 2007. ISSN 1059-1524. doi : 10.1091/mbc.E07-06-0536.
- C L Woodcock, S A Grigoryev, R A Horowitz, and N Whitaker. A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proceedings of the National Academy of Sciences of the United States of America*, 90(19) :9021–5, oct 1993. ISSN 0027-8424. doi : 10.1073/pnas.90.19.9021.

- Christopher L. Woodcock and Stefan Dimitrov. Higher-order structure of chromatin and chromosomes. *Current Opinion in Genetics & Development*, 11(2) :130–135, apr 2001. ISSN 0959437X. doi : 10.1016/S0959-437X(00)00169-6.
- Carl Wu, Y C Wong, and S C Elgin. The chromatin structure of specific genes : II. Disruption of chromatin structure during gene activity. *Cell*, 16(4) :807–14, apr 1979. ISSN 0092-8674. doi : 10.1016/0092-8674(79)90096-5.
- Yuanxin Xi, Jianhui Yao, Rui Chen, Wei Li, and Xiangwei He. Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome research*, 21(5) :718–24, may 2011. ISSN 1549-5469. doi : 10.1101/gr.117101.110.
- Bing Xia, Swapnil Bhatia, Ben Bubenheim, Maisam Dadgar, Douglas Densmore, and J Christopher Anderson. Developer’s and user’s guide to Clotho v2.0 A software platform for the creation of synthetic biological systems. *Methods Enzymol*, 498 :97–135, 2011.
- H Xu and J D Boeke. High-frequency deletion between homologous sequences during retrotransposition of Ty elements in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 84(23) :8553–7, 1987. ISSN 0027-8424. doi : 10.1073/pnas.84.23.8553.
- Charles H. Yang, Eric J. Lambie, John Hardin, Joe Craft, and Michael Snyder. Higher order structure is present in the yeast nucleus : autoantibody probes demonstrate that the nucleolus lies opposite the spindle pole body. *Chromosoma*, 98(2) :123–128, 1989. ISSN 00095915. doi : 10.1007/BF00291048.
- Byung Jo Yu, Bong Hyun Sung, Michael D. Koob, Choong Hoon Lee, Jun Hyoung Lee, Won Sik Lee, Mi Sun Kim, and Sun Chang Kim. Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. *Nature Biotechnology*, 20(10) :1018–1023, 2002. ISSN 10870156. doi : 10.1038/nbt740.
- Gabriel E. Zentner and Steven Henikoff. High-resolution digital profiling of the epigenome. *Nature reviews. Genetics*, 15(12) :814–27, dec 2014. ISSN 1471-0064. doi : 10.1038/nrg3798.
- Zhenhai Zhang and B. Franklin Pugh. High-resolution genome-wide mapping of the primary structure of chromatin. *Cell*, 144(2) :175–86, jan 2011. ISSN 1097-4172. doi : 10.1016/j.cell.2011.01.003.

# **Annexes**



# **Analyse de l'effet de la fixation au formaldéhyde sur les caractéristiques physiques du polymère d'ADN**

# Kinetic signature of cooperativity in the irreversible collapse of a polymer

Vittore F. Scolari,<sup>1,2,\*</sup> Guillaume Mercy,<sup>1,2</sup> Romain Koszul,<sup>1,2</sup> Annick Lesne,<sup>3,4</sup> and Julien Mozziconacci<sup>3,†</sup>

<sup>1</sup>*Spatial Regulation of Genomes, Genomes & Genetics Department, Institut Pasteur, Paris, 75015, France*

<sup>2</sup>*UMR3525, Centre National de la Recherche Scientifique, Paris, 75015, France*

<sup>3</sup>*Laboratoire de Physique Théorique de la Matière Condensée, CNRS UMR 7600,*

*Université Pierre et Marie Curie, Sorbonne Universités, Paris, France*

<sup>4</sup>*IGMM, CNRS, Univ. Montpellier, Montpellier, France*

(Dated: November 28, 2017)

We investigate the kinetics of a polymer collapse due to the formation of irreversible crosslinks between its monomers. Using the contact probability as a function of the linear distance  $s$  along the polymer,  $P(s)$ , as a scale-dependent order parameter, we show in simulations the emergence of a cooperative pearling instability. Namely, the polymer undergoes a sharp conformational transition to a set of absorbing states characterized by a length scale  $\xi$  corresponding to the mean pearl size. This length and the transition time depend on the polymer equilibrium dynamics and the crosslink formation rate. We finally confirm experimentally this transition using a DNA conformation capture experiment.

