

***Marchantia* TCP transcription factor activity correlates
with three-dimensional chromatin structure**

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Ezgi Süheyla Karaaslan

aus Cankaya, Türkei

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 12.02.2021

Stellvertretender Dekan: Prof. Dr. József Fortágh

1. Berichterstatter: Prof. Dr. Thomas Lahaye

2. Berichterstatter: apl. Prof. Dr. Ulrike Zentgraf

3. Berichterstatter: Prof. Dr. Argyris Papantonis

List of Publications

Part of this work have been published:

Chromatin packing and positioning of plant genomes in 3D

Ezgi Süheyla Dogan and Chang Liu[#]

Nature Plants (2018). DOI: 10.1038/s41477-018-0199-5

***Marchantia* TCP transcription factor activity correlates with three-dimensional chromatin structure**

Ezgi Süheyla Karaaslan, Nan Wang, Natalie Faiß, Yuyu Liang, Sean A. Montgomery, Sascha Laubinger, Kenneth Wayne Berendzen, Frédéric Berger, Holger Breuninger, and Chang Liu[#]

Nature Plants (2020). DOI: 10.1038/s41477-020-00766-0

From the work not presented in this thesis the following manuscript is in preparation:

Isolation of Lineage Specific Nuclei Based on Distinct Endoreduplication Levels and Tissue-specific Markers to Study Chromatin Accessibility Landscapes

Ezgi Süheyla Karaaslan, Natalie Faiß, Chang Liu [#], and Kenneth Wayne Berendzen[#]

From the work not presented in this thesis the following manuscripts have been published:

Plant Lamin-like Proteins Mediate Chromatin Tethering at the Nuclear Periphery

Bo Hu*, Nan Wang*, Xiuli Bi*, Ezgi Süheyla Karaaslan, Anna-Lena Weber, Kenneth Wayne Berendzen and Chang Liu#.

Genome Biology (2019). DOI: <https://doi.org/10.1186/s13059-019-1694-3>.

RST1 and RIPR connect the cytosolic RNA exosome to the Ski complex in *Arabidopsis*

Heike Lange, Simon Y. A. Ndecky, Carlos Gomez-Diaz, David Pflieger, Nicolas Butel, Julie Zumsteg, Lauriane Kuhn, Christina Piermaria, Johana Chicher, Michael Christie, Ezgi S. Karaaslan, Patricia L. M. Lang, Detlef Weigel, Hervé Vaucheret, Philippe Hammann & Dominique Gagliardi#

Nature Communications (2019). DOI: <https://doi.org/10.1038/s41467-019-11807-4>

A Role for the F-Box Protein HAWAIIAN SKIRT in Plant microRNA Function

Patricia L. M. Lang*, Michael D. Christie*, Ezgi S. Dogan, Rebecca Schwab, Jörg Hagmann, Anna-Lena van de Weyer, Emanuele Scacchi, Detlef Weigel#

Plant Physiology (2018). DOI: <https://doi.org/10.1104/pp.17.01313>

Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific *Arabidopsis* hybrid.

Wangsheng Zhu*, Bo Hu, Claude Becker, Ezgi Süheyla Dogan, Kenneth Wayne Berendzen, Detlef Weigel# and Chang Liu*#

Genome Biology (2017). DOI: <https://doi.org/10.1186/s13059-017-1281-4>.

* These authors contributed equally to this work. # Corresponding author

“In a city, you can be alone in a crowd, and in fact what makes the city a city is that it lets you hide the strangeness in your mind inside its teeming multitudes.”

Orhan Pamuk, *A Strangeness in My Mind*

Table of Contents

List of Publications	5
Table of Contents	9
Abbreviations	11
Abstract	13
Zusammenfassung	14
1 Introduction	16
1.1 Hierarchical chromatin organization in nucleus	17
1.2 Topologically Associated Domains	19
1.3 Topologically Associated Domains in Plants	21
1.4 <i>Marchantia Polymorpha</i> as an emerging model organism	23
1.5 Objectives of this work	25
2 Three-dimensional chromatin packing and positioning of plant genomes	26
Preamble	26
3 Marchantia TCP transcription factor activity correlates with 3D chromatin structure	38
Preamble	38
4 Discussion	68
4.1 TAD borders in plants	68
4.2 Intrinsic features of plant TADs	69
4.3 Co-expression of genes in TADs	71
4.4 Potential role of TCP in nuclear Liquid-Liquid Phase Separation	73
4.5 Other candidate proteins in <i>Marchantia</i> 3D genome	74
References	76

Acknowledgements **84**

Curriculum Vitae **86**

Abbreviations

3C	Chromosome Conformation Capture
3D	Three Dimensional
BBR	Barley b Recombinant
BPC	Basic Pentacysteine
bZIP	Basic Leucine Zipper
ChIP	Chromatin Immunoprecipitation
CTCF	CCCTC-binding Factor
DI	Directionality Index
DNA	Deoxyreibernucleic acid
ESC	Embryonic Stem Cell
IDR	Intrinsically Disordered Region
KEE	KNOT Engaging Elements
LAD	Lamina Associate Domain
LLPS	Liquid-Liquid Phase Separation
PCA	Principal Component Analysis
PLAAC	Prion Like Amino Acid Composition
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
SINE	Short Interspersed Nuclear Element
sRNA	small RNA
TAD	Topologically Associated Domain
TAK1	Takaragaike-1, male <i>Marchantia polymorpha</i>
TAK2	Takaragaike-2, female <i>Marchantia polymorpha</i>

Abbreviations

TCP	Teosinte branched1 (tb1) in maize, C ycloidea (cyc) in Garden snapdragon and P CF in rice
TE	Transposable Element
TF	Transcription Factor
TSS	Transcription Start Site
TTS	Transcription Termination Site
UTR	Untranslated Region

Abstract

Information of the genome is not only encoded to its sequence or epigenetic modifications but also found in its folding in 3D space. Recent developments in Chromosome Conformation Capture techniques enabled us to unveil spatial positioning of the genome at different scales. The formation of self-interacting genomic regions, named Topologically Associated Domains (TADs), are discovered by Hi-C, as a key feature of genome organization beyond the nucleosomal level. Each TAD is an isolated local packing unit in which intra-TAD interactions are favoured and inter-TAD interactions are insulated. In animals several architectural proteins are shown to contribute the structure and the function of the animal TADs. Unlike those in animals, TAD formation, function and proteins that play a role in these processes in plants are rather unknown.

Our Hi-C analyses show that the genome of *Marchantia polymorpha*, a member of a basal land plant lineage, shares an evolutionary conserved 3D landscape with that of higher plants. The *Marchantia* genome is subdivided into hundreds of TADs and their borders are associated with TCP1 protein binding. Genome-wide epigenetic analysis reveals that a considerable fraction of *Marchantia* TADs represent interstitial heterochromatin and are decorated with repressive epigenetic marks. We also identify a novel type of TAD that we name TCP1-rich TAD, in which genomic regions are highly accessible and densely bound by TCP1 proteins. TCP1-bound genes residing in TCP1-rich TADs exhibit lower gene expression levels compared to the TCP1-bound genes in other locations.

In *tcp1* mutants, TAD patterns in the Hi-C map do not change, indicating that TCP1 protein is not essential for TAD formation and structure. However, we find that in *tcp1* mutants, genes residing in TCP1-rich TADs have a greater extent in expression fold change compared to genes not belonging to these TADs. Our results indicate that, besides standing as spatial chromatin packing modules, plant TADs function as nuclear micro-compartments that correlate transcription factor activities.

Zusammenfassung

Informationen des Genoms werden nicht nur mit der Sequenz oder epigenetischen Modifikation codiert, sondern auch in ihrer Faltung im 3D-Raum gefunden. Jüngste Entwicklungen bei der Konformationserfassung von Chromosomen ermöglichten es uns, die räumliche Positionierung des Genoms in verschiedenen Maßstäben aufzudecken. Die Bildung selbstinteragierender Genomregionen, die als Topologically Associated Domains (TADs) bezeichnet werden, wird von Hi-C als Schlüsselmerkmal der Genomorganisation jenseits der Nukleosomenebene entdeckt. Jedes TAD ist eine isolierte lokale Packungseinheit, in der Intra-TAD-Wechselwirkungen bevorzugt und Inter-TAD-Wechselwirkungen isoliert werden. Bei Tieren wird gezeigt, dass mehrere Architekturproteine zur Struktur und Funktion der tierischen TADs beitragen. Im Gegensatz zu Tieren sind TAD-Bildung, -Funktion und -Proteine, die bei diesen Prozessen in Pflanzen eine Rolle spielen, eher unbekannt.

Unsere vorläufige Hi-C-Analyse zeigte, dass das Genom von *Marchantia polymorpha*, einem Mitglied einer basalen Landpflanzenlinie, eine evolutionär konservierte 3D-Landschaft mit dem höheren Pflanzen teilt. Das *Marchantia*-Genom ist in Hunderte von TADs unterteilt und ihre Grenzen sind mit der TCP1-Proteinbindung verbunden. Eine genomweite epigenetische Analyse ergab, dass ein beträchtlicher Teil der *Marchantia*-TADs interstitielles Heterochromatin darstellt und mit repressiven epigenetischen Markierungen verziert ist. Wir identifizieren auch einen neuartigen TAD-Typ, den wir TCP1-reiches TAD nennen, bei dem genomische Regionen gut zugänglich und durch TCP1-Proteine dicht gebunden sind. TCP1-gebundene Gene, die sich in TCP1-reichen TADs befinden, weisen im Vergleich zu TCP1-gebundenen Genen an anderen Stellen niedrigere Genexpressionsniveaus auf.

In *tcp1*-Mutanten änderten sich die TAD-Muster in der Hi-C-Karte nicht, was darauf hinweist, dass das TCP1-Protein für die TAD-Bildung und -Struktur nicht wesentlich ist. Wir stellen jedoch fest, dass in *tcp1*-Mutanten Gene, die in TCP1-reichen TADs leben, eine größere Veränderung der Expressionsfalte aufweisen als Gene, die nicht

zu diesen TADs gehören. Unsere Ergebnisse zeigen, dass Pflanzen-TADs nicht nur als räumliche Chromatin-Packungsmodule stehen, sondern auch als nukleare Mikrokompartimente fungieren, die die Aktivitäten des Transkriptionsfaktors korrelieren.

1 Introduction

Genomic DNA in the nucleus is under large constraint. If stretched as a very thin thread, the human genome can reach up to 2 meters, however, it has to fit in a 10 μ m diameter nucleus¹. Although DNA tightly folds in hierarchical orders to fit in the nucleus, it has to remain functional for processes like replication and gene expression. Therefore, nowadays we know that, not only its sequence, but also its folding in 3D space is essential for nuclear operations. The spatial organization of the genome has been studied extensively in recent years and several methods have been developed to assess genome-wide chromatin interactions in the nucleus. A considerable fraction of these techniques is based on the principle that spatially close DNA fragments ligate more efficiently than distal ones. The most far-reaching proximity-ligation based method to quantify chromatin interactions is called Hi-C and it combines high throughput sequencing with chromosome conformation capture (3C) technique to catch higher-order chromatin interactions in high resolution² (For further information regarding chromosome conformation capture techniques, please see Chapter 3, page 27 in the manuscript), (Figure 1).

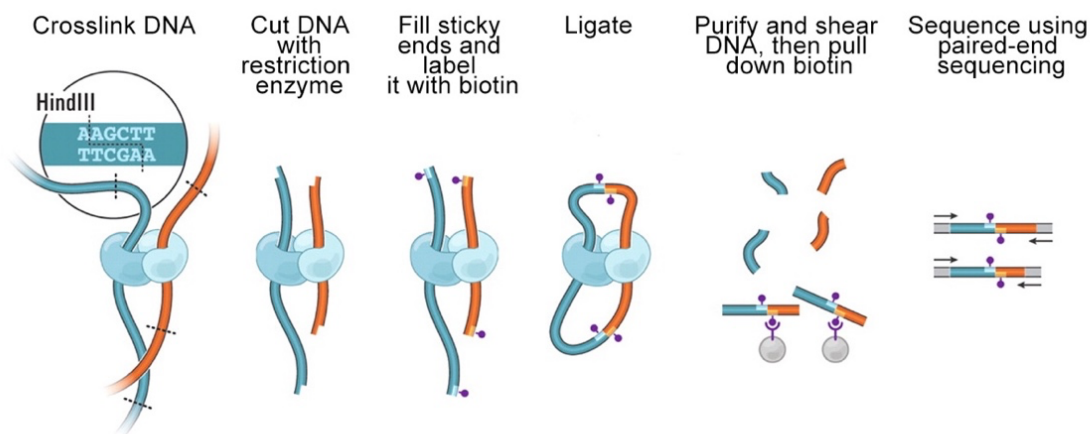


Figure 1: **Overview of Hi-C technique.** In the initial step of Hi-C protocol, cells are fixed with formaldehyde, resulting in crosslink of adjacent chromatin elements. Then, chromatin is digested with restriction enzymes that leave sticky ends (HindIII, DpnII etc.). Sticky ends are filled with biotin labelled nucleotides. Blunt end

ligation is carried out in extremely dilute conditions in order to favour intra-fragmental ligation and prevent cross ligation with other fragments. After ligation, DNA is purified and sheared. Next, biotin labelled fragments are pulled down with streptavidin beads followed by paired-end sequencing to identify chimeric interacting fragments. Figure adapted from³ (Berkum et al. 2010).

1.1 Hierarchical chromatin organization in nucleus

Chromosome conformation experiments revealed that mammalian genomes are hierarchically architected in the confines of the nucleus. At the highest level of this non-random hierarchy, there are chromosome territories, in which each chromosome occupies a discrete space in the nucleus (Figure 2A). These chromosome territories, which can be seen as well-defined squares in the Hi-C map⁴, can be also observed with image-based methods, like in-situ hybridization⁵⁻⁷. Chromosome territories are formed due to the preferred interactions within the same chromosome⁸ and overlaps between two chromosome territories are restricted to their borders⁹.

When we further zoom in a Hi-C, chromosome territories can be further divided into two clusters named A/B compartments according to Principal Component Analysis (PCA) (Figure 2B). Overall, such A/B compartment annotation correlates to chromatin state. Compartment A is enriched with euchromatic and associated with active histone marks like H3K4me3 and H3K27ac; whereas B compartment is enriched with heterochromatic and associated with repressive marks^{4,10,11}. A/B compartments can be further subdivided into smaller compartments according to unique histone modification patterns¹². A/B compartment separation has been shown to be a dynamic process. A chromatin region switching from B to A compartment is associated with increased gene expression, whereas a switch from A to B is associated with lower expression levels¹³.

With the development of Hi-C method, self-associating chromosomal domains called Topologically Associated Domains (TADs) are discovered (Figure 2C). TADs are individual genomic units that restrict chromatin interaction within itself^{14,15}. In other

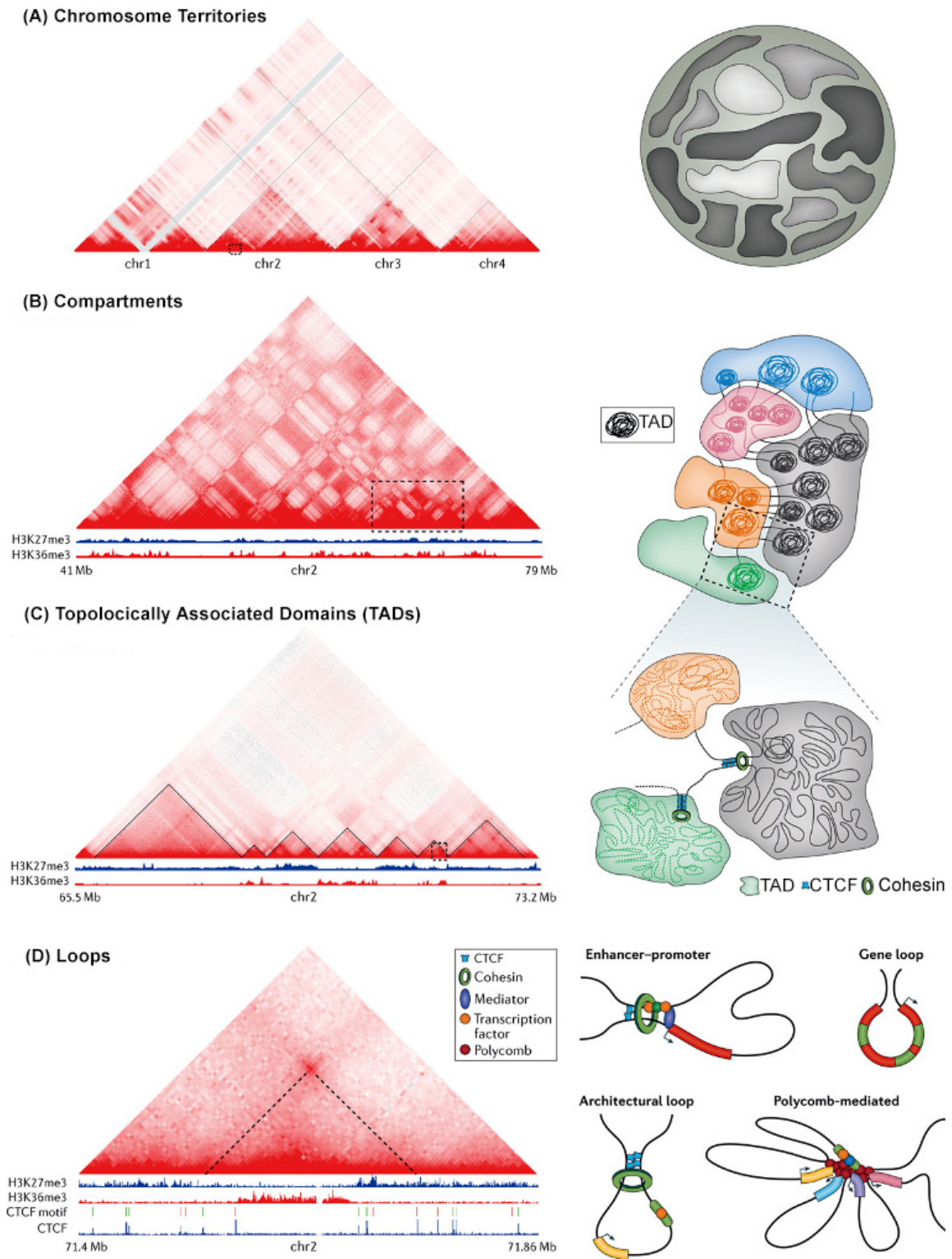


Figure 2: **Hierarchical organization of chromatin.** Panels at the left side are Hi-C heat maps showing interaction frequencies by colour intensity. Panels on the right-side

show models of different hierarchical structures in the nucleus. (A) Each chromosome occupies a subspace in the nucleus. Interactions occur mostly between two loci on the same chromosome. (B) Segregation of the genome into alternation interaction clusters according to epigenetic landscape shows the compartmentalization of A/B domains. (C) TADs are nested interaction units in 3D genome. Due to the enriched intra-TAD interactions TADs appear as high-interaction triangles on the Hi-C map. (D) Several different pairwise chromatin loops can be observed in the genome. Figure adapted from⁸ (Bonev and Cavalli, 2016).

words, intra-TAD interactions are preferred whereas inter-TAD chromatin interactions are insulated¹⁴. In Hi-C maps, the contact frequency of two loci within one TAD is two or three times higher than that of two loci outside of TAD¹⁶. TADs act as functional regulatory units by providing specificity and directionality to gene expression by favouring internal TAD interactions^{17,18}.

At the finest scale, advancements in the methodology are also enabled us to detect chromatin loops as pairwise interactions (Figure 2D). Chromatin loops are critical elements that spatially regulate gene expression by positioning distal regulatory elements of gene expression with proximal elements¹⁹.

1.2 Topologically Associated Domains

In 2012, scientists discovered local chromatin interactions along the map of high-resolution Hi-C and named them as TADs. To systematically identify TADs, Dixon and colleagues used the Directionality Index (DI) which based on the interaction direction shift from upstream to downstream to estimate the boundaries of TADs. At the border of two TADs, chromatin interaction will suddenly shift from downstream to upstream or vice versa. For example, at the right border of a TAD we expect a bias in contact frequency towards the regions on the left side¹⁴. Afterwards, several other methods were applied to approximate the accurate demarcation of TADs. Further investigation revealed that TAD borders are enriched with the binding of the insulator element CTCF (CCCTC-binding factor), which is a highly conserved zinc finger

protein. These boundaries are also associated with accessible chromatin, transcription start sites (TSS), SINE (Short Interspersed Nuclear Element) repeats, housekeeping genes, DNaseI hypersensitive sites, H3K36me3 and H3K4me3 histone modifications¹⁴.

Since its initial discovery, TADs have been broadly studied in a large number of species in the animal kingdom. TADs in animals happen to be stable across cell lines^{12,14} and widely conserved in different species^{20–22}. Another conspicuous characteristic of animal TADs is that they are genuinely insulated regions and their borders are occupied by insulator protein called CTCF^{8,12,14,16,22,23}. The presence of CTCF is very critical for the insulation at the border and TAD structure, the consequence of the loss of these elements is a disruption of spatial genome architecture and misexpression of genes^{24–27}. Cohesin is another important factor that is located at the TAD borders. It mainly plays role in TAD formation according to dynamic “loop extrusion model” where cohesins load chromatin into the loop until it meets CTCF binding sites^{28–31}. The depletion of cohesin perturbs the stability of chromosomal domain architecture³² and eradicates loop domains along with long-range chromatin interactions^{33,34}. It is also shown that, re-introducing cohesin to the genome recovers the loop formation, therefore proving that loop extrusion is an active process³³. Moreover, further research also showed that TADs are not only formed by CTCF/cohesin cooperation but sometimes rather, A/B chromatin state³⁵, transcription^{36–38}, gene density³⁸ and phase separation³⁹ contribute to shape the 3D chromatin landscape⁴⁰. In pluripotent cells, retrotransposon activities are also shown to demarcate TAD borders⁴¹.

TADs are key components of the genome topology at multiple scales, and thus have a central role in gene expression, development, and disease. At higher scales long-range TAD-TAD interactions stabilize heterochromatic B-compartment toward nuclear periphery by forming cliques⁴². Moreover, the insulation at the borders of TADs, which restricts enhancer-promoter activities⁴³, suggests that disruption of the TAD structure might result in ectopic interactions causing mis-expression of genes. Consistently, genetic manipulation of specific TAD borders induced ectopic contacts

and changes in gene expression. After the deletion of a TAD boundary in mouse Embryonic Stem Cells (ESCs), ectopic interactions between adjacent TADs were observed and neighbouring genes were upregulated¹⁵. More specifically, disruption of TAD border insulation in mouse embryo caused an ectopic enhancer-promoter interaction, which result in mis-expression of genes during limb development²⁵. Furthermore, more studies confirmed these initial findings that the disruption of TADs lead to ectopic gene expression in genetic diseases^{44–47} and cancer^{48–52}. Not only inter-TAD interactions but also intra-TAD interactions are critical for productive gene regulation. In ESCs, an asymmetric type of TAD is discovered in which only one border of the TAD has strong intra-TAD interaction. These regions are formed due to the differential enhancer occupancy and are important for cellular identity during differentiation by providing permissive gene expression landscape⁵³. The gene expression regulation of TADs is not limited to the insulator activity at the borders, can also be observed in TAD bodies configured by transcriptional status³⁵. In *Drosophila*, RNA polymerase II (RNAPII) occupancy and actively transcribed genes are the underlying factors of chromatin organization⁵⁴. However, it is not yet clear that whether 3D compartments have an effect on the level of transcription or action of transcription factors inside the particular TAD.

1.3 Topologically Associated Domains in Plants

In the past few years, several 3C derivatives such as Hi-C are conducted in several different plant species, providing us a nice overview of 3D architecture in the kingdom *Plantae*^{55–67}. However, we still don't have a profound understanding of TADs, since these studies remain superficial and have not been followed by in-depth analysis so far. Hence, several important questions remain, such as whether plant TADs have regulatory function or how plant architectural proteins are involved in shaping the genome. Neither the popular model plant *A. thaliana*, nor its close relative *A. lyrata*, possess prominent animal-like TAD structure^{58,68–70}. Nonetheless, in the *Arabidopsis* genome, more than 1000 insulator-like and TAD-like regions are identified, correlating with epigenetic landscape⁷⁰. As the name suggests, “insulator-like” regions in *Arabidopsis* are enriched for accessible chromatin and active histone marks.

Accordingly, genes found in these regions are highly expressed, resembling animal TAD borders⁷⁰.

In this aspect, *Arabidopsis* can be considered as an exception, because unlike *Arabidopsis*, many crop species such as rice, foxtail millet, sorghum, tomato, maize and cotton have genome-wide distinct TADs appear on their Hi-C maps^{55,57,60}. It is speculated that, this discrepancy among different plant species is due to the differences in genome size^{71,72}. *Arabidopsis* has a small genome with high gene density, on the contrary, crop species have larger genome size with low gene density, where the silenced regions between active genes and Transposable Element (TE) rich regions likely to display TAD structures^{71,72}. Although this hypothesis might explain the presence of TADs in larger genomes, it is not sufficient to explain the absence of TADs in *Arabidopsis* and its close relatives, since *Drosophila Melanogaster* genome is 180 Mb⁷³ and displays distinct TAD structures^{74,75}. Moreover, recently, it is suggested that properties of TADs in *Drosophila* 3D landscape might give hints regarding the absence of TADs in *Arabidopsis*⁷⁶. In the *Drosophila* genome, TAD borders are characterized by sudden changes of epigenetic states⁷⁷. Unlike *Drosophila*, sudden changes in epigenetic landscape is not observed frequently in *Arabidopsis*^{70,72}. Further supporting this, the abrupt changes in epigenetic state in *Arabidopsis* genome occurs at the borders of heterochromatic knob structures which show TAD-like features^{76,78}.

Hi-C is also performed in several crop species, revealing distinct TAD structures. Dong and colleagues performed a broad study in which they compared 5 different crop species. TADs in these crops are mostly coincide with heterochromatic B compartments and on a global scale, they have borders that are enriched with active epigenetic marks and gene expression^{57,70,79}. Recently, plant TAD borders in different tissues are shown to be associated with transcription⁷⁹. The genes that are overlapping with tissue-specific TAD borders are found to be upregulated in the particular tissue compared to the other tissues^{79,80}. Together, these findings indicate that transcriptional activity and epigenetic landscape of the region could be the major factors playing roles in TAD demarcation in plants⁸⁰.

In animals, TADs are found to be stable across tissues, developmental stages, and even species¹⁴. Unlike animal TADs, plant TADs in different species but located in syntenic blocks are not found to be conserved⁵⁵. Recent analysis in different maize tissues exhibit a high correlation among TAD borders across tissues of the same plant⁷⁹. Further analysis among non-conserved TAD borders (borders that appear in one tissue) also showed that, even though they cannot be called as TAD borders in one of the tissue, they exhibit TAD-border-like properties compared to a random region⁷⁹. The same weak border-like feature of non-conserved TAD-borders is also observed in rice and foxtail millet, confirming that plant TADs have conserved properties across tissues⁷⁹.

The function of TADs in the animal field is widely studied and discussed in the past decade. As mentioned before, animal TADs are functional regulatory compartments that provide close proximity contact between distal enhancer elements with their corresponding promoters and prevent mis-expression of genes by hindering ectopic interactions with foreign enhancers^{17,25}. On the contrary, long-range inter-TAD contacts have been detected in plants, which shows interaction of putative enhancer-promoter elements can take place across TAD borders⁵⁵. This is plausible, because plants do not have CTCF orthologs, or any other known plant-specific insulator protein that might play a role in TAD border insulation^{81,82}. Studies in *Arabidopsis* TAD-border-like regions and rice TAD borders showed that bZIP (Basic Leucine Zipper) and TCP binding motifs are enriched at TAD borders^{57,71}. Nevertheless, these proteins have not been shown to function as insulator binding factors. It is still not known if they contribute to the TAD structure in plants.

1.4 *Marchantia Polymorpha* as an emerging model organism

The liverwort *Marchantia Polymorpha* is a haploid basal land plant, which has been used for basic research for almost 200 years⁸³. However, compared to *Arabidopsis thaliana*, genomic tools for *Marchantia* fall behind in the last two decades. Nonetheless, due to ease of cultivation and the development of several experimental techniques for *Marchantia*, such as agrobacterium-mediated transformation and CRISPR-CAS9,

researches still worked on it^{84–87}. Recently, its whole genome sequencing is published at 2017^{88,89} and revived *Marchantia* as a modern model organism for high throughput analysis.

Compared to the other land plants, another unique feature of *Marchantia* that makes it a very popular model, is the lack of redundancy in most regulatory transcription factor families. On one hand, this indicates that *Marchantia polymorpha* exhibits genome composition that is predicted for the ancestral land plant, which is a desired characteristic for phylogenetic and evolutionary studies. On the other hand, exhibiting low gene redundancy is advantageous while it provides scientists smooth functional studies to uncover the role of regulatory genes⁸⁸.

Marchantia is a haploid model organism that each individual has only one of the sex chromosomes. In male and female plants, sex-specific V and U chromosomes are present, respectively. The life cycle of *Marchantia* includes both sexual and asexual reproduction. Asexual production takes place through the structures called ‘Gemma Cups’. Each gemmae produced in Gemma cup is emerged from a one single cell and it is a genetically identical copy of the mother plant. In sexual production, male and female plants grow specific umbrella-shaped sexual organs called ‘antheridium’ and ‘archegonium’, respectively. In antheridium, multiple male gametes (sperm cells) are produced and carried to the female archegonium by raindrops. After fertilization, the zygote grows and develops into a sporophyte. Each sporophyte includes numerous spores which can develop into an individual plant (Reviewed in ⁸³). Continuous sexual and asexual reproduction cycles of *Marchantia* and its haploid genome provide an adventitious ground for several popular biological approaches such as CRISPR-CAS9⁹⁰.

In recent years, research related to life cycle⁹¹, epigenome^{89,92}, 3D genome⁹³, signalling pathways^{94,95} and evolution⁹⁶ of *Marchantia* have found a wide audience.

