

**Exploiting native forces  
to capture chromosome conformation  
in mammalian cell nuclei**

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln



vorgelegt von

Lilija Brant

Köln 2017



Promotionskomitee:

Prof. Dr. Björn Schumacher

Prof. Dr. Peter Nürnberg

Prof. Dr. Siegfried Roth

Tag der mündlichen Prüfung:

13.11.2017

*The most beautiful thing we can experience is the mysterious.*

*It is the source of all true art and science.*

*Albert Einstein*

*Für meine Eltern*

# Table of Contents

Abstract .....	9
Zusammenfassung .....	10
1. Introduction .....	12
1.1 <i>Genome architecture in three-dimensional space</i> .....	12
1.1.1 <i>Chromosome conformation capture (3C)-based technologies</i> .....	13
1.1.2 <i>The hierarchy of genome organization: Interpreting chromatin interaction data</i> .....	16
1.2 <i>Evaluation of 3C-based methods and their biases</i> .....	19
1.2.1 <i>In vivo formaldehyde cross-linking: scientists are flying blind</i> .....	20
1.2.2 <i>Experimental pitfalls of conventional 3C technology</i> .....	22
1.3 <i>Genome organization under physiological conditions</i> .....	22
1.4 <i>Aim</i> .....	23
2. Material and Methods .....	25
2.1 <i>Material</i> .....	25
2.1.1 <i>Chemicals</i> .....	25
2.1.2 <i>Buffers and solutions</i> .....	27
2.1.3 <i>Enzymes</i> .....	29
2.1.4 <i>Kits and systems</i> .....	30
2.1.5 <i>Equipment</i> .....	30
2.1.6 <i>General reagents used for molecular biology</i> .....	31
2.1.7 <i>Antibiotics</i> .....	32
2.1.8 <i>Primers</i> .....	32
2.2 <i>Methods</i> .....	34
2.2.1 <i>Molecular cloning</i> .....	34
2.2.1.1 <i>Bacteria strains</i> .....	34
2.2.1.2 <i>Bacterial artificial chromosomes (BAC's)</i> .....	35
2.2.1.3 <i>Plasmids</i> .....	35
2.2.1.4 <i>Generation of transformation-competent E. coli bacteria</i> .....	35
2.2.1.5 <i>Transformation of E. coli</i> .....	36
2.2.1.6 <i>Construction of TALE-DNA methyltransferase (Dam) fusions</i> .....	36
2.2.2 <i>Mammalian cell culture</i> .....	37
2.2.2.1 <i>Cell lines and media composition</i> .....	37

2.2.2.2 Cell line treatments .....	38
2.2.2.3 Cell transfection via nucleofection.....	38
2.2.3 Molecular biology methods .....	39
2.2.3.1 Nucleic acids isolation .....	39
2.2.3.2 Polymerase chain reaction (PCR).....	40
2.2.3.3 Reverse-Transcriptase cDNA Synthesis .....	40
2.2.3.4 Quantitative real-time PCR (qPCR) .....	40
2.2.3.5 Agarose gel electrophoresis .....	42
2.2.4 Chromosome conformation capture technology .....	42
2.2.4.1 Intrinsic 3C (i3C) .....	42
2.2.4.2 Conventional 3C.....	43
2.2.4.3 i4C/4C-seq.....	44
2.2.4.4 iT2C and T2C.....	46
2.2.4.5 Whole-genome iHi-C implementation .....	46
2.2.4.6 TALE-iD .....	47
3. Results.....	49
3.1 Features of intrinsic chromosome conformation capture (i3C) methodology ....	49
3.1.1 Establishment of the i3C protocol.....	49
3.1.2 Chromatin features of the i3C template .....	51
3.1.3 Comparison of i3C and conventional 3C templates .....	53
3.1.4 Capturing native contacts: obtaining the first i3C interaction profile .....	55
3.2 i4C technology: a native one-to-all approach.....	57
3.2.1 Implementation of i4C on the SAMD4A locus .....	57
3.2.1.1 Comparison of i4C and conventional 4C data on the SAMD4A locus.....	58
3.2.1.2 Contact signal versus background noise: a closer look on i4C profiles ...	63
3.2.1.3 SAMD4A locus i4C using various cell types and restriction enzymes.....	65
3.2.2 Implementation of i4C in different loci.....	67
3.2.2.1 i4C implementation in gene-poor and hetero-chromatinized loci.....	67
3.2.2.2 Cis-regulatory interactions are mostly confined within TADs .....	69
3.2.2.3 Capturing inter-chromosomal contacts with i4C .....	71
3.2.3 Enrichment of i4C signal allows reduction of input material .....	72
3.2.4 It is all about controls: different conditions produce different outcomes .....	74
3.2.5 Transcription factories: factory-i4C captures chromatin interactions .....	76
3.3 TALE-iD: an orthogonal approach to verify native chromatin looping events....	77

3.4 <i>iT2C: an intrinsic Targeted Chromatin Capture method</i> .....	79
3.4.1 <i>Implementation of iT2C on a 2.8-Mbp locus of chromosome 14</i> .....	80
3.4.2 <i>Characteristics of iT2C and conventional T2C interaction maps</i> .....	82
3.4.3 <i>Implementation of iT2C in a different cell line</i> .....	84
3.5 <i>Proof of principle iHi-C: generation of intrinsic genome-wide contact maps</i> .....	85
3.5.1 <i>Implementation of iHi-C</i> .....	86
3.5.2 <i>What is lost? iHiC generated in ‘retained’ and ‘lost’ fractions</i> .....	89
3.5.3 <i>Improvement of ligation efficiency leads to better contact enrichment</i> .....	91
3.6 <i>Dynamics of chromatin interactions upon cytokine signalling</i> .....	93
3.6.1 <i>Promoter-enhancer crosstalk after TNF<math>\alpha</math> stimulation using 3C or i3C</i> .....	94
3.6.2 <i>Tracking chromatin interaction dynamics on chromosome 14 using i4C</i> ....	95
3.6.3 <i>Analysis of chromatin dynamics on chromosome 14 using iT2C</i> .....	100
4. Discussion .....	105
4.1 <i>Features of i3C-based techniques</i> .....	105
4.2 <i>Role of i3C-based assays in to date genome organization knowledge</i> .....	110
4.3 <i>Static and dynamic chromatin: what forces shape the genome?</i> .....	112
4.4 <i>Topologically-associated domains: are they key mediators of chromatin folding?</i> .....	114
4.5 <i>Future perspectives</i> .....	115
References .....	117
Figure Index.....	127
Table Index.....	129
List of Abbreviations .....	130
Acknowledgement .....	132
Eidesstattliche Erklärung .....	134
Curriculum vitae.....	135



## Abstract

Mammalian cells contain the same 2 m-long DNA molecule, but different cell fates imply different three-dimensional genome organization that is tightly linked to gene expression and other cellular functions. This can range from shorter-range chromatin looping between genomic elements such as promoters and enhancers to higher-order folding of whole chromosomes within the confines of cell nuclei. In the past two decades, our understanding of this complex genome architecture and its underlying functions has been advanced through microscopy studies and chromosome conformation capture (3C) techniques. Fixation of the spatial chromatin network relies on cell cross-linking using formaldehyde and is a major experimental component of most molecular biological methods, including 3C-based approaches. However, this step remains a 'black box' since it may introduce biases that may skew the resulting data. In fact, such skewing between microscopy and 3C-derived data was recently reported. Here, we address these concerns and provide a novel intrinsic chromosome conformation capture assay (i3C), which allows determination of spatial chromatin interactions without the need for cell cross-linking. We handle mammalian cell nuclei in a close-to-physiological buffer that exploits native forces, such as transcription, to preserve nuclear chromatin folding. We introduce different variations of our intrinsic method that enable us to investigate locus-specific to genome-wide interactions between DNA segments. Using distinct cell lines and chromatin characteristics of the examined loci, we show that most of the native contact signals resemble cross-linked ones and reside within the constraints of topological associated domains (TADs). However, i3C-based tools increase the signal-to-noise ratio and enhance precision of contact determination at regulatory elements and CTCF sites. Moreover, we observed differential contacts by intrinsic techniques and developed an orthogonal validation method called 'TALE-iD'. Finally, we utilized extracellular signaling by the tumor necrosis factor- $\alpha$  to track pre-established and dynamic chromatin looping. Our findings suggest that cross-linking of chromatin may mask such fine-tuned and temporal-specific dynamics of three-dimensional genome organization. Taken together, intrinsic chromosome conformation capture assays improve our understanding of the true nature of chromatin folding stabilization and shed light on the relevance of native forces on nuclear architecture.

## Zusammenfassung

Säugetierzellen beinhalten dasselbe 2-Meter lange DNA Molekül, dessen dreidimensionale Faltung im Zellkern eng mit der Genexpression verknüpft ist und daher während der Differenzierung einer Zelle variiert. Der strukturelle Aufbau des Chromatins beinhaltet spezifische Kontakte zwischen genomischen Elementen wie Promotoren und Enhancer, welche eine Windung des DNA-Moleküls erzeugen, bis hin zu komplexen Faltungsmechanismen von ganzen Chromosomen. Während den letzten zwei Jahrzehnten wurde unser Verständnis über die vielschichtige Architektur des Genoms und dessen einhergehende Funktion dank Mikroskopie- und Chromosomkonformations-Studien (3C) erheblich verbessert. Die Fixierung von Zellen und damit des räumlichen Chromatinnetzwerkes basiert auf der Verwendung von Formaldehyd. Dieses ist Hauptbestandteil von vielen molekularbiologischen Methoden, unter anderem auch konventionelle Chromosomkonformations Techniken. Allerdings kann die Effizienz und Spezifität von chemischer Zellfixierung fehlerhaft sein und somit Ergebnisse beeinträchtigen, wodurch diese Methode eine bislang wenig charakterisierte „Black Box“ darstellt. Die vorliegende Arbeit stellt eine neue, intrinsische Methode (i3C) zur Analyse von nativen Chromatinstrukturen vor, welche nicht auf einer Fixierung mittels Formaldehyd basiert. Dazu behandeln wir Säugetierzellkerne in einem speziellen, physiologischen Puffer, in welchem die Transkription und somit auch die nukleare, drei-dimensionale Faltung des Chromatins erhalten bleiben. Anhand unterschiedlicher Zelllinien und Chromatincharakteristika können wir belegen, dass zahlreiche ähnelnde DNA-Kontakte mittels nativen als auch konventionellen Chromosomkonformationsmethoden detektiert werden und diese sich in separierten Strukturen, auch bekannt als topologisch assoziierte Domänen (TAD), befinden. Außerdem ermitteln wir mittels i3C-Methoden Chromatinkontakte, welche vorallem mit regulatorischen Elementen und CTCF-Regionen assoziiert werden, präzise und mit einer verbesserten Signalrate. Desweiteren haben wir unter nativen Bedingungen Kontaktunterschiede festgestellt und zur Verifizierung dieser eine neue unabhängige Methode namens „TALE-iD“ entwickelt. Abschließend haben wir den extrazellulären Signalweg von dem Tumornekrosefaktor- $\alpha$  stimuliert und die damit zusammenhängenden dynamischen sowie vorher festgesetzten Chromatinstrukturen analysiert. Dabei haben wir festgestellt, dass die chemische Zellfixierung die feinen

Veränderungen des drei-dimensionalen Chromatinnetzwerkes verbergen kann. Daraus schlussfolgern wir, dass unsere intrinsische Chromosomkonformationstechnik unser Verständnis von natürlichen Einwirkungen zur Stabilisierung des nativen Chromatinnetzwerkes in weiterführenden Versuchen verbessern kann.

# 1. Introduction

## 1.1 Genome architecture in three-dimensional space

From the discovery of DNA structure in the 20th century until the completion of the human genome's sequencing in 2001, scientists aimed at understanding how genetic information is encoded in its linear sequence (Lander *et al.*, 2001). Still, unwound human DNA measures ~2m in length and it has now become clear that these long molecules are organized in non-random three-dimensional (3D) structures inside the confined dimensions of mammalian cell nuclei (Cullen *et al.*, 1993; Wijgerde *et al.*, 1995). In order to understand how the genome functions (e.g., how appropriate gene expression is achieved), we need to unravel the mechanisms that drive the structural and physical organization of chromatin within the nucleus.