The multi-stage collapse of a polymeric chain has been subject to multiple theoretical and experimental investigations [1–8]. When the collapse is caused by a reduction in solvent quality, the seminal work of de Gennes has predicted a continuous conformational transition through successive crumpling stages, commonly called the “expanding sausage model” [1]. This model did not consider the effects of topological constraints in the collapse dynamics. In contrast, Grosberg *et al.* proposed a two-stage model, where a fast collapse is followed by a slow unknotting of topological constraints occurring through reptation [2]. The meta-stable intermediate state, called “fractal-globule”, preserves the fractal features of a coil while being compact as a globule. The predicted existence of meta-stability was confirmed by the experiments of Chu *et al.* [3]. Comparing their results to de Gennes’s calculations, they showed the importance of topological constraints and concluded to the emergence of an unexplainable higher viscosity, possibly coming from the process of knots untangling. As another description of polymer collapse, Buguin *et al.* introduced the concept of pearling through the existence of a characteristic size, there explained by nucleation theory, i.e. by the interplay of bulk and surface effects [4]. Pearling has been subsequently studied in different works [5–7]. More recently Bunin and Kardar proposed an effective model of polymer collapse, consisting in a cascading succession of coalescence events of blobs actively compressed in a central potential [8].

In this letter, we study the collapse dynamics when it is caused by the cumulative effect of irreversible crosslinks between monomers of the chain. We complete and clarify the results presented above by introducing a scale-dependent order parameter  $P(s)$ , that is the mean number of crosslink events between two monomers separated by a linear distance  $s$  along the chain. This order parameter has two important advantages: first it reflects the appearance of local structures such as pearls, second

it is a direct experimental observable in the chromosome conformation capture experiments described at the end of this letter.

We first describe the *in silico* collapse dynamics of a population of phantom chains with a focus on the combined effects of aggregation and Rouse diffusion, measuring the effects on  $P(s)$  at different scales. We run a rejection kinetic Monte Carlo simulation [9, 10] reproducing the Rouse phenomenology on a polymer composed of 2048 beads. We add the possibility of making a new link, with a probability  $p$ , each time two non-linked beads are coming into vicinity;  $p$  reflects the crosslinking rate (details in Supplementary materials (SM) §I.A).

In the absence of crosslinking, the correlations of bead positions satisfy the Rouse scaling relations with coeffi-

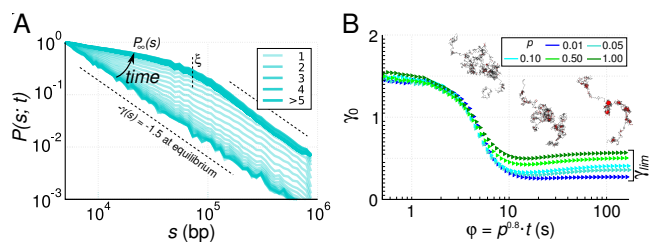


FIG. 1. Kinetics of the pearling transition (simulation) as reflected in the evolution of the  $P(s)$  in time. (A) Evolution of the contact probability curve  $P(s)$  at fixed  $p = 0.1$ , displayed as a superposition of semitransparent plots each obtained at a single simulation time from crosslinking onset; time ordering is indicated by a black arrow and the resulting color density is given in the inset. A crossover at a length  $\xi$  becomes apparent at large enough times. (B) Evolution of  $\gamma_0$ , the slope of the log-log plot of the  $P(s)$  at short distances, as a function of the rescaled time variable  $\phi$ , for different values of the crosslink probability  $p$ . Inset: snapshots of the evolution along time of the polymer conformation ( $p = 0.1$ ).