1.5 Objectives of this work

TADs are a prominent characteristic of both animal and plant genomes. In the animal field, it is known that CTCF insulator proteins are enriched at the borders of TADs and together with the cohesins they are responsible for the formation and proper functioning of animal TADs. However, the main factors related to plant TADs structure are still waiting to be discovered. The major aim of this work was to get further insights into the general 3D organization of chromatin in plants, to discover the structural components that contribute to chromatin organization, and to understand how these spatial features regulate genome function.

Prior to this work, Hi-C analysis on *Oryza sativa* and *Arabidopsis thaliana* have revealed a consensus DNA binding motif at the borders of TADs and TAD-like-regions, respectively. This consensus motif is recognized by the family of conserved transcription factors called TCPs. Therefore, TCP proteins show up as an exciting candidate to contribute TAD structure and function in plants. For higher plants, individual species contain more than 20 members of the TCP protein family with redundant functions. In order to unravel the role of TCP protein in plant genome topology, we exploited the *Marchantia* genome that had low gene-redundancy

The first aim of this study was to identify and characterize functional chromatin domains in *Marchantia polymorpha*, and to investigate their properties in comparison to the higher plants.

The second aim of this study was to gain further insights into the function of *Marchantia* TCP protein in the context of TADs in *Marchantia*, and to examine its potential role in functioning as candidate regulatory protein in chromatin packing.

The results of our experiments shed light on the 3D genome structure of *Marchantia* and fill some blanks with the in-depth functional characterization of plant transcription factor TCP in the framework of spatial genome organization.

2 Three-dimensional chromatin packing and positioning of plant genomes

Preamble

This chapter is published in the journal of Nature Plants:

Ezgi Süheyla Dogan and Chang Liu. Chromatin packing and positioning of plant genomes in 3D. Nature Plants (2018). DOI: 10.1038/s41477-018-0199-5

Ezgi Süheyla Dogan and Chang Liu wrote the manuscript.

Three-dimensional chromatin packing and positioning of plant genomes

Ezgi Süheyla Doğan and Chang Liu *

Information and function of a genome are not only decorated with epigenetic marks in the linear DNA sequence but also in their non-random spatial organization in the nucleus. Recent research has revealed that three-dimensional (3D) chromatin organization is highly correlated with the functionality of the genome, contributing to many cellular processes. Driven by the improvements in chromatin conformation capture methods and visualization techniques, the past decade has been an exciting period for the study of plants' 3D genome structures, and our knowledge in this area has been substantially advanced. This Review describes our current understanding of plant chromatin organization and positioning beyond the nucleosomal level, and discusses future directions.

As the largest molecule in a living cell, the genomic DNA is packed with histones to form chromatin. Chromatin conformation is a critical factor for many regulatory elements to execute biological activity¹. On perceiving environmental and developmental cues, both global and local chromatin rearrangements may occur, along with changes in gene transcription^{2,3}. Early cytological studies have well demonstrated chromatin structure at a global level, showing how chromosomes occupy the nuclear space and how chromosomes interact with each other non-randomly^{4,5}. Newly invented molecular and computational tools enable scientists to unveil chromatin structure at an unprecedented resolution to address a fundamental question in genome biology: how does spatial chromatin organization regulate genome functionality? Three-dimensional (3D) genome organization displays a hierarchical pattern, in which individual chromosomes can be dissected into structural and functional domains at multiple levels⁶. In both the animal and plant fields, many of such chromatin domains have been identified and characterized recently, and range from hundreds of thousands of kb to small chromatin loops with sizes of several kb^{1,7,8}. With recent efforts in unveiling plant genome packing and chromatin positioning patterns in nuclei, we have accelerated our journey to gain a better understanding of plant genomes beyond the DNA sequences. With a focus on chromatin structures above the nucleosomal level, here we compile the results of recent studies of plant 3D genomes and discuss trends in this rapidly expanding area.

State-of-the-art methods to interrogate 3D genomes

Our ever-expanding toolkit enables us to address diverse questions concerning how chromatin structures are formed and regulated, and how the interplay between chromatin structure and proteins (such as transcription and chromatin remodelling factors) contributes to gene expression. From the whole chromosome to a single gene body, newly developed chromosome conformation capture and microscopic techniques allow scientists to interrogate chromatin organization at multiple resolutions. These techniques would be beneficial, particularly in plant sciences, to investigate how multiple sets of genomes interact after species hybridization, which is a common strategy for crop improvement.

Hi-C and its derivatives: a brief history and their applications.

To analyse 3D chromatin interactions in nuclei, Dekker and colleagues developed an approach called chromosome conformation capture (3C), which measures how frequent two genomic loci interact⁹. This method has soon become a standard way to study local chromosome organization. For example, 3C has been widely used to examine juxtaposition between specific transcription units and remote enhancer elements. A limitation of 3C is that it only allows identification of the interaction between two chosen loci. The 3C method was later developed into 4C (circular chromosome conformation capture), which enabled genome-wide detection of chromatin interactions associated with one locus of interest¹⁰. Another powerful 3C derivative is 5C (3C-carbon copy), in which interactions among thousands of selected genomic loci can be studied in a single run^{11,12}. Finally, the Hi-C approach, a 3C-derived method with the most far-reaching impact, allows us to detect interactions at a whole-genome scale¹³. This all-to-all detection is achieved by incorporating a biotin-labelled nucleotide at the ligation junction, thus enabling enrichment of ligation products by affinity purification. The recovered ligation products are subjected to high-throughput sequencing, providing a whole-genome picture of both short- and long-range chromatin interactions¹³. These chromosome conformation capture methods have been adapted for plant research (reviewed in refs ^{14,15}). The approach of identifying the chromatin interaction network can be tailored for different purposes. For instance, through combining it with chromatin immunoprecipitation (named ChIA-PET and HiChIP) or hybridization capture (named Capture-C) approaches, one can obtain comprehensive chromatin interaction networks of genomic regions bound by a protein (for example, transcription factors) or genomic regions belonging to a certain annotation category (for example, promoters), respectively^{16–20}. Additionally, the Hi-C method has been scaled down to a single-cell level to study cell-to-cell variability in chromatin structures^{21,22}, as well as dynamic chromatin organization during cell differentiation^{23,24} and cell-cycle progression²⁵. Still, there are many Hi-C-related cutting-edge techniques under development (summarized in ref. ²⁶), such as the newly developed genome architecture mapping method, which does not require any chromatin digestion or ligation steps²⁷. Such an array of powerful methods

Center for Plant Molecular Biology (ZMBP), University of Tübingen, Tübingen, Germany. *e-mail: chang.liu@zmbp.uni-tuebingen.de

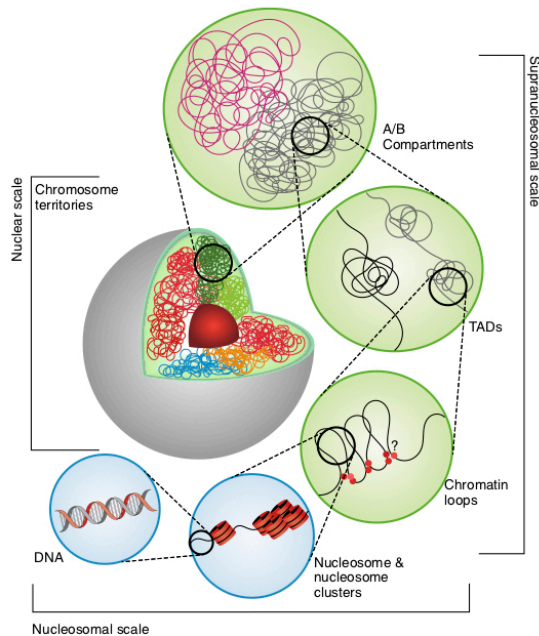


Fig. 1 | Schematic representation of hierarchical chromatin organisation in plants. Individual chromosomes occupy a subspace in the nucleus called chromosome territories. Chromosome territories can be further partitioned to distinct A and B compartments, which are enriched for active and repressed chromatin, respectively. Genomic regions within TADs display increased interactions, while their interactions with neighbouring regions outside of the TADs are rather limited. A limited number of chromatin loops connecting regulatory elements to their target loci have been described in plants (summarized in ref. ⁹), the molecular mechanisms driving plant chromatin loop formation (especially in crop species) are largely unknown (question mark).

enables us to address many biological questions concerning chromatin structures at a remarkable resolution.

Visualizing chromatin regions of interest. As a complementary approach to 3C-based experiments, methods that allow tracing and visualizing chromatin regions of interest are considered to be crucial tools for data validation^{28–30}. The fluorescent in situ hybridization (FISH) method, which is the ‘gold standard’ technique for detecting chromatin localization, has constantly been improved to achieve higher sensitivity, specificity and resolution³¹. Today, a wide variety of FISH methods are used to examine chromatin organization, such as 3D-FISH (refs ^{32,33}) and FISH using Oligopaint^{34,35} or molecular beacon probes³⁶. By combining FISH with super-resolution microscopy, folding and positioning of structural chromatin domains can be characterized in detail, providing a complementary insight into chromatin organization^{37,38}.

Apart from FISH, which deals with fixed materials, several recently developed tools offer means of live imaging to document the dynamics of chromatin with minimum perturbation to chromatin itself. Previously, the resolution of chromatin live imaging was limited to repetitive regions. However, with advanced signal detection measures, such as those using modified dCRISPR–Cas9 or TALE proteins labelled with quantum dots, it is now feasible to visualize the real-time behaviour of low-repetitive regions or even a single genomic locus^{39–42}. In the dCRISPR–Cas9 labelling system,

a ~20 nt RNA sequence guides a single fluorescent dCRISPR–Cas9 reporter protein to the locus of interest, which is technically challenging to observe. This difficulty has been much alleviated with a recent design that combines tandem tags and bimolecular fluorescence complementation, which significantly improves signal-to-noise ratio⁴³. Other than labelling single genomic loci, a strategy using GFP-tagged m6A-tracer proteins allows tracking of chromatin domains specifically localized in a certain nuclear subcompartment (for example, the nuclear periphery)⁴³.

Global and local levels of plant chromatin organisation

Eukaryotic genomes are organized in a hierarchical fashion. At the nucleosomal scale, genomic DNA is packed around histones to form nucleosomes. Beyond that, chromatin-looping gives rise to topologically associated domains (typically tens of kilobase-pairs large) and A/B compartments (typically hundreds to thousands of kilobase-pairs large). The A/B compartments are generally euchromatic and heterochromatic regions, respectively. On top of these domains and compartments, chromosome territories are formed (reviewed recently in refs ^{6,44}) (Fig. 1). Interestingly, structural features of different genomes in 3D turn out to be highly diversified. For example, by comparing *Arabidopsis* and rice, without looking into the details of their local chromatin organization patterns, researchers can already reveal drastic differences at the chromosomal level with cytological analysis (Fig. 2a).

Chromosome conformation. In an interphase nucleus, chromosomes occupy distinct nuclear spaces, or compartments, that are referred to as chromosome territories. The way plant chromosomes fold to occupy chromosome territories is diverse across different species¹⁵. *Arabidopsis* chromosomes adopt an overall ‘Rosette’ configuration¹⁵. *Arabidopsis* chromosomes adopt an overall ‘Rosette’ configuration, in which the centromere of a chromosome is highly condensed with its flanking pericentromeric heterochromatin to form a so-called chromocenter (CC)⁴⁶. In a Rosette configuration, euchromatin emanates from CCs as megabase-size chromatin loops⁴⁶. Distinct from *Arabidopsis*, plants such as wheat, rye, barley and oats, adopt a ‘Rabl’ chromosome configuration, where centromeres and telomeres are located at opposite poles of the nucleus^{47–49}. Rice chromosomes in xylem vessel cell nuclei show a Rabl conformation⁵⁰. On the other hand, recent Hi-C analyses of rice genome organization in leaf tissues (with the majority material coming from mesophyll cells) indicate a non-Rabl chromosome organization, due to the absence of strong interactions among centromeres in Hi-C maps, which is a characteristic feature of the Rabl conformation⁵¹. Furthermore, a ‘Bouquet’ chromosome conformation has been described in meiotic maize, wheat and rice cells, in which telomeres cluster at the small area beneath the nuclear envelope while the rest of the chromosomes spread throughout the nucleoplasm^{52–54}. The presence of different chromosome conformations in the same species (for example, in rice, Rabl in xylem vessel cells; non-Rabl in mesophyll cells; and Bouquet in meiotic cells) indicates that they are possibly correlated to different cell identities. Chromosome conformation can also be flexible in the same cell type. For example, on perceiving light, etiolated seedlings switch their developmental program from skotomorphogenesis to photomorphogenesis. During this transition, decondensed heterochromatin becomes highly condensed and forms CCs^{55,56}. Another example highlights the decondensation of CCs during heat shock response in *Arabidopsis*⁵⁷. Thus, having flexible chromosome conformation might be an integral part of plants’ adaptability to ever changing environmental conditions.

Inside chromosome territories there exists spatial chromatin compartmentalization. With principal component analysis (PCA) on Hi-C maps, chromatin regions can be arbitrarily annotated as ‘A’ or ‘B’ spatial compartments, according to the first component¹³ (see Fig. 2b for example). Spatial separation of A and B compartments in nuclei has been confirmed with FISH at the single-cell level¹⁸. In principle,

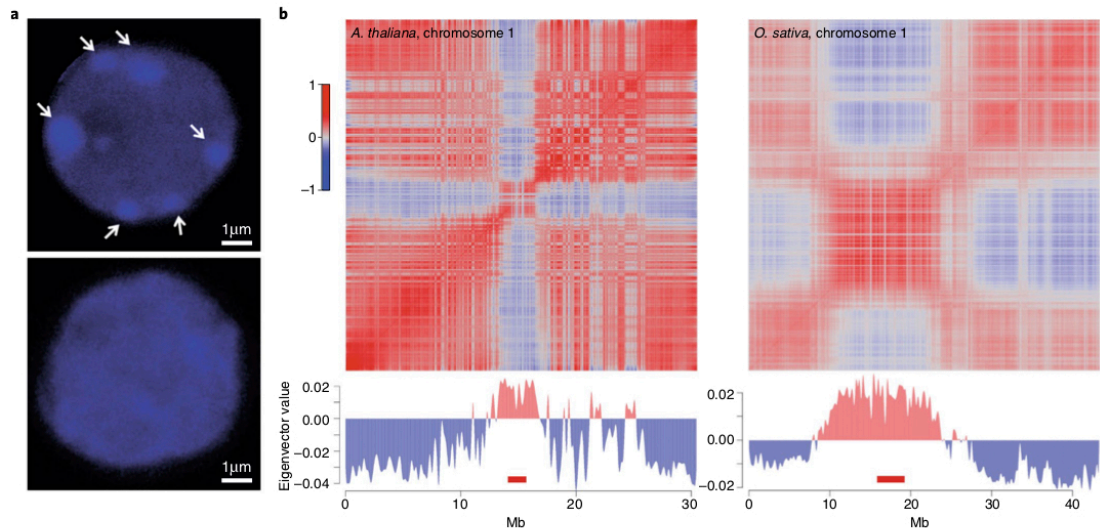


Fig. 2 | Comparison of *Arabidopsis* and rice chromatin organization at chromosomal level. **a**, *Arabidopsis* (top) and rice (bottom) nuclei stained with DAPI. Arrows depict chromocentres. **b**, Correlation matrices of *Arabidopsis* (left; bin size, 20 kb) and rice (right; bin size, 10 kb) chromosome 1 Hi-C maps. The plot below each matrix denotes the eigenvector of the first principal component, correlating to chromosomal domain compartmentalization. For each eigenvector plot, the compartment bearing the centromere is coloured red. The red bars denote centromeric regions^{60,62}.

this binary classification of chromatin regions can be applied to all Hi-C maps (reviewed recently in ref. 43). Notably, similar to animals, the A/B compartment annotation in *Arabidopsis* and many crop species is largely in accordance with the euchromatin/heterochromatin landscapes of the genome^{51,58–62}. Moreover, the annotated plant A/B compartment can be further divided into a mixture of smaller compartments, and this lower level compartment annotation is highly correlated with euchromatin/heterochromatin marks as well^{51,58}. Similar to that in animals, *Arabidopsis* chromatin compartmentalization patterns show a strong correlation to DNA replication timing, in which genomic regions replicated at the same pace tend to physically interact more strongly than do those replicated at different time phases⁶³.

Specific intra- and inter-chromosomal interactions can further shape the folding of plant chromosomes. In *Arabidopsis* seedlings, highly reproducible, strong intra- and inter-chromosomal interactions among several interstitial heterochromatic regions are detected^{59,64}. These regions, named IHIs (interactive heterochromatic island) or KEEs (KNOT engaged element), are highly enriched with heterochromatic histone marks and transposons. Neither the biological meaning nor molecular mechanism underlying strong interactions among IHIs and KEEs is clear. Nevertheless, clustering of these regions does not seem to be involved in transcriptional regulation, as attenuated interactions among them do not affect local gene expression⁶⁵. It was proposed that IHIs and KEEs represent preferred transposon element insertion sites in the genome⁵⁹.

Topologically associated domains. Topologically associated domains (TADs) have received the most attention among all the structural features revealed by Hi-C. By definition, each TAD is a chromatin region showing suppressed interactions with its flanking regions. Meanwhile, chromatin interactions within the same TAD are favoured. Thus, having chromatin organized in the form of TADs allows long-range chromatin contacts with spatial constraints, conferring target specificity of cis-regulatory elements. With super-resolution microscopy techniques applied to single

cells, it has been shown that besides being a reflection of statistical frequencies of chromatin interactions, TADs are genuine structural units that are separated from each other in 3D (ref. 66). Remarkably, TADs have been observed in many organisms, such as human⁶⁷, mouse⁶⁸, fruit fly^{69,70}, roundworm (*Caenorhabditis elegans*)⁷¹, fission yeast⁷² and the bacterium *Caulobacter crescentus*⁷³. On a megabase scale, TAD patterns in mammals are largely conserved in different cell lines and even across species⁶⁷; whereas on a sub-megabase scale, subdomains within a TAD could become merged or disconnected in orchestration with developmentally regulated events^{67,74,75}. It is worth noting that even for organisms showing TADs, TADs can be absent in certain circumstances, such as during mitosis or before *Drosophila* zygotic genome activation^{66,77}.

Even before being associated to Hi-C maps, the CCCTC-binding factor (CTCF) has been known as an essential protein regulating gene transcription and genome organization in higher eukaryotes (reviewed in ref. 78). With the initial finding that CTCF is highly enriched at mammalian TAD borders⁶⁷, efforts to understand how it contributes TAD formation have surged. In brief, the consensus has been that CTCF-dependent TAD formation is mediated by CTCF-cohesin protein complexes that stall at TAD borders, together with a ‘loop extrusion’ mechanism executed by free cohesins sliding inside TADs^{67,79–85}. At a molecular level, the loop-extrusion model is supported by the discovery of how condensin, another structural protein thought to play a role similar to that of cohesin in TAD formation, translocates along DNA in an ATP-hydrolysis-dependent manner^{86,87}. Moreover, condensin has been found to directly manipulate chromatin loops, transforming interphase chromatin to highly condensed chromosomes during cell division⁸⁸. On the other hand, the cohesin-unloading factor Wings apart-like (Wapl) has been found to play an antagonizing role, to prevent the association of cohesins with the genomic loci where CTCF binds; thereby attenuating contacts between TAD borders^{89–91}. Together, these findings indicate that the recruitment of cohesin to various loci is a balanced procedure, enabling a correctly folded genome.

REVIEW ARTICLE

NATURE PLANTS

Table 1 | Genome sizes of various plant species and the occurrence of TADs

	Genome size (Mb)	TADs	References
<i>Arabidopsis thaliana</i>	120	No	59,64,99
<i>Arabidopsis lyrata</i>	230	No	65
<i>Oryza sativa</i> (rice)	430	Yes	51,58
<i>Setaria italica</i> (foxtail millet)	490	Yes	58
<i>Sorghum bicolor</i> (sorghum)	770	Yes	58
<i>Solanum lycopersicum</i> (tomato)	900	Yes	58
<i>Gossypium hirsutum</i> (cotton)	2,300	Yes	61
<i>Zea mays</i> (maize)	2,500	Yes	58

As expected, targeted degradation of the CTCF–cohesin complex disturbs many TAD boundaries^{92–94}. Interestingly, even in the absence of cohesin, a large number of TADs are still present. Detailed analyses indicate that these CTCF–cohesin-independent TADs correlate to the compartmentalization of chromatin regions into active and inactive spatial compartments, which is probably driven by the transcriptional state of chromatin itself^{92–94}. In a recent study, Rowley and colleagues showed that transcriptional state is a major predictor of Hi-C map patterns in both animals and *Arabidopsis*⁹⁵. Several lately published case studies support this notion that local gene expression *per se* can act as a driving force shaping chromatin domain formation. For instance, the establishment of TADs in *Drosophila* zygotes coincides with local gene activation⁷⁶. In another example, transcriptional activation of a noncoding RNA is found to trigger chromatin compartmentalization during T cell differentiation, repositioning this noncoding RNA locus and its flanking chromatin from the nuclear lamina to the nuclear interior, which further leads to enhancer–promoter interaction within this translocated chromatin domain⁹⁶. Furthermore, investigation in yeast favours a model in which the self-interaction domains (similar to TADs) can be explained by transcription-induced DNA supercoiling⁹⁷.

In summary, there are two major independent mechanisms that contribute to TAD formation in animals: the action of loop extrusion by cohesin and the pause of extrusion action by CTCF proteins, and the spatial chromatin compartmentalization in accordance with epigenomic landscape and transcriptional activity.

Chasing TADs in plants. The first Hi-C experiment in plants was conducted on *Arabidopsis thaliana* to show how CCs were differentially packed in wild type and mutant specimens⁹⁸. Later on, detailed analyses of *Arabidopsis* Hi-C maps were reported by three research groups, independently^{29,64,99}. Surprisingly, TADs can hardly be found in *Arabidopsis* chromosome arms, despite appearing to be a prevalent structural feature of genome packing in many other species. A possible explanation for this was the absence of canonical insulators, such as CTCF, in plants¹⁰⁰. Another possible reason for not observing TADs in *Arabidopsis* could be technical—because non-homogeneous, unsynchronized nuclei with different cell types, endopolyploidy levels and cell cycle stages were harvested as the input material for Hi-C, TADs that specifically form in one type of nuclei might have been masked by the Hi-C patterns of the rest. Nevertheless, investigation of a high-resolution *Arabidopsis* Hi-C map (as the average of such a nuclear mixture) led to identification of over 1,000 TAD-boundary-like and insulator-like regions⁹⁹. These regions possess similar properties to those of animal TAD borders; where chromatin contacts over insulator-like regions are much attenuated, and they are enriched for open chromatin and highly expressed genes, which is in line with findings showing

a strong correlation between active gene transcription and TAD boundaries^{99,101}.

With efforts in expanding studies of plant Hi-C maps to multiple crop species, one can conclude that TADs are common structures in plants, while *Arabidopsis* is an exception. So far, Hi-C maps of rice, foxtail millet, sorghum, tomato, cotton and maize have been reported, and these genomes all display conspicuous TADs^{51,58,61,102}. Like those in animals, plant TAD borders are enriched for active genes and euchromatic histone marks associated with open chromatin^{51,58,103}. In mammals, TAD distribution patterns are highly conserved among different species^{67,104}, yet this is not the case for plants⁵⁸. As pointed out by Dong and colleagues, the lack of TAD pattern conservation across plant species might be due to the absence of CTCF protein, which is highly associated with borders of conserved TAD boundaries in mammals⁵⁸.

Interestingly, whether a plant species manifests TADs seems to be related to its genome size (Table 1). Among all plants known for having TAD structures, rice has the smallest genome size (roughly 430 Mb), which is more than three times larger than that of *Arabidopsis thaliana* (roughly 120 Mb). Additionally, based on our recent Hi-C experiments on *Arabidopsis lyrata* (genome size: ~230 Mb), a close relative of *A. thaliana*, we did not observe TADs either⁶⁵. Notably, TADs identified from high-resolution rice Hi-C maps tend to be depleted from protein-coding genes and enriched for DNA methylation⁵¹. As a negative correlation holds between genome size and gene density, we speculate that plant genomes with lower gene density (that is, larger genome size) are likely to display TADs, reflecting the spatial separation of chromatin regions with contrasting gene densities and/or epigenetic marks. It should be pointed out that here we refer to an insulated plant chromatin region as ‘TAD’ if it has a size comparable to TADs identified in animals. In fact, the pericentromeric chromatin in each *Arabidopsis* chromosome can be considered as a ‘megaTAD’ (>5Mb), which is gene-poor and heavily decorated with heterochromatic marks. Another possible factor explaining the absence of TADs in *Arabidopsis* might be its smooth transcription density pattern in the linear genome. Rowley and colleagues reported a strong correlation between high transcription density and the occurrence of TAD borders in different species⁹⁵. Interestingly, *Arabidopsis* chromosome arms display a uniform transcription density. However, a few TADs can be found inside the pericentromeric regions, and their borders overlap with sparse transcriptionally active loci.

Chromatin structure at the gene-body level. The formation of a chromatin loop allows physical interaction between a regulatory element and its target gene regardless of their distance in the genome. A large number of case studies in animals have demonstrated the importance of having proper chromatin loop formation for precise transcriptional control. Accordingly, recent work demonstrated the involvement of various plant chromatin loops in diverse developmental processes (reviewed in refs. ^{84,105}). In addition, analyses of a high-resolution *Arabidopsis* Hi-C map revealed the presence of many chromatin loops connecting the 5′ end of a gene to the corresponding gene body¹⁰⁵ resembling the ‘gene globule’ model reported in yeast¹⁰⁶. *Arabidopsis* genes with such self-looping structures tend to be more actively expressed than those without¹⁰⁵. It is not clear if a causal relationship exists between gene loops and active transcription.

An additional chromatin feature, called R-loops, has been described in plants. R-loops consist of one-stranded DNA and a DNA:RNA hybrid, and are functional structural units of chromatin with wide distribution in the *Arabidopsis* genome¹⁰⁷. Pattern analyses of the distribution of *Arabidopsis* R-loops suggest that they play diverse roles in genome organization and gene regulation¹⁰⁷. For example, an R-loop structure at the *SEPALLATA3* locus was recently shown to regulate splicing of the cognate mRNA (ref. ¹⁰⁸).

Certainly, nucleosome positioning and occupancy, dynamics of histone variants, post-translational modifications of histones and DNA methylation are all crucial factors influencing plant chromatin structure and accessibility, as well as the expression of individual genes (reviewed in refs ^{109–112}). Due to space constraints and the focus of this work, however, how these factors are linked with local chromatin structure is not discussed here.

Chromatin positioning in the nuclear space

A growing body of evidence shows that gene expression and transcriptional activity is not only linked with local chromatin modifications (for example, chromatin accessibility and epigenetic marks), but also the spatial localization within the nucleus (reviewed in refs ^{113–115}). Different from the previous section that covered various aspects of plant chromatin organizations, discussing how chromatin conformation affects itself (for example, information derived from Hi-C), this section focuses on recent progress in understanding plant chromatin positioning in the nuclear space, analysing chromatin organizations with respect to different types of nuclear compartments. Here, we discuss the specific association of plant chromatin with two distinct regions in the nucleus: the nucleolus and nuclear periphery.

Nucleolus. As a prominent organelle in the nuclear space, the nucleolus has long been well known as a central hub of processing ribosome biogenesis. Nucleoli also play roles in the 3D organization of chromatin in the nucleus by serving as docking sites for certain chromatin, such as ribosomal RNA (rRNA) genes¹¹⁶. In plants, nucleoli are strongly associated with rRNA genes and the spatial localisation of these genes with respect to nucleoli correlates to their transcriptional status^{117,118}. Furthermore, *Arabidopsis* telomeres physically associate with nucleoli, which are required for telomere maintenance^{46,119,120}.

In addition to rRNAs, other chromatin regions termed NADs (nucleolus-associated domains) have been identified using genome-wide approaches in mammals^{121,122}. These NADs are enriched with chromatin containing AT-rich sequence elements, low gene density and transcriptionally repressed genes¹²². A recent study reporting genome-wide identification of NADs in *Arabidopsis* has largely broadened our view on the role of plant nucleoli in chromatin organization¹⁹. Pontvianne and colleagues isolated intact nucleoli from *Arabidopsis* seedlings with a fluorescence-activated cell sorting-based approach and analysed the associated DNA. In addition to rRNA genes and telomeric regions, *Arabidopsis* NADs were also highly enriched with transposable elements and inactive protein-coding genes¹⁹. These findings suggest that plant nucleoli are involved in transcriptional regulation of many other genomic loci in addition to rRNA genes. Proteome analysis of plant nucleoli might further expand our understanding of the nucleoli's role by determining the identities of unknown proteins^{123,124}.

Nuclear periphery. The nuclear periphery (NP) not only serves as a physical barrier separating nuclear content from the cytoplasm, but also plays critical roles in modulating 3D chromatin structure. At the NP, the nuclear lamina (consisting of lamins and other interacting proteins) is a meshwork layer beneath the nuclear envelope. In animals, nuclear lamins selectively tether lamina-associated domains (LADs), most of which are heterochromatic, to the NP to modulate gene expression (reviewed in refs ^{113,125}). In contrast, research of plant LADs is rather limited due to lack of knowledge of the lamin counterparts in plants¹²⁶. Over the past few years, a group of plant-specific proteins named CROWDED NUCLEI (CRWN) have emerged as 'plant lamina' components^{127,128}. In addition, with several newly discovered plant-specific proteins, such as KAKU4 (ref. ¹²⁹) and NEAPs (ref. ¹³⁰), the list of plant lamina candidates has expanded. Apart from the nuclear lamina, nuclear pore com-

plexes (NPCs) at the NP have been shown to participate in regulating gene expression by interacting with chromatin (reviewed in refs ^{131,132}). Plant NPCs are hypothesized to have comparable functions in transcriptional regulation to NPCs in animals^{133,134}. This is supported by a recent study showing that tethering genes artificially to plant NPCs can influence gene expression¹³⁵. Additionally, an *Arabidopsis* NPC member, NUP85, was reported to modulate plants' response to abscisic acid and salt stress, by interacting with several mediator subunits required for gene transcription by RNA polymerase II (ref. ¹³⁶).