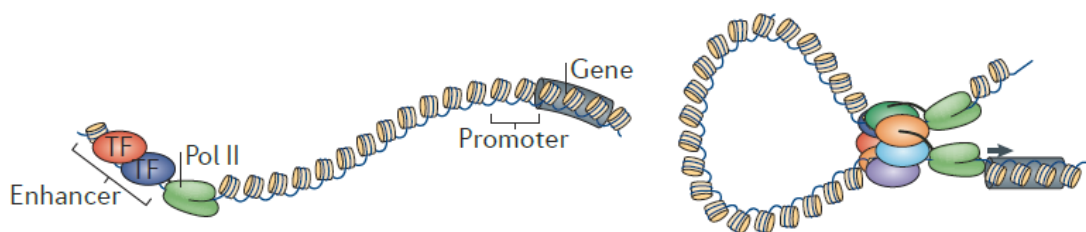
The organizational features of genomic architecture apply to multiple levels: First, the DNA is wrapped around histones and the resulting chromatin fiber from multiple physical contacts of genomic regions that are non-adjacent on the linear sequence by looping out the intervening DNA sequence (Tolhuis *et al.*, 2002). Then, these multi-loop formations provide the basis of the discrete territories of interphase chromosomes, where preferential proximity between transcriptionally-active or between silenced chromatin segments is observed (Cremer *et al.*, 1982; Boyle *et al.*, 2001). Such organization hints to a tight link between 3D chromosome folding and gene regulation.

Multiple genomics and microscopy approaches now allow for the investigation of 3D nuclear organization. For example, chromatin immunoprecipitation (ChIP) coupled to high throughput sequencing provides comprehensive datasets on transcription factor binding and post-translational histone modifications, hence enhancing our knowledge of structural gene regulation from an epigenetic point of view (Barski *et al.*, 2007). At the single-cell level, DNA/RNA fluorescence *in situ* hybridization (FISH) is a powerful tool for analyzing gene/RNA positioning; however, such approaches are limited in throughput and resolution (Wijgerde *et al.*, 1995; Schermelleh *et al.*, 2010). Thus, in order to study individual interactions between genomic elements (e.g., between promoters and enhancers; Fig. 1) called for the development of more sophisticated

methods that would allow for the study of 3D chromosomal organization at the molecular level.

### 1.1.1 Chromosome conformation capture (3C)-based technologies

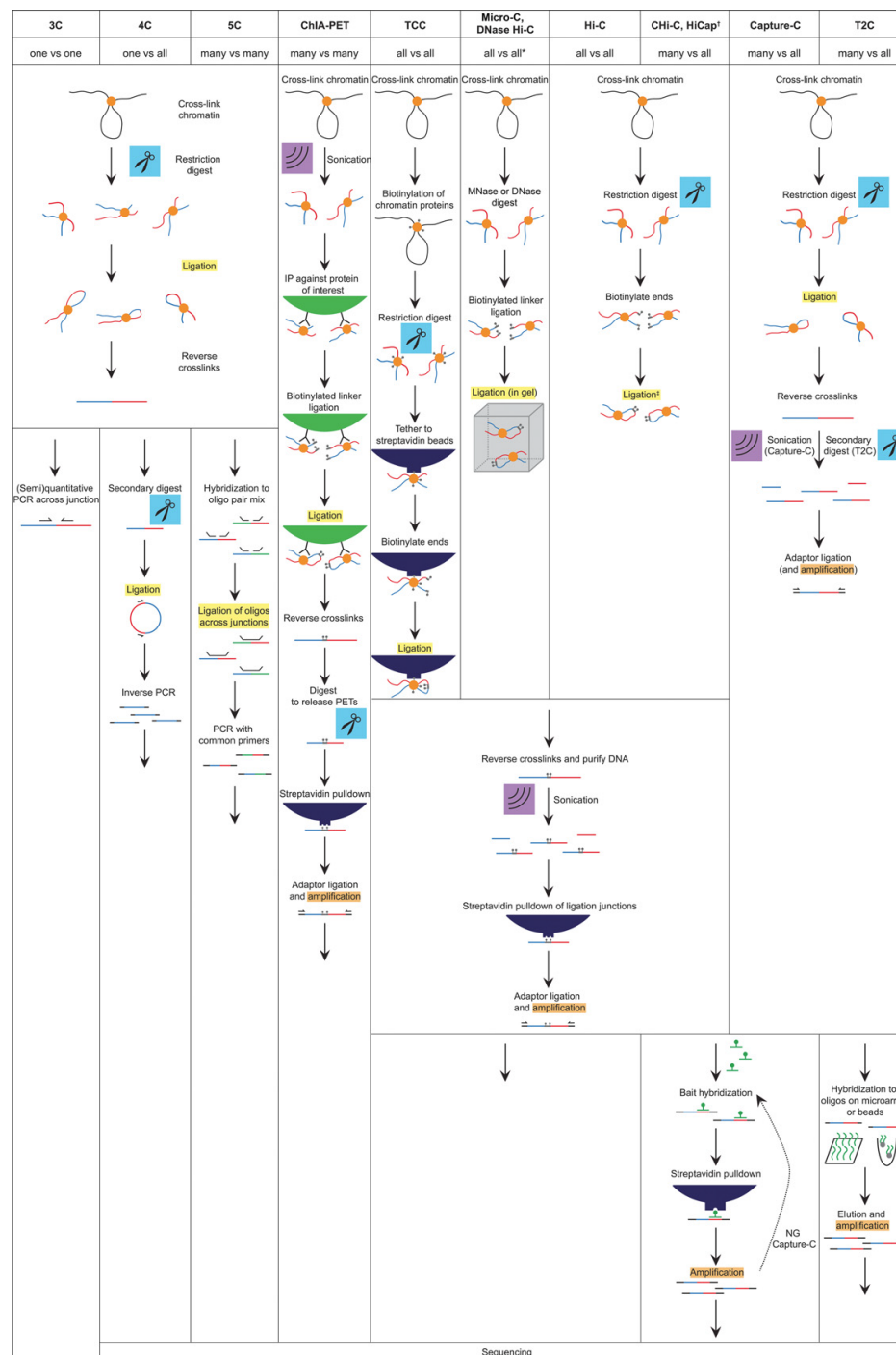
In 1993, Cullen et al. devised a “nuclear ligation” assay to identify looping interactions between enhancer and promoter regions of a rat gene locus (Cullen *et al.*, 1993). This approach is the first described method that relies on natural interactions to keep DNA sequences in close proximity, employing basic key principles of the ensuing chromosome conformation capture (3C) technologies.



**Figure 1| Chromatin looping.** Promoters directly interact with distal enhancers through looping out intervening DNA sequences. Binding of transcription factors and RNA polymerase II (Pol II) activates transcription of genes (adapted from Pombo & Dillon, 2015).

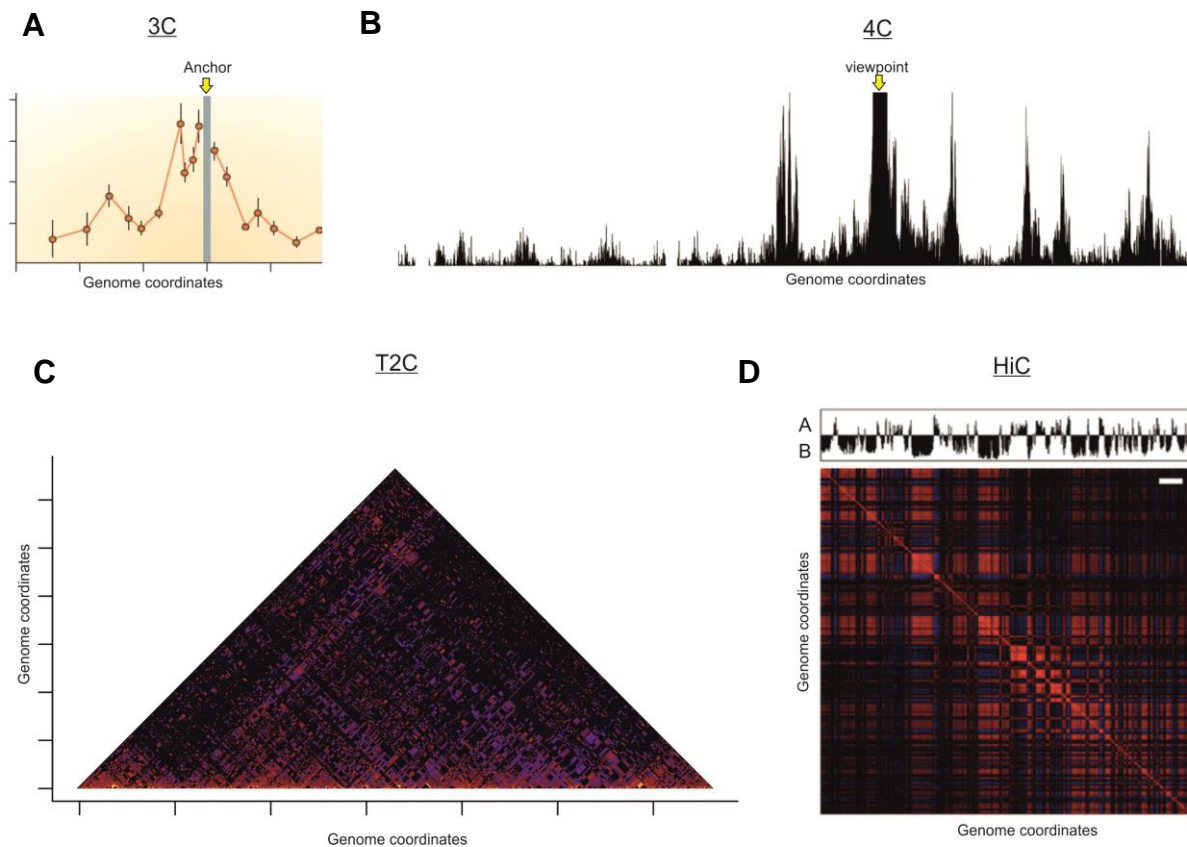
More than a decade ago, Dekker et al. introduced 3C, an approach that allowed scientists to study chromosome folding at high resolution by determining contact probability among a given pair of genomic loci (Dekker *et al.*, 2002). Several modifications of this basic 3C idea, like circularized chromosome conformation capture (4C; Stadhouders *et al.*, 2013), carbon copy chromosome conformation capture (5C; Dostie *et al.*, 2006), Hi-C (Lieberman-Aiden *et al.*, 2009), chromatin interaction analysis coupled to paired-end tagging (ChIA-PET; Fullwood *et al.*, 2009), targeted conformation capture (T2C; Kolovos *et al.*, 2014) and many other variants now allow for the high-throughput analysis and mapping of chromatin looping events at a genome-wide scale. Without exception, 3C-based assays rely on chemical cell fixation by formaldehyde to preserve DNA-protein interactions, thus “fixing” a snapshot of the *in vivo* genomic structure (Fig. 2). DNA is then digested by a restriction enzyme of choice, and cross-linked fragments are ligated under dilute conditions to favor fusion between fragments originally in close proximity in nuclear 3D space. The resulting “3C template” contains linear and circular pieces of ligated DNA fragments that were assembled according to their relative spatial positions at

the time of fixation. Interaction frequencies between different genomic regions may consequently be quantified and correlated to studies of regulatory elements so as to gain insight into chromatin looping models.



**Figure 2 | Overview of 3C-derived methods.** Basic steps of all standard 3C-derived protocols imply cell cross-linking using formaldehyde, followed by chromatin fragmentation with either restriction endonucleases, MNase, DNase or sonication. Subsequent ligation of cross-linked fragments that were close in 3D space yields a 3C template, which serves as input for all 3C-based techniques. Ligation junctions are detected and quantified according to the specific 3C strategy (adapted from Denker & de Laat, 2016).

All 3C-based methods rely on the 3C template as input, but differ in their strategy to detect and quantify ligation products, thus enabling different questions to be addressed each time (Fig. 3). In classical 3C, single ligation hybrids of genomic regions of interest are amplified and quantified by PCR using primers targeting two selected fragments. The interaction frequency of these two fragments can be measured by PCR product intensity after gel electrophoresis or by quantitative real-time PCR (Dekker *et al.*, 2002). 4C identifies all regions in contact with one locus of interest, generating a genome-wide interaction profile for a selected “viewpoint”. For this, the 3C template is processed in a second round of digestion and subsequent ligation, which yields DNA circles containing the viewpoint of interest and its hitherto unknown contacting sequences. By inverse PCR using adapter-containing primers in the viewpoint fragment, captured sequences can be amplified and identified by next-generation-sequencing (Stadhouders *et al.*, 2013). Contrary to the above described methods, Hi-C offers whole-genome contact maps in an “all-versus-all” manner (Lieberman-Aiden *et al.*, 2009). The approach relies on ligation products that are tagged with biotinylated nucleotides and are pulled-down using streptavidin-coated beads to enrich for ligation hybrids. Then, paired-end massively parallel sequencing is used to uncover as many ligation junctions as possible, before the data is processed computationally to draw 2D interaction heat maps revealing pairwise information on 3D chromatin organization at varying resolutions (relying on the combination of restriction enzyme choice and sequencing depth). More recently, CaptureC/T2C/CaptureHi-C were introduced (Sexton *et al.*, 2012; Kolovos *et al.*, 2014; Martin *et al.*, 2015). These very similar approaches produce pairwise interaction maps for preselected loci of interest (e.g., for a multi-Mbp region of one mammalian chromosome). Here, after digestion of the 3C template with a second enzyme and sonication, interacting sequences are captured by a set of oligonucleotide probes that are immobilized on beads and were specifically designed against each restriction fragment in the locus of interest. Hybridized DNA fragments are isolated, paired-end sequenced, and mapped to the relevant reference genome to generate 2D interaction maps as in Hi-C. Since such capture-based approaches are more cost-effective than Hi-C, they serve as alternative tools to study chromatin interactions at a scale just under that the whole genome.