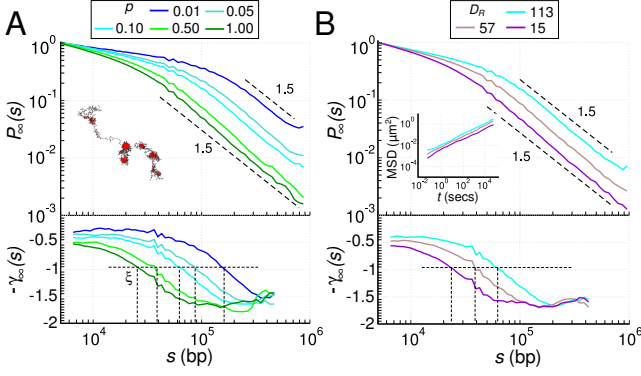


FIG. 2. Average quantitative features of the pearled absorbing states (simulation). (A) (Upper panel) asymptotic curve  $P_\infty(s)$  and (Lower panel) its local slope  $\gamma_\infty(s)$  and pearling length  $\xi$ , for different crosslink probabilities  $p$ . Inset: example of a pearled state ( $p = 0.1$ ). (B) (Upper panel) asymptotic curve  $P_\infty(s)$  and (Lower panel) its local slope  $\gamma_\infty(s)$  for different polymer dynamics, parameterized by the Rouse coefficient  $D_R$ . Inset: monomer mean square displacement (MSD) as a function of time, the slope of the Rouse dynamics permits a measurement of  $C_t$  (eq. 1).

cients  $C_t$  and  $C_s$  [11]:

$$\begin{aligned} \left\langle \left| \vec{R}(0, t_0) - \vec{R}(0, t + t_0) \right|^2 \right\rangle &\sim C_t \cdot t^{1/2}, \\ \left\langle \left| \vec{R}(s_0, 0) - \vec{R}(s + s_0, 0) \right|^2 \right\rangle &\sim C_s \cdot s. \end{aligned} \quad (1)$$

After thermal equilibration, crosslinking is introduced as a succession of irreversible and configuration-dependent changes in the chain topology. As a proxy for steric interactions, we limit the crosslink events to a maximum number  $N_{max}$  per bead and stop the simulation once this number is reached for all the beads.

Given this dynamics, the contact probability  $P(s; p, C_s, C_t, t)$ , is a function of  $s$ , the crosslink probability  $p$ , the Rouse coefficients and the elapsed time  $t$  from the crosslinking onset. At constant  $p$ , the time evolution of this curve displays a transition from the equilibrium contact probability, scaling as  $\propto s^{-\gamma}$  with  $\gamma = 3/2$ , [12], to an asymptotic shape  $P_\infty(s)$  displaying a crossover between two different scaling behaviors at short and long linear distances (fig. 1A). This shape and the crossover length  $\xi$  reflects the population average of the absorbing states reached by the polymer at crosslink saturation. The exponent  $\gamma_0(t)$ , corresponding to the value at short distances of the local exponent  $\gamma(s; t)$  defined from the discrete differential

$$\gamma(s; t) = -\frac{\Delta \ln[P(s; t)]}{\Delta \ln[s]}, \quad (2)$$

presents a sharp decrease in time (fig. 1B, cyan symbols).

In order to understand the dependence of these features on the kinetic parameters, we first monitored the