Similarly to animals, the plant NP has been shown to be a docking station primarily for repressed chromatin. In *Arabidopsis*, heterochromatic CCs are preferentially located at the NP (refs ^{46,137}). In addition to CCs, many repressed loci in *Arabidopsis* chromosome arms are also tethered specifically to the NP. As reported in a recent study, *Arabidopsis* LADs are enriched with silenced protein-coding genes, transposable element genes and heterochromatic marks¹³⁸. Interestingly, in many cases where CC structures are disrupted, such as during mesophyll cell dedifferentiation, in response to low light or with mutations causing heterochromatin decondensation, chromatin regions belonging to CCs still show an NP-enriched distribution pattern^{98,139–141}. This implies that local genomic and epigenomic features of pericentromeric chromatin can contribute NP tethering, regardless the formation of CCs. In animals, methylation marks on the histone H3K9 residue guide chromatin tethering to the NP (refs ^{142–144}). It would be interesting to test in plants if this histone mark, or other heterochromatic marks, are involved in this process. Tethering chromatin to the NP in animals also requires lamins^{145,146}, and emerging evidence suggests that plant lamins play a similar role. In *Arabidopsis*, loss of function in CRWNs causes pleiotropic changes in nuclear morphology and genome organization, one of which causes attenuated association between CCs and the NP (refs ^{147–150}). Additionally, the CC–NP association also requires members of the linker of nucleoskeleton and cytoskeleton (LINC) complex¹⁴⁸.

It should be noted that specific chromatin positioning at the plant NP might also be linked to active transcription. Artificial recruitment of a reporter gene to the NP with an NPC member causes transcriptional upregulation¹³⁵. In another example, the *Arabidopsis* *CAB* gene-cluster locus and several other light-inducible genes, including *RBCS1A*, *PC* and *GUN5*, relocated from the nuclear interior towards the NP, along with transcriptional activation, in response to light stimuli¹⁵¹. It seems that transcriptional regulation at the plant NP is more complex than simply be either suppressing or promoting gene expression. Future work aiming to identify potential transcription factors, chromatin remodelling factors and other chromatin-associated factors that interact with the NP will be critical to solving this puzzle.

In conclusion, plant chromatin organization at the local scale turns out to be highly variable (for example, having or not having TADs), and characterizing local chromatin structures in detail in more plant species will provide useful insights into understanding how these differences occur. On the other hand, similar to animals, plant chromatin organization at the chromosomal scale seems to be driven by the euchromatic and heterochromatic landscapes of the genome, which give rise to patterns such as A/B chromatin compartmentalization.

Future directions

Understanding how plant genomes function in 3D is an exciting question that attracts many plant scientists. At the moment, much of the work concerning plant 3D genomes purely describes structural patterns and correlations. We envisage that in the near future, integrating various plant 3D chromatin structural features into research addressing fundamental biological questions of plant growth and development will be the main theme. Nevertheless, we

REVIEW ARTICLE

NATURE PLANTS

still lack a comprehensive view of how plants' 3D genomes are built and how plant chromatin packing patterns interact with different genomic and epigenomic features. Thanks to the development of various techniques, nowadays we can easily adapt many of them to plant materials. We expect that over the next few years there will be a wave of datasets (especially for crop species) describing plant epigenomes and transcriptomes in the 3D context, serving as crucial resources to identify key regulators of plant chromatin folding and positioning.

In our view, the discovery of plant TADs has at least three immediate implications. First, from a structural point of view, a TAD-containing genome possesses a stronger extent of spatial chromatin compartmentalization than those without TADs (such as *Arabidopsis*). This spatial property can promote the 'directionality' and 'specificity' of a genomic locus when it interacts with distant neighbouring regions. Hence, observing a plant genome with TADs would substantiate the demand of identifying remote cis-regulatory elements to achieve a better understanding of gene regulation¹⁵². Secondly, detailed functional studies on plant TAD boundaries might help us understand how plants achieve chromatin insulation, which remains largely unknown. Our recent attempt at analysing sequences enriched around rice TAD boundaries revealed a motif recognized by plant-specific TCP transcription factors (named after TEOSINTE-LIKE1, CYCLOIDEA, and PROLIFERATING CELL FACTOR1)⁵¹. In addition, further motif analyses imply that plant bZIP proteins can potentially contribute to rice TAD formation as well (Supplementary Fig. 1). Thirdly, on increasing the sequencing depth, thousands of long-range chromatin loops that connect TAD borders can be spotted from 1 kb resolution mammalian Hi-C maps¹⁵³. So far, similar high-resolution Hi-C maps have not been established in plants. To unveil details of how plant TAD borders interact, as discussed by Dong and colleagues, it is worth increasing sequencing depth specifically in these regions⁵⁸.

In addition to transcription factors, here we highlight another two emerging groups of factors involved in shaping plant 3D genomes: polycomb repressive complexes (PRCs) and RNAs. PRCs establish H3K27me3 patterns across the genome. Hi-C analyses in *Arabidopsis* reveal a strong association of the H3K27me3 mark with long-range chromatin interactions^{6,4,65,99,105,154,155}. In animals, removal of polycomb repression causes not only transcriptional dysregulation, but also an altered chromatin interaction network, in which physical chromatin interactions among H3K27me3-marked loci are attenuated at a genome-wide scale (recently reviewed in ref. ¹⁵⁶). Physical clustering of H3K27me3-marked chromatin, correlated with gene silencing, has been showcased at specific loci in *Arabidopsis*⁵. Considering the potential role of the plant PRCs in shaping chromatin folding, it would be interesting to pursue this direction to unveil mechanisms of PRCs in crop species, which show rather different chromatin organization patterns to that of *Arabidopsis*.

The effect of RNA molecules in scaffolding genome organization has also received increasing attention^{157,158}. The work of genome-wide characterization of R-loops in *Arabidopsis* might reflect a possible wide occurrence of chromatin-RNA interactions across plant genomes¹⁰⁷. Drastic changes in heterochromatin organization have been documented in a wide range of plant growth, development and stress response processes (reviewed in refs ^{2,159}). During such chromatin rewiring events, we expect an intense reprogramming of RNA transcriptional regulation (with the majority at non-coding RNA loci in heterochromatin). As proposed recently in a 'cat's cradling' model¹⁶⁰, are local RNA transcriptional profile changes part of the driving forces that determine how heterochromatin is reorganized? Furthermore, which RNAs act as trans-regulatory elements to modulate chromatin structures? What are the identities of the chromatin remodelling factors and/or other structural proteins involved? Answering these questions will greatly advance our knowledge of

nuclear structure, genome architecture and gene regulation, paving the way towards understanding 3D plant genomes.

Code availability. Scripts generating plots in Fig. 2 and Supplementary Figure 1 are available on request.

Received: 7 December 2017; Accepted: 11 June 2018;
Published online: 30 July 2018

References

- Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
- Probst, A. V. & Mittelsten Scheid, O. Stress-induced structural changes in plant chromatin. *Curr. Opin. Plant Biol.* **27**, 8–16 (2015).
- Rosa, S. et al. Physical clustering of FLC alleles during Polycomb-mediated epigenetic silencing in vernalization. *Genes Dev.* **27**, 1845–1850 (2013).
- Paweletz, N. Walther Flemming: pioneer of mitosis research. *Nat. Rev. Mol. Cell Biol.* **2**, 72–75 (2001).
- Flemming, W. Zellsubstanz, kern und zelltheilung. (Vogel, Leipzig, 1882).
- Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).
- Grob, S. & Grossniklaus, U. Chromosome conformation capture-based studies reveal novel features of plant nuclear architecture. *Curr. Opin. Plant Biol.* **36**, 149–157 (2017).
- Liu, C. & Weigel, D. Chromatin in 3D: progress and prospects for plants. *Genome Biol.* **16**, 170 (2015).
- Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
- Zhao, Z. et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341–1347 (2006).
- Dostie, J. et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
- Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **4**, 895–901 (2007).
- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Grob, S. & Cavalli, G. Technical review: a hitchhiker's guide to chromosome conformation capture. *Methods Mol. Biol.* **1675**, 233–246 (2018).
- Sotelo-Silveira, M., Chávez Montes, R. A., Sotelo-Silveira, J. R., Marsch-Martínez, N. & de Folter, S. Entering the next dimension: plant genomes in 3D. *Trends Plant Sci.* <http://doi.org/cqwt> (2018).
- Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
- Li, G. et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* **11**, R22 (2010).
- Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
- Jäger, R. et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* **6**, 6178 (2015).
- Li, G. et al. Chromatin interaction analysis with paired-end tag (ChIA-PET) sequencing technology and application. *BMC Genom.* **15**, S11 (2014).
- Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
- Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
- Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
- Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
- Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
- Dekker, J. et al. The 4D nucleome project. *Nature* **549**, 219–226 (2017).
- Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
- Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biol.* **17**, 215 (2016).
- Fudenberg, G. & Imakaev, M. FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat. Methods* **14**, 673–678 (2017).
- Williamson, I. et al. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence *in situ* hybridization. *Genes Dev.* **28**, 2778–2791 (2014).
- Cui, C., Shu, W. & Li, P. Fluorescence *in situ* hybridization: cell-based genetic diagnostic and research applications. *Front. Cell Dev. Biol.* **4**, 89 (2016).

32. Solovei, I. et al. Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence *in situ* hybridization (3D-FISH). *Exp. Cell Res.* **276**, 10–23 (2002).
33. Cremer, M. et al. Multicolor 3D fluorescence *in situ* hybridization for imaging interphase chromosomes. *Methods Mol. Biol.* **463**, 205–239 (2008).
34. Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
35. Beliveau, B. J. et al. Single-molecule super-resolution imaging of chromosomes and *in situ* haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).
36. Ni, Y. et al. Super-resolution imaging of a 2.5 kb non-repetitive DNA *in situ* in the nuclear genome using molecular beacon probes. *eLife* **6**, e21660 (2017).
37. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
38. Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
39. Qin, P. et al. Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nat. Commun.* **8**, 14725 (2017).
40. Ye, H., Rong, Z. & Lin, Y. Live cell imaging of genomic loci using dCas9-SunTag system and a bright fluorescent protein. *Protein Cell* **8**, 853–855 (2017).
41. Ma, Y. et al. Live cell imaging of single genomic loci with quantum dot-labeled TALEs. *Nat. Commun.* **8**, 15318 (2017).
42. Hong, Y., Lu, G., Duan, J., Liu, W. & Zhang, Y. Comparison and optimization of CRISPR/dCas9/gRNA genome-labeling systems for live cell imaging. *Genome Biol.* **19**, 39 (2018).
43. Kind, J. et al. Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178–192 (2013).
44. Yu, M. & Ren, B. The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289 (2017).
45. Pecinka, A. et al. Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. *Chromosoma* **113**, 258–269 (2004).
46. Franz, P., De Jong, J. H., Lysak, M., Castiglione, M. R. & Schubert, I. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl Acad. Sci. USA* **99**, 14584–14589 (2002).
47. Tiang, C.-L., He, Y. & Pawlowski, W. P. Chromosome organization and dynamics during interphase, mitosis, and meiosis in plants. *Plant Physiol.* **158**, 26–34 (2012).
48. Rodriguez-Granados, N. Y. et al. Put your 3D glasses on: plant chromatin is on show. *J. Exp. Bot.* **67**, 3205–3221 (2016).
49. Dong, F. & Jiang, J. Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. *Chromosome Res.* **6**, 551–558 (1998).
50. Prieto, P., Santos, A. P., Moore, G. & Shaw, P. Chromosomes associate premeiotically and in xylem vessel cells via their telomeres and centromeres in diploid rice (*Oryza sativa*). *Chromosoma* **112**, 300–307 (2004).
51. Liu, C., Cheng, Y.-J., Wang, J.-W. & Weigel, D. Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*. *Nat. Plants* **3**, 742–748 (2017).
52. Bass, H. W. et al. Evidence for the coincident initiation of homolog pairing and synapsis during the telomere-clustering (bouquet) stage of meiotic prophase I. *J. Cell Sci.* **113**, 1033–1042 (2000).
53. Schwarzscher, T. Three stages of meiotic homologous chromosome pairing in wheat: cognition, alignment and synapsis. *Sex. Plant Reprod.* **10**, 324–331 (1997).
54. Zhang, F. et al. The F-box protein ZYG1 mediates Bouquet formation to promote homologous pairing, synapsis, and recombination in rice meiosis. *Plant Cell* **29**, 2597–2609 (2017).
55. Bourbousse, C. et al. Light signaling controls nuclear architecture reorganization during seedling establishment. *Proc. Natl Acad. Sci. USA* **112**, 2836–2844 (2015).
56. Pecinka, A. et al. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*. *Plant Cell* **22**, 3118–3129 (2010).
57. Wang, L.-C., Wu, J.-R., Hsu, Y.-J. & Wu, S.-J. *Arabidopsis* HIT4, a regulator involved in heat-triggered reorganization of chromatin and release of transcriptional gene silencing, relocates from chromocenters to the nucleolus in response to heat stress. *New Phytol.* **205**, 544–554 (2015).
58. Dong, P. et al. 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* **10**, 1497–1509 (2017).
59. Grob, S., Schmid, M. W. & Grossniklaus, U. Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol. Cell* **55**, 678–693 (2014).
60. Wang, J. et al. Genome-wide analysis of the distinct types of chromatin interactions in *Arabidopsis thaliana*. *Plant Cell Physiol.* **58**, 57–70 (2017).
61. Wang, M. et al. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat. Plants* **4**, 90–97 (2018).
62. Grob, S., Schmid, M. W., Luedtke, N. W., Wicker, T. & Grossniklaus, U. Characterization of chromosomal architecture in *Arabidopsis* by chromosome conformation capture. *Genome Biol.* **14**, R129 (2013).
63. Concia, L. et al. Genome-wide analysis of the *Arabidopsis thaliana* replication timing program. *Plant Physiol.* **176**, 2166–2185 (2018).
64. Feng, S. et al. Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in *Arabidopsis*. *Mol. Cell* **55**, 694–707 (2014).
65. Zhu, W. et al. Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific *Arabidopsis* hybrid. *Genome Biol.* **18**, 157 (2017).
66. Szabo, Q. et al. TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Sci. Adv.* **4**, eaar8082 (2018).
67. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
68. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
69. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
70. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
71. Crane, E. et al. Condensin-driven remodeling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
72. Mizuguchi, T. et al. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* **516**, 432–435 (2014).
73. Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**, 731–734 (2013).
74. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
75. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
76. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228 (2017).
77. Naumova, N. et al. Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).
78. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
79. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
80. Zain, J. et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111**, 996–1001 (2014).
81. Sofueva, S. et al. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–3129 (2013).
82. de Wit, E. et al. CTCF binding polarity determines chromatin looping. *Mol. Cell* **60**, 676–684 (2015).
83. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
84. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
85. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, 6456–6465 (2015).
86. Terakawa, T. et al. The condensin complex is a mechanochemical motor that translocates along DNA. *Science* **358**, 672–676 (2017).
87. Ganji, M. et al. Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).
88. Gibcus, J. H. et al. A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).
89. Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707 (2017).
90. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
91. Busslinger, G. A. et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503–507 (2017).
92. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
93. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
94. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).

95. Rowley, M. J. et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* **67**, 837–852 (2017).
96. Isoda, T. et al. Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. *Cell* **171**, 103–119 (2017).
97. Benedetti, F., Racko, D., Dorier, J., Burnier, Y. & Stasiak, A. Transcription-induced supercoiling explains formation of self-interacting chromatin domains in *S. pombe*. *Nucleic Acids Res.* **45**, 9850–9859 (2017).
98. Moissiard, G. et al. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* **336**, 1448–1451 (2012).
99. Wang, C. et al. Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res.* **25**, 246–256 (2015).
100. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl Acad. Sci. USA* **109**, 17507–17512 (2012).
101. Ulianov, S. V. et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70–84 (2015).
102. Wang, M. et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* **49**, 579–587 (2017).
103. Dong, Q. et al. Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice. *Plant J.* **94**, 1141–1156 (2018).
104. Vietri Rudan, M. et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309 (2015).
105. Liu, C. et al. Genome-wide analysis of chromatin packing in *Arabidopsis thaliana* at single-gene resolution. *Genome Res.* **26**, 1057–1068 (2016).
106. Hsieh, T.-H. S. et al. Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* **162**, 108–119 (2015).
107. Xu, W. et al. The R-loop is a common chromatin feature of the *Arabidopsis* genome. *Nat. Plants* **3**, 704–714 (2017).
108. Conn, V. M. et al. A circRNA from *SEPALLATA3* regulates splicing of its cognate mRNA through R-loop formation. *Nat. Plants* **3**, 17053 (2017).
109. Franz, P. & de Jong, H. From nucleosome to chromosome: a dynamic organization of genetic information: Dynamic organization of genetic information. *Plant J.* **66**, 4–17 (2011).
110. Weber, C. M. & Henkoff, S. Histone variants: dynamic punctuation in transcription. *Genes Dev.* **28**, 672–682 (2014).
111. Vergara, Z. & Gutierrez, C. Emerging roles of chromatin in the maintenance of genome organization and function in plants. *Genome Biol.* **18**, 96 (2017).
112. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
113. van Steensel, B. & Belmont, A. S. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169**, 780–791 (2017).
114. Nguyen, H. Q. & Bosco, G. Gene positioning effects on expression in eukaryotes. *Annu. Rev. Genet.* **49**, 627–646 (2015).
115. Groves, N. R., Biel, A. M., Newman-Griffis, A. H. & Meier, I. Dynamic changes in plant nuclear organization in response to environmental and developmental signals. *Plant Physiol.* **176**, 230–241 (2018).
116. Németh, A. & Längst, G. Genome organization in and around the nucleolus. *Trends Genet.* **27**, 149–156 (2011).
117. Durut, N. et al. A duplicated NUCLEOLIN gene with antagonistic activity is required for chromatin organization of silent 45S rDNA in *Arabidopsis*. *Plant Cell* **26**, 1330–1344 (2014).
118. Pontvianne, F. et al. Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev.* **27**, 1545–1550 (2013).
119. Pontvianne, F. et al. Identification of nucleolus-associated chromatin domains reveals a role for the nucleolus in 3D organization of the *A. thaliana* genome. *Cell Rep.* **16**, 1574–1587 (2016).
120. Armstrong, S. J., Franklin, F. C. & Jones, G. H. Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*. *J. Cell Sci.* **114**, 4207–4217 (2001).
121. Németh, A. et al. Initial genomics of the human nucleolus. *PLoS Genet.* **6**, e1000889 (2010).
122. van Koningsbruggen, S. et al. High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell* **21**, 3735–3748 (2010).
123. Montacié, C. et al. Nucleolar proteome analysis and proteasomal activity assays reveal a link between nucleolus and 26S proteasome in *A. thaliana*. *Front. Plant Sci.* **8**, 1815 (2017).
124. Pendle, A. F. et al. Proteomic analysis of the *Arabidopsis* nucleolus suggests novel nucleolar functions. *Mol. Biol. Cell* **16**, 260–269 (2005).
125. Harr, J. C., Gonzalez-Sandoval, A. & Gasser, S. M. Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. *EMBO Rep.* **17**, 139–155 (2016).
126. Ciska, M. & Moreno Diaz de la Espina, S. The intriguing plant nuclear lamina. *Front. Plant Sci.* **5**, 166 (2014).
127. Zhou, X., Graumann, K. & Meier, I. The plant nuclear envelope as a multifunctional platform LINCed by SUN and KASH. *J. Exp. Bot.* **66**, 1649–1659 (2015).
128. Meier, I., Richards, E. J. & Evans, D. E. Cell biology of the plant nucleus. *Annu. Rev. Plant Biol.* **68**, 139–172 (2017).
129. Goto, C., Tamura, K., Fukao, Y., Shimada, T. & Hara-Nishimura, I. The novel nuclear envelope protein KAKU4 modulates nuclear morphology in *Arabidopsis*. *Plant Cell* **26**, 2143–2155 (2014).
130. Pawar, V. et al. A novel family of plant nuclear envelope-associated proteins. *J. Exp. Bot.* **67**, 5699–5710 (2016).
131. Ibarra, A. & Hetzer, M. W. Nuclear pore proteins and the control of genome functions. *Genes Dev.* **29**, 337–349 (2015).
132. Strambio-De-Castilla, C., Niepel, M. & Rout, M. P. The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 490–501 (2010).
133. Parry, G. The plant nuclear envelope and regulation of gene expression. *J. Exp. Bot.* **66**, 1673–1685 (2015).
134. Yang, Y., Wang, W., Chu, Z., Zhu, J.-K. & Zhang, H. Roles of nuclear pores and nucleo-cytoplasmic trafficking in plant stress responses. *Front. Plant Sci.* **8**, 574 (2017).
135. Smith, S. et al. Marker gene tethering by nucleoporins affects gene expression in plants. *Nucleus* **6**, 471–478 (2015).
136. Zhu, Y. et al. An *Arabidopsis* nucleoporin NUP85 modulates plant responses to ABA and salt stress. *PLoS Genet.* **13**, e1007124 (2017).
137. Fang, Y. & Spector, D. L. Centromere positioning and dynamics in living *Arabidopsis* plants. *Mol. Biol. Cell* **16**, 5710–5718 (2005).
138. Bi, X. et al. Nonrandom domain organization of the *Arabidopsis* genome at the nuclear periphery. *Genome Res.* **27**, 1162–1173 (2017).
139. Tessoro, F. et al. Large-scale dissociation and sequential reassembly of pericentric heterochromatin in dedifferentiated *Arabidopsis* cells. *J. Cell Sci.* **120**, 1200–1208 (2007).
140. van Zanten, M. et al. Photoreceptors CRYTOCHROME2 and phytochrome B control chromatin compaction in *Arabidopsis*. *Plant Physiol.* **154**, 1686–1696 (2010).
141. Soppe, W. J. J. et al. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J.* **21**, 6549–6559 (2002).
142. Towbin, B. D. et al. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**, 934–947 (2012).
143. Bian, Q., Khanna, N., Alvikas, J. & Belmont, A. S. β -Globin cis-elements determine differential nuclear targeting through epigenetic modifications. *J. Cell Biol.* **203**, 767–783 (2013).
144. Gonzalez-Sandoval, A. et al. Perinuclear anchoring of H3K9-methylated chromatin stabilizes induced cell fate in *C. elegans* embryos. *Cell* **163**, 1333–1347 (2015).
145. Harr, J. C. et al. Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J. Cell Biol.* **208**, 33–52 (2015).
146. Solovei, I. et al. LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152**, 584–598 (2013).
147. Wang, H., Dittmer, T. A. & Richards, E. J. *Arabidopsis* CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. *BMC Plant Biol.* **13**, 200 (2013).
148. Poulet, A. et al. The LINC complex contributes to heterochromatin organisation and transcriptional gene silencing in plants. *J. Cell Sci.* **130**, 590–601 (2017).
149. Dittmer, T. A., Stacey, N. J., Sugimoto-Shirasu, K. & Richards, E. J. LITTLE NUCLEI genes affecting nuclear morphology in *Arabidopsis thaliana*. *Plant Cell* **19**, 2793–2803 (2007).
150. Sakamoto, Y. & Takagi, S. LITTLE NUCLEI 1 and 4 regulate nuclear morphology in *Arabidopsis thaliana*. *Plant Cell Physiol.* **54**, 622–633 (2013).
151. Feng, C.-M., Qiu, Y., Van Buskirk, E. K., Yang, E. J. & Chen, M. Light-regulated gene repositioning in *Arabidopsis*. *Nat. Commun.* **5**, 3027 (2014).
152. Oka, R. et al. Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* **18**, 137 (2017).
153. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
154. Veluchamy, A. et al. LHP1 Regulates H3K27me3 spreading and shapes the three-dimensional conformation of the *Arabidopsis* genome. *PLoS ONE* **11**, e0158936 (2016).
155. Chica, C., Louis, A., Roest Crolius, H., Colot, V. & Roudier, F. Comparative epigenomics in the Brassicaceae reveals two evolutionarily conserved modes of PRC2-mediated gene regulation. *Genome Biol.* **18**, 207 (2017).
156. Xiong, J., Zhang, Z. & Zhu, B. Polycomb 'polypacks' the chromatin. *Proc. Natl Acad. Sci. USA* **113**, 14878–14880 (2016).

157. Hall, L. L. & Lawrence, J. B. RNA as a fundamental component of interphase chromosomes: could repeats prove key? *Curr. Opin. Genet. Dev.* **37**, 137–147 (2016).
158. Marchese, F. P., Raimondi, I. & Huarte, M. The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* **18**, 206 (2017).
159. Benoit, M., Layat, E., Tourmente, S. & Probst, A. V. Heterochromatin dynamics during developmental transitions in *Arabidopsis* - a focus on ribosomal DNA loci. *Gene* **526**, 39–45 (2013).
160. Melé, M. & Rinn, J. L. 'Cat's Cradling' the 3D genome by the act of lncRNA transcription. *Mol. Cell* **62**, 657–664 (2016).
161. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
162. Wu, J. et al. Physical maps and recombination frequency of six rice chromosomes. *Plant J.* **36**, 720–730 (2003).

Acknowledgements

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement

No. 757600). We apologize to our colleagues whose work was not included or discussed sufficiently in this manuscript due to space constraints.

Author contributions

E.D. and C.L. wrote the manuscript.

Competing interests

The authors declare no competing interests.

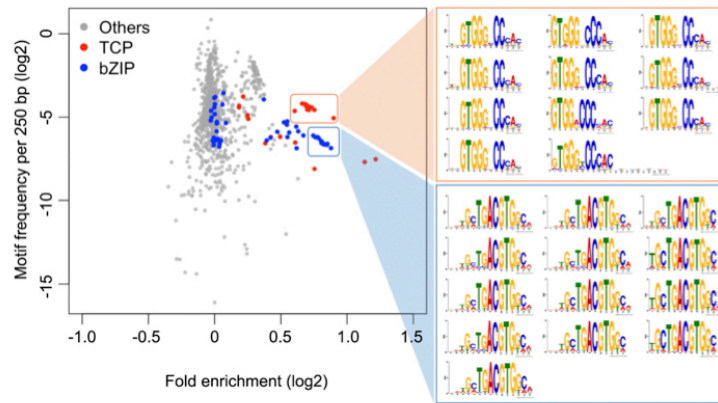
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41477-018-0199-5>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to C.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Supplementary Fig. 1 Motif enrichment analysis of rice TAD borders. Scatter plot shows enrichment of motifs, which are recognized by various plant transcription factors (position weight matrices are according to the *Arabidopsis* DAP-seq dataset¹⁶³), at rice TAD boundary regions. The fold enrichment of a motif is calculated as the relative density of this motif in a 5 kb region overlapping with TAD borders compared to that in 100 kb region flanking TAD borders. The presence of motifs in query DNA sequences was determined by the `matchPWM` function in the 'Biostrings' package in R (ref. ¹⁶⁴), with the search stringency set to 85%. To assess the statistical significance of the fold enrichment of a selected motif, a Monte Carlo procedure with 1000 simulations was performed. In each simulation, a pool of 2000 5 kb regions, together with their 100 kb flanking regions, were picked up randomly as 'TAD borders', which were subsequently scanned for the motif. All the dots with motifs shown on right have empirical p-values less than 0.001.

Supplementary References

163. O'Malley, R. C. et al. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **166**, 1598 (2016).
164. Pages, H., Abouyou, P., Gentleman, R. & DeRoy, S. Biostrings: String objects representing biological sequences, and matching algorithms. v.2.24.1 (R package, 2012).

3 Marchantia TCP transcription factor activity correlates with 3D chromatin structure

Preamble

This chapter is published in the journal of Nature Plants:

Ezgi Süheyla Karaaslan, Nan Wang, Natalie Faiß, Yuyu Liang, Sean A. Montgomery, Sascha Laubinger, Kenneth Wayne Berendzen, Frédéric Berger, Holger Breuninger and Chang Liu. Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure.

Nature Plants (2020) doi:10.1038/s41477-020-00766-0

C.L. conceived and designed the experiments. **E.S.K.**, N.W., N.F. and H.B. established and characterized transgenic lines. **E.S.K.** performed ChIP-seq, ATAC-seq and RNA-seq experiments. N.W. performed FISH and immunostaining experiments. Y.L. performed coexpression analysis. S.A.M. and F.B. performed epigenomic profiling. K.W.B. performed nuclei sorting. C.L., **E.S.K.** and S.L. performed Hi-C experiments. **E.S.K.** and C.L. wrote the manuscript with contributions from other authors. All authors read and accepted the final version of the manuscript.



Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure

Ezgi Süheyla Karaaslan¹, Nan Wang¹, Natalie Faiß¹, Yuyu Liang¹, Sean A. Montgomery², Sascha Laubinger³, Kenneth Wayne Berendzen¹, Frédéric Berger¹, Holger Breuninger¹ and Chang Liu^{1,4}✉

Information in the genome is not only encoded within sequence or epigenetic modifications, but is also found in how it folds in three-dimensional space. The formation of self-interacting genomic regions, named topologically associated domains (TADs), is known as a key feature of genome organization beyond the nucleosomal level. However, our understanding of the formation and function of TADs in plants is extremely limited. Here we show that the genome of *Marchantia polymorpha*, a member of a basal land plant lineage, exhibits TADs with epigenetic features similar to those of higher plants. By analysing various epigenetic marks across *Marchantia* TADs, we find that these regions generally represent interstitial heterochromatin and their borders are enriched with *Marchantia* transcription factor TCP1. We also identify a type of TAD that we name ‘TCP1-rich TAD’, in which genomic regions are highly accessible and are densely bound by TCP1 proteins. Transcription of TCP1 target genes differs on the basis gene location, and those in TCP1-rich TADs clearly show a lower expression level. In *tcp1* mutant lines, neither TCP1-bound TAD borders nor TCP1-rich TADs display drastically altered chromatin organization patterns, suggesting that, in *Marchantia*, TCP1 is dispensable for TAD formation. However, we find that in *tcp1* mutants, genes residing in TCP1-rich TADs have a greater extent of expression fold change as opposed to genes that do not belong to these TADs. Our results suggest that, besides standing as spatial chromatin-packing modules, plant TADs function as nuclear microcompartments associated with transcription factor activities.