**Figure 3 | Examples of 3C, 4C, T2C and Hi-C interaction profiles.** **A** | 3C-qPCR contact profile. Contact frequencies (orange dots) are plotted along the genomic distance from the viewpoint. **B** | 4C-seq contact profile. Sequenced reads are mapped to the genome, with peaks indicating contact frequencies around the viewpoint fragment (yellow arrow). **C** | T2C profile for a specific locus of interest. Different colors from red to blue indicate high or low contact frequency for each read pair. **D** | Genome-wide Hi-C heat map for a whole chromosome. Different colors from red to blue indicate high or low contact frequency for each read pair. Switch of active A- and inactive B-compartments is used to calculate the directionality Index (DI) and call TADs.

### 1.1.2 The hierarchy of genome organization: Interpreting chromatin interaction data

The genome is shaped at many consecutive hierarchical layers, ranging from higher-order chromatin organization at the level of chromosomal arrangements into distinct chromosome territories (CT; Cremer *et al.*, 2006), down to dynamic smaller-scale looping events between distal loci (Fig. 4). In the past decade, a combination of microscopy studies and 3C-based approaches allowed us to unravel molecular processes that govern three-dimensional chromatin folding and its impact on the regulation of genes.



At the single-cell level, DNA FISH is a powerful tool to visualize entire chromosomes occupying their own territory in the interphase nucleus (Bolzer *et al.*, 2005), which was also confirmed by genome-wide Hi-C analysis demonstrating that intra-chromosomal interactions occur more frequently than interchromosomal ones (Lieberman-Aiden *et al.*, 2009). Although CT's are spatially separated, gene-rich regions tend to dynamically intermingle with chromatin loops from neighboring CT's, indicative of a correlation between chromatin organization and transcriptional activity (Meaburn & Misteli, 2007). The positions of CT's themselves appear static in interphase nuclei, but vary from cell to cell due to chromatin reshuffling during mitosis. Despite these population-level variations, genome-wide 3C approaches do reveal many of the principal 3D chromatin folding characteristics. Genomic regions correlating with indicators of transcriptional activity, such as chromatin accessibility, gene density, GC content, replication timing and "active" histone marks, preferentially interact with other regions showing similar transcriptional activity characteristics (Brown *et al.*, 1997; Croft *et al.*, 1999). On the other hand, regions carrying heterochromatic features such as gene-poor and transcriptionally-inactive loci tend to interact with similar inactive regions. These subnuclear regions are referred as active "A" and inactive "B" compartments, respectively (Lieberman-Aiden *et al.*, 2009) and are generally distinctly located in nuclear space. Studies of the nuclear periphery using the DamID method (van Steensel and Henikoff, 2000; Kind *et al.*, 2013) uncovered a link between transcriptional silencing and the nuclear lamina, contrasting active regions that group together closer to the nucleus center. These observations highlight the influence of nuclear localization on transcriptional activity and vice-versa, but more as general trends rather than deterministic rules of higher-order chromatin organization.

Nonetheless, a model whereby chromosomal regions with similar transcriptional activity can colocalize in space at specific subnuclear sites such as nucleoli, transcription factories, lamina or Polycomb bodies is now very well established (Pederson, 2011; Entrevan *et al.*, 2016; Spector & Lamond, 2011). For instance, genes that are transcribed by RNA polymerase II associate together at foci known as transcription factories (Papantonis & Cook, 2013). These nucleoplasmic sites contain at least two active RNA polymerases engaged on different templates, and Papantonis *et al.* demonstrated congregation of cytokine-responsive genes upon stimulation to discrete 'NF $\kappa$ B-rich factories' (Papantonis *et al.*, 2012). According to

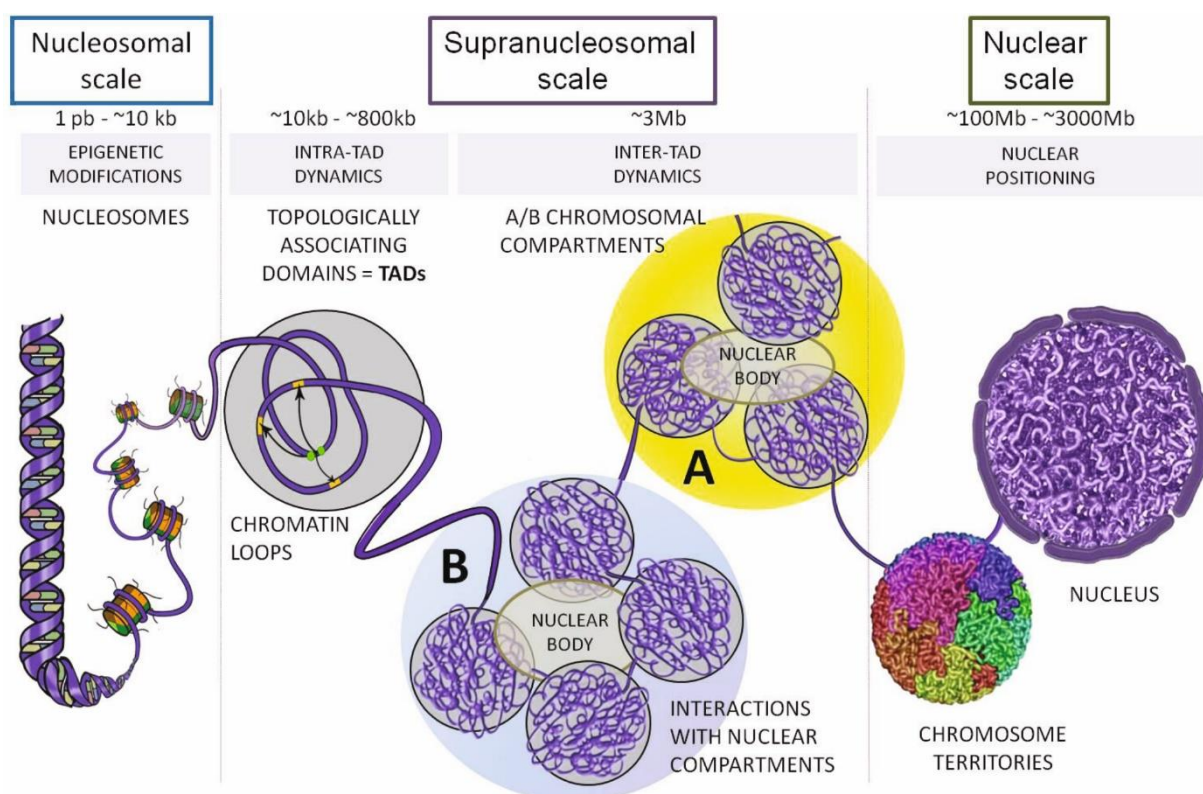
this increasingly accepted model, >95% of cell transcription occurs on the surface of factories, and a normal primary human cell contains ~2500 such entities.

At the sub-compartment scale, chromatin is organized into spatially separated domains ranging from a few hundred kilobases (kbp) up to a few megabases (Mbp) in size. These structural units are known as topologically associating domains (TADs), and are marked by significantly higher-frequency chromatin interactions within one domain compared to interactions across domain boundaries (Dixon *et al.*, 2012; Nora *et al.*, 2012). Genome-wide Hi-C studies identified approximately 2000 TADs, covering 90 % of the human and mouse genomes with an average size of 880 kb (Dixon *et al.*, 2012). TADs are highly conserved between different cell types and were also described with an average size of ~100 kb in the small genome of *Drosophila* larvae (Sexton *et al.*, 2012). This leads to the assumption that TADs might function as chromatin organization units by regulating gene looping events under particular spatial constraints. Due to the advantage of recent higher-resolution Hi-C and 5C contact maps, TADs can be sub-divided into sub-TADs with a median size of approximately 185 kbp, pointing to cell-type specific features of chromatin folding at the sub-megabase scale (Phillips-Cremins *et al.*, 2013). In addition, TADs are found within both the active A- and the inactive B-compartments, thus allowing for large-scale inter-domain interactions and the formation of “metaTADs” (Fraser *et al.*, 2015).

How the separation between consecutive TADs along chromosomes is achieved still remains elusive, but some characteristics of TAD boundaries have been revealed. They appear enriched for house-keeping protein-coding, tRNA genes, and binding sites for architectural proteins such as the CCCTC-factor (CTCF) and cohesin (Dixon *et al.*, 2012). Moreover, CTCF-sites occur in a convergent orientation at the anchor of chromatin loops (Rao *et al.*, 2014). However, CTCF binding site conversion and deletion at TAD boundaries or global CTCF and cohesin depletion studies have shown increased interaction frequencies between adjacent domains, but not a complete loss of boundaries (Nora *et al.*, 2017; Zuin *et al.*, 2014). Nevertheless, the contribution of CTCF and other factors in TAD boundary formation remaining obscure, as they also bind multiple sites within TADs.

The vast majority of long-range interactions between genes and their distal regulatory elements are observed confined within the boundaries of their respective TAD. Pioneering 3C-based work on chromatin contacts and DNA looping was performed

on the  $\beta$ -globin gene and its 50 kb distant locus control region (LCR). Looping between the gene promoter and its complex enhancer element will only occur in cell lines with high  $\beta$ -globin expression (Tolhuis *et al.*, 2002). This work highlights the importance of promoter-enhancer contacts creating a complex regulated genome via both dynamic chromatin conformation changes and pre-established contacts. The mechanisms that drive long-range regulatory interactions remain unclear, but it is very well known that transcription factors and RNA polymerases are key components in mediating the clustering of loci, while looping out intervening DNA (Papantonis *et al.*, 2012; Ghavi-Helm *et al.*, 2014).



**Figure 4 | Hierarchical genome organization.** The DNA fiber is wrapped around histones, which are epigenetically modified. Long-range interactions between *cis*-regulatory elements such as promoters and enhancers form chromatin loops by excluding intervening DNA. Topologically associated domains (TADs) harbor frequent chromatin looping events and are separated by boundaries demarcated with less interaction frequencies. At larger scale, the genome is divided into active A- and inactive B compartments. Chromosomes occupy distinct chromosome territories in the nucleus (Ea *et al.*, 2015).

## 1.2 Evaluation of 3C-based methods and their biases

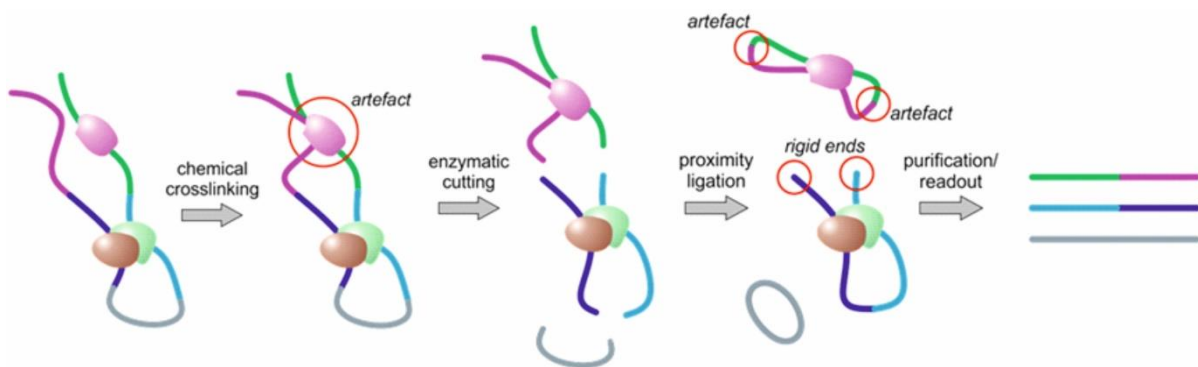
In recent years, 3C technology and its derivatives have been widely used to study the close relationship between the shape of the genome and cellular processes such as gene expression. Several studies have proved that regulatory elements communicate

with distant target genes through physical looping interactions, thereby controlling transcription regulation (Tolhuis et al., 2002; Palstra et al., 2003). Although 3C technology is a very useful tool for expanding our understanding of genome architecture, one must consider its intricacies and experimental biases in order to minimize data misinterpretation.