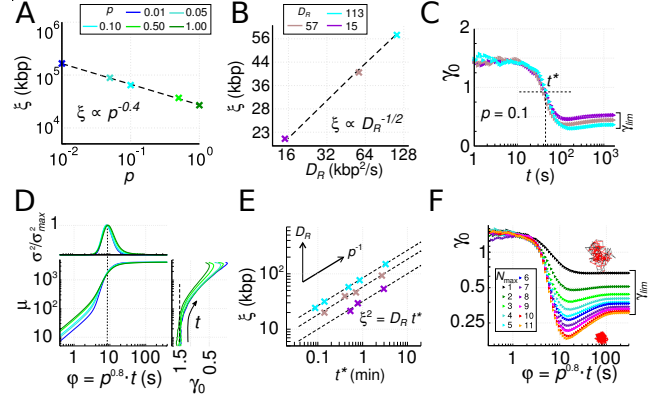


FIG. 3. Dependency of the transition dynamics (simulation) on the kinetic parameters (A) Variation of the pearling length  $\xi$  with the crosslink probability  $p$ . (B) Variation of the pearling length  $\xi$  with the Rouse coefficient  $D_R$ . (C) Evolution of  $\gamma_0$  at different  $D_R$  in function of time  $t$  at fixed  $p = 0.1$ . (D) (Lower panel) Mean cumulative number  $\mu$  of crosslink events and (Upper panel) its normalized variance  $\sigma^2/\sigma_{max}^2$  as a function of  $\phi$ , and (Right panel) scatterplot of  $\mu$  and  $\gamma_0$ . (E) Scatter plot of the pearling length  $\xi$  and the transition time  $t^*$ ; dashed lines are plotted using Eq. 8. (F) Evolution of  $\gamma_0$  as a function of  $\phi$  for different values of  $N_{max}$ .

effect of the crosslink probability  $p$  on the asymptotic curve  $P_\infty(s)$  (fig. 2A, upper panel). The crossover length  $\xi$  can be precisely estimated as the middle point in the transition from the short-distance to large-distance values of the asymptotic exponent  $\gamma_\infty(s)$  (fig. 2A, lower panel). This length  $\xi$  corresponds to the average length of the polymer segments captured in the pearls formed in long enough simulations, and will hereafter be referred as the pearling length. Between the pearls (i.e. for  $s > \xi$ )  $\gamma_\infty(s) = 3/2$ , consistent with the initial equilibrium state of the polymer, whereas  $\gamma_\infty(s)$  tends inside the pearls (i.e. for  $s < \xi$ ) to an asymptotic value  $\gamma_{lim} < 1$ .

Interestingly, the length  $\xi$  scales with the crosslink probability  $p$  as  $\xi \propto p^{-\delta}$ , with  $\delta = 0.4$  (fig. 3A), indicating that the extent along the chain of the crosslinking-induced collapse is more prominent for small  $p$  (i.e., low crosslinking rate). We explain this seemingly paradoxical behavior as follows: polymer loops of size greater than  $\xi$  do not have time to relax and their conformation change is thus diffusion-limited, while for smaller loops, Rouse diffusion is faster than the crosslinking reaction and many conformational fluctuations and contacts can occur and be fixed by crosslinks. Pearls with a mean size  $\xi$  form in this latter reaction-limited regime. Based on this qualitative picture, we propose a mean-field calculation of the dependence of  $\xi$  in  $p$ . The relaxation time for a fixed loop of size  $s$  scales as:

$$\tau_R(s) = D_R^{-1} \cdot s^2, \quad \text{with} \quad D_R = \frac{\pi^3}{4} \left( \frac{C_t}{C_s} \right)^2, \quad (3)$$

(detailed derivation in SM §II.C.4) while the average

duration  $\tau_{xl}$  needed to crosslink contacting beads is inversely proportional to the crosslink probability:

$$\tau_{xl} \propto p^{-1}. \quad (4)$$

Writing that the pearling length  $\xi$  emerges from the competition between these two dynamical processes yields:

$$\xi(p) \propto p^{-\delta}, \quad (5)$$

with  $\delta = 1/2$  correctly recapitulating the decrease of  $\xi$  at increasing  $p$ . In this argument, we assumed that the dynamics is consistent with Rouse diffusion during the pearling formation and collapse. However, Rouse diffusion is not expected to apply to the mesh into what the initially linear polymer is transformed after enough crosslinks, which may explain the different value  $\delta = 0.4$  measured in the simulations (fig. 3A). With the same argument we also predict that the value of  $\xi$  varies with the dynamical properties of the polymer. Simulations actually show that variation of the Rouse diffusion coefficient  $D_R$  has a dramatic effect on  $\xi$  (fig. 2B). For small  $D_R$ ,  $\xi$  is small and most of the crosslink effect occurs locally. On the other hand, when  $D_R$  increases, longer polymer segments can reach their equilibrium conformation between two crosslink events so that  $\xi$  becomes larger. In the line of the above calculation, we expect a scaling

$$\xi(D_R) \propto D_R^{1/2}, \quad (6)$$

which is well reproduced in the simulations (fig. 3B).