The development of many high-resolution imaging and high-throughput sequencing methods in the last decade has enabled the discovery of key features of three-dimensional (3D) genome organization^{1–3}. Of these, topologically associated domains (TADs), revealed by using high-throughput chromosome conformation capture (Hi-C), stand as the most prominent ones^{4–7}. TADs are architectural genomic interaction components in the genome, which can be found as squares at the diagonal line in a symmetric Hi-C map, or as triangles if only half of the map is shown^{8,9}. A characteristic of animal TADs is that they are insulated chromatin regions whose borders are often occupied by insulator proteins called CTCF^{8,10–13}.

TADs in animal genomes are not only chromatin-packing compartments, but also are involved in regulating critical processes, including DNA replication and gene expression (reviewed in ref. ¹⁴). In mammals, TAD boundaries demarcate the range of enhancer activity, suppressing non-specific inter-TAD chromatin interactions that can potentially lead to enhancer misregulation and aberrant gene expression^{15,16}. In some cases, disrupting TAD structure causes ectopic expression of genes and spread of active and repressed chromatin onto each other^{16,17}. Besides having enhancer–promoter interaction specificity, chromatin regions in the same TAD are also associated with synchronized replication timing^{18,19} and gene expression co-regulation^{8,20–25}.

On the basis of Hi-C analyses of many plant species, plant TADs are not as predominant in the species analysed to date, compared

with animal genomes. TADs are absent in *Arabidopsis thaliana* and its close relative, *Arabidopsis lyrata*; however, two other cruciferous plants, *Brassica rapa* and *Brassica oleracea*, show TADs along their genomes²⁶. Although TADs are absent in *Arabidopsis*, Hi-C analysis revealed heterochromatic small interacting units that are enriched with repressive histone marks, termed ‘TAD interior-like’ regions²⁷. Unlike *Arabidopsis*, genomes of many crop species do possess TADs^{28–32}. This discrepancy between plant species has been speculated to be linked to differences in genome size^{33,34}. *Arabidopsis* has a small genome with high gene density; on the contrary, crop species, such as rice, tomato, maize and wheat, have larger genomes with more evenly distributed, repetitive and longer intergenic regions. These silenced regions located between active genes are likely to display TAD structures^{33,34}. In a recent study by Dong and colleagues on five different crops, plant TADs were classified into four categories according to their distinct epigenetic features: active (accessible chromatin), repressive (enriched in DNA methylation), polycomb silenced (enriched in H3K27me3 mark) and an intermediate type that lacks specific features²⁸. In hexaploid wheat, TADs feature depletion of genes in TAD bodies but enrichment of active genes at TAD boundaries; moreover, genes located at wheat TAD borders tend to form chromatin loops³². Apart from annotating TADs in various plant species, how plant TADs interact with transcription factors is poorly studied. In our previous analyses of sequence motifs associated with rice TADs, those recognized by bZIP and TCP (TEOSINTE BRANCHED 1, CYCLOIDEA and PCF1)

¹Center for Plant Molecular Biology (ZMBP), University of Tübingen, Tübingen, Germany. ²Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Vienna, Austria. ³Institute for Biology and Environmental Science, University of Oldenburg, Oldenburg, Germany. ⁴Institute of Biology, University of Hohenheim, Stuttgart, Germany. ✉e-mail: chang.liu@uni-hohenheim.de

transcription factors were found to be enriched at TAD borders^{29,33}. However, it is unknown whether these plant transcription factors play a role in TAD formation or function.

To gain more insight into the biological importance of plant TADs, we investigated 3D chromatin organization patterns of a recently assembled plant genome, *Marchantia polymorpha*^{35,36}. We chose this plant species for three reasons. First, *Marchantia* has a relatively small genome (~220 megabases (Mb)), allowing the generation of high-resolution Hi-C maps with affordable sequencing costs. Second, comparing Hi-C maps of this non-vascular plant with those of vascular plants may reveal consensus features of chromatin organization in the kingdom Plantae. Third, the well-known low gene redundancy in the *Marchantia* genome can largely facilitate functional studies on regulators of plant chromatin organization.

In this manuscript, we present detailed analyses of TADs at autosomes of the *Marchantia* genome. We show that they can be classified into different types according to their epigenetic status. We report a type of TAD decorated with *Marchantia* TCP1 transcription factors, which we name 'TCP1-rich TADs'. Loss of function of *TCP1*, however, does not lead to drastic alteration in 3D chromatin organization or TCP1-rich TAD structures. Nevertheless, in *tcp1* mutant, genes located in TCP1-rich TADs exhibit larger changes in expression compared to genes located outside these TADs. Our work reveals a type of plant TAD that is heavily loaded with transcription factors, defining functional nuclear compartments that regulate gene expression.

Results

Characterization of TADs in *Marchantia polymorpha*. We generated high-resolution Hi-C maps of male *Marchantia* thalli from two biological replicates (Supplementary Table 1). In a recent study, we showed that at a chromosomal level, the sex and autosomes of *Marchantia* had remarkably distinct features in linear genomic structure, epigenomic landscape and 3D chromosome organization³⁶. Thus, it is noteworthy that all analyses in this work were focused on autosomes. Upon zooming into the diagonals of *Marchantia* Hi-C maps, we found that a considerable fraction of the genome clearly exhibited TADs (Fig. 1a). Visual inspection through all autosomes indicated that *Marchantia* TADs are not directly next to each other, which is a characteristic of animal TADs⁶. Instead, similarly to those in higher plants, there are non-TAD regions between TAD structures. We applied a previously established 'arrowhead' method to scan the Hi-C maps^{15,29} and annotated 4,013 TADs (Supplementary Table 2). These were scattered throughout all chromosomes, collectively covering 40% of the genome (Supplementary Fig. 1). Upon checking the average genomic and epigenomic features across these TADs, we found that they clearly differed from their neighbouring chromatin regions. In general, heterochromatic and euchromatic marks were highly enriched inside and in flanking regions of TADs, respectively (Fig. 1b). Such features of chromatin states at *Marchantia* TADs and TAD borders are similar to TADs annotated in many other plant genomes, and we could also detect similarities to the 'TAD interior-like' and 'TAD border-like' regions in *Arabidopsis thaliana*²⁷. In addition, *Marchantia* TADs were depleted with genes; while their boundaries preferentially overlapped with the transcription start sites (TSSs) or transcription termination sites (TTSs) of genes (Fig. 1c,d). Accordingly, TAD bodies were enriched with various types of repeats and showed a high level of DNA methylation (Fig. 1e and Supplementary Fig. 2). In summary, *Marchantia* TADs were mainly scattered heterochromatin and/or long intergenic regions along chromosomes, and they appeared as self-organized structural modules in the nucleus.

Albeit that they were heterochromatic from their average profile, *Marchantia* TADs were not a homogenous population. By looking at the average DNA methylation ratio in individual TADs, we found that they could be classified into at least two categories, in which members were highly methylated throughout (mCG ratio \approx 0.8) or

methylation-free (mCG ratio \approx 0) (Fig. 1f). To gain more insight into these subcategories of TADs, we arbitrarily annotated them as 'mCG-poor' (mCG ratio lower than 0.2), 'mCG-rich' (mCG ratio higher than 0.6) and 'mCG-intermediate' (mCG ratio between 0.2 and 0.6) (Fig. 1f). As expected, besides being depleted with euchromatic marks, mCG-rich TADs were densely decorated with classic heterochromatin marks (for example, H3K9me1 and H3K27me1) (Extended Data Fig. 1a). By computing interaction decay exponents, which describe how fast chromatin interaction strengths drop with increasing genomic distance, we found that the heterochromatic mCG-rich TADs had more condensed chromatin organization than the other two types (Fig. 1g). Interestingly, mCG-poor TADs showed moderate depletion of active epigenetic marks and enrichment of H3K27me3, suggesting that overall they were not highly active in transcription (Extended Data Fig. 1a). Compared to whole-genome gene expression profile, genes residing inside mCG-rich TADs showed lower expression levels (Extended Data Fig. 1b). Notably, both mCG-rich and mCG-poor TADs showed enrichment of accessible chromatin at their borders, implying that the formation of *Marchantia* TAD boundaries was linked to interactions between boundary regions and unknown chromatin-interacting factors (Extended Data Fig. 1c). As *Marchantia* TADs were not homogeneous, we decided to classify them according to recently published datasets describing various *Marchantia* histone marks³⁶. Using this approach, we classified TADs into eight clusters, among which cluster 4 (enriched for H3K9ac and H2A.Z) and clusters 5 and 6 (enriched for H3K9me1 and H3K27me1) collectively included over 60% of the annotated TADs (Fig. 1h and Extended Data Fig. 1d). These three clusters account for the majority of a TAD group annotated on the basis of DNA methylation level. Taken together, *Marchantia* TADs are a mixed population of individual chromatin regions with distinct epigenetic profiles.

Next, we asked if the TAD layout was associated with gene co-expression, as one would expect to find more frequent gene co-regulation if a gene pair was located in the same TAD that promoted the promoter contacts. In this regard, we generated coexpression matrices by integrating transcriptome data of different *Marchantia* tissues at various developmental stages (Extended Data Fig. 2a, see Methods). Due to physical linkage and the sharing of *cis*-elements^{37–40}, genes abutting each other in the linear genome (separated by less than 5 kilobases (kb)) showed a higher level of positive expression correlation (Extended Data Fig. 2b). Apart from this, we found that TADs contained more coexpressed gene pairs than expected (Extended Data Fig. 2c,d), suggesting that *Marchantia* TAD demarcation is a part of the mechanisms underlying coexpression networks.

TCP1 protein is enriched at many TAD borders but dispensable for TAD formation. In animals, it is well established that the CTCF insulator protein, along with cohesin, contributes TAD formation through interacting with chromatin at TAD boundaries (reviewed recently in refs. ^{11,41,42}). Previously, we showed that TAD borders in the rice genome were enriched with a motif recognized by plant-specific class I TCP transcription factors²⁹. The *Marchantia* genome encodes two *TCP* genes, *TCP1* and *TCP2*, which belong to class I and II clade, respectively^{35,43,44} (Extended Data Fig. 3a). As each *Marchantia* TCP has a highly conserved DNA-binding domain compared to founding members in the same clade^{45,46} (Extended Data Fig. 3b), we expected that they should recognize known consensus DNA sequences identified in higher plants^{47–49}. Thus, hypothesizing that plant TCP proteins are involved in TAD-boundary formation, we examined how the TCP-binding motifs were distributed across *Marchantia* TAD borders. Both analyses with the motif similarity measure or with the text-matching search indicated that sequences recognized by TCP1 were clearly enriched in TAD border regions; whereas those recognized by TCP2 were marginally enriched (Extended Data Fig. 3c–e and Supplementary Table 3).

NATURE PLANTS ARTICLES

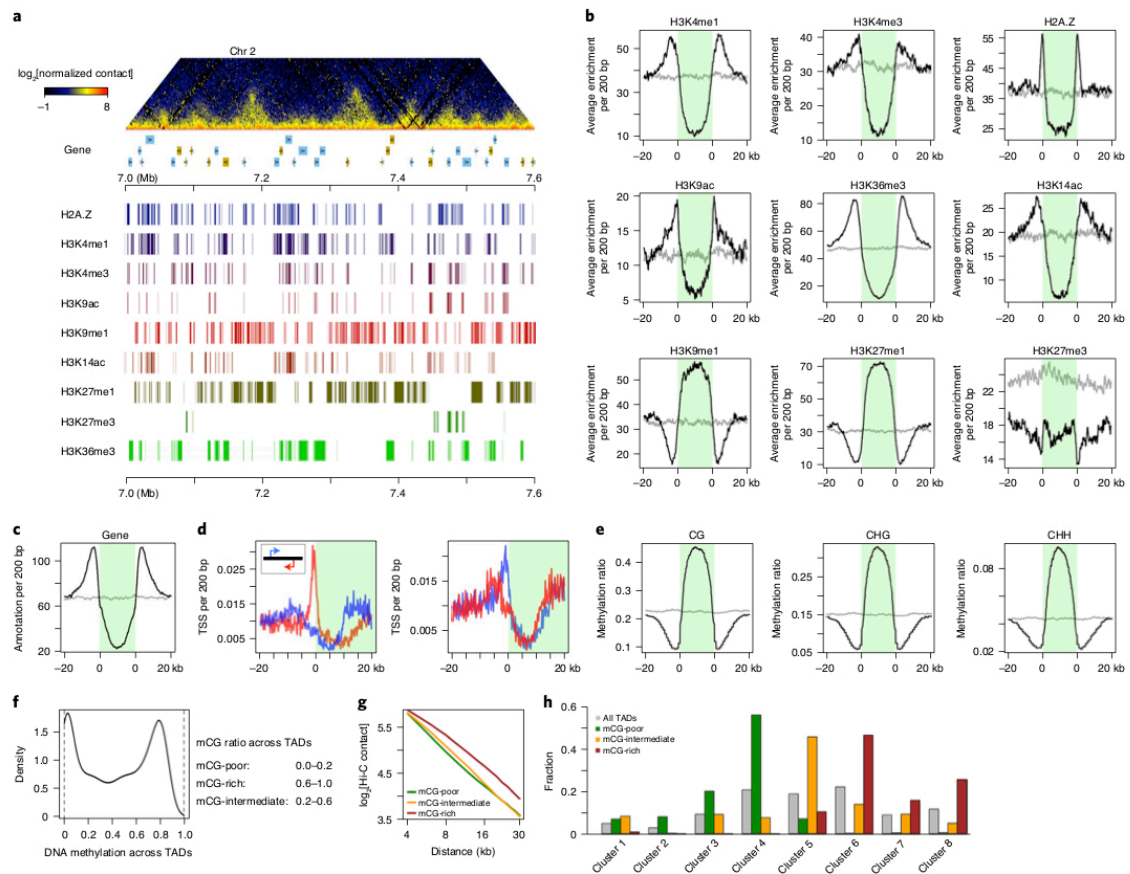


Fig. 1 | Topologically associated chromatin domains at *Marchantia* autosomes. **a**, A snapshot of a Hi-C map of a 600-kb region at chromosome 2. Tracks below the Hi-C map depict gene layout and chromatin domains enriched with various epigenetic marks. **b–e**, Distribution of average epigenetic features across *Marchantia* TADs (green block). **b**, Distribution of histone marks across TADs. **c**, Protein-coding gene annotation around TADs. **d**, Distribution of TSS and TTS around TAD borders. Inset: transcription directions of genes. **e**, Distribution of CG, CHG and CHH methylation across TADs. TADs (black curves) shown in panels **b**, **c** and **e** were linearly transformed to align their borders. Grey curves mean background, which was estimated by distributing TADs randomly throughout the genome. Blue and red curves in **d** depict genes encoded on the Watson and Crick strand, respectively. Both 5' and 3' borders of all TADs are aligned at the point '0'. Note that features at the 3' borders are flipped, so that their 'inside TAD' and 'outside TAD' regions match those of the 5' borders. **f**, Distribution of CG DNA methylation levels of the TADs. **g**, Comparison of chromatin compactness in different types of TADs using their interaction decay exponents. **h**, Classification of TADs according to epigenetic clustering (according to histone marks) and DNA methylation. See Extended Data Fig. 1d for profiles of various histone marks in each cluster.

In addition, expression analyses suggested that *TCP1* was constitutively transcribed in thalli, whereas *TCP2* was not detectable (Supplementary Fig. 3). These data suggested *TCP1* as a potential regulator of TADs, which prompted us to generate a *TCP1* antibody to perform ChIP-seq experiments for examining in vivo *TCP1*-chromatin interactions (Extended Data Fig. 4a). *TCP1* showed extensive interactions with the *Marchantia* genome. In total, about 11,600 regions among autosomes were identified as *TCP1* peaks, which collectively covered around 11.8 Mb (~5%) of the genome (Extended Data Fig. 4 and Supplementary Table 4). Chromatin bound by *TCP1* was highly enriched with H2A.Z but depleted in heterochromatic marks, indicating that this transcription factor did not interact with heterochromatin (Extended Data Fig. 4d). We performed de novo sequence analysis

to reveal motifs enriched in *TCP1* peaks. As expected, of all the motifs identified, the one resembling the canonical sequence motif recognized by class I TCP proteins in higher plants showed the most robust enrichment at the centre of the *TCP1* peak regions (Supplementary Fig. 4 and Supplementary Table 5). Upon examining the *TCP1* peak distribution at TADs, we found that *TCP1* was enriched at TAD borders (Fig. 2a). Among all the 4,013 identified TADs, 1,164 (29%) had a *TCP1* peak nearby (within 1 kb) or overlapping with it; while 499 (12%) had both borders associated with *TCP1*. In addition, more *TCP1* ChIP-seq peaks than expected were associated with TAD borders (Fig. 2b). These observations revealed noticeable overlap between demarcation of *Marchantia* TADs and *TCP1*. Regarding individual genes, *TCP1* could interact with both highly and lowly expressed genes, suggesting a

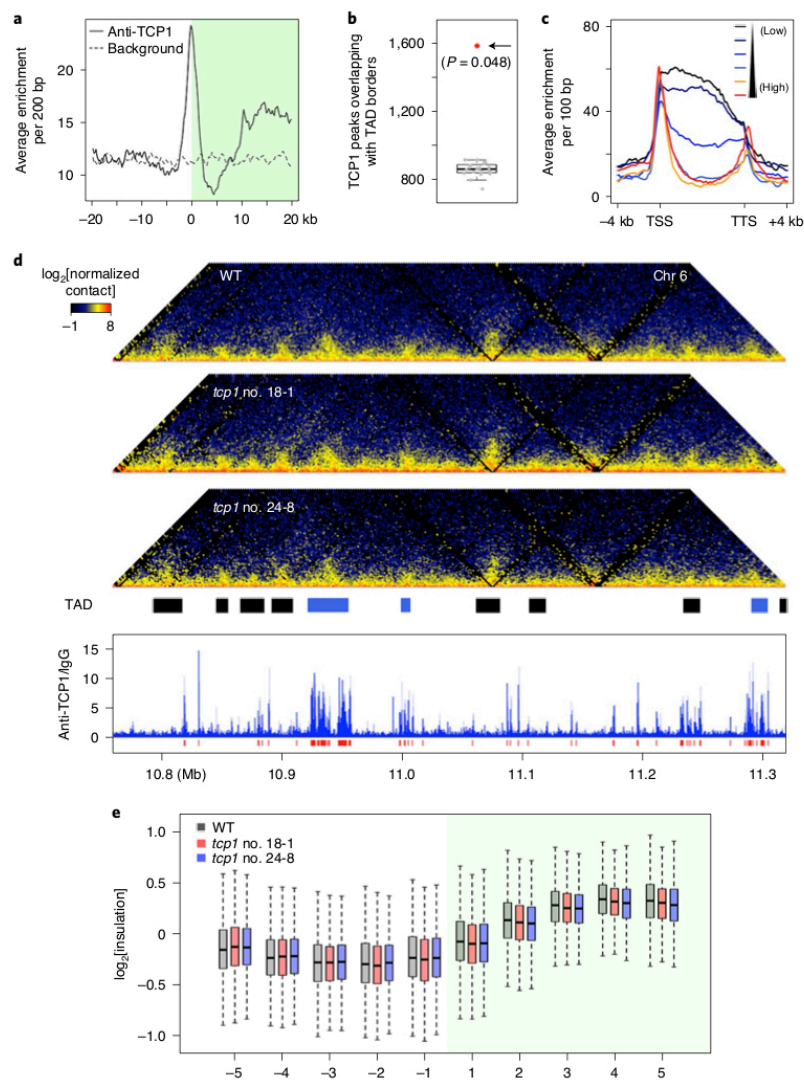


Fig. 2 | *Marchantia* TCP1 is dispensable for TAD patterns. **a**, Metagenome plot of TCP1 ChIP-seq peaks at TAD borders. The green area indicates TADs. The dotted grey curve depicts background, in which TADs are randomly assigned to the genome. **b**, More TCP1 ChIP-seq peaks than expected are associated with TAD borders (2-kb wide). The box plot with 20 grey data points denotes 20 Monte Carlo simulation results in which TADs were randomly assigned to the genome. The red dot, pointed to by an arrow, indicates the observed value. The P value is an empirical P value calculated on the basis of 20 simulations. **c**, Interaction patterns of TCP1 with its target genes grouped according to different expression levels. Genes were divided equally into six groups according to published thallus transcriptome data⁷⁹. In this plot, genes are linearly transformed so that their TSSs and TTSs are aligned. **d**, Comparison of Hi-C maps between wild-type and *tcp1* mutant thalli. Both the black and blue segments indicate TADs, among which the blue ones depict a special type ('TCP1-rich') that are described later in this manuscript. Here, the wild-type (WT) Hi-C map is from an independent experiment to that shown in Fig. 1. The red segments beneath the TCP1 ChIP-seq data depict enriched regions. **e**, Insulation scores of regions flanking TAD borders with TCP1 binding. The green area depicts TADs, and the numbers below the box plots mean bin position relative to TAD boundaries, which are labelled as $\pm 1, \pm 2, \pm 3, \dots$. Only TCP1-bound TAD borders were included in this plot. In this regard, we annotated a TAD border as 'TCP1-bound' if there was at least one TCP1 ChIP-seq peak within ± 1 kb. In total, $n=1,180$ such regions fulfilled this criterion and were included in the box plots. For all groups of box plots, the P values of the comparison between wild type and mutant are larger than 0.01 according to two-sided Mann-Whitney U -tests. The box plots in **b** and **e** indicate the median (line within the box), the lower and upper quartiles (box), margined by the largest and smallest data points that are still within the interval of 1.5 times the interquartile range from the box (whiskers); outliers are not shown.

complex association between TCP1 and gene expression (Fig. 2c). Interestingly, for TCP1 target genes that were moderately to highly expressed, TCP1 primarily bound to their TSSs and/or TTSs; while

for those that had low expression or were silenced, TCP1 tended to interact through their gene bodies (Fig. 2c), implying different modes of transcriptional regulation by TCP1.

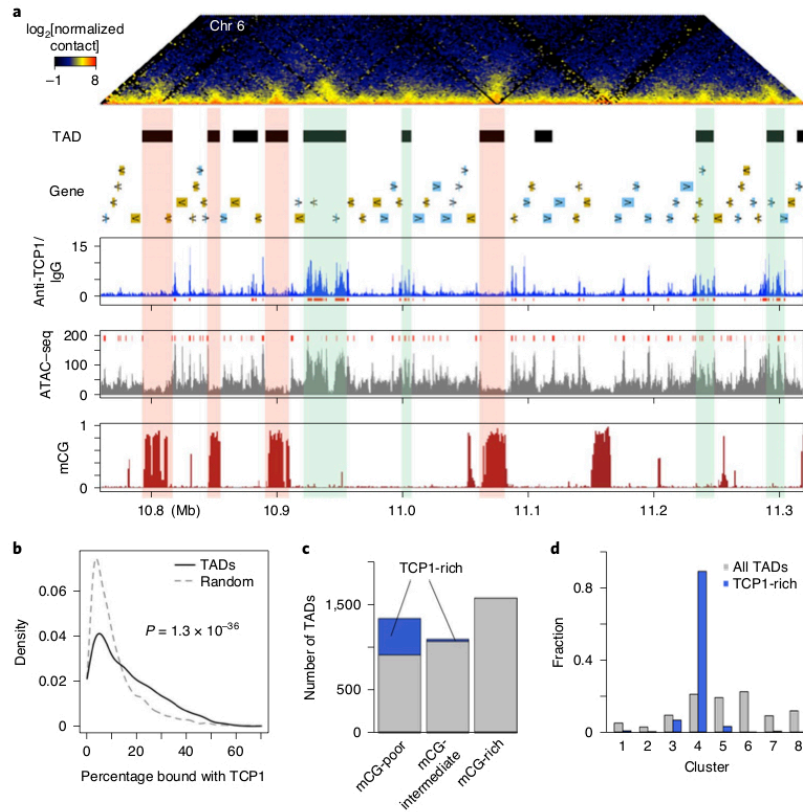


Fig. 3 | Some *Marchantia* TADs having intensive interactions with TCP1. **a**, Different types of TADs revealed by DNA methylation and TCP1 binding. The Hi-C map shows a ~0.5-Mb genomic segment at chromosome 6, below which the black segments indicate annotated TAD regions. Among these TADs, those with conspicuous strong TCP1 binding ('anti-TCP1/IgG', blue track) or DNA methylation in CG sequence context ('mCG', red track), which appear mutually exclusive, are highlighted in green or red, respectively. This Hi-C map is derived from the same dataset shown in Fig. 1. The identified peaks of TCP1 ChIP-seq data and accessible chromatin regions from ATAC-seq data are shown as red segments in each track accordingly. DNA methylation data was from Schmid et al.⁶⁹. **b**, Density plot of the percentage of regions in TADs bound by TCP1. The grey curve depicts background, in which TADs are randomly assigned to the genome. The P value indicates the two-sided Mann-Whitney U -test result. **c**, Distribution of TCP1-rich TADs (blue) among TAD categories defined according to DNA methylation. **d**, Distribution of TCP1-rich TADs among TAD clusters according to histone marks. Details of the epigenetic profiling of each TAD cluster are shown in Extended Data Fig. 1d.

Next, we used CRISPR-Cas9 gene editing to assess the impact of loss of *TCP1* on TAD structure. Consistent with a recent report, our *tcp1* knockout mutants showed a reduced growth rate and gradually developed curly thallus lobes⁵⁰ (Supplementary Fig. 5). All these growth defects could be fully complemented with a 6.8-kb genomic fragment containing the *TCP1* locus (Supplementary Fig. 5). Next, to assess potential roles of TCP1 in 3D genome organization, we generated two Hi-C datasets corresponding to two independent *tcp1* lines and one more Hi-C dataset from Tak-1 plants, which were cultured together with the mutants. At a chromosomal scale, *tcp1* mutant Hi-C maps displayed highly similar A/B compartment patterns to those of wild-type plants, suggesting the absence of marked changes in genome organization (Supplementary Fig. 6). Regarding local chromatin interactions, we examined whether the *tcp1* Hi-C map showed changes in TADs having TCP1 associated with their borders. However, manual inspection of Hi-C maps of two independent *tcp1* mutant lines did not reveal conspicuous changes in their TAD patterns (Fig. 2d). A quantitative comparison of insulation scores⁵¹, which measure the degree of chromatin insulation at

a given genomic region, showed that loss of *TCP1* did not result in a systematic change in insulation scores in these TAD borders (Fig. 2e). Nevertheless, when changes in insulation scores for TCP1-bound TAD borders were compared with those of TCP1-free TAD borders, statistical nuances were found (Extended Data Fig. 5a). However, visual inspection of average chromatin contact patterns around TCP1-bound TAD borders did not reveal any clear alteration in the mutants (Extended Data Fig. 5b). We speculated that changes in the transcription of TCP1-bound genes at TAD borders contributed to the subtle differences in the Hi-C maps, as gene transcription state per se has recently been demonstrated as a strong predictor of Hi-C patterns⁵². Overall, neither of the two *tcp1* mutant lines exhibited drastic changes in TCP1-bound TAD borders, implying that TCP1 is dispensable for TAD structure in *Marchantia*.

Some mCG-poor TAD bodies have intensive interactions with TCP1. Manual inspection of TCP1 ChIP-seq peaks and Hi-C maps revealed a type of TAD that displayed dense TCP1-chromatin interactions (Fig. 3a). Across the *Marchantia* genome, this type of TAD,

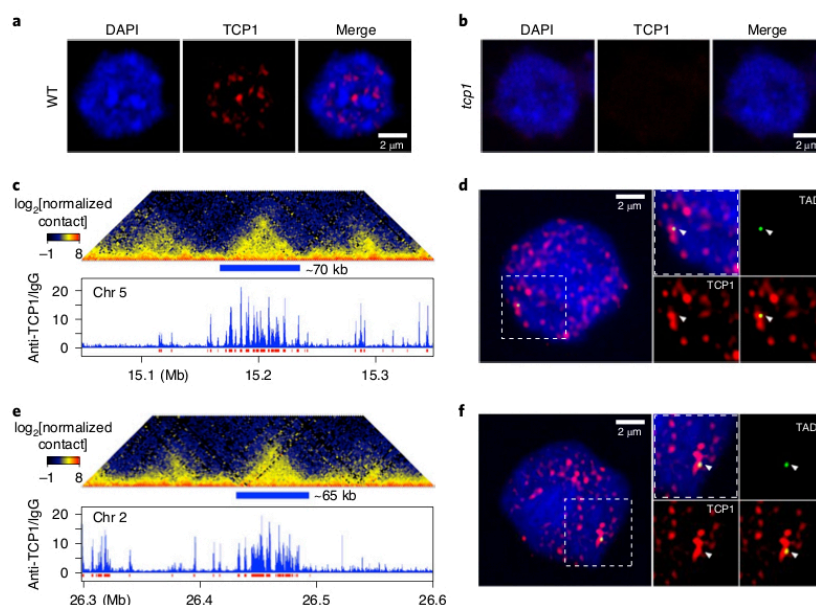


Fig. 4 | TCP1-rich TADs are part of TCP1 protein speckles in the nucleus. a, b, Immunostaining of TCP1 with anti-TCP1 antibodies: wild-type nucleus (Tak-1) (**a**); TCP1 knockout nucleus (**b**). **c–f,** Spatial localization of TCP1 proteins and selected TCP1-rich TADs. The Hi-C maps, TCP1 ChIP-seq signals and combined immunostaining and FISH data are shown for one TCP1-rich TAD at chromosome 5 (**c,d**) and another at chromosome 2 (**e,f**), in which the TAD region marked with a blue segment is labelled for FISH. Arrowheads in **d** and **f** point to FISH signals. Figure labels in **c** and **e** are the same as in Fig. 3a. Note that the Hi-C maps are derived from the same dataset shown in Fig. 1. Images in **a, b, d** and **f** are representatives from two independent experiments with similar patterns.

which showed considerable overlap with TCP1 ChIP-seq peaks, occurred more often than by chance (Fig. 3b). To further characterize these TADs, we arbitrarily named those with at least 20% of a TAD body covered by TCP1 peaks as ‘TCP1-rich’ TADs. In total, 456 TCP1-rich TADs were identified (Supplementary Table 2). Notably, TCP1-rich TADs almost exclusively belonged to the mCG-poor TAD category according to DNA methylation, or to ‘cluster 4’ according to histone marks (Fig. 3c,d). Compared to other TADs, members in cluster 4 had higher levels of H3K9ac and H2A.Z across TAD bodies (Extended Data Fig. 1d). Next, we used immunostaining of TCP1 coupled with FISH (fluorescence in situ hybridization) to examine the spatial localization of TCP1 proteins and TCP1-rich TADs. In the nucleus, TCP1 proteins were not evenly distributed throughout the nucleoplasm; instead, they exhibited a speckled pattern (Fig. 4a,b). In addition, we observed that TCP1-rich TADs were localized in TCP1 protein speckles (Fig. 4c–f). Compared with other members in the mCG-poor TAD category, TCP1-rich TADs had different epigenetic landscapes but comparable chromatin accessibility and other genomic features (Extended Data Fig. 6). In particular, TCP1-rich TADs were depleted with active euchromatin marks, including H3K4me1, H3K4me3 and H3K36me3; however, they were also depleted with H3K27me3, which was a hallmark of silenced protein-coding genes (Extended Data Fig. 6a). These patterns suggest that TCP1-rich TADs define a type of chromatin domain with different transcriptional regulatory regimes in comparison to other TADs.