### **1.2.1 *In vivo* formaldehyde cross-linking: scientists are flying blind**

The key steps of the 3C protocol are based on the prevailing assumption that *in vivo* formaldehyde fixation of cells, generating crosslinks between DNA-protein, RNA-protein and protein-protein complexes, produces a faithful snapshot of the cell's biopolymer networks (Solomon & Varshavsky, 1985). This is an obsolescent idea that is impossible to be universally verified. As a very reactive chemical, formaldehyde (HCHO) reacts with amino-groups of proteins as well as of nucleic acids creating an intermediate product linked through a Schiff base, which subsequently condenses with a second amino group (Orlando *et al.*, 1997). Although the chemistry of formaldehyde fixation is well understood, its *in vivo* effects remain unknown. Chemical fixation of cells, integral to all 3C derivatives and chromatin immunoprecipitation (ChIP) approaches, can introduce spurious cross-links between randomly colocalizing but functionally-unrelated proteins and DNA sequences, thus increasing the background of random ligation events not representing meaningful and persistent interactions (Fig. 5). Moreover, we now know that proteins (e.g., transcription factors) not directly bound to DNA are difficult to cross-link using formaldehyde and complexes involving such proteins often produce false-positive signals (Solomon & Varshavsky, 1985; Nowak *et al.*, 2005). In addition, fixation protocols need to be elaborated carefully for individual experiments by adapting temperature, incubation time and pH conditions (Schmiedeberg *et al.*, 2009). The effects of minor variations in specificity and efficiency of *in vivo* cross-linking remain unclear and therefore experimentalists constantly run the risk of predominantly dealing with an abundant subset of fixed complexes just because of them being less labile, which can severely compromise data interpretation. There are several examples where interactions captured by 3C could not be reconciled to those obtained by microscopy and vice versa (e.g., Williamson *et al.*, 2014). Other reports have revealed different interaction profiles between euchromatin- and heterochromatin-residing loci, revealing higher contact frequencies in open

chromatin, and loss of TADs in condensed and mitotic chromatin (Sanyal *et al.*, 2012; Naumova *et al.*, 2013). These observations raise the question if some fractions of 3C contacts are not detected due to cross-linking biases in heterochromatic regions. In addition, formaldehyde treatment was shown to induce widespread polyADP-ribosylation of proteins, alter nuclear protein composition, and thus affect their degree of cross-linking to chromatin (Beneke *et al.*, 2012).



**Figure 5 | Cross-linking biases in 3C technologies.** Cell fixation using chemical reagents such as formaldehyde might result in artefactual cross-linking of bystander sequences, rigid chromatin networks and random self-ligation events of digested fragments that are still present in the sample. These experimental biases can cause false data outcome.

More and more evidence highlights limitations that come along with formaldehyde cell cross-linking and demands for improvements and alternatives (Beneke *et al.*, 2012; Williamson *et al.*, 2014; Teytelman *et al.*, 2013). There is a great variety of cell fixation protocols and chemical compounds that can be used as fixative reagents such as ethanol, methanol, acetone, glutaraldehyde, osmium tetroxide and the most common used formaldehyde, as well as ultra-violet (UV) light cross-linking or cryopreservation. However, all these methods carry some bias affecting the efficient maintenance of native subnuclear 3D chromatin architecture. For instance, coagulant fixation agents like ethanol, methanol and acetone cause protein denaturation and cell dehydration, resulting in chromatin conformation changes (Kumarasinghe *et al.*, 1997). Fixation agents like formaldehyde, glutaraldehyde and osmium tetroxide react with proteins and lead to a better retention of nuclear compartments. Nevertheless, unclear *in vivo* effects still raise concerns over the applicability of these reagents in molecular biology approaches, despite their widespread and long-standing use (Fox *et al.*, 1985; Guillot *et al.*, 2004). Furthermore, UV light irradiation triggers DNA damage in cells, while immobilization of cellular components during cryopreservation

tends to be insufficient and needs to be combined with chemical fixation in order to sustain the subsequent heating steps implemented in the 3C protocol (Ripper et al., 2008). Taken together, these observations call for an alternative procedure for the conservation of native nuclear organization.

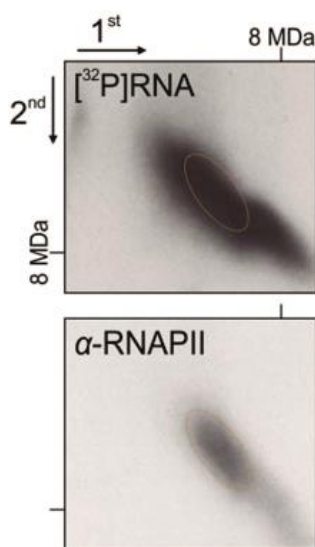
### **1.2.2 Experimental pitfalls of conventional 3C technology**

Due to the flexibility of the long chromatin fibers, DNA sequences are assembled in random collisions. As described above, cross-linking produces a rigid network of chromatin fibers, which is subjected to sodium dodecyl sulfate treatment and restriction enzyme digestion in order to ideally solubilize chromatin complexes and favor ligation of DNA fragments that have been joined via protein bridges. Instead, it was shown that ligation happens within the rigid chromatin network and not all detected ligation hybrids necessarily represent meaningful interactions (Gavrilov *et al.*, 2013). Therefore, 3C is a quantitative assay and functional interaction frequencies need to be assessed together with high background noise (Dekker, 2006). As mentioned, partial decompaction of chromatin is performed using sodium dodecyl, although it can introduce variable restriction and ligation efficiencies in each individual experiment, hence altering the data outcome. Additionally, the conventional 3C protocol implies harsh manipulation of the chromatin structure by long incubations while heating and shaking at more than 900 rpm, and the usage of buffers with unphysiological salt concentrations. Still some (not clear exactly how many) cross-linked complexes persist via this protocol, pointing to a rather non-uniform stabilization of the *in vivo* chromatin structure. Such biases and other still-unaccounted for effects of cross-linking render the need for studies using native, non-cross-linked, chromatin preparations more pressing.

### **1.3 Genome organization under physiological conditions**

Revolutionary studies in the 1970s suggested transcription as a major force tethering DNA loops to subnuclear structures, engaged through immobilized and active RNA polymerases functioning as molecular ties (Cook & Brazell, 1975). But these results were criticized, because chromatin extractions were performed in 2 M NaCl which is likely to cause artefactual aggregation of the transcription complexes. Generally, methods to isolate and treat nuclei with buffers containing non-physiological ion concentrations involve swelling of cells, activation of nucleases and aggregation of

chromatin. Therefore, Peter Cook developed a protocol to isolate and handle cell nuclei under “mild” conditions using a “physiological buffer”, whose composition is approximately cytoplasmic (Cook, 1984). A combination of this buffer with a detergent was then used to permeabilize cells, treat chromatin with nucleases to remove most unrelated DNA and unveil active transcription foci by tracking [<sup>3</sup>H]uridine labelled, nascent RNA. This so called “run-on” experiment provided the first evidence for active RNA polymerases being fixed at subnuclear sites during transcription, today known as transcription factories, supporting a model widespread DNA looping (Fig. 6; Jackson *et al.*, 1981).



**Figure 6 | Transcriptional activity is preserved in PB.** HUVECs were permeabilized in PB supplemented with saponin and used for a ‘nuclear run-on’ experiments. Hence, engaged polymerases were allowed to produce nascent transcripts by incorporating [<sup>32</sup>P] UTP in PB (top). Isolated RNA polymerase II (RNAP II) complexes were run in 2D native gel electrophoresis, followed by immunoblotting of the largest subunit, RPB1, of RNAP II (bottom). Dotted circles indicate gel regions that are enriched for nascent transcripts RPB1. Detected RNAP II factory complexes have a greater mass than 8-MDa (adapted from Caudron-Herger *et al.*, 2015).

Hence, approaches using this near-physiological buffer do preserve nuclear functionality of transcription and with it much (if not all) of the 3D nuclear structure. Consequently, chromatin preparation methods using this buffer could serve as a suitable tool for functional and structural studies of native chromatin complexes.

#### 1.4 Aim

In recent years, chromosome conformation capture based techniques have been used to assess chromatin looping between genomic elements and its relation to genome function during cell differentiation and disease progression. The conventional 3C protocol relies on the assumption that DNA-protein complexes can be fixed by formaldehyde cross-linking to capture a snap-shot of the spatial chromatin interactome. Instead, chemical cell fixation yields a rigid and poorly characterized chromatin network that implies biases regarding specificity and

efficiency of *in vivo* cross-linking. Moreover, 3C assays are based on harsh manipulations of the chromatin fiber to ensure efficient restriction enzyme digestion by applying sodium-dodecyl-sulfate and extended incubation times under high temperatures as well as mechanical rotation. However, these steps may involve hazards and produce false data outcome, which was recently observed in differential microscopy and 5C studies. Here, we aim to draw attention to limitations that come along with conventional 3C-derived methods and introduce a novel intrinsic chromosome conformation capture approach (i3C). We wish to exploit native forces that stabilize spatial chromatin interactions by processing mammalian cell nuclei in a close-to-physiological buffer, which was shown to preserve transcriptional activity. Furthermore, we will compare features of conventional and native contact profiles and outline their advantages in chromosome folding studies. As a model, we aim to investigate native chromatin dynamics upon extracellular signaling in a locus-specific and genome-wide manner, and gain information on principles that link three-dimensional chromatin architecture to gene expression.



## 2. Material and Methods

### 2.1 Material

#### 2.1.1 Chemicals

Table 1 | Chemicals

Reagent	Company	Catalog no.
1,4-Dithiothreitol (DTT)	Sigma-Aldrich	D9779
2-Propanol	Sigma-Aldrich	33539
2-Mercaptoethanol	Gibco	21985-023
3-(N-Morpholino)propanesulfonic acid (MOPS)	Sigma-Aldrich	M1254
4', 6-Diamidin-2-phenylindol (DAPI)	Sigma-Aldrich	D9542
6-amminocarpoic acid	Sigma-Aldrich	A2504
Adenosine 5'triphosphate disodium (Na <sub>2</sub> ATP)	Sigma-Aldrich	A2383
Agar	Sigma-Aldrich	05039
Agarose	Sigma-Aldrich	A9539
Bovine serum albumin (BSA)	Sigma-Aldrich	A9418
Bromochloropropane (BCP)	Sigma-Aldrich	B9673
Calcium chloride	Sigma-Aldrich	C3306
Caspase inhibitor III	Calbiochem	218745
Cyclohexamide	Sigma-Aldrich	C7698
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	D8418
Disodium hydrogen phosphate	Sigma-Aldrich	S3264
Ethanol	Sigma-Aldrich	24102
Ethylenediaminetetraacetic acid disodium (EDTA)	Sigma-Aldrich	E5134
EGTA	Sigma-Aldrich	E3889
Fetal calf serum (FCS)	Gibco	
Glycine	Sigma-Aldrich	G8898
Glycerol	Sigma-Aldrich	G5516

Hydrochloric acid (HCl)	Sigma-Aldrich	258148
IGEPAL CA-630 (NP-40)	Sigma-Aldrich	I8896
LB-Broth	Sigma-Aldrich	L3022
Magnesium chloride hexahydrate	Sigma-Aldrich	M2670
Manganese(II) chloride	Sigma-Aldrich	31422
MEM Non-essential amino acid solution	Sigma-Aldrich	M7145
Paraformaldehyde (PFA)	Electron Microscopy Sciences	15710
Phenol:Chloroform:Isoamyl Alcohol	Sigma-Aldrich	77617
Poly(ethylene glycol) (PEG)	Sigma-Aldrich	89510
Potassium acetate	Sigma-Aldrich	P1190
Potassium chloride	Sigma-Aldrich	P9541
Potassium phosphate dibasic trihydrate	Sigma-Aldrich	P9666
Protease inhibitor cocktail (PIC)	Roche	11836153001
RiboLock RNase Inhibitor	Thermo-Scientific	#E00384
Sodium acetate	Sigma-Aldrich	S8750
Sodium chloride	Sigma-Aldrich	S3014
Sodium dodecyl sulfate (SDS)	Roth	CN30.3
Sodium fluoride	Sigma-Aldrich	30105
Sodium orthovanadate	Sigma-Aldrich	S6508
Sodium pyruvate solution	Sigma-Aldrich	S8636
Sucrose	Sigma-Aldrich	S0389
Tris	Sigma-Aldrich	T1503
Triton X-100	Sigma-Aldrich	T8787
Trizol	Life Technologies	15596018
Tween-20	Sigma-Aldrich	P9416
Water (Ultrapure)	Milli-Q	
$\beta$ -Glycerophosphate	Sigma-Aldrich	G9422