Our simulation moreover shows that the collapse happens suddenly. The short-distance exponent  $\gamma_0$  presents a sharp decrease at a time  $t^*$ , which we call the pearling time. Before this transition,  $t \ll t^*$ ,  $\gamma_0$  coincides with the exponent at long distances,  $3/2$ , as expected for an equilibrium state; only after the transition a smaller exponent is observed, with a limiting value  $\gamma_{lim} < 1$  depending on the simulation parameters.  $t^*$  depends on the crosslink probability with a scaling  $t^* \propto p^{-0.8}$  prompting to define a rescaled variable  $\phi = p^{0.8} \cdot t$ . The evolution of  $\gamma_0$  as a function of  $\phi$  rescales at any  $p$  to a single transition curve (fig. 1B). The scaling of the transition time  $t^*$  can also be explained with the above mean-field calculation: as  $t^*$  emerges from pearling (see polymer snapshots along the transition curve in fig. 1B), it is equal to the relaxation time of pearls of mean size  $\xi$ :  $t^* = \tau_R(\xi)$ . From Eq. 3,

$$t^* \propto p^{-2\delta}, \quad (7)$$

and  $\phi^* = p^{2\delta} \cdot t$ . As predicted by the above argument and confirmed in the simulation, the transition time does not depend on the Rouse diffusion coefficient  $D_R$  (fig. 3C). The pearling transition is the result of the cooperative effect of multiple crosslinks, that takes place only after relaxation of loops with length  $s < \xi$ . This effect is highlighted in fig. 3D, lower panel, that shows the acceleration of crosslink events at the transition. This process

is accompanied by the decrease of  $\gamma_0$  (3D, right panel) and a large increase of crosslink number variability, due to the fluctuation in the size and time of pearl formation and consistent with a phase transition (3D, upper panel). Collecting the results from simulations performed at various values of crosslink probability  $p$  and fixed Rouse diffusion coefficient  $D_R$ , the transition points in the plane defined by transition time  $t^*$  and pearling length  $\xi$  (fig. 3E) satisfy the Rouse scaling relation:

$$t^* = D_R^{-1} \cdot \xi^2; \quad (8)$$

that fully recapitulates the interplay between the pearling time, the pearling length and the polymer dynamics. We finally determine the influence of steric constraints on the final state by changing the maximum number of crosslinks,  $N_{max}$ , allowed for each bead. While  $\xi$  and  $t^*$  do not depend on  $N_{max}$ , the pearl formation and final internal conformation do, as shown by the time behavior of  $\gamma_0$ . After a transition in  $t^*$ , this short-distance exponent transiently goes toward 0 for large enough values of  $N_{max}$  before plateauing to an asymptotic value  $\gamma_{lim}$  varying from 0.3 to 0.7 when  $N_{max}$  varies (see fig. 3F and suppl. fig. 5). A careful examination of the conformational trajectories shows that this behavior can be explained by a two-stage dynamics taking place after the transition in  $t^*$ : first the formation of densely connected pearls (in red on the snapshots of fig. 1A), followed with a diffusion-limited crumpling of the linkers between adjacent pearls toward a single spherically shaped globule (see the snapshots in fig. 3F) still internally segregated into distinct pearls.