Approximately 22.6% of TCP1 target regions were located inside TCP1-rich TADs (Fig. 5a). By checking the TCP1 ChIP-seq data across TCP1-rich TADs, we found a sharp transition of TCP–chromatin interaction strength at these TAD borders (Fig. 5b), implying possible feedback between TCP1–chromatin interactions and the

establishment of TCP1-rich TADs. TCP1–chromatin interactions located inside TCP1-rich TADs appeared to be different from those outside. Among all identified TCP1 ChIP-seq peaks, those located in TCP1-rich TADs were much broader, and they were separated by smaller distances (calculated on the basis of the midpoints of peaks) (Fig. 5c,d), suggesting that TCP1-rich TADs defined chromatin domains with enhanced TCP1–chromatin interactions. In both *tcp1* knockout lines, we found a systematic shift of arrowhead scores¹³ (that is, a quantitative measurement to identify TADs on Hi-C maps) of TCP1-rich TADs toward lower values; as a control, such changes were not observed among ‘TCP-free’ TADs belonging to the mCG-poor TADs category (Fig. 5e). However, this statistically significant change in *tcp1* did not manifest structural alteration of TCP1-rich TADs, as their chromatin interaction patterns still highly resembled that of Tak-1 (Fig. 5f). Thus, albeit that it was linked to many TAD borders (Fig. 2) and TCP1-rich TADs (Fig. 5), TCP1 appears to be dispensable for TADs formation in *Marchantia*.

TCP1-rich TADs provide a repressive environment for gene expression. As mentioned above, TCP1 could directly bind to both active and repressed genes (Fig. 2c). Interestingly, TCP1 target genes located in TCP1-rich TADs were significantly less actively expressed than those distributed elsewhere in the genome (Fig. 6a,b). This correlated with our observation that TCP1-rich TADs were depleted with H3K36me3 modification (Extended Data Fig. 6a), which has recently been shown as a histone mark that correlates positively to gene expression¹⁶, suggesting that TCP1 proteins are associated with a repressive environment in TCP1-rich TADs.

Next, we compared the *tcp1* mutant transcriptome to that of wild type to gain insights into how gene expression, and particularly with respect to TCP1-rich TADs, was affected. We found that genes

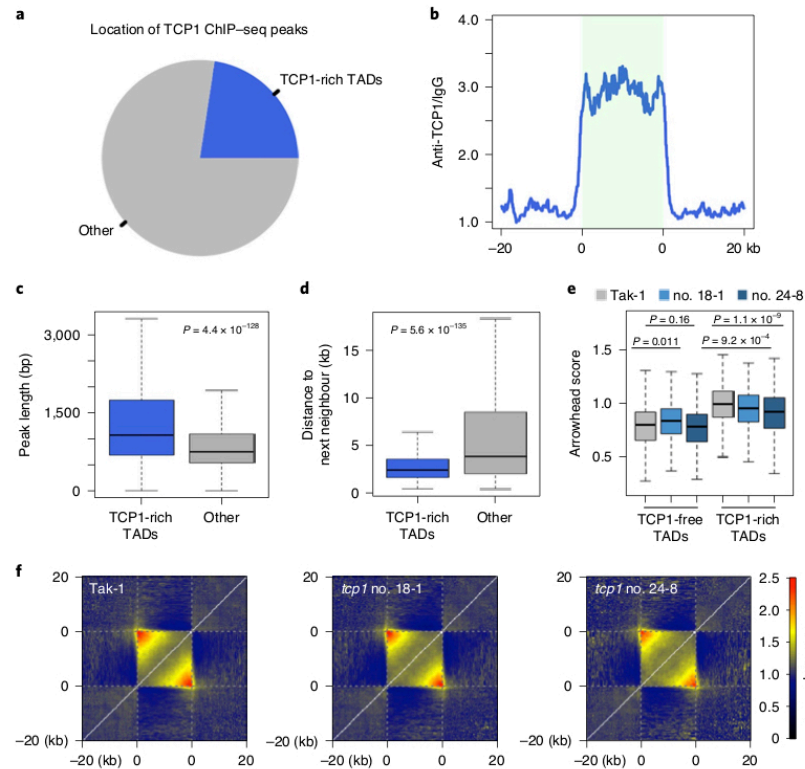


Fig. 5 | Characteristics of TCP1-rich TADs concerning TCP1-chromatin interactions. **a**, Pie chart showing the distribution of TCP1 ChIP-seq peaks. **b**, Profiling of TCP1-chromatin interactions across TCP1-rich TADs. The light green area depicts TADs, borders of which are aligned and marked as '0'. **c,d**, Comparisons of TCP1 ChIP-seq peak width (**c**) and density (**d**). For **d**, the distance between adjacent peaks was calculated as the distance between their midpoints, which was independent from peak width. *P* values in **c** and **d** indicate results of two-sided Mann-Whitney *U*-tests: *n* = 2,610 for 'TCP1-rich TADs'; *n* = 8,940 for 'other'. **e**, Comparison of arrowhead scores of wild-type and *tcp1* mutant TADs. 'TCP-free TADs' refer to the subset of mCG-poor TADs that do not overlap with any TCP1 ChIP-seq peaks. The numbers above box plots denote *P* values of two-sided Mann-Whitney *U*-tests between *tcp1* mutant lines (that is, no. 18-1 and no. 24-8) and Tak-1: *n* = 377 for 'TCP-free TADs'; *n* = 456 for 'TCP-rich TADs'. **f**, No drastic changes in 'TCP1-rich' TADs are found in *tcp1* mutants. These metagenome plots illustrate relative chromatin contact strengths in areas annotated as 'TCP1-rich' TADs plus 40-kb flanking regions. Due to size difference of TADs, linear transformation was applied for aligning their boundaries (labelled as '0'). The box plots in **c-e** indicate the median (line within the box), the lower and upper quartiles (box), margined by the largest and smallest data points that are still within the interval of 1.5 times the interquartile range from the box (whiskers); outliers are not shown.

having extensive interactions with TCP1 (that is, a considerable fraction of a gene body overlapped with TCP1 peaks) tended to become upregulated in *tcp1* (Extended Data Fig. 7a). At a genomic level, TCP1 target genes showed a larger extent in expression change than did genes not bound by TCP1 (Extended Data Fig. 7b). Interestingly, the extent of gene expression change in TCP1 target genes was also associated with the 3D genomic location. Upon grouping TCP1 target genes according to whether or with which type of TAD they overlapped, we found that those residing in TCP1-rich TADs showed the largest variance in gene expression change (Fig. 6c). It should be noted that a large number of TCP1 non-target genes had differential expression in *tcp1* thalli. In fact, the majority of differentially expressed genes in *tcp1* mutants were TCP1 non-targets. Among 1,595 and 760 up- and downregulated genes in the mutants, only 37.9% (605/1,595) and 29.2% (222/760) were bound by TCP1, respectively (Supplementary Table 6), indicating that changes in the *tcp1* transcriptome mainly reflected indirect effects of loss of TCP1. Not only TCP1 target genes in TCP1-rich

TADs, but also TCP1 non-target genes residing in these TADs were affected by the absence of TCP1. For TCP1 non-targets, we found more differentially expressed genes located in TCP1-rich TADs than by chance (Extended Data Fig. 7c,d). Additionally, TCP1 non-target genes overlapping with TCP1-rich TADs showed a larger change in expression than the rest (Fig. 6d). Thus, regardless of whether or not they are bound by TCP1, genes in TCP1-rich TADs had a larger degree of expression change in *tcp1* than genes elsewhere in the genome.

Discussion

In this study, we provide an overall look at the 3D organization of autosomes in *Marchantia*. Our *Marchantia* TADs annotation, like all other TAD identification done on various animal and plant genomes, is solely based on a structural perspective. Although appearing visually identical on the Hi-C maps, different types of *Marchantia* TADs were identified on the basis of their distinct epigenetic profiles (Extended Data Fig. 1), suggesting that they

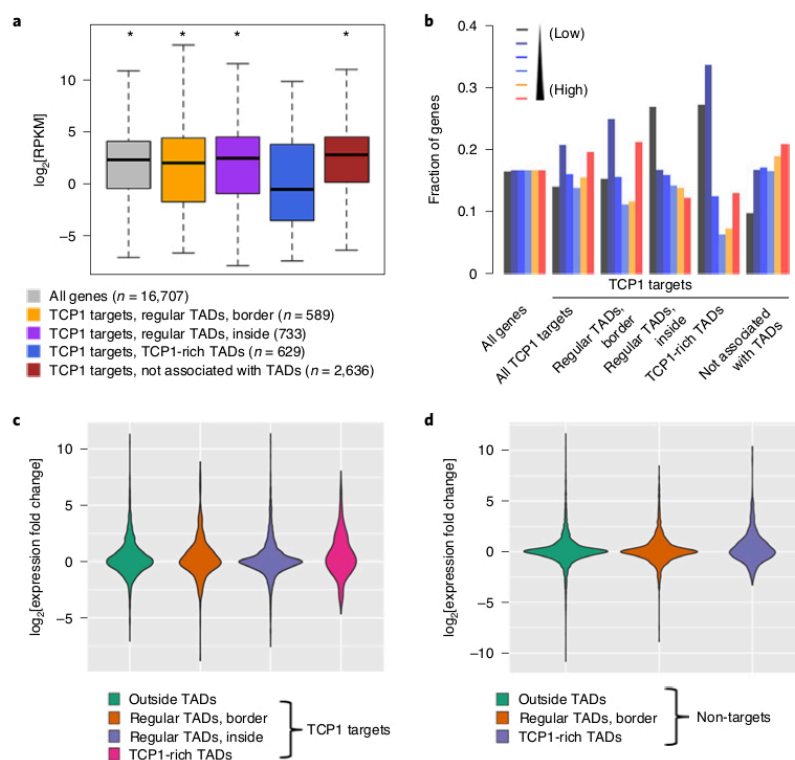


Fig. 6 | Impact of loss of TCP1 on gene expression. **a**, Expression of genes grouped according to their location relative to TADs. All the “*” signs on top of box plots indicate statistical significance from two-sided Mann-Whitney *U*-tests when compared to the blue box plot (TCP1 target genes residing in TCP1-rich TADs); from left to right: 1.6×10^{-26} , 7.6×10^{-9} , 6.3×10^{-15} and 1.4×10^{-31} . The term ‘regular TADs’ in these two panels refers to TADs that are not annotated as TCP1-rich. The box plots indicate the median (line within the box), the lower and upper quartiles (box), margined by the largest and smallest data points that are still within the interval of 1.5 times the interquartile range from the box (whiskers); outliers are not shown. **b**, Association between location and expression of TCP1 target genes. Genes are divided equally into six groups according to the Tak-1 transcriptome data generated in this study. **c,d**, Distribution of gene expression changes in *tcp1* knockout mutants shown in violin plots. Genes with (**c**) and without (**d**) TCP1 binding are categorized according to their location regarding TAD annotation. The term ‘regular TADs’ in these two panels is the same as in panel **a**; the term ‘outside TADs’ refers to genes not overlapping with any TADs.

have different underlying mechanisms of functional regulation. Identification of different groups of *Marchantia* TADs is in correlation with recent work from Dong and colleagues, who pointed out that structural domains (named ‘TAD-like’ regions in their work) in several crop genomes could be classified into groups, each bearing distinct epigenomic decorations²⁸. In our view, it is recommended that the term ‘plant TAD’ is only used for describing chromatin domains exclusively from a structural point of view. As plant genomes presumably contain diverse TADs, we think that it is necessary to categorize them first rather than taking all TADs as a whole for downstream pattern analysis.

In this study, we report a new type of TAD (TCP1-rich) (Fig. 3). Although these TADs are rather depleted of repressive epigenetic marks, genes targeted by TCP1 and located in TCP1-rich TADs show lower expression levels compared to the other TCP1 target genes (Fig. 6a,b). Furthermore, among differentially expressed TCP1 target genes located in TCP1-rich TADs, 80% (114 out of 141) are upregulated in *tcp1*, which is higher than the percentage of the rest TCP1 target genes (Supplementary Table 6). These data indicate that the TCP1 proteins in TCP1-rich TADs predominantly

function as repressors. Probably, extensive TCP1–chromatin interactions over gene bodies influence how a TCP1 target gene is accessible to other transcription factors (TFs) or transcription cofactors, resulting in transcriptional suppression. Besides, the formation of these TADs containing highly concentrated TCP1 underpins a dosage-dependent transcriptional regulatory mechanism²³. A similar example in plants was reported on the *WUSCHEL* (*WUS*) gene, by which higher and lower *WUS* concentration at its target gene (*CLAVATA3*) promoter resulted in gene repression and activation, respectively²⁴. We speculate that, in *tcp1*, the absence of extensive TCP1–chromatin interactions eventually resulted in an overall upregulation of genes located inside TCP1-rich TADs.

At the moment, we do not know if the formation of TCP1 protein speckles is dependent on TCP1-rich TADs. Notably, the N-terminal region of TCP1 protein contains a long stretch of intrinsically disordered region (IDR) that is rich in glycine and proline residues. IDRs are known to mediate multivalent interactions, which promote the coalescence of IDR-containing proteins⁵⁵. We speculate that in some preformed TAD structures, chromatin folding brings multiple TCP1 target regions into physical proximity, and, together

with the help of IDRs, the locally concentrated TCP1 proteins are prone to form protein bodies. In higher plants, many TCP proteins, such as *Arabidopsis* TCP8, TCP14 and TCP24, and rice TCP9, have been shown to display nuclear speckles^{56–59}. According to our recent survey of IDRs in plants, all three of these *Arabidopsis* TCP proteins possess IDRs (ref. ⁶⁰), while the presence of IDRs in the rice TCP9 can be revealed with online IDR prediction⁶¹. It would be intriguing to study how IDRs in these proteins can potentially modulate protein distribution and protein–chromatin interaction patterns.

In this study, we demonstrate that the sole homologue of TCP class I in *Marchantia* occupies many TAD borders and ‘TCP1-rich’ TADs. In the *tcp1* mutants, the absence of TCP1 did not lead to drastic changes in TAD structures (Figs. 2e and 5f). Still, in comparison with the regions without TCP1 binding, changes in chromatin interactions in TCP1-bound TADs or TAD borders were observed, although these changes were subtle (Fig. 5e and Extended Data Fig. 5). According to studies on various animal models, TAD formation is contributed to by multiple factors, such as architectural protein complexes, epigenetic decoration on chromatin and gene expression (recently reviewed in ref. ¹⁴). In our opinion, *tcp1* mutant Hi-C maps do not support our hypothesis that TCP1 functions as an architectural protein for TAD formation; instead, the subtle changes of chromatin interaction patterns in TCP1-bound regions largely reflect changes in gene expression. Due to the following two arguments, however, we cannot completely rule out the possibility that TCP1 functions as a structural protein. First, TCP2 might function as a redundant factor. However, our ongoing work shows that the *tcp1 tcp2* double mutants do not display synergistic growth defects, and they phenocopy *tcp1* single mutant at the thallus growth stage (E.S.K., manuscript in preparation). If TCPs play redundant roles in TAD formation, we expect that *Marchantia* double mutant *tcp1 tcp2*, devoid of TCP genes, develops distinct phenotypes compared to each single mutant. Also, TCP2 (*Mp1g19590*) was not upregulated by *tcp1*, implying that there was no compensation on a transcriptional level (Supplementary Table 6). Second, it is possible that collective activities of TCP1 and other TFs contribute to TAD structure. Our motif analysis revealed that, like TCP1-binding motifs, additional motifs recognized by other TFs were enriched at TAD borders as well (Extended Data Fig. 8a and Supplementary Table 3). Most of these additionally enriched motifs correlated to members from the BBR/BPC and bHLH TF family (Extended Data Fig. 8a).

Among the mCG-poor TADs, only around 30% were classified as TCP1-rich (Fig. 3c). Importantly, the remaining 70% non-TCP1-rich, mCG-poor TADs have similar chromatin accessibility, like TCP1-rich TADs (Extended Data Fig. 6b). These observations suggest a possibility that other types of TFs or regulatory factors have intensive interactions with these TADs. Supporting this point, motifs bound by BBR/BPC and C2C2-Gata TFs showed higher densities within mCG-poor TADs (Extended Data Fig. 8b and Supplementary Table 3). Notably, the BBR/BPC motifs stood out from our analyses on both TAD borders and mCG-poor TAD bodies (Extended Data Fig. 8). BBR/BPC proteins are plant-specific TFs recognizing (GA/TC)_n repeats⁶². This motif has been shown to be enriched in promoter regions of the *Arabidopsis* genome⁶³. The single corresponding *Marchantia* BBR/BPC homologue in the Tak-1 genome (*MpVg00350*) appears to be a good candidate for investigation to uncover new ‘TF-rich’ TADs via examining chromatin interaction patterns.

It is noteworthy that ‘TF-rich’ TADs have not been reported in animals; this phenomenon might be plant-specific (note that the TCP proteins are only found in the plant kingdom). It is unknown whether certain TADs in higher plants are heavily loaded with TFs, and therefore would be analogous to the *Marchantia* TCP1-rich TADs described in this study. We envisage that, in the near future, this point will be clarified by association studies that integrate TF–chromatin binding and 3D genome organization patterns.

Methods

Plant material. *Marchantia polymorpha* Tak-1 gemmae were cultured on half-strength B5 medium supplemented with 1% sucrose. The condition for growing gemmae was set to be long-day (16h light and 8h dark, 3,000 lux), and the temperature was maintained at 22 °C. Two-week-old thalli were used for all experiments performed in this study. *Agrobacterium*-mediated spore transformation was done as previously described⁶⁴.

Cloning and plasmids. The established CRISPR–Cas9-mediated mutagenesis system in *Marchantia*⁶⁵ was used to generate TCP1 mutant lines. Two different single guide RNAs (sgRNAs) were used in a single vector: sgRNA1: 5′-GACAGGCACACGAAGGTCGA-3′ and sgRNA2: 5′-GATTGGTTAAATGATAAGCG-3′, which targeted PAM sites upstream of TCP1 DNA-binding domain and downstream of TCP1 stop codon, respectively. To genotype the mutant lines with deletion, the following primer sequences were used: 5′-AGGCACAGCAGGAAGTATG-3′ and 5′-TGCCTCTCGTCTCTCTACTTCTC-3′. Among the transgenic plants, only mutants showing a deletion between the two sgRNA sites and in Tak-1 background were selected for further experiments.

To construct the TCP1 complementation construct, a genomic fragment DNA containing the TCP1 locus plus its 4.7-kb upstream region was amplified with oligonucleotides 5′-ACGAGACTGATTGGTTCTT-3′ and 5′-TTACTGCGAGC TAGTGGGATCGT-3′. Next, the DNA fragments were mutagenized with oligonucleotides 5′-TCCATCCACCTTGGTATGTCGGTCTTCGTAGACGATC GTT-3′ and 5′-CGACATACCAAGGTGGATGGAAGGGGCAGGAGGAT-3′, by which silent mutations were introduced to prevent it from being recognized by sgRNA1. Finally, this DNA fragment was cloned into *pMpGWB301* (ref. ⁶⁶) and was used to rescue *tcp1* knockout mutants.

To construct the *pTCP:GFP* lines, promoters were amplified with the following oligonucleotides: for TCP1, 5′-CACGAGACTGATTGGTTCTT-3′ and 5′-TCAGTATCCAGTGTTCATT-3′, for TCP2, 5′-TACTTTGATTGCTGACTGGAT-3′ and 5′-CACATGGCATATGAGCC GGA-3′. Then, these promoter fragments, together with a GFP fragment, were cloned into *pMpGWB101* and transformed into the spores.

In situ Hi-C and data processing. The in situ Hi-C library preparation was performed essentially as described⁶⁷. For each replicate, around 0.5 g of fixed thalli were homogenized for nuclei isolation. The libraries were sequenced on an Illumina HiSeq 3000 instrument with 2 × 150-base pair (bp) reads. Reads mapping, removal of PCR duplicates and reads filtering were performed as described⁶⁷. Hi-C reads of each sample are summarized in Supplementary Table 1. Hi-C map normalization was done according to our previous rice Hi-C study by using an iterative matrix correction function in the ‘HITC’ package in R (refs. ^{26,68}). For wild-type Hi-C data presented in Figs. 1, 3 and 4, and their associated extended data and supplementary figures, the two replicates were merged and used to generate Hi-C maps. For Hi-C maps presented in Figs. 2 and 5, and their associated extended data and supplementary figures, Hi-C maps were generated from one replicate. For all Hi-C maps, the iterative normalization process was stopped when the eps value, which reflected how similar the matrices in two consecutive correction steps were, dropped below 1 × 10⁻⁴. Normalization was performed on each Tak-1 autosome separately at 2-kb resolution for all samples. In this study, three sets of Tak-1 Hi-C data were generated. Two of them were used for TAD annotation and characterization; the third was from plants grown along with *tcp1* mutants, and this single Tak-1 dataset was used for comparisons with the mutant Hi-C data.

TAD calling and categorization. Our visual inspection of *Marchantia* TAD patterns indicated that they were not distributed in a side-by-side manner. Thus, we chose to use the ‘arrowhead’ algorithm¹³ for TAD calling, which suited with the rice genome that exhibited a similar TAD layout⁴⁹. As described previously, changes in parameters for filtering the TAD score matrix and for selecting potential TAD borders could affect TAD-calling sensitivity²⁹. In this study, we set the cut-off for the TAD score matrix as 0.95, the minimum number of filtered pixels belonging to a potential TAD as 8 and the minimum TAD score of pixels at TAD borders as 1.05. For TAD annotation, the CG methylation ratio of individual TADs and the fraction of regions overlapping with TCP1 ChIP-seq peaks were calculated. Recently reported whole-genome bisulfite sequencing data of Tak-1 thalli⁶⁹ were used to calculate DNA methylation ratios in TADs. We arbitrarily used CG methylation ratios of 0.2 and 0.6 to classify TADs, with TADs having average CG methylation ratios less than 0.2 or more than 0.6 being termed ‘mCG-poor’ and ‘mCG-rich’, respectively. The rest of the TADs having CG methylation ratios between 0.2 and 0.6 were termed ‘mCG-intermediate’. The computation of interaction decay exponents of each type of TAD was performed as described²⁹. The *k*-means clustering of TADs was performed essentially as described by taking TAD regions as the input⁶⁸.

Independent from its epigenetic profiling, a TAD was annotated as ‘TCP1-rich’ if over 20% of its chromatin region overlapped with TCP1 ChIP-seq peaks. Details of TAD coordinates and their annotation concerning DNA methylation and TCP1 binding can be found in Supplementary Table 2.

ARTICLES

NATURE PLANTS

Motif analysis. De novo motif analysis was performed with MEME software v.4.11.2 (ref. ⁷⁰). We randomly selected 2,000 of the TCP1-enriched regions and used their central 300-bp sequences as the input for motif identification. The motif search was conducted using a 'zoo' model (zero or one occurrence per sequence), and the motif length was set between 4 and 12bp. The search was stopped when the *E* value (number of expected motifs found by chance) of an identified motif was above 0.05. Details of the identified motifs can be found in Supplementary Table 5.

Motif scan of enquiry DNA sequences for a given motif was done by using the 'matchPWM' function in the 'Biostrings' package in R. This approach was applied to reveal the occurrence of hundreds of known plant transcription factor-binding sites around TAD borders (Extended Data Fig. 3c), in which motif position weight matrices were determined previously⁴⁸. For the analyses shown in Supplementary Fig. 4, the motif position weight matrices were obtained from the de novo motif search mentioned above, details of which can be found in Supplementary Table 7.

RNA-seq library preparation and analysis. Total RNA was isolated from wild-type and *tcp1* mutant thalli with RNeasy Plant Mini Kit (Qiagen). Libraries of each genotype were prepared with three biological replicates as described¹. RNA-seq reads were aligned against the Tak-1 v5 genome using TopHat 2 (v.2.1.1) with default parameters⁷², and were further assigned to genes using the GenomicAlignments package⁷³ in R. Differentially expressed genes were identified with the DESeq2 package⁷⁴ in R. We used criteria of false discovery rate smaller than 0.01 and expression fold change more than 3 to call up- and downregulated genes. Details of the reads count table, gene expression measurement (in reads per kilobase per million mapped reads) and differentially expressed genes can be found in Supplementary Table 6.

Coexpression analysis. Coexpression correlation was computed essentially according to Contreras-López et al.⁷⁵ In addition to the RNA-seq data of Tak-1 thalli that we generated in this study, a list of public transcriptomic datasets of various wild-type *Marchantia* tissues under normal growth conditions were included (see Data availability). Pearson's correlation coefficients of all gene pairs in each chromosome were calculated in R, and those with *q* values smaller than 0.05 were identified as coexpressed gene pairs.

ChIP-seq library preparation and analysis. Anti-TCP1 antibodies were raised using a synthesized peptide SKGGIRKRARPGSS that corresponded to *Marchantia* TCP1 protein sequence. The peptide was injected into rabbits, from which the antisera were verified and affinity purified (BioGenex).

For ChIP-seq experiment, tissue fixation and nuclei isolation were performed as in the in situ Hi-C protocol, and the subsequent steps were performed mainly according to our previous study¹. The isolated chromatin was sonicated with a Covaris E220 instrument to achieve an average fragment size of 400bp. After sonication, chromatin was divided into two equal aliquots, and was incubated with 1 µg ml⁻¹ of rabbit IgG (Abcam, ab37415, lot no. GR3219601-1) or affinity-purified anti-TCP1 antibody, respectively. After incubation, 10 µl protein A/G magnetic beads (Pierce) were added to each tube to recover chromatin associated with antibodies. DNA recovery and library preparation were as described¹.

For ChIP-seq analysis, reads were aligned against the Tak-1 v5 reference genome using Bowtie 2 v.2.2.4 (ref. ⁷⁶) with a 'very sensitive' mapping mode. ChIP-seq peak calling was done with MACS2 v.2.1.1 (ref. ⁷⁷) using default settings and the reads from IgG pulldown were used as control. The enriched chromatin regions obtained from two biological replicates of wild-type Tak-1 thalli were compared, after which overlapping regions were annotated as TCP1 ChIP-seq peaks. Details of ChIP-seq peaks, as well as peak-calling results of individual replicates, can be found in Supplementary Table 4.

ATAC-seq library preparation and analysis. ATAC-seq was performed with two biological replicates. For each replicate, 20,000 DAPI-stained nuclei were collected via a MoFlo XDP FACS instrument (Beckman Coulter) as described¹. The nuclei were processed to generate ATAC-seq libraries according to our established pipeline for *Arabidopsis*⁷¹. After sequencing, ATAC-seq reads were aligned against the Tak-1 v5 reference genome with Bowtie 2 v.2.2.4 (ref. ⁷⁶) with a 'very sensitive' mapping mode. An ATAC-seq coverage file was generated by using the bedtools multicov command in BEDTools v.2.26.0 (ref. ⁷⁸), with which the sorted bam mapping files were used as input. ATAC-seq peaks were identified with MACS2 (ref. ⁷⁷) using settings as 'nomodel --shift -50 --extsize 100 --keep-dup=1', and peaks with a false discovery rate smaller than 0.05 were retained. ATAC-seq peaks can be found in Supplementary Table 8.

Combined FISH and immunohistostaining. The genomic sequences of selected TADs were amplified as tiling fragments, each with length around 8kb, with oligonucleotides as listed in Supplementary Table 9. The PCR products corresponding to one TAD were mixed equally and labelled with digoxigenin-11-dUTP using the Nick Translation Kit (Roche Applied Science) according to the manufacturer's instructions.

Combined FISH and immunohistostaining were performed according to Hu et al. with minor changes¹. Briefly, for antigen retrieval, slides with nuclei were treated for 10 min in a microwave oven at 700 W in 10 mM sodium citrate

at pH 6.0. For DNA probe hybridization, the nuclei were incubated in 5 µl hybridization mix containing 25 ng DIG-labelled probe at 37 °C overnight. On the next day, the nuclei were blocked with blocking buffer (5% BSA in 4× SSC with 0.2% Tween-20) for 10 min at room temperature, which was followed by primary antibody incubation for 1 h at 37 °C. The primary antibody solution consisted of anti-DIG Alexa Fluor 488 (R&D System, catalogue no. IC7520G, 1:10 dilution) and anti-TCP1 (1:100 dilution) diluted in blocking buffer. The nuclei were washed and then incubated with goat anti-rabbit Alexa Fluor 546-conjugated antibody (ThermoFisher Scientific, catalogue no. A-11035, 1:150 dilution) for 1 h at 37 °C. Washing was performed with TBS-T (100 mM Tris-HCl at pH 7.5, 150 mM NaCl, 0.1% Tween) for 10 min three times after primary and secondary antibody binding. Finally, nuclei were counterstained with DAPI (1 µg ml⁻¹) in 2× SSC buffer.