**2.1.2 Buffers and solutions****Table 2 | Buffers and solutions**

Buffer/Solution	Composition
0.5x TBE	45 mM Tris-borate 1 mM EDTA
1x PBS (Sigma-Aldrich)	0.01M Phosphate buffer 0.154M sodium chloride (pH 7.4)
LB-Agar	10 g/l LB Broth 15 g/l Agar Autoclave
LB-Media	10 g/l LB Broth Autoclave
Lysis Buffer	10mM Tris-HCl (pH 8) 10mM Sodium chloride 0.4% NP-40 1x PIC
Native Lysis Buffer	40mM Tris acetate (pH 7.4) 2M 6-amminocaproic acid 0.2M Sucrose 50 Units/ml RiboLock RNase inhibitor 1/1000 PIC
Permeabilization Buffer	15mM Tris-HCl (pH 7.6) 60mM Potassium chloride 15mM Sodium chloride 4mM Calcium chloride 0.5mM EGTA 300mM Sucrose 0.2% NP-40 0.5mM $\beta$ -Mercaptoethanol
Physiological Buffer (PB)	100mM Potassium acetate 30mM Potassium chloride 10mM Disodium hydrogen phosphate

	<p>1mM Magnesium chloride</p> <p>1mM Na<sub>2</sub>ATP</p> <p>1mM DTT</p> <p>10mM β –Glycerophosphate</p> <p>10mM Sodium fluoride</p> <p>0.2mM Sodium orthovanadate</p> <p>1/1000 PIC</p> <p>25U/ml RiboLock RNase inhibitor</p> <p>Adjust pH to 7.4 with 100mM potassium phosphate dibasic.</p>
Quenching Buffer	<p>0.125 mM Glycine</p> <p>1x PBS</p>
TE	<p>10mM Tris-HCl (pH 8)</p> <p>0.1mM EDTA</p>
TFB I	<p>50mM Manganese(II) chloride</p> <p>100mM Potassium chloride</p> <p>10mM Calcium chloride</p> <p>30mM Potassium acetate (pH 6)</p> <p>15% (v/v) Glycerol</p> <p>Adjust pH to 6.1 with Acetic acid and autoclave. Stored at 4°C.</p>
TFB II	<p>75mM Calcium chloride</p> <p>10mM Potassium chloride</p> <p>10mM MOPS</p> <p>15% (v/v) Glycerol</p> <p>Adjust pH to 7.0 with potassium hydroxide.</p>

### 2.1.3 Enzymes

Table 3 | Enzymes

Enzyme	Company	Catalog no.
Restriction enzymes		
ApoI	New England Biolabs	R0566L
BsaBI	New England Biolabs	R0537S
DpnII	New England Biolabs	R0543T
HaeIII	New England Biolabs	R0108T
MseI	New England Biolabs	R0525M
MspI	New England Biolabs	R0106S
NlaIII	New England Biolabs	R0125L
PacI	New England Biolabs	R0547S
Ligases		
T4 DNA Ligase	Invitrogen	15224041
T4 DNA Ligase	New England Biolabs	M0202T
Polymerases		
DreamTaq	Thermo Fisher Scientific	EP0711
Expand High Fidelity	Sigma-Aldrich (Roche)	04743733001
Klenow Fragment	New England Biolabs	M0210
O5 High Fidelity	New England Biolabs	M0491S
SYBR Green Jumpstart	Sigma-Aldrich	S4438
T4 DNA Polymerase	New England Biolabs	M0203L
Others		
Accutase	Sigma-Aldrich	A6964
Caspases Group III	PromoKine	PK-RP577-K243
DnaseI	Worthington	LS006333
Proteinase K	Thermo Fisher Scientific	AM2548
RNase A	AppliChem	A3832,0050
Trypsin	Sigma-Aldrich	T3449

### 2.1.4 Kits and systems

**Table 4 | Kits and systems**

Company	Kit/System	Catalog no.
Lonza	Amaya Cell Line Nucleofector Kit V	VCA-1003
Zymo Research	Direct-zol RNA MiniPrep	R2052
Zymo Research	DNA Clean & Concentrator (5, 25)	D4014, D4034
Qiagen	Large Construct Prep	12462
Invitrogen	Lipofectamine RNAiMax	13778-100
Invitrogen	Qubit dsDNA BR Assay	Q32850
Invitrogen	Qubit dsDNA HS Assay	Q32851
Zymo Research	RNA Clean & Concentrator 25	R1018
Invitrogen	SuperScript II RT	18064-071
Zymo Research	Zymoclean Gel DNA Recovery	D4008
Zymo Research	ZymoPURE Plasmid Maxiprep	D4203
Zymo Research	Zyppy Plasmid Miniprep	D4019

### 2.1.5 Equipment

**Table 5 | General equipment**

Application	Type	Company
Bacteria incubator	Ecotron	Infors HT
Cell counting	Neubauer chamber	Optik Labor
Cell culture	CO <sub>2</sub> Incubator	Sanyo
Centrifuge	221.12	Hermle
Centrifuge	220.87	Hermle
Centrifuge	75005719	Thermo Scientific
Centrifuge	JA-14	Beckman Coulter

Centrifuge	Scanspeed mini	Labogene
Electrophoresis	Mini-Sub Cell GT System	Bio-Rad
Electrophoresis	Wide Mini-Sub Cell GT System	Bio-Rad
Fluorometer	Qubit 2.0	Invitrogen
Imaging System	ChemiDoc MP	Bio-Rad
Microscope	DMIL LED	Leica
Microscope	Inverted Fluorescence IX 81	Olympus
pH meter	PB-11	Sartorius
pH meter	H138 minilab	Hach
Realtime thermal cycler	Rotor-Gene Q	Qiagen
Sequencing		
Sonicator	Bioruptor Plus	Diagenode
Spectrophotometer	BioPhotometer	Eppendorf
Spectrophotometer	NanoDrop ND-1000	Peqlab
Thermal Cyclers	C1000 Touch	Bio-Rad
Transfection	Nucleofector	Lonza
Transfection	GenePulser Xcell	Bio-Rad
Vortexor	VTX-3000L	Harmony

### **2.1.6 General reagents used for molecular biology**

**Table 6 | General reagents used for molecular biology approaches**

Reagent	Company	Catalog no.
1 kb DNA Ladder	New England Biolabs	N0552G
100 bp DNA Ladder	New England Biolabs	N0551G
Biotin-14-ATP	Invitrogen	19524-016
DNA SYBR Safe	Invitrogen	S33102
Gel Loading Dye	New England Biolabs	B7024S

### 2.1.7 Antibiotics

**Table 7 | Antibiotics used for cloning experiments**

Antibiotic	Working concentration ( $\mu\text{g/ml}$ )	Company	Catalog no.
Ampicillin	100	Sigma-Aldrich	A0166
Chloramphenicol	25	Sigma-Aldrich	C0378
Kanamycin	50	Sigma-Aldrich	K1377

### 2.1.8 Primers

**Table 8 | 3C-qPCR primers**

Genomic region	Sequence (5'→3')	Chromosome coordinates
*Viewpoint <i>EDN1</i> _TSS_	AATCAGAAGAGGGGACTCCAG	chr6:12398145-12398895
<i>EDN1</i> _downstr._enhancer	TGCTCACAGTCCAAATCCAG	chr6:12648802-12649007
<i>HIVEP</i> _TSS	TGGTGGTGGAGTATGTTTCC	chr6:12116454-12117979
<i>ADTRP</i> _upstr._enhancer	GCCCAGGATTTCAAGTTACG	chr6:11760065-11762575
Negative control_1	AACTGGAGGAAAGCAGGAAAC	chr6:12571009-12571640
Negative control_2	GGCCTTAGAGACCCCTAAAATG	chr6:12762201-12763363
Negative control_3	TCTCCCGTGTTCAGTATTCAAAG	chr6:12285966-12286094
Negative control_4	AGGATGAGAGAGGAGGATTGTG	chr6:11795742-11796613
Loading control_For	AGGCCTGGCAGTGAACCTTATTT	chr14:54203315-54205816
Loading control_Rev	ACTGTCATTAACCCCTCTGTGG	chr14:54203315-54205816



Table 9 | 4C primers

Viewpoint	Primer pair sequence (5'→3')	Restriction enzyme	Location (hg19/mm9)
<i>SAMD4A</i>	GACGGGTCCGGGTGAATTT	ApoI	chr14:55034216-55034641
TSS	CGCAGCCGAACCTTTCTTTG	DpnII	
<i>SAMD4A</i>	TCTGTAGACCGAGGGCGGC	NlaIII	chr14:55032924-55034637
TSS	CAACTCGGACCCTTCACG	DpnII	
<i>EDN1</i>	TTGTTGTGTGCGGGGAATTT	ApoI	chr6: 12290178-12290759
TSS	GCACTTGGGCTGAAGGATC	DpnII	
<i>BMP4</i>	ACGTGCGGAGGTACTAGAAAG	NlaIII	chr14: 54422893-54423841
TSS	GTCGTTGGGAAAACTGTGG	DpnII	
<i>CDKN3</i>	CGACACCACCGCTGTCAC	NlaIII	chr14: 54863838-54864286
TSS	ACCCTGCTCCTTCGTCTCTC	DpnII	
<i>CNIH</i>	GCTCCCCGCTCCTCCTCC	NlaIII	chr14: 54908039-54908769
TSS	AAGTGCAAGACAGTGGTGAGAC	DpnII	
<i>TBX5</i>	GACTGAGGTCTCTTGCATAAAG	NlaIII	chr12: 114846093-114846250
TSS	TGAAGAGTTCCTCCTCCTCC	DpnII	
<i>Nanog</i>	CCACCAGCCCTGTGAATTC	ApoI	chr6: 122657477-122657871
TSS	GGCTCACTTCCTTCTGACTTC	DpnII	
<i>Sox2</i>	CCCAGAAAAATTGTGGTAAAG	ApoI	chr3: 34547092-34548290
CTCF	TCTTTACGTCTGGACAATGG	DpnII	
<i>ZFPM2</i>	GGTCAACTTTTCTTGGCTTGG	ApoI	chr8:106329924-106332167
TSS	AAGAGTAGTCCCACGTCAATCG	MseI	
Illumina adapter sequences (5'-3')			
Reading primer	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC GCTCTTCCGATCT		
Non-reading primer	CAAGCAGAAGACGGCATAACGA		

**Table 10 | TALE-iD cloning primers**

Oligonucleotide		Sequence (5' → 3')
Dam from pTuner- IRES	Dam_BsaBI_For	AGTTCAGATCGACATCGATGAAGAAAAATCGCGC
	Dam_Pacl_Rev	CGACCTTTAATTAATTATTTTTTCGCGGGTG
TALE_Nterm_InsCtrl_F2		TGCTGAAGATCGCGAAGAG
TALE_Cterm_InsCtrl_R		GACAGCTGGGCCACAATG
Dam_Cterm_InsCtrl_F		CGCAAGCATTTGATGACG
Dam_pA_InsCtrl_R		GGACAAACCACAACACTAGAATGC

## 2.2 Methods

### 2.2.1 Molecular cloning

#### 2.2.1.1 Bacteria strains

**Table 11 | Bacterial strains used for cloning experiments**

Strain	Genotype
Top10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str <sup>R</sup> ) endA1 λ <sup>-</sup>
Dam-/dcm- K12	<i>ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44</i> <i>galK2 galT22 mcrA dcm-6 hisG4 rfbD1</i> <i>R(zgb210::Tn10) Tet<sup>S</sup> endA1 rspL136</i> <i>(Str<sup>R</sup>) dam13::Tn9 (Cam<sup>R</sup>) xylA-5 mtl-1</i> <i>thi-1 mcrB1 hsdR2</i>

### 2.2.1.2 Bacterial artificial chromosomes (BAC's)

Table 12 | List of bacterial artificial chromosomes

BAC ID	Chr. Position start	Chr. Position end	Chromosome
RP11-453F19	53899754	54062955	14
RP11-1107N7	54074915	54282534	14
RP11-242P6	54261556	54465831	14
RP11-663M15	11287269	11482099	6
RP11-1070E20	11687255	11895575	6
RP11-845M8	12098076	12292331	6
RP11-338L10	12350362	12513884	6
RP11-689G9	12498112	12672329	6
RP11-1033M8	113232348	113424812	2
RP11-65205	41044981	41233201	1

### 2.2.1.3 Plasmids

Table 13 | List of plasmids used for TALE-iD cloning

Plasmid	TALE Target	Reference
TALE_LSD1 # 28	TAAGTCTACATATAGTATCC	Mendenhall <i>et al.</i> , 2013
TALE_LSD1 # Scramble	CCAGTCCCTGGCTCCCAT	Mendenhall <i>et al.</i> , 2013
TALE_mCherry	Fluorescence control	Mendenhall <i>et al.</i> , 2013
pTUNER_Ires2	DNA methyltransferase	Kind <i>et al.</i> , 2013