To recapitulate our simulation results, we have shown how the interplay between the polymer Rouse dynamics and the rate at which crosslinks are made induces a cooperative phase transition to pearled conformations with characteristic size  $\xi$ . We reproduced a two-stage kinetics described previously with some significant differences in the interpretation of results and the underlying mechanisms. In our irreversible model we can exclude nucleation: in the nucleation-inspired model of Buguin *et al.* [4] pearls created with a minimal size of  $\xi$  grow continuously till overall polymer collapse. This behavior, reproduced also in simulations [13], requires reversible monomer-monomer interactions. Our crosslinks can be seen as an effective model for interactions of duration longer than  $t^*$ , and we observe a different kind of pearling which happens through a sharp transition in both space and time. We can also exclude knotting effects: Grosberg *et al.* [2] focused on the role of knots in the conformational relaxation and predicted a dense globule with a fractal dimension of 3 and a relaxation through reptation. In contrast, we neglect steric interactions which is a necessary element for knot stability. Also, the first step of the crosslink dynamics fixes loops in the thermally equilibrated samples, including the long range contacts with probability of  $s^{-3/2}$ . Once fixed, these loops would

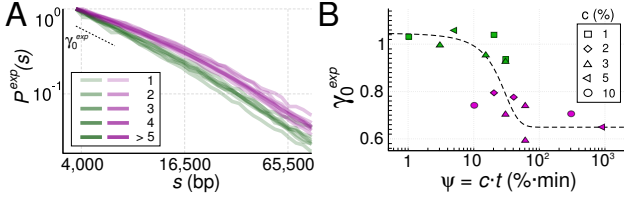


FIG. 4. (A) Experimental contact probability curves  $P^{exp}(s)$  for various crosslinker concentrations  $c$ , displayed as a superposition of semitransparent plots (see Fig. 1A). (B) Evolution of the experimental slope  $\gamma_0^{exp}$  as a function of the rescaled time variable  $\psi = c \cdot t$ . (see Fig. 1B). The color discriminates the experiments belonging to the two  $\gamma_0$  modalities, the dashed line is for guiding the eyes.

impede reptation and make an insurmountable barrier to knot untying. The emergence of the characteristic length  $\xi$  also excludes fractality of the absorbing conformations. Finally, the two-stage kinetics observed in our simulation reflects the presence of two dynamics: reaction-limited pearling at short distances along the chain, diffusion-limited collapse at large distances.

Experimental approaches in chromosome biology have been recently renewed by chromosome conformation capture (3C) that uses a succession of crosslinking, restriction, religation and sequencing steps to measure contact probabilities along a DNA molecule *in vivo*. This technique centrally exploits the unique opportunity offered by the DNA heteropolymer to have a single sequence identifier at each loci (for long enough identifiers).

In the seminal paper in which this technique was introduced, Lieberman-Aiden *et al.* [14] measured the contact probability,  $P(s)$ , as a function of the genomic distance  $s$ . In the range between 1 and 10 millions base pairs (bp), these authors fitted the resulting curve with a scaling relation  $P(s) \propto s^{-\gamma}$ , with a value of  $\gamma$  close to 1 compatible with the value expected for the fractal-globule state. However, an exponent of 0.75 has also been reported at shorter scale in two human-cell studies, and other out-of-equilibrium mechanisms were invoked to explain this alternative exponent: the tension globule [15] or the extrusion of loops by molecular motors such as condensins [15–17]. While these models can have important implications on the role of chromosome folding in cells, they do not take explicitly into account the potential distortion that the polymer can undergo during the experiment. Moreover, since the initial step of the 3C technique consists in chemically crosslinking the DNA polymer with formaldehyde, we reasoned that this experiment could be used to check the collapse scenario described in our simulations.

We thus performed Hi-C experiment on Yeast chromosomes at different formaldehyde concentrations  $c$  and exposure times  $t$  in order to observe the evolution of the polymer conformation during the crosslink-induced collapse (detailed description of the methods in SM §I.C).