Fluorescence microscopy and image processing. All confocal image acquisitions of nuclei were performed on an LSM 880 confocal laser scanning microscope (Zeiss) Airyscan system equipped with a ×63/1.4 numerical aperture water objective and processed using ImageJ (NIH).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Short read data of in situ Hi-C, ChIP-seq, ATAC-seq and RNA-seq are publicly available at NCBI Sequence Read Archive under accession number PRJNA597314. Large datasets, including Hi-C matrices (2-kb bin size for individual chromosomes), integrated epigenetic marks, ATAC-seq and ChIP-seq track files in 200-bp bin size are available from figshare repository, which are accessible with the following link: https://figshare.com/articles/dataset/Marchantia_TCP_transcription_factor_activity_correlates_with_3D_chromatin_structure/11309657. All figures presented in this manuscript are associated with these data. The following public datasets were downloaded for coexpression analysis (with their accession numbers from the NCBI Sequence Read Archive): 11-day thalli (DRR050343, DRR050344, DRR050345), Archegoniophore (DRR050351, DRR050352, DRR050353), Antheridiophore (DRR050346, DRR050347, DRR050348), Antheridia (DRR050349, DRR050350), apical cell (SRR1553294, SRR1553295, SRR1553296), 13d-Sporophyte (SRR1553297, SRR1553298, SRR1553299), Sporelings 0 hr (SRR4450262, SRR4450261, SRR4450260), 24hr-Sporeling (SRR4450266, SRR4450265, SRR4450259), 48hr-Sporeling (SRR4450268, SRR4450264, SRR4450263), 72hr-Sporeling (SRR4450267, SRR4450258, SRR4450257), 96hr-Sporeling (SRR4450256, SRR4450255, SRR4450254), mock-inoculated plants, 1dpi (SRR7977545, SRR7977546, SRR7977548), mock-inoculated plants, 2dpi (SRR7977547, SRR7977550, SRR7977549), mock-inoculated plants, 3dpi (SRR7977552, SRR7977551, SRR7977554), mock-inoculated plants, 4dpi (SRR7977553, SRR7977556, SRR7977555), 1-month thallus (SRR6685782, SRR6685783, SRR6685784), Tak-1-1_Mp (SRR772758, SRR772757, SRR772756, SRR772755), Tak-1-2_Mp (SRR772759, SRR772761, SRR772760, SRR772762), Tak-1-3_Mp (SRR772763, SRR772764, SRR772765, SRR772766), and Mp-mock (SRR5905098, SRR5905099, SRR5905100). Source data are provided with this paper.

Code availability

All scripts used for pattern analysis are available upon request.

Received: 24 March 2020; Accepted: 7 August 2020;
Published online: 07 September 2020

References

- Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
- Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).
- Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).
- Gomez-Diaz, E. & Corces, V. G. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol.* **24**, 703–711 (2014).
- Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
- Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 772 (2016).

11. Merckenschlager, M. & Nora, E. P. CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.* **17**, 17–43 (2016).
12. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
13. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
14. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
15. Andrey, G. et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
16. Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
17. Narendra, V. et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015).
18. Eser, U. et al. Form and function of topologically associating genomic domains in budding yeast. *Proc. Natl Acad. Sci. USA* **114**, E3061–E3070 (2017).
19. Pope, B. D. et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).
20. Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
21. Ramirez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
22. Shen, Y. et al. A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
23. Zhan, Y. et al. Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Res.* **27**, 479–490 (2017).
24. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
25. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
26. Xie, T. et al. Biased gene retention during diploidization in *Brassica* linked to three-dimensional genome organization. *Nat. Plants* **5**, 822–832 (2019).
27. Wang, C. et al. Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res.* **25**, 246–256 (2015).
28. Dong, P. et al. 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* **10**, 1497–1509 (2017).
29. Liu, C., Cheng, Y. J., Wang, J. W. & Weigel, D. Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*. *Nat. Plants* **3**, 742–748 (2017).
30. Wang, M. et al. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. *Nat. Genet.* **49**, 579–587 (2017).
31. Wang, M. et al. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat. Plants* **4**, 90–97 (2018).
32. Concia, L. et al. Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol.* **21**, 104 (2020).
33. Dogan, E. S. & Liu, C. Three-dimensional chromatin packing and positioning of plant genomes. *Nat. Plants* **4**, 521–529 (2018).
34. Stam, M., Tark-Dame, M. & Fransz, P. 3D genome organization: a role for phase separation and loop extrusion? *Curr. Opin. Plant Biol.* **48**, 36–46 (2019).
35. Bowman, J. L. et al. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287–304.e15 (2017).
36. Montgomery, S. A. et al. Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. *Curr. Biol.* **30**, 573–588.e7 (2020).
37. Krom, N. & Ramakrishna, W. Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, *Arabidopsis*, and *Populus*. *Plant Physiol.* **147**, 1763–1773 (2008).
38. Reimegard, J. et al. Genome-wide identification of physically clustered genes suggests chromatin-level co-regulation in male reproductive development in *Arabidopsis thaliana*. *Nucleic Acids Res.* **45**, 3253–3265 (2017).
39. Ren, X. Y., Stiekema, W. J. & Nap, J. P. Local co-expression domains in the genome of rice show no microsynteny with *Arabidopsis* domains. *Plant Mol. Biol.* **65**, 205–217 (2007).
40. Williams, E. J. & Bowles, D. J. Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res.* **14**, 1060–1067 (2004).
41. Braccioli, L. & de Wit, E. CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.* **63**, 157–165 (2019).
42. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
43. Danisman, S. TCP transcription factors at the interface between environmental challenges and the plant's growth responses. *Front. Plant Sci.* **7**, 1930 (2016).
44. Li, S. The *Arabidopsis thaliana* TCP transcription factors: a broadening horizon beyond development. *Plant Signal. Behav.* **10**, e1044192 (2015).
45. Busch, A. & Zachgo, S. Flower symmetry evolution: towards understanding the abominable mystery of angiosperm radiation. *Bioessays* **31**, 1181–1190 (2009).
46. Navaud, O., Dabos, P., Carnus, E., Tremousaygue, D. & Herve, C. TCP transcription factors predate the emergence of land plants. *J. Mol. Evol.* **65**, 23–33 (2007).
47. Franco-Zorrilla, J. M. et al. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl Acad. Sci. USA* **111**, 2367–2372 (2014).
48. O'Malley, R. C. et al. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**, 1280–1292 (2016).
49. Kosugi, S. & Ohashi, Y. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* **30**, 337–348 (2002).
50. Busch, A. et al. MpTCP1 controls cell proliferation and redox processes in *Marchantia polymorpha*. *New Phytol.* **224**, 1627–1641 (2019).
51. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
52. Rowley, M. J. et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* **67**, 837–852 (2017).
53. Birchler, J. A., Bhadra, U., Bhadra, M. P. & Auger, D. L. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev. Biol.* **234**, 275–288 (2001).
54. Perales, M. et al. Threshold-dependent transcriptional discrimination underlies stem cell homeostasis. *Proc. Natl Acad. Sci. USA* **113**, E6298–E6306 (2016).
55. Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).
56. Kubota, A. et al. TCP4-dependent induction of CONSTANS transcription requires GIGANTEA in photoperiodic flowering in *Arabidopsis*. *PLoS Genet.* **13**, e1006856 (2017).
57. Mazur, M. J. et al. *Arabidopsis* TCP transcription factors interact with the SUMO conjugating machinery in nuclear foci. *Front. Plant Sci.* **8**, 2043 (2017).
58. Mukhopadhyay, P. & Tyagi, A. K. OsTCP19 influences developmental and abiotic stress signaling by modulating ABI4-mediated pathways. *Sci. Rep.* **5**, 9998 (2015).
59. Valsecchi, I. et al. The intrinsically disordered C-terminal region of *Arabidopsis thaliana* TCP8 transcription factor acts both as a transactivation and self-assembly domain. *Mol. Biosyst.* **9**, 2282–2295 (2013).
60. Pontvianne, F. & Liu, C. Chromatin domains in space and their functional implications. *Curr. Opin. Plant Biol.* **54**, 1–10 (2019).
61. Meszaros, B., Erdos, G. & Dosztanyi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).
62. Theune, M. L., Bloss, U., Brand, L. H., Ladwig, F. & Wanke, D. Phylogenetic analyses and GAGA-motif binding studies of BBR/BPC proteins lend to clues in GAGA-motif recognition and a regulatory role in brassinosteroid signaling. *Front. Plant Sci.* **10**, 466 (2019).
63. Santi, L. et al. The GA octadecanucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene *Bkn3*. *Plant J.* **34**, 813–826 (2003).
64. Ishizaki, K., Chiyoda, S., Yamato, K. T. & Kohchi, T. Agrobacterium-mediated transformation of the haploid liverwort *Marchantia polymorpha* L., an emerging model for plant biology. *Plant Cell Physiol.* **49**, 1084–1091 (2008).
65. Sugano, S. S. et al. CRISPR/Cas9-mediated targeted mutagenesis in the liverwort *Marchantia polymorpha* L. *Plant Cell Physiol.* **55**, 475–481 (2014).
66. Ishizaki, K. et al. Development of gateway binary vector series with four different selection markers for the liverwort *Marchantia polymorpha*. *PLoS ONE* **10**, e0138876 (2015).
67. Liu, C. et al. Genome-wide analysis of chromatin packing in *Arabidopsis thaliana* at single-gene resolution. *Genome Res.* **26**, 1057–1068 (2016).
68. Servant, N. et al. HiTC: exploration of high-throughput 'C' experiments. *Bioinformatics* **28**, 2843–2844 (2012).
69. Schmid, M. W. et al. Extensive epigenetic reprogramming during the life cycle of *Marchantia polymorpha*. *Genome Biol.* **19**, 9 (2018).
70. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
71. Hu, B. et al. Plant lamin-like proteins mediate chromatin tethering at the nuclear periphery. *Genome Biol.* **20**, 87 (2019).
72. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
73. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
74. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
75. Contreras-Lopez, O., Moyano, T. C., Soto, D. C. & Gutierrez, R. A. Step-by-step construction of gene co-expression networks from high-throughput *Arabidopsis* RNA sequencing data. *Methods Mol. Biol.* **1761**, 275–301 (2018).
76. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
77. Liu, T. Use Model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein–DNA interactions in embryonic stem cells. *Methods Mol. Biol.* **1150**, 81–95 (2014).

ARTICLES

NATURE PLANTS

78. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
79. Lin, P. C. et al. Identification of miRNAs and their targets in the liverwort *Marchantia polymorpha* by integrating RNA-seq and degradome analyses. *Plant Cell Physiol.* **57**, 339–358 (2016).

Acknowledgements

We thank S. Czernel from the Center for Quantitative Biology (University of Tübingen) for their assistance with sequencing. We are grateful for inspiring discussions with members of the COST Action CA1612 INDEPTH network. We acknowledge computing support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no. INST 37/935-1 FUGG. This work was supported by the Deutsche Forschungsgemeinschaft (LI 2862/4) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 757600).

Author contributions

C.L. conceived and designed the experiments. E.S.K., N.W., N.F. and H.B. established and characterized transgenic lines. E.K. performed ChIP-seq, ATAC-seq and RNA-seq experiments. N.W. performed FISH and immunostaining experiments. Y.L. performed

coexpression analysis. S.A.M. and E.B. performed epigenomic profiling. K.W.B. performed nuclei sorting. C.L., E.S.K. and S.L. performed Hi-C experiments. E.S.K. and C.L. wrote the manuscript with contributions from other authors. All authors read and accepted the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41477-020-00766-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41477-020-00766-0>.

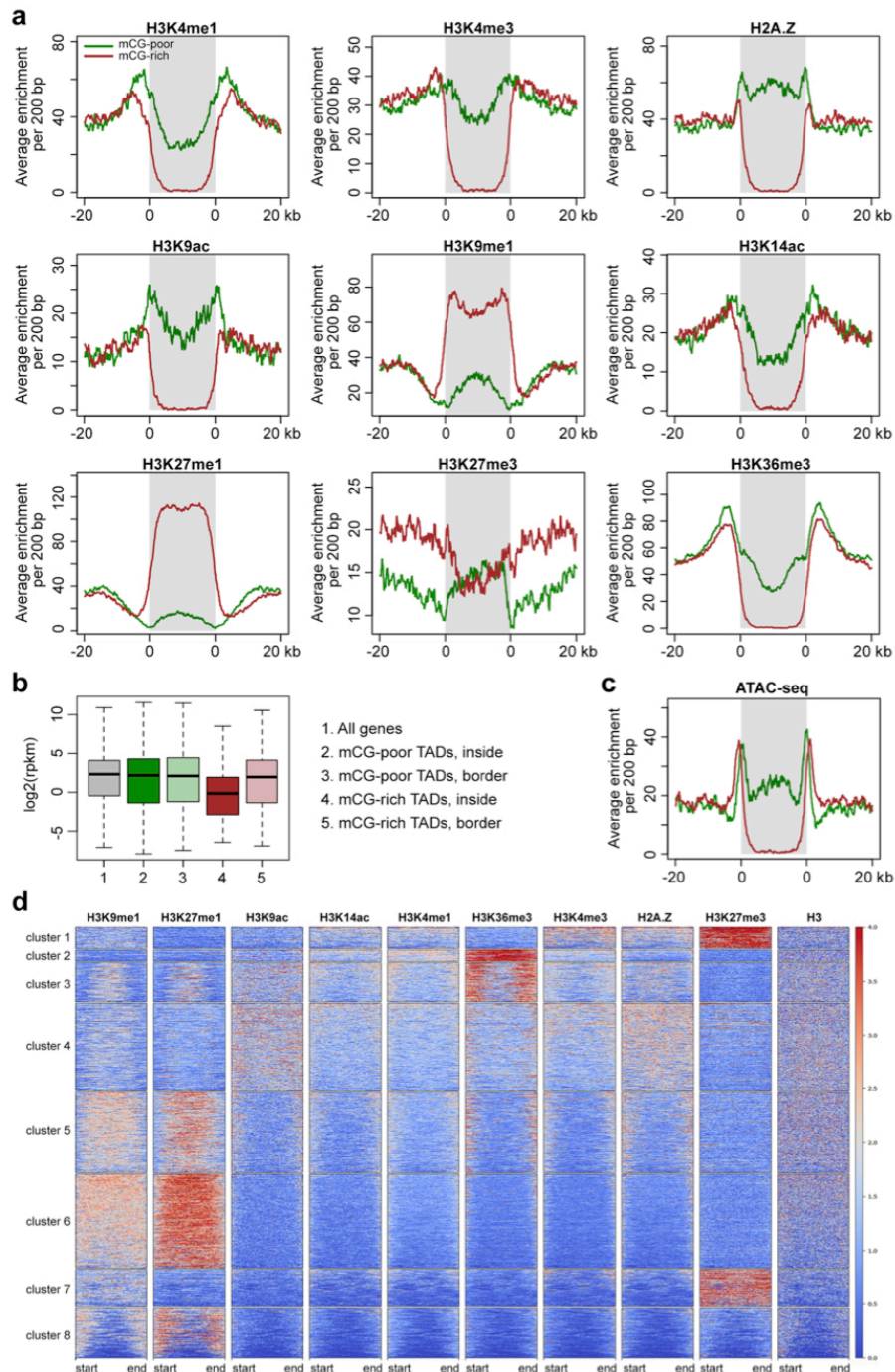
Correspondence and requests for materials should be addressed to C.L.

Peer review information *Nature Plants* thanks Stefan de Folter, Stefan Grob and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

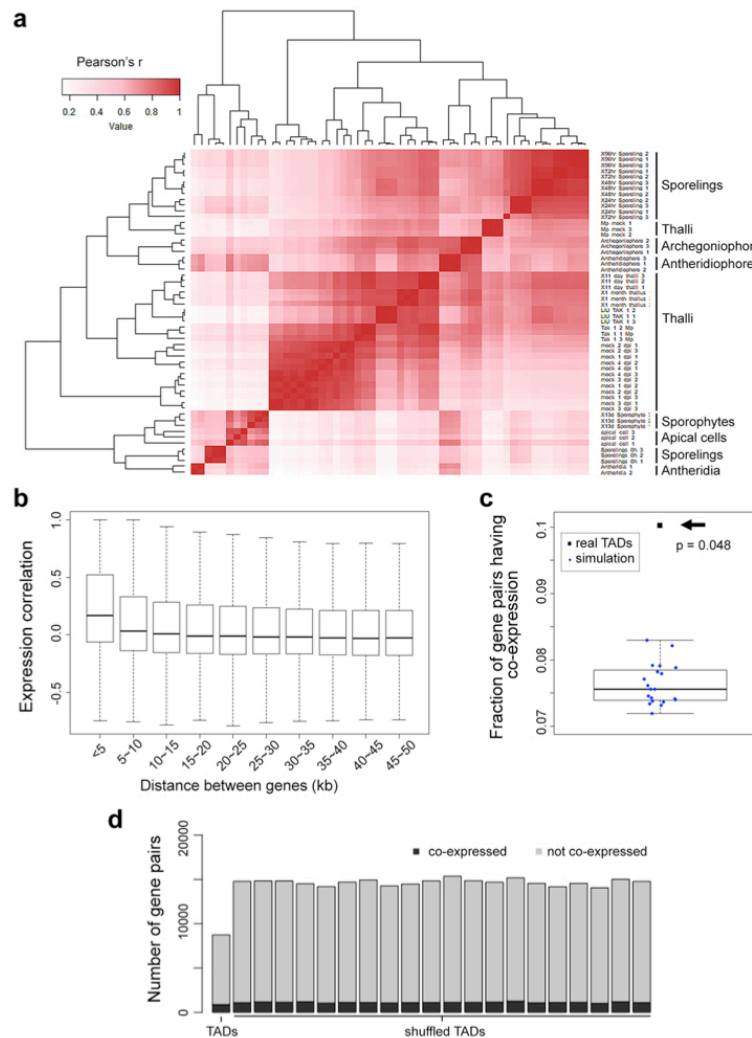
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

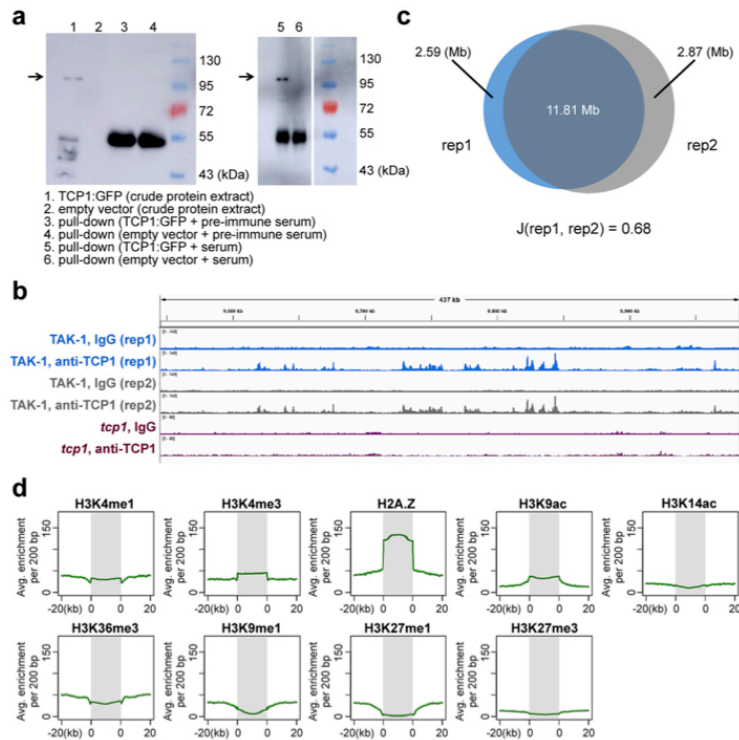


Extended Data Fig. 1 | The *Marchantia* genome has different types of TADs. **a**, Epigenetic marks across *Marchantia* TADs. ‘mCG-poor’ and ‘mCG-rich’ TADs are shown in green and brown curves, respectively. **b**, Comparison of gene expression according to gene locus location. *Thalli* transcriptome data was from (doi: 10.1093/pcp/pcw020). Boxplots from left to right: n = 16707, 2264, 972, 215 and 575. **c**, Chromatin accessibility across ‘mCG-poor’ (green) and ‘mCG-rich’ (brown) TADs. **d**, Clustering analysis of TADs according to histone marks. The epigenetic profiling of various histone marks was from our previous study (doi: 10.1016/j.cub.2019.12.015).

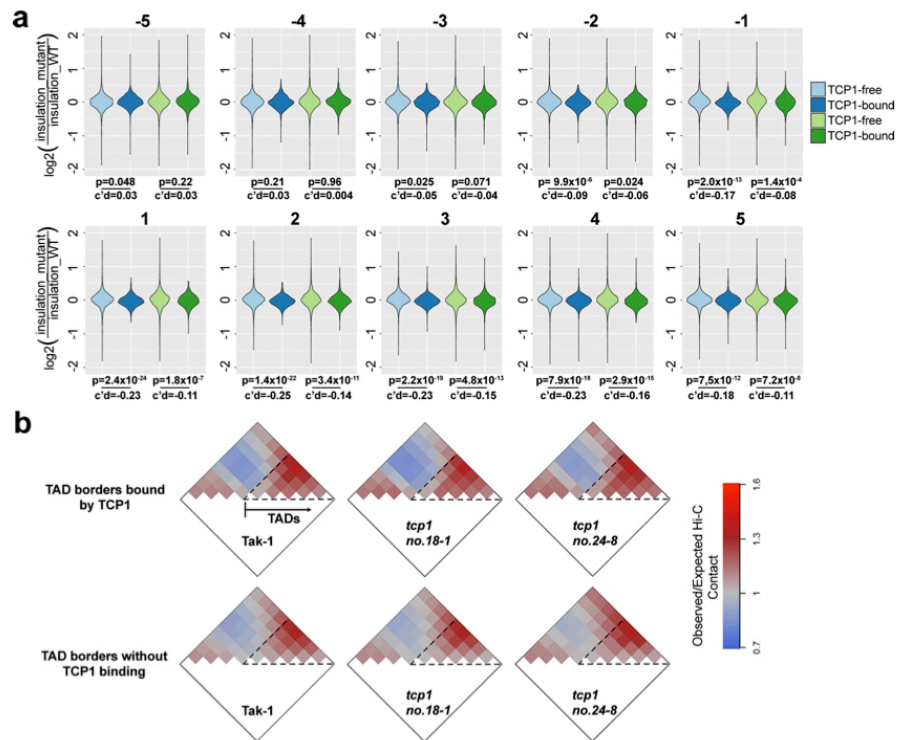


Extended Data Fig. 2 | Co-expression analysis. **a**, Clustering of transcriptome datasets used for calculating gene co-expression. The dendrogram shows hierarchical clustering based on Euclidean distance. **b**, Distribution of expression correlation coefficients as a function of distance. The distance of a given gene pair was determined according to their annotated TSSs. Boxplots from left to right: $n = 6455, 9068, 9238, 9165, 8992, 8931, 8832, 9053, 8736,$ and 8793 . **c, d**, TADs contain more co-expressed genes than expected. **c**, For all the gene pairs located in the same TAD, the fraction of co-expressed gene pairs ($q < 0.05$) was computed and indicated by a black dot at the top of the panel (highlighted with an arrow). The boxplot with twenty blue data points denotes results in which TADs were randomly assigned to the genome. The p-value is an empirical p-value calculated based on twenty simulations. **d**, Numbers of co-expressed and not co-expressed gene pairs in each round of shuffled TADs (simulation) are shown. As TADs are gene-poor (Fig. 1c), randomly shuffling TADs leads more genes (hence gene pairs) to overlap with TADs. The boxplots in **b** and **c** indicate the median (line within the box), the lower and upper quartiles (box), margined by the largest and smallest data points which are still within the interval of 1.5 times the interquartile range from the box (whiskers); outliers are not shown.

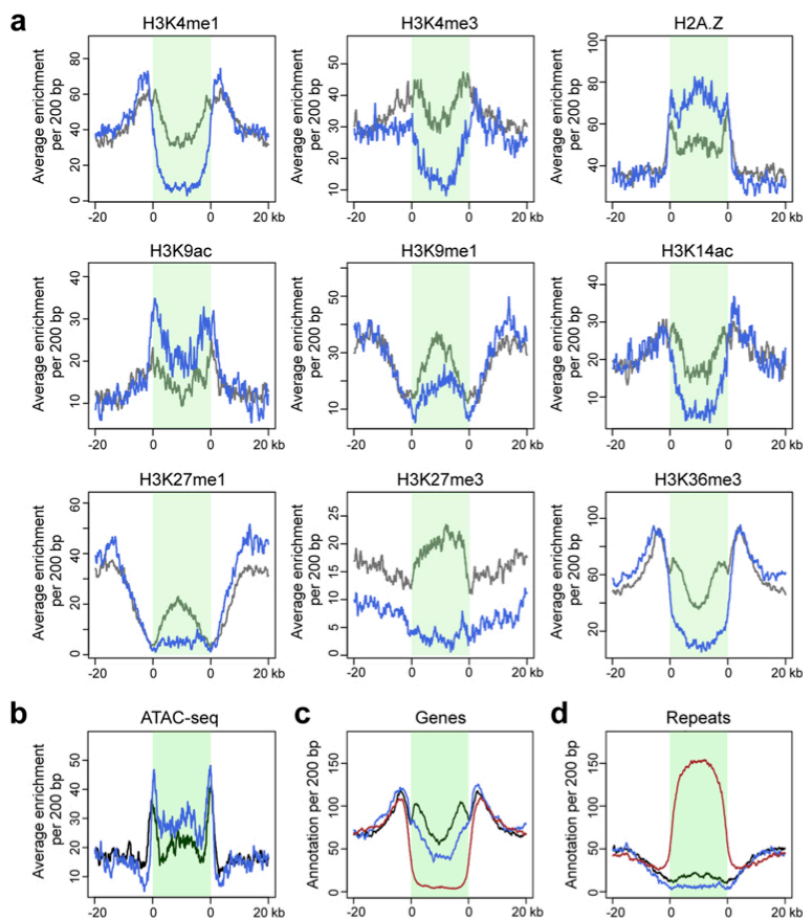
ARTICLES NATURE PLANTS



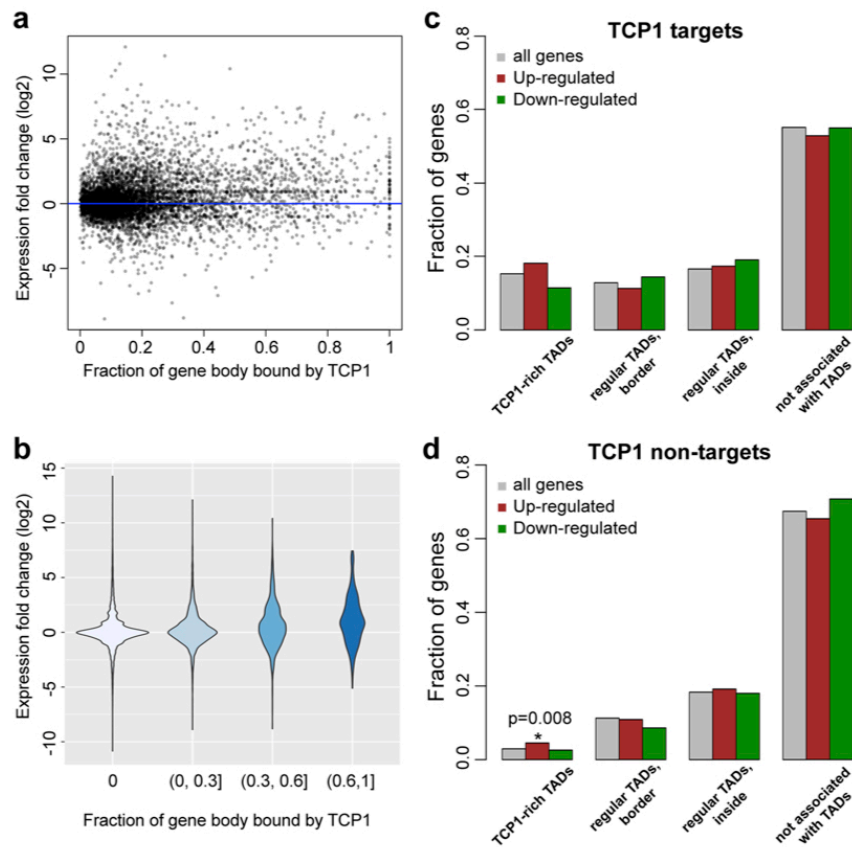
Extended Data Fig. 4 | Genome-wide identification of TCP1 target regions. **a**, Verification of anti-TCP1 antibody for immunoprecipitation. TCP1:GFP fusion proteins under the control of 35S promoter were expressed transiently in *Nicotiana benthamiana* leaves. The presence of TCP1:GFP in each sample was examined by using anti-GFP antibody. Similar results were observed from two independent experiments. **b**, Snapshot showing reads distribution of ChIP-seq reads in different samples. *tcp1* represents a *TCP1* knock-out line (in Tak-1 background). See 'methods' for details of knock-out line generation. **c**, Venn diagram of genomic regions enriched in each biological replicate. Below this Venn diagram, the $J(\text{rep1}, \text{rep2})$ indicates the Jaccard index. **d**, Epigenetic marks across TCP1-bound chromatin regions (grey block).