### 2.2.1.4 Generation of transformation-competent *E. coli* bacteria

5 ml of LB medium were inoculated with the *E. coli* strain and incubated overnight with agitation (250 rpm) at 37 °C. On the next day 900 µl of the overnight culture were used to inoculate 150 ml of LB medium. The culture was grown until an OD<sub>600</sub> value of 0.45-0.55 was reached. The bacteria suspension was incubated on ice for 10 min and then harvested at 2.000 rpm at 4°C for 10 min. The supernatant was removed and the bacteria were resuspended in 30 ml TFBI. After a 10-min incubation on ice, the bacteria were centrifuged and the pellet was resuspended in 6 ml ice-cold

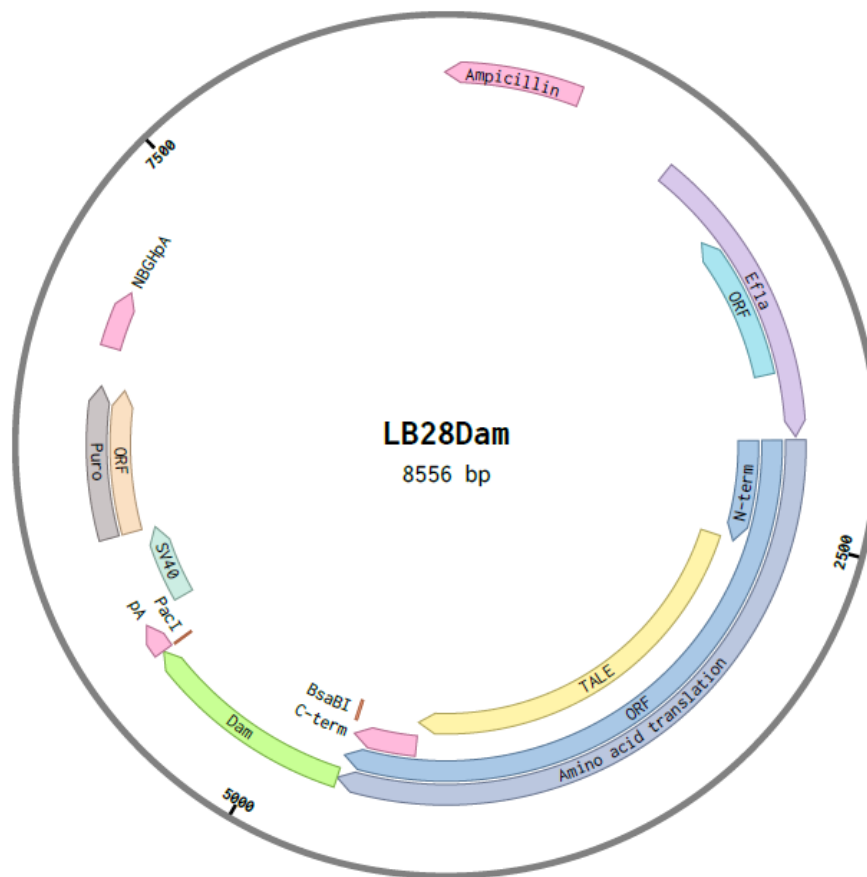
TFBII. 50µl aliquots of this bacteria suspension were transferred to 1.5 ml tubes and flash frozen with liquid nitrogen. The competent bacteria were stored at -80 °C until usage.

#### **2.2.1.5 Transformation of *E. coli***

Competent bacteria were thawed on ice and mixed with plasmid DNA (1-10 ng) or a ligation reaction (5-10 µl). The transformation sample was incubated on ice for 30 min, followed by a heat-shock at 42 °C for 45 sec and subsequent replacement on ice for 2 min. Then 500 µl of LB medium were added and the suspension was incubated at 37 °C under agitation (250 rpm) for 45 min. Afterwards the bacteria were spread onto LB Agar plates containing an appropriate antibiotic and incubated at 37°C overnight. On the next day transformants were analysed by Colony PCR. 5 ml LB overnight cultures were inoculated with positive clones by taking up a few colony cells with a pipette tip. For large scale plasmid amplification, 150 ml of LB medium supplemented with appropriate antibiotics were then inoculated with 500 µl of overnight culture and grown until an OD<sub>600</sub> value of 0.45-0.55 was reached. Plasmids were purified using the Zymo Research Maxi Preparation Kit.

#### **2.2.1.6 Construction of TALE-DNA methyltransferase (*Dam*) fusions**

We fused a custom TAL-effector DNA-binding domain of interest with a bacterial adenosine methylase (*Dam*). The *Dam* open reading frame was amplified from the pTuner-IRES2 plasmid (Kind *et al.*, 2013) using a proofreading polymerase to minimize the error rate during nucleotide incorporation. TAL effector vectors fused to a lysine-specific demethylase 1 (LSD1) were provided by EM Mendenhall (Mendenhall *et al.*, 2013). The PCR product was cloned into a TAL effector expression vector such that *Dam* is fused to the C-terminal end of the TAL effector. For this, TAL effector vectors were transformed into *dam*-/*dcm*- bacteria and LSD1 was replaced with *Dam* using *Bsa*BI and *Pa*cl. These assembled TAL effector *Dam* vectors were verified by enzyme digestion and sequencing. The mCherry control vector carries a mCherry open reading frame in place of the TAL effector. Control TAL effectors lacking *Dam* were constructed by incorporating a frame shift in the *Dam* sequence.



**Figure 7 | TALE-iD plasmid.** TAL effector targeting a strong enhancer in the first intron of ZFP2 was fused to Dam. TALE-iD plasmids were used for nucleofection of K562 cells. Similar cloning strategies were performed for non-targeting, scrambled control plasmids.

### 2.2.2 Mammalian cell culture

All cell lines were cultured and processed under sterile conditions, using laminar flow cell culture hoods as well as sterile solutions and materials. Cells were grown in their appropriate medium (see table 14) in an incubator at 37 °C and 5% CO<sub>2</sub>.

#### 2.2.2.1 Cell lines and media composition

**Table 14 | List of mammalian cell lines and corresponding culture media**

Cell type	Description	Growth media composition
HUVEC	human umbilical vein endothelial cells	500ml Endopan 3 3% FBS Supplements ( <i>Heparin, R3-IGF-1, FGF-2, Gentamicin sulfate amphotericin B,</i>

		<i>EGF, Asorbic acid, VEGF, Hydrocortisone)</i>
IMR90	human fetal lung fibroblasts	500ml MEM 20% FBS 1x NEAA
K562	human myelogenous leukemia cells	500ml DMEM 10% FBS 1% Pen-Strep
mESC	mouse embryonic stem cells	DMEM GlutaMax 15% FBS (heat-inactivated) 1x NEAA 1x $\beta$ -Mercaptoethanol 1mM Sodium pyruvate 6 ml LIF (purified supernatant from LIF-1ca COS cells)

### **2.2.2.2 Cell line treatments**

All cells were grown to ~90% confluency before harvesting for further experimental procedures. HUVEC and IMR90 cells were treated with TNF $\alpha$  (10 ng/ml) for 30 or 60 minutes at 37 °C/5 % CO<sub>2</sub>. To ensure a maximal identification of TNF $\alpha$ - responsive genes, HUVEC cells were serum-starved for not more than 16 hours and IMR90 cells were treated with cycloheximide (5  $\mu$ g/ml) prior to cytokine treatment to sustain signaling (Jin et al, 2013). To block transcription elongation by RNA Polymerase II, cells were treated with 50  $\mu$ M DRB for 1 h at 37°C.

### **2.2.2.3 Cell transfection via nucleofection**

The Nucleofector-Technology is an improved electroporation based method, which allows transfection of primary cells and other cell lines that are difficult to transfect with higher efficiencies. Approximately 10<sup>6</sup> K562 cells were used for each nucleofection reaction. Cells were washed twice with PBS and cell pellet was resuspended in 100  $\mu$ l Nucleofector solution. Then 100  $\mu$ l of cell suspension were combined with 1-5  $\mu$ g of highly concentrated plasmid DNA and transferred into

certified cuvette without leaving air bubbles. The appropriate Nucleofector program T-016 was selected and applied to the cuvette containing the cell/DNA suspension. Immediately 500 µl of pre-equilibrated culture medium were added to the cuvette and the whole sample was gently transferred into a well of a 12-well plate containing 2ml of pre-warmed medium. Cells were incubated at 37 °C and 5% CO<sub>2</sub> for 24 hours. Medium was changed on the next day.

### **2.2.3 Molecular biology methods**

#### **2.2.3.1 Nucleic acids isolation**

*Phenol-Chloroform purification of DNA and Ethanol precipitation:* Isolation of genomic DNA from cells was performed after incubation of cells in permeabilisation buffer including 20 µl of Proteinase K (10 mg/ml) and 20 µl of RNase A (10 mg/ml) over night at 37 °C. DNA from 3C-based experiments was isolated after Proteinase K and RNase A treatment of the samples. One volume of Phenol-Chloroform was added to the DNA suspension and sample was mixed by inverting the tube several times. After centrifugation, the upper aqueous phase was transferred into a fresh tube. DNA was precipitated by adding 1/10 volume of Sodium acetate and 2.5 volume of ice-cold Ethanol (100%). Prior to centrifugation, samples were kept at -80 °C for 30 min to improve DNA precipitation. DNA pellet was washed once with diluted Ethanol (70%) and resuspended in ultrapure water or TE buffer. To reduce co-precipitating DTT in 3C-based samples, the volume of the aqueous phase was doubled with water before adding sodium acetate and pure ethanol.

*Trizol-based purification of RNA and 2-Propanol precipitation:* TRIzol combines phenol, guanidine isothiocyanate, and other proprietary components in one monophasic solution which facilitate the isolation of RNA. Cells that grow in one well of a 6-well plate are lysed by adding 1 ml of TRIzol. The homogenate is then transferred to a fresh tube and separated into aqueous and organic phases by adding 100 µl of BCP. After centrifugation, RNA from the upper aqueous phase is precipitated in a fresh tube with 500 µl of ice-cold 2-propanol. The RNA pellet is washed once with 1ml of diluted Ethanol (75%) and resuspended in ultrapure water. The RNA is depleted from DNA by DNase I treatment for 15 min at room temperature and purified using the RNA Clean & Concentrator Kit.

### **2.2.3.2 Polymerase chain reaction (PCR)**

In order to determine single contacts made between long-distance regulatory elements that are close in space, PCR was used to amplify small amounts of ligation junctions from i3C and 3C templates. Locus-specific primers were designed using Primer3Plus close to restriction sites to have amplicons between 70 and 250 bp. Typically, 100 ng of i3C/3C template DNA was incubated with 10 pmol of 3C primer pairs, 1 % DMSO and 'ready-to-load'-DreamTaq PCR Mix in a total volume of 25  $\mu$ l. PCR was carried out in a BioRad Thermal Cycler as follows: one cycle at 95°C for 2 min (initial denaturation), followed by 34 cycles at 94°C for 40 s (cyclic denaturation)/ 55-61°C for 40 s (cyclic annealing)/ 72°C for 1 min (cyclic elongation), before a final extension at 72°C for 2 min. Afterwards, PCR products were analysed by Agarose gel electrophoresis. A list of 3C primer sequences is provided in table 8.

### **2.2.3.3 Reverse-Transcriptase cDNA Synthesis**

Isolated RNA was used to generate cDNA in a reverse transcription reaction. Therefore 1  $\mu$ g RNA was incubated with 1  $\mu$ l of random primers (50 ng/ $\mu$ l), 1  $\mu$ l of dNTP Mix (10 mM) and sterile water in a total volume of 12  $\mu$ l for 5 min at 65° and quick-chilled on ice for 5 min. Afterwards the master mix including the SuperScript II Reverse Transcriptase was added and primer annealing was allowed for 10 min at 25 °C followed by synthesis reaction for 50 min at 42 °C. The reaction was inactivated by heating up to 70 °C for 15 min.

### **2.2.3.4 Quantitative real-time PCR (qPCR)**

Quantitative real-time PCR is a sensitive method for detection and quantification of nucleic acids in different samples. This methodology was on one hand used to detect DNA abundance of 3C ligation products and make conclusions on interaction frequencies between loci of interest. On the other hand this technique was used to discriminate between different RNA levels in gene knock-down experiments. During this study, the SYBR Green I method was performed, which is based on PCR coupled to fluorescence emission from a reporter molecule. The SYBR Green I dye intercalates with double-stranded DNA and fluorescence increase is detected with accumulation of the PCR product during each amplification cycle at real time. The qRT measurements were performed in technical triplicates using the 72-well format of the Rotor-Gene Q system (Qiagen). For this, 5-10 ng of i3C/3C template DNA or



cDNA were mixed with SYBR Green JumpStart Taq ReadyMix and 0.16  $\mu\text{M}$  of the appropriate primer pair in a total volume of 15  $\mu\text{l}$ . The PCR reaction was incubated according to manufacturer's instructions described in table 15.