Not knowing the reaction order, we cannot establish an exact mapping between  $k_{on}$  and  $c$ , so we used a simple ansatz,  $\psi = c \cdot t$ , for the rescaled time variable. The experimental curves  $P^{exp}(s)$  cluster around two different mean-curves differing by their slope at short distances  $\gamma_0^{exp}$  (Fig. 4A). When plotting this exponent as a function of  $\psi$ , we observe a sharp transition (Fig. 4B) as predicted by the simulations. Two differences are nevertheless worth discussing. Before the transition, the short-distance exponent of yeast chromosomes is not equal to 1.5 as in the simulations (fig. 1B), but to 1 ( $\pm 0.05$ ). This value might either correspond to an effect of steric interactions during the early phases of pearling collapse or to an *in-vivo* special organization of the DNA in chromosomes, potentially induced by the regular wrapping of DNA around the nucleosomal protein spools. For distances above 10kbp these constraints weaken and the chain follows a more typical random walk with exponent closer to 1.5. After the transition,  $\gamma_0^{exp}$  is equal to 0.7 ( $\pm 0.06$ ), corresponding to  $N_{max} = 1$  in our simulations. This value is likely explained by strong steric constraints preventing a crosslinked loci to contact other loci. The precise estimation of  $\xi$  in various conditions was impaired by the higher biological, experimental and statistical noise on  $P(s)$  at increasing distance  $s$  so that we could not measure experimentally the dependency of  $\xi$  on the crosslinker concentration. Nevertheless, the experiment clearly demonstrates that a collapsing polymer undergoes a sudden transition. It also confirms that inside pearls, at length scales lower than  $\xi$ , the polymer conformation in the absorbing state is very compact, with an exponent  $\gamma_0$  lower than 1, whereas the polymer topology remains unchanged at longer length scales.

We believe that our work will settle the ground for further studies, also in the case where multiple chains are involved such as gelification and aggregation of concentrated and semi-diluted solutions.

We thanks Madan Rao, John Marko, Jean-Marc Victor, Axel Cournac, Benjamin Audit, Marco Cosentino-Lagomarsino, Maxim Dolgushev, Kirill Polovnikov and Daniel Jost for the extremely useful discussions and suggestions, and Véronique Legrand and the DSI of Institut Pasteur for the computational power and assistance.

\* vittore.scolari@gmail.com

† mozziconacci@lptmc.jussieu.fr

- [1] P. De Gennes, Journal de Physique Lettres **46**, 639 (1985).
- [2] A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, Journal de physique **49**, 2095 (1988).
- [3] B. Chu, Q. Ying, and A. Y. Grosberg, Macromolecules **28**, 180 (1995).
- [4] A. Buguin, F. Brochard-Wyart, and P. De Gennes, Comptes rendus de l'Académie des sciences. Série II,

- Mécanique, physique, chimie, astronomie **322**, 741 (1996).
- [5] E. Pitard and J.-P. Bouchaud, The European Physical Journal E: Soft Matter and Biological Physics **5**, 133 (2001).
  - [6] N. V. Dokholyan, E. Pitard, S. V. Buldyrev, and H. E. Stanley, Physical Review E **65**, 030801 (2002).
  - [7] A. Halperin and P. M. Goldbart, Physical Review E **61**, 565 (2000).
  - [8] G. Bunin and M. Kardar, Physical review letters **115**, 088303 (2015).
  - [9] A. Cacciuto and E. Luijten, Nano Letters **6**, 901 (2006).
  - [10] V. F. Scolari and M. C. Lagomarsino, Soft matter **11**, 1677 (2015).
  - [11] M. Doi and S. F. Edwards, *The theory of polymer dynamics*, Vol. 73 (oxford university press, 1988).
  - [12] P.-G. De Gennes, *Scaling concepts in polymer physics* (AIP, 1980).
  - [13] C. F. Abrams, N.-K. Lee, and S. Obukhov, EPL (Europhysics Letters) **59**, 391 (2002).
  - [14] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, *et al.*, Science **326**, 289 (2009).
  - [15] A. L. Sanborn, S. S. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, *et al.*, Proceedings of the National Academy of Sciences **112**, E6456 (2015).
  - [16] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, Cell reports **15**, 2038 (2016).
  - [17] T. Terakawa, S. Bisht, J. M. Eeftens, C. Dekker, C. H. Haering, and E. C. Greene, Science **358**, 672 (2017).