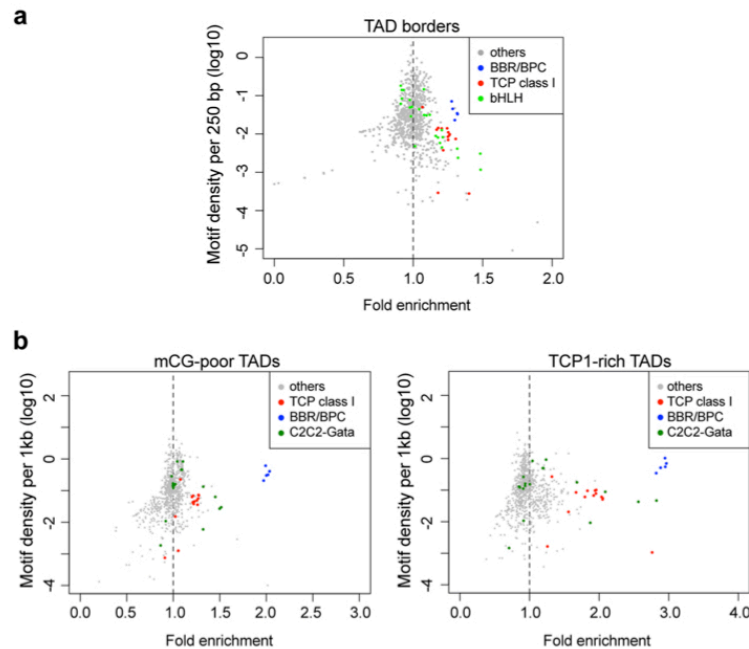
Extended Data Fig. 5 | Comparison of insulation scores of chromatin regions around wild-type and *tcp1* mutant TAD borders. a, Comparison of insulation scores in wild-type TAD borders with those in *tcp1* no. 18-1 (blue plots) or in *tcp1* no. 24-8 (green plots). See Fig. 2e legend for the definition of ‘TCP1-bound’ TAD borders. The titles above these plots, which indicate bin positions, are as those under boxplots in Fig. 2e. Assuming that TCP1 plays a structural role on TCP1-bound TAD borders, we expect that the removal of TCP1 results in specific changes in insulation scores of these regions compared to regions not bound by TCP1. The violin plots in this panel show distribution of changes in insulation scores in the mutant Hi-C maps. For each pair of comparison (that is, changes in insulation scores of TCP1-bound TAD borders vs. TCP1-free TAD borders), its p-value from the two-sided Mann-Whitney U test is given. To assess effect size, the cohen’s d (c’d) is also given below each p-value note. In general, the difference between two populations is considered ‘trivial’ or ‘negligible’ when the absolute value of cohen’s d is less than 0.2. **b**, Metagene plots showing chromatin contacts around TAD border regions. Pixels in the plots stand for 2-kb bins in the Hi-C matrices. For each plot, TAD borders are aligned and indicated with a dotted triangle. With careful inspection, we conclude that the differences of chromatin organization between TCP1-bound TAD borders and TCP1-free TAD borders are comparable in Tak-1 and *tcp1*, and loss-of-TCP1 does not led to drastic structural changes in TCP1-bound TAD boundaries.



Extended Data Fig. 6 | Epigenetic and transcriptional profiling of TCP1-rich TADs. Comparison of various epigenetic marks (a) and chromatin accessibility (b) between TCP1-rich TADs (blue curves) and the rest TADs (gray curves) belonging to the 'mCG-poor' category. c, d, Comparison of genes (c) and repeats (d) in different TADs. Same as those in panel b, the blue and gray curves denote TCP1-rich TADs and rest TADs in the 'mCG-poor' category, respectively. The brown curves denotes 'mCG-rich' TADs. Labels are the same as in Fig. 1b.



Extended Data Fig. 7 | Changes in gene expression in relation to TCP1 binding, and the distribution of differentially expressed genes in *tcp1* in relation to their location. **a**, Changes in expression of genes bound by TCP1. Only genes with their gene bodies (defined as their transcribed region plus 0.5 kb flanking regions) overlapping with TCP1 ChIP-seq peaks are included in this plot. **b**, Distribution of gene expression changes in *tcp1* mutants. All the genes from the genome are divided into four groups according to the extent to which they overlap with TCP1 ChIP-seq peaks. **c, d**, Differentially expressed genes bound (**c**) and not bound (**d**) by TCP1 are divided into different groups. The p value indicates two-sided Fisher's exact test result. The term 'regular TADs' in these two panels refers to TADs that are not annotated as TCP-rich.



Extended Data Fig. 8 | Motif analysis of *Marchantia* TAD borders and TADs. **a**, Motif analysis of *Marchantia* TAD borders. This plot is the same as Extended Data Fig. 3c, but highlighting motifs of a few transcription factor families. **b**, Motif analysis of *Marchantia* mCG-poor and TCP1-rich TAD bodies. Fold enrichment was calculated as the ratio of motif density in TADs over that in 30 kb flanking genomic regions. Other than that, the motif search was performed as for Extended Data Fig. 3c.

natureresearch

<https://doi.org/10.1038/s41477-020-00766-0>

Supplementary information

***Marchantia* TCP transcription factor activity correlates with three-dimensional chromatin structure**

In the format provided by the authors and unedited

1 Supplemental Information

2

3

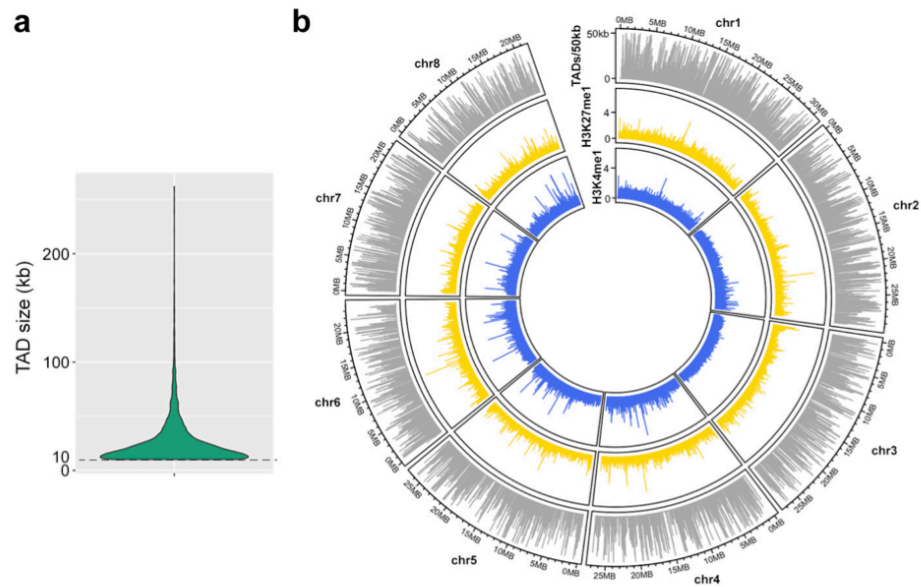
4 ***Marchantia* TCP transcription factor activity correlates with 3D chromatin structure**

5

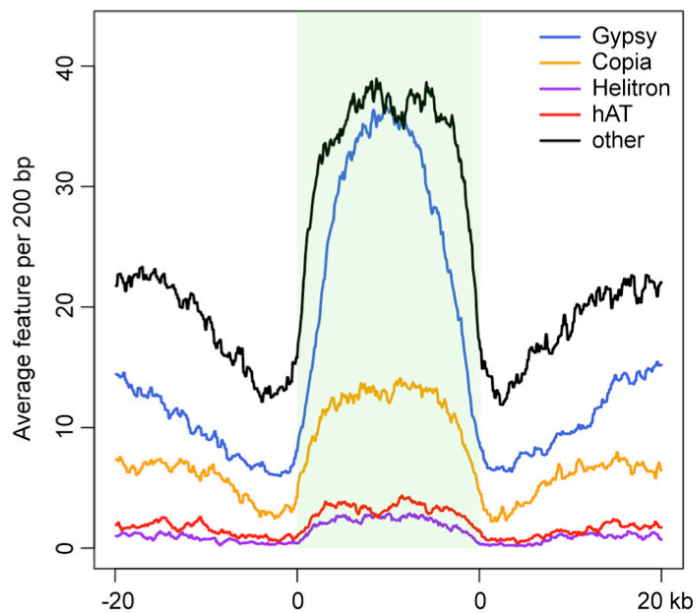
6 Ezgi Süheyla Karaaslan, Nan Wang, Natalie Faiss, Yuyu Liang, Sean A. Montgomery,

7 Sascha Laubinger, Kenneth Wayne Berendzen, Frédéric Berger, Holger Breuninger,

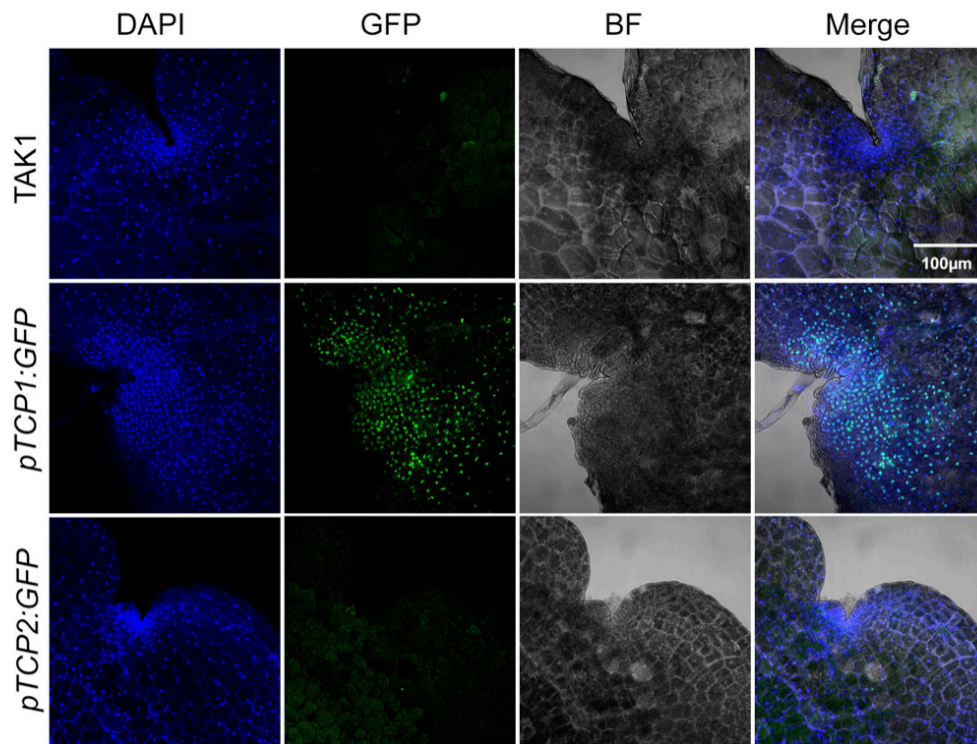
8 and Chang Liu



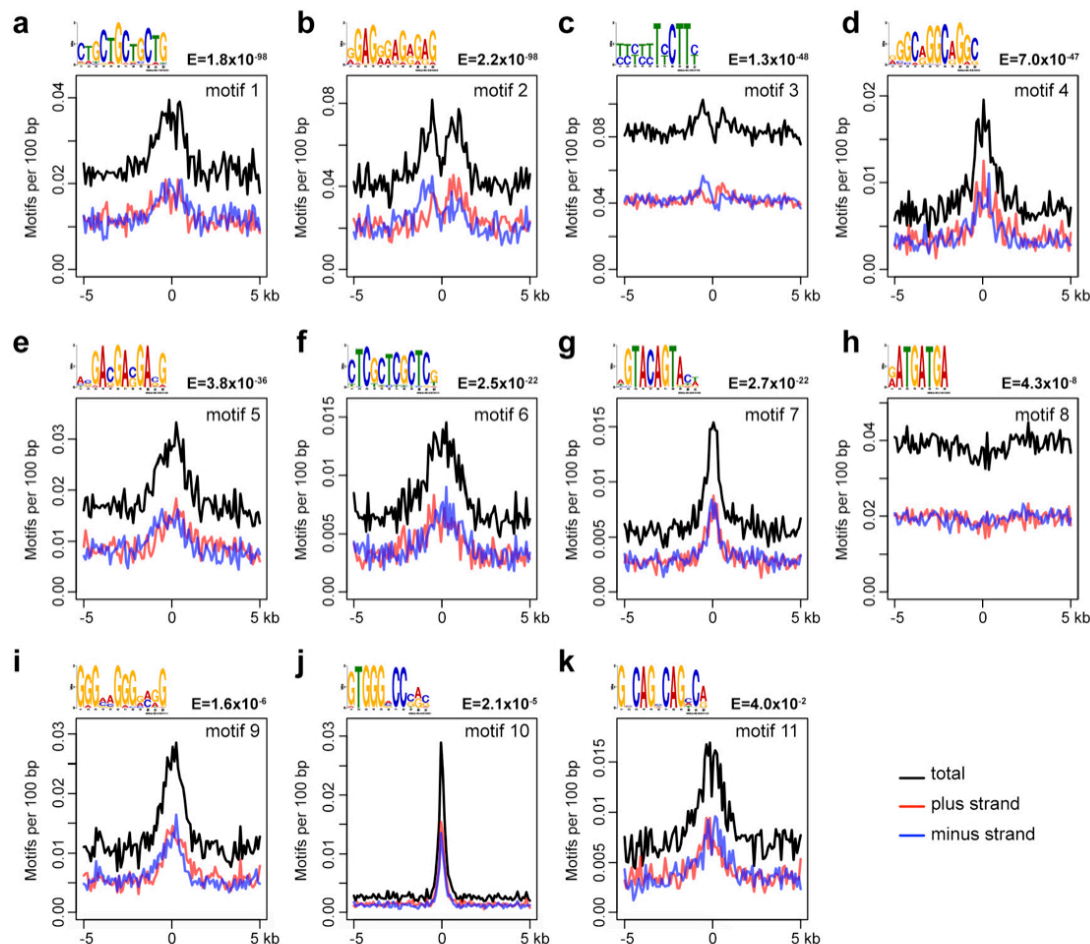
9
10 **Supplemental Figure 1 | Distribution of *Marchantia* TADs in the genome. a**, Size
11 distribution of TADs, the dotted line denotes the lower size limit that we applied for
12 calling TADs from 2 kb Hi-C maps. **b**, Distribution of genomic regions covered with
13 TAD annotation. The plot was generated with 50 kb windows. The H3K27me1 and
14 H3K4me1 (both are normalized against H3) tracks were from our previous study¹.
15



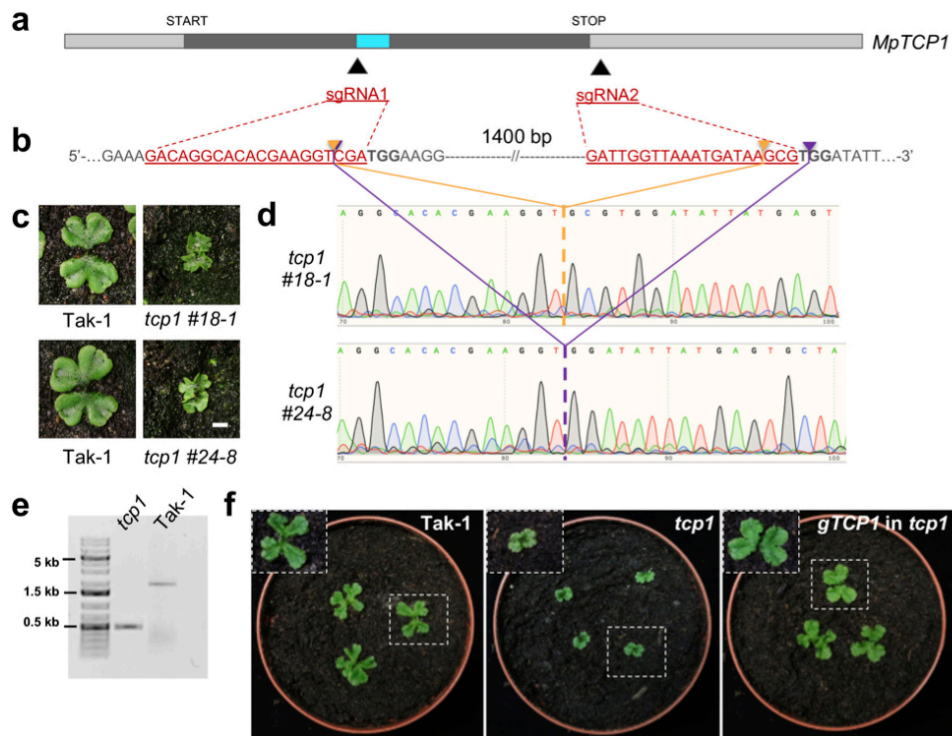
16
17 **Supplemental Figure 2 | Repetitive regions across *Marchantia* TADs.** Repeats
18 annotation was retrieved from the v5.1 Tak-1 reference¹. Features in TADs were
19 linearly transformed as in Fig. 1b.



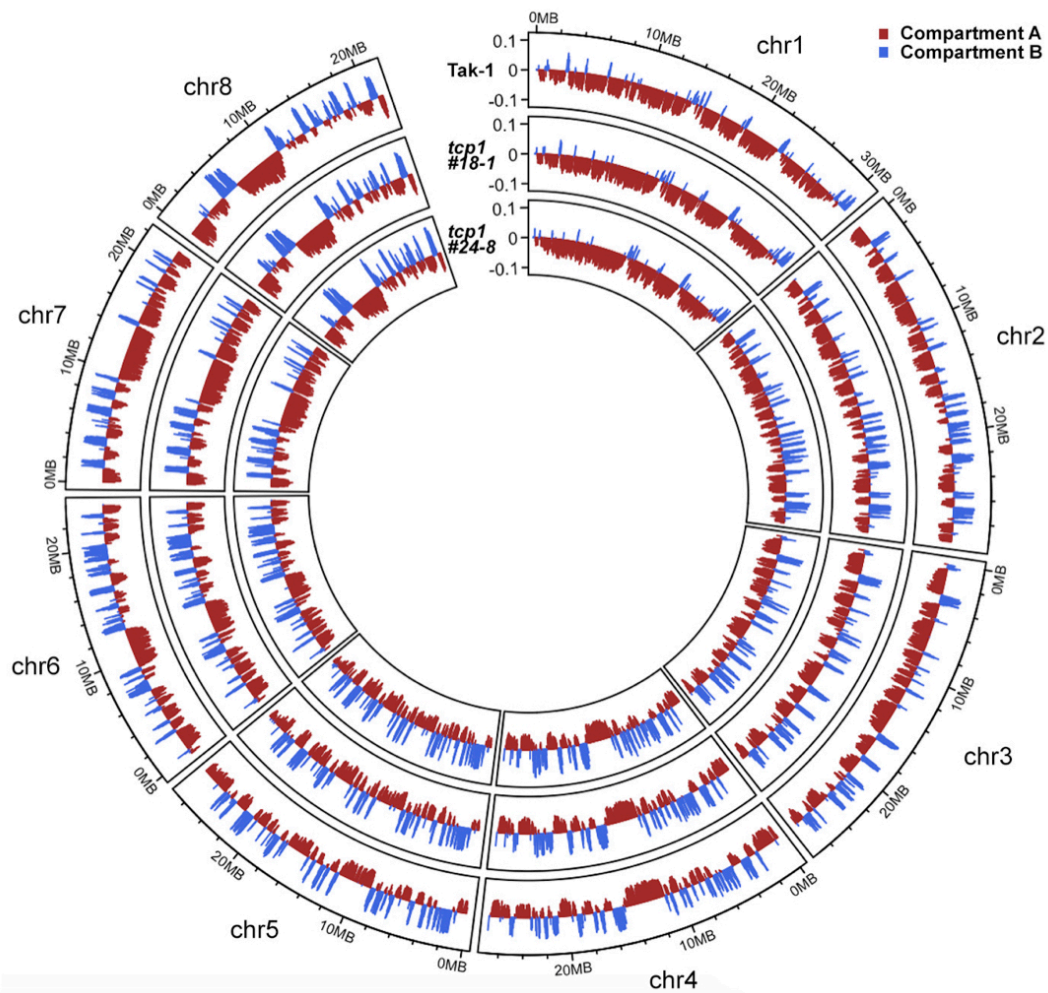
20
21 **Supplemental Figure 3 | Expression of *Marchantia* TCP genes.** Representative
22 confocal images of localization of DAPI stained 9 days old wild type Tak-1 and
23 transgenic thalli of *pTCP* driving GFP. Images are representatives from more than
24 fifteen independent lines with similar patterns.



25 **Supplemental Figure 4 | Motif analyses of chromatin regions bound by TCP1. a-**
 26 **k**, Distribution of individual motifs around TCP1-enriched regions. In each panel, all
 27 TCP1-enriched regions were piled with their central nucleotide aligned up (defined as
 28 "0"). Each motif was identified via *de novo* motif search, from which its position weight
 29 matrix was used for motif calling in query sequences. For motif calling, the
 30 "matchPWM" function (with the search stringency set to 90%) in the "Biostrings"
 31 package in R was used. Along with the motif logo, each motif's E value, which is the
 32 expected number of motifs found by chance, is given (see "methods" for details). By
 33 observing **a-k**, we concluded that the motif shown in **j** is the one directly cognized by
 34 TCP1, because it exhibited a much narrower peak centering at the mid-point of all
 35 enriched regions, as well as the highest peak relative to flanking regions, which
 36 collectively best explained sequence-dependent interactions between TCP1 and DNA.
 37 Other enriched motifs showing broader enrichment peaks likely reflect their co-
 38 existence with that in **j**, or the recruitment of TCP1 to chromatin via other interacting
 39 partners.



40 **Supplemental Figure 5 | Generation of *MpTCP1* knock-out mutant.** **a**, Structure of
 41 the *MpTCP1* locus. The dark segment is the CDS. The region corresponding to TCP1's
 42 DNA binding domain is shown as a cyan block. Black triangles indicate the locations
 43 of two sgRNAs used for CRISPR-CAS9 targeting. **b**, Sequences of two sgRNAs. Bold
 44 letters show the PAM (protospacer adjacent motif) sequences. The distance between
 45 these two sgRNAs is around 1400 bp. **c**, Phenotypes of two-week old *tcp1* mutants
 46 #18-1 and #24-8 compared to Tak-1. Scale bar is 5 mm. **d**, Sequencing results of 2
 47 different *tcp1* deletion alleles in which the deletion sites are linked to the sgRNA
 48 binding sites shown in panel **b**. **e**, DNA gel comparison of *TCP1* fragment in mutant and Tak-
 49 1 plants. Primers used for genotyping are indicated in Materials and Methods part. In
 50 total, five independent mutant lines were identified with similar genotyping results. **f**,
 51 Morphological comparisons of 2-week-old Tak-1, *tcp1* and a *tcp1* complementation
 52 line.



53 **Supplemental Figure 6 | Global A/B compartment analysis.** The A/B compartment
54 annotation of each genotype was based on 50-kb resolution Hi-C maps. For every
55 chromosome, we annotated the compartment bearing centromere¹ as “Compartment
56 B”.

57 **Reference**

58

- 59 1 Montgomery, S. A. *et al.* Chromatin Organization in Early Land Plants Reveals an
60 Ancestral Association between H3K27me3, Transposons, and Constitutive
61 Heterochromatin. *Curr Biol* **30**, 573-588 e577, doi:10.1016/j.cub.2019.12.015
62 (2020).
63

4 Discussion

Our knowledge regarding the non-random organization of interphase chromatin is significantly enhanced in the last two decades, thanks to the development in the imaging-based tools and chromosome conformation capture techniques⁹⁷. Interaction maps revealed hierarchical organization of the chromatin which ensures precise regulation of the gene expression in the nucleus⁹⁷. The initial findings regarding systematic folding of the genome have led to many fundamental questions of how the spatial positioning of the chromatin is established and how this domain organization contributes to the function of the genome.

Although imaging-based techniques provide us a colorful overview of how chromatin is spatially segregated in the nucleus, computational analysis of chromosome conformation techniques enabled us to study the detailed properties of 3D genome associated with DNA sequence. These high-throughput data empower us to analyze the dynamics of epigenetic landscape, TF activities, and gene expression in a 3D context.

4.1 TAD borders in plants

A notable feature of most mammalian TAD borders is the CTCF occupancy together with Cohesin complex^{12,14,22,98}. In plants, the absence of an insulator protein like CTCF suggests a divergent mechanism of TAD establishment in plants^{81,82}.

In our previous studies, we demonstrated that DNA binding motifs of TCP and bZIP family proteins are enriched at the borders of plant TADs^{57,71}. In this thesis, I showed TCP1 protein binding at *Marchantia* TAD borders by ChIP-seq⁹³. In order to analyze TCP1 function at TAD borders, we created *Marchantia tcp1* mutant. Hi-C analysis of this mutant demonstrated that the absence of TCP protein did not change the overall chromatin contact landscape, it rather induced a slight decrease in the insulation score of TCP1-bound TAD borders⁹³. Hence, these results suggest that missing TCP1 protein alone is not enough to alter the *Marchantia* TAD landscape.

Animal TAD borders are not only defined by CTCF binding. CTCF-independent TAD boundaries are found to be associated with transcription^{14,36} or the demarcation of A/B chromatin compartments^{12,35}. Moreover, TAD boundary region displays mainly euchromatic properties and host mainly active genes¹⁴. Our study confirm that *Marchantia* TAD borders are also enriched with euchromatic histone marks and TSSs of the genes⁹³. Consistently, in rice, active histone marks that are typically found at TSSs of the genes are also enriched at TAD borders⁵⁷. In other crops, TSS occupancy at TAD borders or compartmental domains has not been studied so far. This knowledge gap in plant chromatin topology should be filled urgently, because in a recent study in animals, it has been shown that the insertion of contact domain boundary regions can alter the genome topology⁹⁹. In animals, scientist inserted a 2kb DNA sequence containing CTCF binding site and/or TSS into the several parts of the genome, such as TAD bodies and TAD borders. These ectopic boundary insertions into the genome, demarcated new domains or strengthened the pre-existing boundaries. Inserting a fragment that solely contains TSS, to a region of interest, formed a new domain with A-compartment features⁹⁹. As plant compartmental domain borders are mainly associated with euchromatic landscape and lack insulator proteins, it is critical to unravel whether insertion of ectopic boundary sequences containing TSS induces changes in plant chromatin topology.

4.2 Intrinsic features of plant TADs

In their study, Dong and colleagues classify TADs into 4 categories according to their distinct epigenetic features: active (accessible chromatin), repressive (DNA methylation), polycomb silenced (enriched in the H3K27me3 mark) and intermediate type which lacks specific features⁵⁵. In the same study, a tight association between all these categories of TADs and A/B compartmental domains is found⁵⁵. These TAD domains mainly follow the epigenetic landscape of the higher hierarchical level in the nucleus.

Overall, *Marchantia* TAD bodies exhibit a heterochromatic landscape with heterochromatic histone marks and considerable DNA methylation. Nonetheless,

Marchantia TADs can be divided into 3 subgroups according to their DNA methylation status with arbitrarily set cut-offs. The first group of TADs (“mCG-rich”) have high CG, CHG and CHH methylation at their body. The second group (“mCG-poor”), consists of methylation-poor TADs, which has very low mCG at their TAD body. The last group with the fewest member is the intermediate group that exhibits intermediate methylation levels. Although mCG-poor TADs do not bear DNA methylation at their bodies, they are also depleted of active histone marks. This diversity of *Marchantia* TADs signifies that there exist distinct structure and regulatory regimes of plant TADs.

We also discovered a novel kind of TADs that belonged to the mCG-poor group. They are highly bound by TCP1 proteins and named as “TCP1-rich” TADs. Although TCP1-rich TADs are not decorated with repressive histone marks, the overlapping genes still display lower expression levels. TCP1-rich TADs provide a repressive environment for genes reside in them. Moreover, in the absence of TCP1 proteins, not only genes directly targeted by this TF, but also genes that are not directly targeted but found in TCP1-rich TADs are more differentially expressed, suggesting a 3D localization dependent expression control.

By inserting reporter genes to distinct places in the mouse genome, a study revealed that a reporter gene's expression correlates with the TAD landscape⁴³. Similarly, in order to identify effect of the intrinsic properties of plant TADs on gene expression, we can design reporter genes and observe their expression patterns across distinct type of TADs in *Marchantia* and other crops. It is plausible to estimate when the characteristic of the TAD is determined by the methylation status, newly introduced fragment might also acquire similar methylation patterns with the TAD of residence, therefore, an expression pattern parallel with innate features of TADs might be observed. In TCP1-rich TADs, this approach might be greatly enlightening to observe how the expression of a foreign fragment is regulated and changed.

Moreover, in *Arabidopsis*, Grob and Grossniklaus show that 3D chromatin interactions are linked to transgene silencing^{76,100}. In *Arabidopsis*, KNOT is a chromatin interaction

hub comprising 10 KNOT ENGAGED ELEMENTs (KEEs) and is mainly enriched with transposable elements⁶⁹. In this recent study, it is demonstrated that transgenes had ectopic contacts with specific KNOT regions in the genome, which promoted transgene silencing. High KNOT interaction frequency of transgenes correlated with their silencing and KNOT-linked silencing does not require DNA methylation¹⁰⁰. As TCP1-rich TADs display repressive properties without high levels of DNA methylation, these specific TADs in *Marchantia* might also acquire a non-canonical silencing pathway. To illustrate such function of the 3D chromatin compartments, it is important to initiate further functional analysis. Initially, it is critical to explore whether post-transcriptional silencing induced by small RNA (sRNA) is associated with TCP1-rich TAD dependent gene repression. To identify possible roles of sRNAs in gene repression in TCP1-rich TADs, we can conduct sRNA sequencing. We then should compare the abundance of sRNAs associated with genes found in TCP1-rich TADs with sRNAs associated with other genes. Moreover, to reveal the mechanisms of gene repression within the TCP1-rich TADs, it is important to find interaction partners of TCP1. Therefore, identifying interaction partners of TCP1 by mass spectrometry analysis might give us hints regarding possible pathways.

4.3 Co-expression of genes in TADs

Clusters of co-expressed genes in many different tissues and conditions in higher eukaryotes are observed prior to our expanding knowledge regarding the importance of spatial chromatin organization¹⁰¹. It was observed that co-expressed genes span a limited genomic distance, and the driving force behind gene co-expression is speculated to be uncharacterized cis-acting elements¹⁰¹. Nowadays, we appreciate mammalian TADs contributes to the regulation of gene expression by insulating the interaction between proximal enhancer-promoter elements. Detailed analysis of each TAD revealed that, genes in the same TAD are particularly co-expressed¹⁵. TAD landscape is associated with gene co-expression as genes found in the same TADs tend to share same cis-regulatory elements.