**Table 15 | qRT-PCR cycling conditions**

Step	Temperature	Time	Cycle number
Hold	95 °C	5 min	1
Denaturation	95 °C	15 sec	35
Annealing	60 °C	40 sec	
Extension	72 °C	20 sec	
Hold	4 °C	~	~

Primers were designed using Primer3Plus to have a length of 18-23 nucleotides, a  $T_m$  of 58-62°C, and to yield amplicons of 70-150 bp.  $i3C/3C$  PCR product levels were normalized to both a "loading" primer pair (for equiloading), and to control 3C templates including bacterial artificial chromosomes (BACs) spanning the studied loci of interest (see table 12). Gene expression studies were performed with primers spanning exon-exon borders to specifically detect cDNA and normalized to amplicon levels of housekeeping genes that are unperturbed throughout all measurements independent of sample condition (control vs. knockdown). Data were analyzed according to the  $\Delta\Delta C_t$  method (Livak and Schmittgen, 2001). The  $C_T$  (cycle threshold) value is defined as an intersection of the threshold line and the exponential reaction curve. The numbers of cycles required to reach this spot above fluorescence background represent the real PCR signal. The following formula was used to calculate the fold change:

$$\Delta C_T1 = C_{t(\text{Target in explorative population})} - C_{t(\text{Housekeeping/Loading in explorative population})}$$

$$\Delta C_T2 = C_{t(\text{Target in control population})} - C_{t(\text{Housekeeping/Loading in control population})}$$

$$\Delta\Delta C_T = \Delta C_T1 - \Delta C_T2$$

$$\text{Fold change} = 2^{(-\Delta\Delta C_T)}$$

### **2.2.3.5 Agarose gel electrophoresis**

To determine efficiency of chromatin digestion with restriction enzymes and subsequent ligation while preparing i3C and 3C templates, approximately 500 ng DNA aliquots were mixed in a ratio of 1:6 with loading dye and run at 80 V for 2 h in 0.5x TBE electrophoresis buffer, using 1% (when six bp-cutting enzymes were used) or 1.5% (when four bp-cutting enzymes were used) agarose gels. 4C inverse PCR and regular PCR products were run using 1.5 % agarose gels.

### **2.2.4 Chromosome conformation capture technology**

#### **2.2.4.1 Intrinsic 3C (i3C)**

An adapted close-to-physiological, isotonic, buffer is prepared fresh every time in nuclease-free water (Millipore MilliQ), supplemented with 25 U/ml RiboLock (Thermo Scientific) and protease inhibitors (Roche), and kept on ice throughout the procedure. The buffer composition is approximately cytoplasmic (pH 7.4) and contains 22 mM Na<sup>+</sup>, 130 mM K<sup>+</sup>, 1 mM Mg<sup>2+</sup>, 132 mM Cl<sup>-</sup>, 11 mM phosphate, 1 mM ATP and 1 mM dithiothreitol. Initially, levels of Cl<sup>-</sup> seem unphysiologically high, but *in vivo* proteins function as counter-ions. In order to prevent activation of nucleases and protect DNA integrity, ATP was added as chelating agent for Mg<sup>2+</sup> (Jackson *et al.*, 1988). Typically, 5x10<sup>6</sup> cells are used per experiment, harvested in 5 ml of ice-cold PB from 15-cm culture plates using a soft rubber cell scraper (Roth) on ice. To lyse cells and release nuclei, harvested cells are spun at 160 x g (4°C, 5 min), resuspended in 10 ml of ice-cold PB supplemented with 0.4% NP-40 and incubated on ice for 15 min, followed by centrifugation of nuclei at 600 x g (4°C, 5 min). Depending on cell type, this step is usually repeated one or two times with incubation times on ice between 5-10 min and PB/0.4% NP-40 volumes between 5-10 ml. To check nuclei integrity, 10 µl of this suspension is mount on a hemocytometer. Isolated nuclei are collected via centrifugation at 600 x g (4°C, 5 min), gently resuspended in 500 µl of ice-cold PB/0.4% NP-40 with 500 units units of *ApoI* or *NotI* (New England Biolabs), and transferred to 2-ml round-bottom, low-retention tubes. Next, Chromatin is digested without shaking at 33°C for 30 min. Aliquots of 10 µl are put aside right before and after digestion as “uncut” and “cut” chromatin controls and kept on ice. Treated nuclei are then spun at 600 x g (4°C, 5 min) to separate cut, unattached chromatin fragments that are released into the supernatant from chromatin fragments that are

associated to sub-nuclear structures and remain in the nucleus. After one wash step in 500  $\mu$ l ice-cold PB, nuclei are respun, followed by resuspension in 1ml ice-cold PB with 20 $\mu$ l T4 DNA Ligase (5U/ $\mu$ l; Invitrogen) and 10  $\mu$ l BSA (10 mg/ml), and incubation at 16°C for 8 h or over-night without shaking. During this step, spatially-proximal chromatin ends are ligated together in intact nuclei, an idea based on the original “proximity ligation” assay by Cullen *et al* (1993) and supported by recent findings that, even under crosslinked conditions, ligations predominantly occur within the “chromatin cage” of intact nuclei. Finally, 30  $\mu$ l proteinase K (10 mg/ml) are added to the samples, which are kept at 42°C for 8-12 h. Then samples are treated with 30  $\mu$ l RNase A (10 mg/ml) at 37 °C for 1 h and purified by phenol/ chloroform extraction (pH 8.0) and ethanol precipitation. i3C templates were used for detection of long-range contacts by PCR and qRT-PCR or processed further to create 4C templates. Aliquots of “uncut” and “cut” chromatin controls were mixed with 200  $\mu$ l of 1 x PBS and incubate with 5  $\mu$ l of proteinase K (10 mg/ml) at 42 °C for 2 hours. Then 5  $\mu$ l of RNase A (10 mg/ $\mu$ l) are added and controls are incubated at 37 °C for 30 minutes. DNA was purified as described in 2.2.3.1 and samples were analyzed using agarose gel electrophoresis.

#### **2.2.4.2 Conventional 3C**

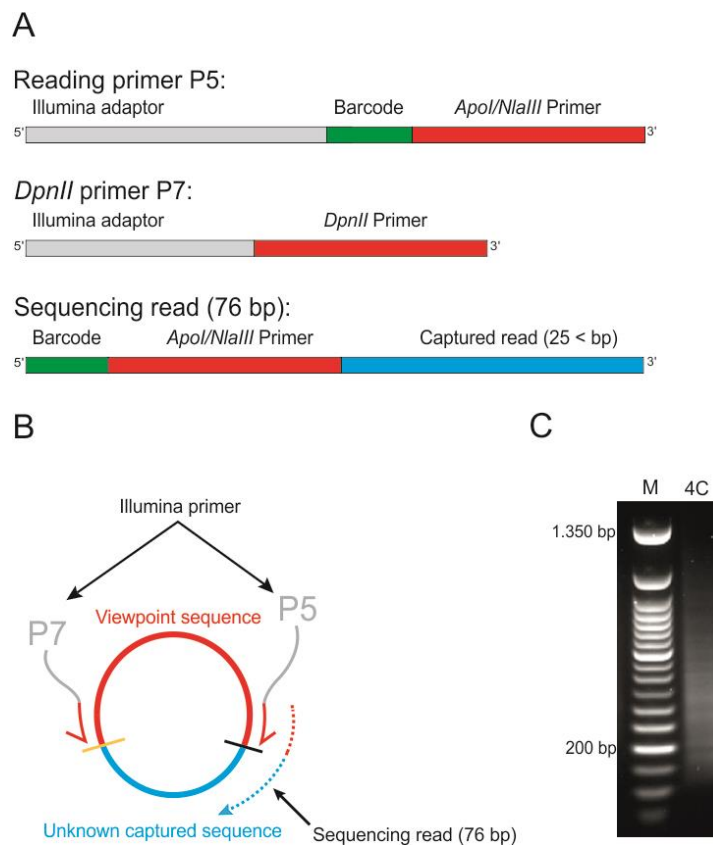
Conventional 3C/4C experiments were performed following a previously established protocol as described (Stadhouders *et al*, 2013). In brief,  $10^7$  cells are fixed in appropriate cell culture medium containing 1% para-formaldehyde for 10 min at room temperature on a rocking plate, followed by quenching in ice-cold 0.125 M glycine/PBS for 5 min at room temperature and subsequent harvesting of cells using a soft rubber cell scraper on ice. Cell nuclei are isolated in lysis buffer and pelleted at 600 x g (4°C; 5 min). The nuclei pellet is incubated in 500 $\mu$ l of 1.2 x appropriate restriction enzyme buffer supplemented with 0.3% SDS for 1h at 37°C with agitation (900 rpm). After sequestering SDS with addition of 2% Triton X-100 (shaking at 900 rpm for 1h; 37°C), nuclei are digested overnight with 400 units of *ApoI* or *NotI* at 37 °C while shaking (900 rpm). Aliquots of 10  $\mu$ l are put aside right before and after digestion as “uncut” and “cut” chromatin controls and kept on ice. Next day, the enzyme is heat inactivated by adding 1.6% SDS for 20-25 min at 65°C. Nuclei are then diluted to 7 ml in 1.15 x ligation buffer supplemented with 1% Triton X-100 to sequester SDS for 1h at 37°C, followed by ligation in the presence of 100 units of T4

DNA ligase 6-8 h at 16°C. Crosslinks are reversed in the presence of 30 µl proteinase K (10mg/ml) at 65°C overnight and then 30 µl RNase A (10 mg/ml) at 37°C for 1 h. DNA is isolated according to standardized phenol/chloroform extraction and ethanol precipitation described in 2.2.3.1. This serves as template in 3C-qPCR or 4C-seq, using the same primers as for i3C/i4C. Aliquots of “uncut” and “cut” chromatin controls were mixed with 200 µl of 1 x PBS and incubate with 5 µl of proteinase K (10 mg/ml) at 42 °C for 2 hours. Then 5 µl of RNase A (10 mg/µl) are added and controls are incubated at 37 °C for 30 minutes. DNA was purified as described in 2.2.3.1 and samples were analyzed using agarose gel electrophoresis.

### **2.2.4.3 i4C/4C-seq**

i4C-/4C-seq are based on template circularization and inverse-PCR and were performed as previously described (Stadhouders *et al*, 2013). For this, ~25 µg of i3C/3C template were digested with 25 units of *DpnII*, *MseI* or *MspI* at 37°C overnight. Next day, restriction enzyme was heat inactivated at 65°C for 25 min and DNA was diluted in 1.15 x ligation buffer to a volume of 7 ml supplemented with 20 µl T4 DNA ligase (5 U/µl). After incubation at 16°C for 6-8 h, i4C/4C template DNA was purified as described in 2.2.3.1. Then, ~100-200 ng of the circularized i4C/4C template was mixed with 4C Illumina primer pair, DMSO and reagents from the Expand long template HF DNA Polymerase PCR system (Roche; see table 3). Inverse-PCR reactions were performed according to settings described by Stadhouders *et al.* using a BioRad Thermal Cycler (Stadhouders *et al.*, 2013). Typically, 8 such PCR reactions are pooled, purified using the DNA Clean & Concentrator kit (Zymo Research), and amplicons checked by electrophoresis on a 1.5% (wt/vol) agarose gel (see figure 8C). The rest of the sample was directly sequenced on a HiSeq2500 platform (Illumina) as the primers used in inverse-PCR reactions carry the P5/P7 Illumina adapters as overhangs. For each viewpoint fragment of interest two illumina primers (P5 and P7) were designed (see figure 8 A&B). The reading primer, from which library sequencing is initiated, needs to be located as close as possible and maximum 5 bp separated from the *ApoI/NlaIII* restriction site. This primer contains a P5 illumina adapter sequence (5'-AATGATACGGCGACCACCGAACACTC-TTTCCTACACGACGCTCTTCCGATCT-3'), followed by a 6 bp-long barcode to tag parallel sequenced samples for multiplexing purposes and a fragment specific annealing primer. A 76 bp long read

length is used to obtain sufficient information from the unknown captured sequence (25 bp <). The second primer is not required for sequencing and can therefore be located within 100 bp from the restriction site of the second enzyme (*DpnII*, *MseI* or *MspI*). This primer contains a P7 adapter sequence (5'-CAAGCAGAAGACGGCATACGA-3'; Illumina), followed by a fragment specific annealing primer. All annealing primers have a length of 18-23 nucleotides, a  $T_m$  of 54-59°C, and were designed using Primer3Plus. Illumina oligonucleotide sequences are copyright of 2007-2012 Illumina. A list with all 4C primers is provided in table 9.