In maize, analysis by Dong and colleagues detected no significant co-regulation activity among genes residing in the same genomic compartments compared to random control gene pairs, which was not the case in mammalian TADs⁷⁹. On the contrary, our co-expression analysis based on different tissues in *Marchantia* revealed that, there are more co-expressed genes found in TADs than expected. Our results suggest that TADs across the *Marchantia* genome contributes to the co-regulation of the genes that are spatially adjacent.

Such a disparity between maize and *Marchantia* might be explained by high frequency of cis- and trans-interactions within the maize genome. Besides interaction along the diagonal of the maize Hi-C map, there are increased interactions between different chromosome arms which can be seen as dots at the end of chromosome (Figure 3A- upper circle). Moreover, in maize there are also recurrent trans interaction among centromeric regions of different chromosomes, which can be observed as a dot in the middle of the X shaped pattern (Figure 3A- lower circle). These results indicate that other than compartmental domains, there are frequent trans-interactions in the large

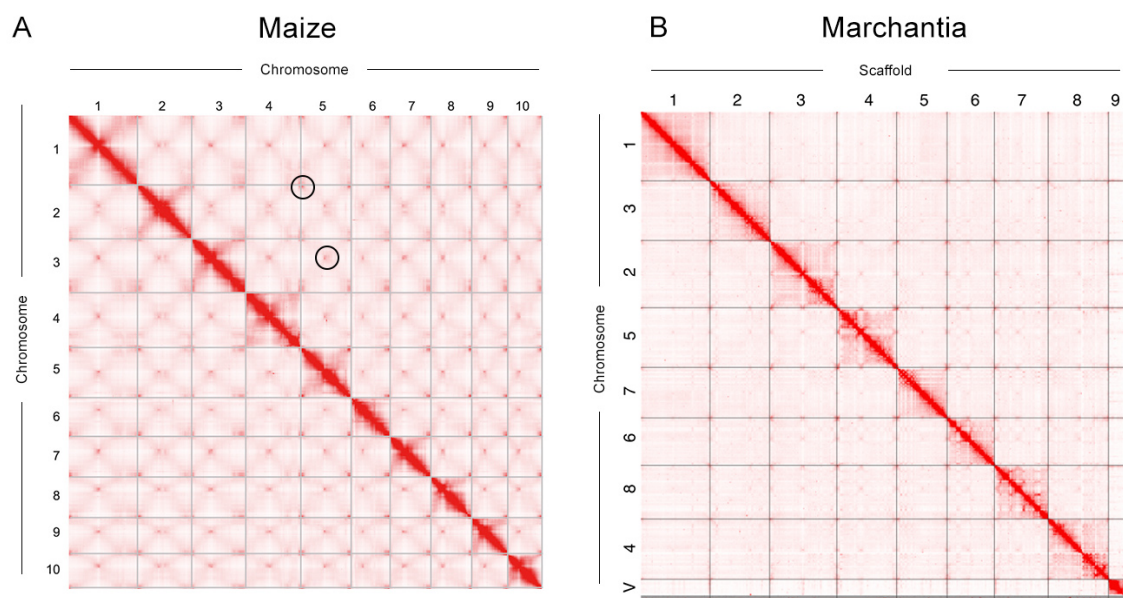


Figure 3: **Hi-C contact maps of Maize and Marchantia.** Genome wide contact matrix of maize (A) and *Marchantia* (B). Maize figure is adapted from⁵⁵ (Dong et al. 2017), and *Marchantia* map is adapted from⁸⁹ (Montgomery et al. 2020).

maize genome. Therefore, not only the genes found in intra-domain regions but also those found in inter-domain regions might also display comparable co-expression patterns due to the frequent interactions, resulting no detection of significance. In *Marchantia*, main chromatin interactions take place along the diagonal of Hi-C map (Figure 3B) suggesting inter-domains have less interaction compared to the maize. Therefore, in order to have more detailed overview regarding co-expression events in plant TADs, we need to analyse different plant species.

4.4 Potential role of TCP in nuclear Liquid-Liquid Phase Separation

Liquid-liquid phase separation (LLPS) can be described as assembly of dense droplet-like bodies separated from dilute phased environment, according to their biochemical properties¹⁰². Phase separation is not a new concept for nuclear organization as the membraneless compartment nucleolus is delineated in 1830s¹⁰³. Recently, it has been proposed that, biomolecular attractions between heterochromatin in nucleus that result in LLPS is a major driver of compartmentalization in genome organization^{39,102,104}.

Intrinsically disordered regions (IDR) are protein domains that grant flexibility to the 3D structure of the protein, promoting the coalescence of the proteins. TCP1 protein has an IDR domain close to its N-terminal region (Figure 4). Moreover, our cellular localization experiments of TCP protein confirmed that it forms speckles. Therefore, it is intriguing to show whether TCP containing speckles are phase separated liquid condensates. Moreover, it will be also interesting if the IDR domain of TCP contributes to bring repressed regions (e.g. TCP1-rich TADs) to close proximity to facilitate compartmental domain formation. Not only *Marchantia* TADs, but also A/B compartments in higher plants phase separation model might explain euchromatin/heterochromatin separation. Further supporting this hypothesis, temperature has been shown to be an important factor for interaction potential of proteins due to the temperature-dependent solvent-mediated interactions of each type of amino acid¹⁰⁵ and heat has been shown to weaken chromatin compartmentation¹⁰⁶.

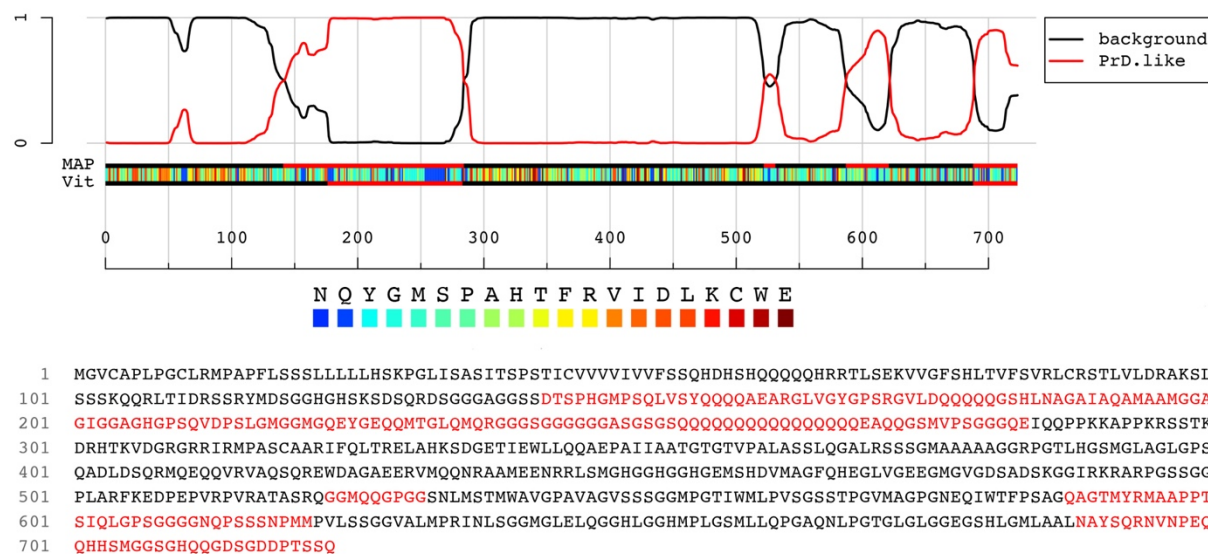


Figure 4: **IDR Domain of Marchantia TCP1 protein.** Upper panel shows the IDR domain annotation according to the PLAAC (Prion-Like Amino Acid Composition) server, black line indicates the background and red line prion-like domain¹⁰⁷. In the middle, there is a colour key for one letter codes of amino acids. Bottom panel is the protein sequence of MpTCP1, red letters indicate amino acids in prion-like domain.

4.5 Other candidate proteins in Marchantia 3D genome

Among mCG-poor TADs in *Marchantia*, less than half of them can be categorized as TCP1-rich TADs. Therefore, the remaining mCG-poor TADs might be associated with other transcription factors. In order to unravel this fact, we performed further motif analysis at *Marchantia* TAD bodies and borders. Our motif analysis showed that not only TCP binding sites, but also GAGA-binding motif is enriched at TAD boundary regions and the bodies of mCG-poor TADs⁹³. GAGA binding motif is recognized by BBR/BPC family transcription factors in plants¹⁰⁸. In the *Marchantia* genome, there are two BBR/BPC protein ortholog, each resides in one sex chromosome. Therefore, per *Marchantia* individual there is only one BPC gene present.

Our phylogenetic analysis showed that *Marchantia* BPC has a very conserved DNA binding domain (Figure 5), strengthening our assumption that *Marchantia* BBR/BPC will recognize a similar GAGA-binding motif and is, therefore, enriched at *Marchantia* TAD borders. Of course, in order to confirm this hypothesis, it is required to show the sites of BPC protein binding with ChIP experiments. If BBR/BPC protein is indeed enriched at the border of the TADs and at the body of mCG poor TADs, this could indicate an interplay of multiple transcription factors in the TAD establishment in *Marchantia*. Double mutant of *TCP/BPC* will be also further needed to examine potential contact map changes in the *Marchantia* genome.

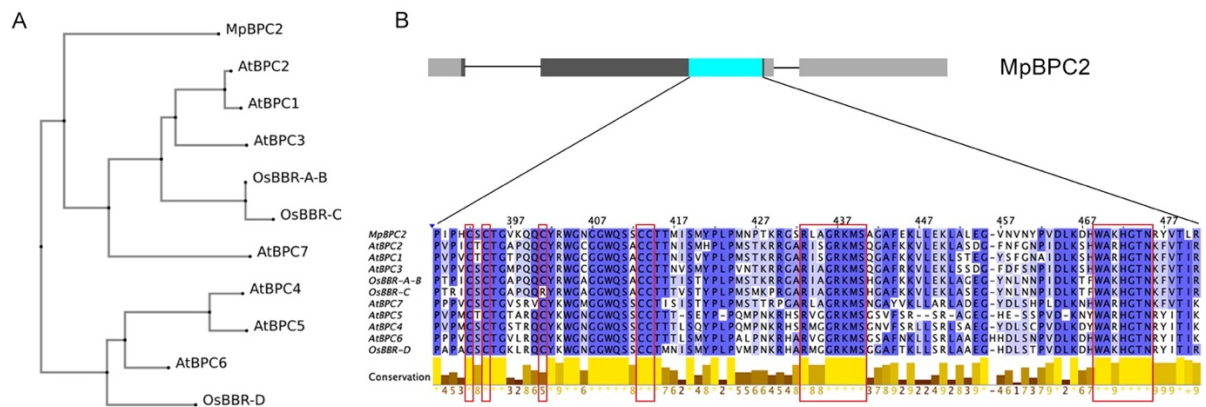


Figure 5. **Marchantia BBR/BPC family protein.** (A) Phylogenetic tree of *Marchantia* BPC2 (BBR/BPC member in TAK1 genotype) with *Arabidopsis* and rice family proteins. (B) Conserved DNA binding domain of BPC2. Cyan block indicates DNA binding domain at the *Marchantia* BPC2 gene locus. Dark grey bars indicate exons, straight lines indicate introns and light grey bars indicate UTRs. Red boxes indicate the important amino acids for GAGA-binding motif recognition¹⁰⁸.

References

1. Annunziato, A. DNA packaging: nucleosomes and chromatin. *Nature Education* **1**, 26 (2008).
2. Han, J., Zhang, Z. & Wang, K. 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. *Mol. Cytogenet.* **11**, 21 (2018).
3. Van Berkum, N. L. *et al.* Hi-C: a method to study the three-dimensional architecture of genomes. *JoVE (Journal of Visualized Experiments)* e1869 (2010).
4. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
5. Pinkel, D. *et al.* Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9138–9142 (1988).
6. Lichter, P., Cremer, T., Borden, J., Manuelidis, L. & Ward, D. C. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum. Genet.* **80**, 224–234 (1988).
7. Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157 (2005).
8. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
9. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, e138 (2006).
10. Wijchers, P. J. *et al.* Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Mol. Cell* **61**, 461–473 (2016).
11. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
12. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
13. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
14. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

15. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
16. Merkenschlager, M. & Nora, E. P. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genomics Hum. Genet.* **17**, 17–43 (2016).
17. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci Adv* **5**, eaaw1668 (2019).
18. Ji, X. *et al.* 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* **18**, 262–275 (2016).
19. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
20. Gómez-Marín, C. *et al.* Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7542–7547 (2015).
21. Krefting, J., Andrade-Navarro, M. A. & Ibn-Salem, J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol.* **16**, 87 (2018).
22. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309 (2015).
23. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
24. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–910 (2015).
25. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
26. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930–944.e22 (2017).
27. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 996–1001 (2014).
28. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6456–65 (2015).
29. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.* **82**, 45–55 (2017).

30. Murayama, Y., Samora, C. P., Kurokawa, Y., Iwasaki, H. & Uhlmann, F. Establishment of DNA-DNA Interactions by the Cohesin Ring. *Cell* **172**, 465–477.e15 (2018).
31. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
32. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–3129 (2013).
33. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24 (2017).
34. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
35. Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* (2017) doi:10.1016/j.molcel.2017.07.022.
36. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557–572.e24 (2017).
37. Cook, P. R. & Marenduzzo, D. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res.* **46**, 9895–9906 (2018).
38. Fishman, V. *et al.* 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Res.* **47**, 648–665 (2019).
39. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
40. Rada-Iglesias, A., Grosveld, F. G. & Papanonis, A. Forces driving the three-dimensional folding of eukaryotic genomes. *Mol. Syst. Biol.* **14**, e8214 (2018).
41. Zhang, Y. *et al.* Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51**, 1380–1388 (2019).
42. Paulsen, J. *et al.* Long-range interactions between topologically associating domains shape the four-dimensional genome during differentiation. *Nat. Genet.* **51**, 835–843 (2019).
43. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).

44. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
45. Redin, C. *et al.* The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.* **49**, 36–45 (2017).
46. Ibn-Salem, J. *et al.* Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* **15**, 423 (2014).
47. Lupiáñez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet.* **32**, 225–237 (2016).
48. Akdemir, K. C. *et al.* Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat. Genet.* **52**, 294–305 (2020).
49. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
50. Taberlay, P. C. *et al.* Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res.* **26**, 719–731 (2016).
51. Weischenfeldt, J. *et al.* Pan-cancer analysis of somatic copy-number alterations implicates *IRS4* and *IGF2* in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2017).
52. Achinger-Kawecka, J., Taberlay, P. C. & Clark, S. J. Alterations in Three-Dimensional Organization of the Cancer Genome and Epigenome. *Cold Spring Harb. Symp. Quant. Biol.* **81**, 41–51 (2016).
53. Barrington, C. *et al.* Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. *Nat. Commun.* **10**, 2908 (2019).
54. Ulianov, S. V. *et al.* Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* (2015) doi:10.1101/gr.196006.115.
55. Dong, P. *et al.* 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* **0**, (2017).
56. Dong, Q. *et al.* Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice. *Plant J.* (2018) doi:10.1111/tpj.13925.
57. Liu, C., Cheng, Y.-J., Wang, J.-W. & Weigel, D. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. *Nat Plants* (2017) doi:10.1038/s41477-017-0005-9.

58. Zhu, W. *et al.* Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific Arabidopsis hybrid. *Genome Biol.* **18**, 157 (2017).
59. Hu, L. *et al.* The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nat. Commun.* **10**, 4702 (2019).
60. Hu, Y. *et al.* Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* **51**, 739–748 (2019).
61. Jibrán, R. *et al.* Chromosome-scale scaffolding of the black raspberry (*Rubus occidentalis* L.) genome based on chromatin interaction data. *Hortic Res* **5**, 8 (2018).
62. Maccaferri, M. *et al.* Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.* **51**, 885–895 (2019).
63. Mascher, M. *et al.* A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427–433 (2017).
64. Raymond, O. *et al.* The Rosa genome provides new insights into the domestication of modern roses. *Nat. Genet.* **50**, 772–777 (2018).
65. Shi, J. *et al.* Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nat. Commun.* **10**, 464 (2019).
66. VanBuren, R. *et al.* Desiccation Tolerance Evolved through Gene Duplication and Network Rewiring in Lindernia. *Plant Cell* **30**, 2943–2958 (2018).
67. Zhang, L. *et al.* Improved Brassica rapa reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Hortic Res* **5**, 50 (2018).
68. Feng, S. *et al.* Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. *Mol. Cell* **55**, 694–707 (2014).
69. Grob, S., Schmid, M. W. & Grossniklaus, U. Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. *Mol. Cell* **55**, 678–693 (2014).
70. Wang, C. *et al.* Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. *Genome Res.* **25**, 246–256 (2015).
71. Doğan, E. S. & Liu, C. Three-dimensional chromatin packing and positioning of plant genomes. *Nature Plants* (2018) doi:10.1038/s41477-018-0199-5.
72. Stam, M., Tark-Dame, M. & Fransz, P. 3D genome organization: a role for phase separation and loop extrusion? *Curr. Opin. Plant Biol.* **48**, 36–46 (2019).

73. dos Santos, G. *et al.* FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.* **43**, D690-7 (2015).
74. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
75. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
76. Pontvianne, F. & Grob, S. Three-dimensional nuclear organization in *Arabidopsis thaliana*. *J. Plant Res.* (2020) doi:10.1007/s10265-020-01185-0.
77. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).
78. Grob, S. & Grossniklaus, U. Chromosome conformation capture-based studies reveal novel features of plant nuclear architecture. *Curr. Opin. Plant Biol.* **36**, 149–157 (2017).
79. Dong, P. *et al.* Tissue-specific Hi-C analyses of rice, foxtail millet and maize suggest non-canonical function of plant chromatin domains. *J. Integr. Plant Biol.* **62**, 201–217 (2020).
80. Dong, P., Tu, X., Liang, Z., Kang, B.-H. & Zhong, S. Plant and animal chromatin three-dimensional organization: similar structures but different functions. *Journal of Experimental Botany* vol. 71 5119–5128 (2020).
81. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17507–17512 (2012).
82. Heger, P. & Wiehe, T. New tools in the box: an evolutionary synopsis of chromatin insulators. *Trends Genet.* **30**, 161–171 (2014).
83. Shimamura, M. *Marchantia polymorpha*: Taxonomy, Phylogeny and Morphology of a Model System. *Plant Cell Physiol.* **57**, 230–256 (2016).
84. Ishizaki, K., Chiyoda, S., Yamato, K. T. & Kohchi, T. Agrobacterium-mediated transformation of the haploid liverwort *Marchantia polymorpha* L., an emerging model for plant biology. *Plant Cell Physiol.* **49**, 1084–1091 (2008).
85. Ishizaki, K., Johzuka-Hisatomi, Y., Ishida, S., Iida, S. & Kohchi, T. Homologous recombination-mediated gene targeting in the liverwort *Marchantia polymorpha* L. *Sci. Rep.* **3**, 1532 (2013).

86. Kubota, A., Ishizaki, K., Hosaka, M. & Kohchi, T. Efficient Agrobacterium-mediated transformation of the liverwort *Marchantia polymorpha* using regenerating thalli. *Biosci. Biotechnol. Biochem.* **77**, 167–172 (2013).
87. Sugano, S. S. *et al.* CRISPR/Cas9 Mediated Targeted Mutagenesis in the Liverwort *Marchantia polymorpha* L. *Plant Cell Physiol.* cu014 (2014).
88. Bowman, J. L. *et al.* Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell* **171**, 287–304.e15 (2017).
89. Montgomery, S. A. *et al.* Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Curr. Biol.* **30**, 573–588.e7 (2020).
90. Ishizaki, K., Nishihama, R., Yamato, K. T. & Kohchi, T. Molecular Genetic Tools and Techniques for *Marchantia polymorpha* Research. *Plant Cell Physiol.* **57**, 262–270 (2016).
91. Kato, H., Yasui, Y. & Ishizaki, K. Gemma cup and gemma development in *Marchantia polymorpha*. *New Phytol.* (2020) doi:10.1111/nph.16655.
92. Aguilar-Cruz, A., Grimanelli, D., Haseloff, J. & Arteaga-Vázquez, M. A. DNA methylation in *Marchantia polymorpha*. *New Phytol.* **223**, 575–581 (2019).
93. Karaaslan, E. S. *et al.* *Marchantia* TCP transcription factor activity correlates with three-dimensional chromatin structure. *Nature Plants* (2020) doi:10.1038/s41477-020-00766-0.
94. Hirakawa, Y. *et al.* Control of proliferation in the haploid meristem by CLE peptide signaling in *Marchantia polymorpha*. *PLoS Genet.* **15**, e1007997 (2019).
95. Aki, S. S. *et al.* Cytokinin Signaling Is Essential for Organ Formation in *Marchantia polymorpha*. *Plant Cell Physiol.* **60**, 1842–1854 (2019).
96. Linde, A.-M., Sawangproh, W., Cronberg, N., Szövényi, P. & Lagercrantz, U. Evolutionary History of the *Marchantia polymorpha* Complex. *Front. Plant Sci.* **11**, 829 (2020).
97. Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289 (2017).
98. Phillips-Cremins, J. E. & Corces, V. G. Chromatin insulators: linking genome organization to cellular function. *Mol. Cell* **50**, 461–474 (2013).
99. Zhang, D. *et al.* Alteration of genome folding via contact domain boundary insertion. *Nat. Genet.* (2020) doi:10.1038/s41588-020-0680-8.

100. Grob, S. & Grossniklaus, U. Invasive DNA elements modify the nuclear architecture of their insertion site by KNOT-linked silencing in *Arabidopsis thaliana*. *Genome Biol.* **20**, 120 (2019).
101. Purmann, A. *et al.* Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality. *Genomics* **89**, 580–587 (2007).
102. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
103. Valentin, G. *Repertorium für Anatomie und Physiologie*. (books.google.com, 1837).
104. Larson, A. G. *et al.* Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
105. Dignon, G. L., Zheng, W., Kim, Y. C. & Mittal, J. Temperature-Controlled Liquid-Liquid Phase Separation of Disordered Proteins. *ACS Cent. Sci.* **5**, 821–830 (2019).
106. Sun, L. *et al.* Heat stress-induced transposon activation correlates with 3D chromatin organization rearrangement in *Arabidopsis*. *Nat. Commun.* **11**, 1886 (2020).
107. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502 (2014).
108. Theune, M. L., Bloss, U., Brand, L. H., Ladwig, F. & Wanke, D. Phylogenetic Analyses and GAGA-Motif Binding Studies of BBR/BPC Proteins Lend to Clues in GAGA-Motif Recognition and a Regulatory Role in Brassinosteroid Signaling. *Frontiers in Plant Science* vol. 10 (2019).

Acknowledgements

First, I would like to thank Jun.-Prof. Chang Liu for his supervision and the opportunity to do my doctoral studies in his laboratory. It was a great pleasure to explore puzzles of plant genetics under your continuous patience, support and trust.

I also would like to thank my TAC committee members Prof. Ulrike Zentgraf and Prof. Thomas Lahaye for their valuable advices on the way and evaluation of my thesis. I also would like to thank Dr. Holger Breuninger for introducing me to a new model organism and helping me to overcome intricacies of the work.

I would like to thank my bench-mates in the lab: Natalie, Nan, Ingrid. Special thanks to Natalie for easing my work substantially, German lessons and very pleasant lunch breaks. I was very lucky to have a bestie in the same office; Nan. I want to thank you for your friendship, all the laughter, interesting stories, enlightening discussions about Chinese food and history. You were my bubble of joy during these years.

I would like to thank both former and current members of General Genetics department for a wonderful work atmosphere: Kyrylo, Mayank, Maria, Betty, Stefan, Erin, Trang, Danalyn, Suayib, Paul, Lena, Theresa, Jorinde, Rotem, Guillaume... To Charley, for helping me tons with German bureaucracy. Swathi and Moka for horror stories around the camp fire and sharing their painting and cooking skills. Special thanks to Tine for securing me on the rope, cheering me on the wall, swimming in Steinlach, for wedding dances, camping trips, Schelling parties, bartending experiences, for all the wine and strawberries.

I would like to thank my Teletübies Bilge, Efe and Caner. Thank you for all the jokes, games, beer, white football, bike trips, beach, barbecue, and brownies. To Efe, for his endless support on my YouTube channel. To Caner, for answering all the R questions and helping me tons. To my sweet Bilge, for climbing, pedalling, roller blading, very good food and shopping tips.. You are my adventure-mate. For Catan and Carcassonne nights to Ali and Direnc. For spontaneous encounters to Sergio, Elena and Gülüm. For chill couples' evenings and Swiss trips to Mustafa and Selen.

Thanks to my Stuttgart squad: Pinar, Gözde, Efe, Sitar, Berkan, Ekin.. Thank you, guys, for all the breakfasts, cheesecakes, parties and sleepovers. Special thanks to Pinar for being such an amazing friend for many years, not-that amazing vacation planner, most importantly for sharing the life, tea and comfort. To my Kölsch lady Nur, for camps, hikes, carnivals, concerts, all the tequila and red lipstick. To my Izmir girl Melike for being by my side all these years, for pop-up Amsterdam trips, for turquoise vacations.

To my family, my parents-in-law, grandparents and parents for love and support. Thank you all for raising me in a way that I love the life and all the colours of it. Thank you for teaching me the importance of hard-work and commitment. To my dad showing me spring was there and trees were blossoming in one bitter day of mine. To my mom, for passing me her passion about nature and cats.

Finally, I would like to thank to the person who I share life with: Emre Baris. Your amazing spirit and endless jokes can make me laugh and cheer under any circumstances. Thank you for renovating one empty apartment with me, painting it, decorating it and more importantly filling it with lots of love, music and flowers to make it our home. Also, for all the sassy song-writing, delicious barbecue, football games, bike trips, HDfilmcehennemi and chill sessions, yoga and balcony nights. I am so lucky that we share the same travel enthusiasm for road trips, weekend getaways, adventures on other continents... Thank you for sunsets over the sea, swimming in the oceans, surfing, diving, sunburns, port wine, exploring new food, new people with me. Life is amazing with you. And with our fluffy new family member Rocky... Thank you for your unfailing love, support and understanding during my pursuit of PhD degree. I am looking forward to a future with you...

Ezgi

Curriculum Vitae

Ezgi Süheyla Karaaslan

Education

- 2017–current **PhD Thesis**, *General Genetics*, ZMBP University Tübingen, Tübingen.
Supervisor: Chang Liu
- 2015–2016 **Master's Thesis**, *Max Planck Institute for Developmental Biology*, Tübingen.
Supervisor: Detlef Weigel
- 2014–2016 **Master of Science in Molecular Life Sciences**, *Friedrich-Schiller-University*, Jena,
1.50/1.00.
- 2009–2014 **Bachelors of Science in Molecular Biology and Genetics**, *Middle East Technical University*, Ankara, 3.15/4.00 (German grading 1.85).

Publications

Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure, *Nature Plants* (2020), DOI: 10.1038/s41477-020-00766-0.

Ezgi Süheyla Karaaslan, Nan Wang, Natalie Faiss, Yuyu Liang, Sean A. Montgomery, Sascha Laubinger, Kenneth Wayne Berendzen, Frédéric Berger, Holger Breuning, and Chang Liu.

Chromatin packing and positioning of plant genomes in 3D, *Nature Plants* (2018), DOI: 10.1038/s41477-018-0199-5.

Ezgi Süheyla Dogan and Chang Liu.

Plant Lamin-like Proteins Mediate Chromatin Tethering at the Nuclear Periphery, *Genome Biology* (2019), DOI: 10.1186/s13059-019-1694-3.

Bo Hu*, Nan Wang*, Xiuli Bi*, Ezgi Süheyla Karaaslan, Anna-Lena Weber, Kenneth Wayne Berendzen and Chang Liu.

RST1 and RIPR connect the cytosolic RNA exosome to the Ski complex in Arabidopsis, *Nature Communications* (2019), DOI: 10.1038/s41467-019-11807-4.

Heike Lange, Simon Y. A. Ndecky, Carlos Gomez-Diaz, David Pflieger, Nicolas Butel, Julie Zumsteg, Lauriane Kuhn, Christina Piermaria, Johana Chicher, Michael Christie, Ezgi Süheyla Karaaslan, Patricia L. M. Lang, Detlef Weigel, Herve Vaucheret, Philippe Hammann and Dominique Gagliardi.

A Role for the F-Box Protein HAWAIIAN SKIRT in Plant microRNA Function, *Plant Physiology* (2018), DOI:10.1104/pp.17.01313.

Patricia L. M. Lang*, Michael D. Christie*, Ezgi Süheyla Dogan, Rebecca Schwab, Jörg Hagmann, Anna-Lena van de Weyer, Emanuele Scacchi, Detlef Weigel.

Weißdornweg 14/183 – 72076, Tübingen – Germany

☎ +49 (157) 374 52662 • ✉ ezgi.dogan@zmbp.uni-tuebingen.de

01.09.1991, Turkish

1/2

Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific *Arabidopsis* hybrid, *Genome Biology* (2017), DOI: 10.1186/s13059-017-1281-4.

Wangsheng Zhu*, Bo Hu, Claude Becker, [Ezgi Süheyla Dogan](#), Kenneth Wayne Berendzen, Detlef Weigel and Chang Liu*.

Languages

Turkish **Native speaker**
English **Advanced**
German **Intermediate**

TOEFL IBT: 102
B2

Awards

2009 University Entrance Exam-7233th among 1.5 million participants
2012-2014 METU Honor Student
2014-2016 DAAD Scholarship

Additional Skills

- Confocal microscopy
- MS Office, Adobe Photoshop, Adobe Illustrator, Fiji
- Unix, R, Python (basic), Java (basic)

References

Supervisors	e-mail address
○ Chang Liu	○ chang.liu@zmbp.uni-tuebingen.de
○ Detlef Weigel	○ detlef.weigel@tuebingen.mpg.de
○ Thomas Lahaye	○ thomas.lahaye@zmbp.uni-tuebingen.de