**Figure 8 | Strategy for 4C primer design. A** | The reading primer contains a P5 illumina adaptor sequence, followed by a 6 bp long barcode and a fragment specific annealing primer. The non-reading primer contains a P7 illumina adaptor sequence, followed by a fragment specific annealing primer. **B** | Schematic illustration of 4C inverse PCR. **C** | 4C inverse PCR amplicons were run on a 1.5 % (wt/vol) agarose gel by electrophoresis. Smearly band profile indicates captured and amplified sequences.

The bioinformatic analysis of high throughput sequencing data from i4C-/4C-seq experiments was carried out by Theodore Georgomanolis using the fourSig (Williams *et al.*, 2014) or FourCSeq (Klein *et al.*, 2015) packages. In brief, resulting 76-bp long reads were trimmed using the appropriate homerTools command in order to remove the viewpoint primer sequences and receive captured, unknown reads (Citation?).

These reads were then mapped to the reference genome (hg19) using the short read aligner BWA-MEM36 according to following parameters: `bwa mem -t 8 -k 15 -r 1 -B 1 -M`) and further processed via fourSig/FourCSeq. 4C/i4C-profiles were visualized on the UCSC genome browser (<https://genome.ucsc.edu/>; hg19) and by uploading BedGraph files with editable smoothing options.

#### **2.2.4.4 iT2C and T2C**

Many 3C-derived approaches, like 4C, 5C or Hi-C, come with limitations, such as being restricted to only one viewpoint or requiring massive and expensive sequencing efforts. The Targeted chromosome conformation capture technique (T2C/iT2C) provides local interaction maps of selected loci of interest at restriction fragment resolution. For this 3C/i3C templates were prepared as described in 2.2.4.1 and 2.2.4.2 using *ApoI*, followed by secondary enzyme digestion with *DpnII*, overhang end repair and sequencing adapter ligation. These libraries are then hybridized to a set of oligonucleotide probes that are immobilized on beads and were specifically designed close to the restriction site of every *ApoI* restriction fragment in a 2.8-Mbp locus of interest. The captured DNA, which contains all interactions that are made with the region of interest, was then pair-end sequenced to >50 million reads on an Illumina HiSeq2000, mapped to the reference genome (hg19) and analyzed with custom R scripts by Milos Nikolic (see below). Finally, data were binned and visualized via 2D interaction maps at 10-, 5-, and 1.5-kbp resolution. Interaction enrichments for CTCF and different histone modifications were visualized in PE-SCAN plots at highest resolution (de Wit *et al.*, 2013). TAD/subTAD boundaries were called by calculating the “directionality index” as described in previous published literature (Dixon *et al.*, 2012). Note that i3C/3C templates were generated in our laboratory and then processed as described above in a collaboration-based project with the laboratory of Frank Grosveld (Erasmus MC, NL).

#### **2.2.4.5 Whole-genome iHi-C implementation**

iHi-C was performed using 20 million HUVECs. Cell nuclei were isolated in the normal flow of the i3C protocol as described in x. Chromatin was digested using 2000 U of *DpnII* at 33°C for 30 min without agitation. After centrifugation of cut nuclei at 600 g and 4°C for 5 min, cohesive DNA fragment ends were filled-in with dGTP (30 mM), dCTP (30 mM), dTTP (30 mM) and biotin-dATP (30 mM) using 150 U of

Klenow Fragment in a total volume of 500  $\mu$ l PB/0.4% NP-40 at 33°C for 1h. Then, nuclei are spun and blunt DNA ends are ligated in intact nuclei in a total volume of 1 ml PB supplemented with either 0.1 mg/ml BSA or 20 % PEG and 200 U of T4 DNA Ligase at 16 °C over-night. Next, DNA was isolated according to standardized protocols (see 2.2.3.1) and biotin-dATP was removed from non-ligated DNA ends using the exonuclease activity of T4 DNA polymerase. For this, 5  $\mu$ g of iHi-C library DNA was incubated with 1  $\mu$ l BSA (10 mg/ml), 10  $\mu$ l 10 x NEBuffer2, 1  $\mu$ l dGTP (10 mM) and 5 U of T4 DNA polymerase in a total volume of 100  $\mu$ l at 12 °C for 2 h. DNA was subsequently purified as described in 2.2.3.1 and sonicated to ~ 800 bp on a Bioruptor. Finally, biotinylated ligation junctions are selected on streptavidin coated beads, washed to remove non-captured DNA using a magnetic rack, amplified for 6 PCR cycles to add sequencing linkers, and sequenced to 300 million read pairs (100 bp in length) on a HiSeq4000 platform (Illumina). In order to provide maximum recovery of uniquely mapped pairs, resulting reads were mapped to the reference genome (hg19) reiteratively using Burrows-Wheeler Aligner (BWA; Li H. & Durbin R., 2010). BAM files containing mapped reads were merged, duplicates were removed (<http://picard.sourceforge.net/>), and the output converted into BEDPE format (Quinlan & Hall, 2010). Then, reads were assigned to equally-sized genome bins (25-250 kbp) by applying custom R scripts and read counts were normalized to library size. Next, the HiTC package (Servant *et al.*, 2012) was used to annotate and correct matrices for biases in genomic features and visualize 2D heat maps. Bioinformatic analysis were performed by Milos Nikolic.

#### **2.2.4.6 TALE-iD**

We developed TALE-iD to investigate native chromatin interactions in another independent manner. For this, we used constructs encoding a Transcription activator-like effector (TALE) repeat domain fused to the LSD1 histone demethylase and replaced LSD1 with the open reading frame of the bacterial DNA adenine methyltransferase (Dam) gene by molecular cloning as described in 2.2.1.6. Dam methylates adenines near your region of interest and allows to study long-range chromatin interactions upon restriction enzyme digestion at these sites (Vogel *et al.*, 2007). All TALE constructs have been previously proofed to bind at their target DNA binding site by CHIP studies (Mendenhall EM. *et al.*, 2013). One TALE candidate was shown to bind an enhancer region downstream of the ZFPM2 TSS that is active in

K562 cells. Here, we transfected 3  $\mu\text{g}$  of this construct in  $10^6$  K562 cells using the Nucleofector kit V as described in 2.2.2.3. In parallel, we also introduced 3 $\mu\text{g}$  of two TALE-DamID control constructs respectively, one with a scrambled non-specific DNA-binding domain and one construct carrying a disrupted Dam ORF to inactivate the methylase. Cells were incubated 48 h post-transfection with a permeabilization buffer at 37 °C over-night and total genomic DNA was isolated as described in 2.2.3.1. Then 5  $\mu\text{g}$  genomic DNA were digested with 20U of *DpnI* in a total volume of 100 $\mu\text{l}$  1 x CutSmart Buffer (NEB) at 37 °C over-night and purified using the Zymo Research Clean & Concentrator Kit 5. In this step *DpnI* will ideally only cut at GATC sites that were methylated by Dam. Finally, qPCR was performed to quantify Dam methylation using primers that encompass *DpnI* sites in the ZFPM2 locus. Resulting low amplification levels indicate efficient *DpnI* digestion and hence high methylation at the respective site of interest. Fold change was calculated after normalizing for the primer amplification efficiency on genomic K562 DNA, as well as for random background *DpnI* digestion of genomic K562 DNA. All primers used in this experiment are provided in table 10.



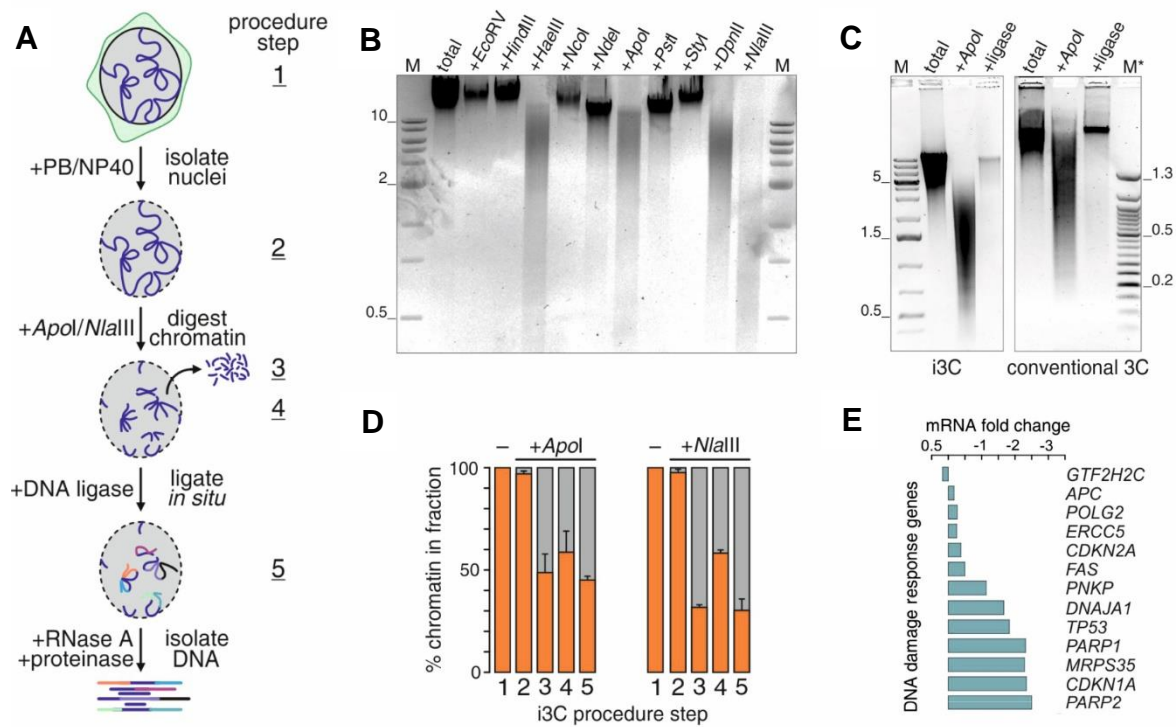
## 3. Results

### **3.1 Features of intrinsic chromosome conformation capture (i3C) methodology**

Chromosome conformation capture assays are broadly used to study chromosome folding in 3D space. In order to eliminate biases arising from key steps of the 3C protocol, such as formaldehyde cross-linking and treatment with non-physiological buffers, here we introduce a modified “physiological buffer” that best approximates nucleoplasmic conditions and retains most transcriptional activity to satisfactorily preserve nuclear structure (Jackson *et al.*, 1981). The presented i3C methodology exploits native interactions to stabilize DNA molecules that were in close spatial proximity, thus comprising an alternative method for the analysis of intrinsic genome organization.

#### **3.1.1 Establishment of the i3C protocol**

Intact nuclei are isolated from living, uncross-linked cells in a “physiological buffer” to maintain chromatin structure (overview in Fig. 9A). Both outer-cell and nuclear membranes are permeabilized under mild conditions using a non-ionic detergent to render the nuclear compartment accessible to enzymes. The next step involves chromatin digestion in intact nuclei using a restriction endonuclease that recognizes either 6- or 4-bp palindromes. The choice of the restriction enzyme influences digestion efficiency and subsequent resolution; not all commercial restriction enzymes work equally well in isotonic buffer at the sub-optimal temperature of 33°C, and we found that *ApoI* and *NlaIII* cut chromatin most efficiently with the majority of cut fragments spanning 2-7 kb or 0.5-1.5 kb, respectively (Fig. 9B). Then, still-whole nuclei are spun to allow removal of cut chromatin not associated with any nuclear substructure. Critically, this step removes >40% of chromatin that is not attached (or only weakly attached), thus enriching for the remaining associations (Fig. 9D). The signal-to-noise ratio, a key consideration of yielding information-rich interactomes, is then further improved by proximity ligation of cohesive ends *in situ* in intact nuclei, rather than in dilute solutions (Gavrilov *et al.*, 2013; Rao *et al.*, 2014). Preliminary work using this much simplified protocol, in which chromatin is efficiently cut and re-ligated, is shown for *ApoI*-digested human primary endothelial cells (Fig. 9C).



**Figure 9 | Features of the i3C approach.** **A** | Strategic overview of the i3C protocol. Cells are harvested in PB (step1). Nuclei are isolated in PB supplemented with 0.4 % NP-40 on ice (step 2). Chromatin is digested with ApoI or NlaIII, nuclei are spun and unattached fragments are released to the supernatant (step 3) and leave cut chromatin that is associated with subnuclear structure in the nucleus (step 4). *In situ* ligation is performed (step 5). **B** | Digestion efficiencies of commercially available restriction endonucleases in PB at 33 °C for 30 minutes. **C** | Chromatin digestion and re-ligation efficiency demonstrated for i3C and conventional 3C protocols. **D** | Chromatin quantification in percentage during different steps of the i3C protocol. **E** | Bar plots show fold changes in mRNA levels of DNA damage response genes comparing whole-cell RNA-seq data to that from nuclei digested with DNase I in PB (Caudron-Herger *et al.*, 2015).

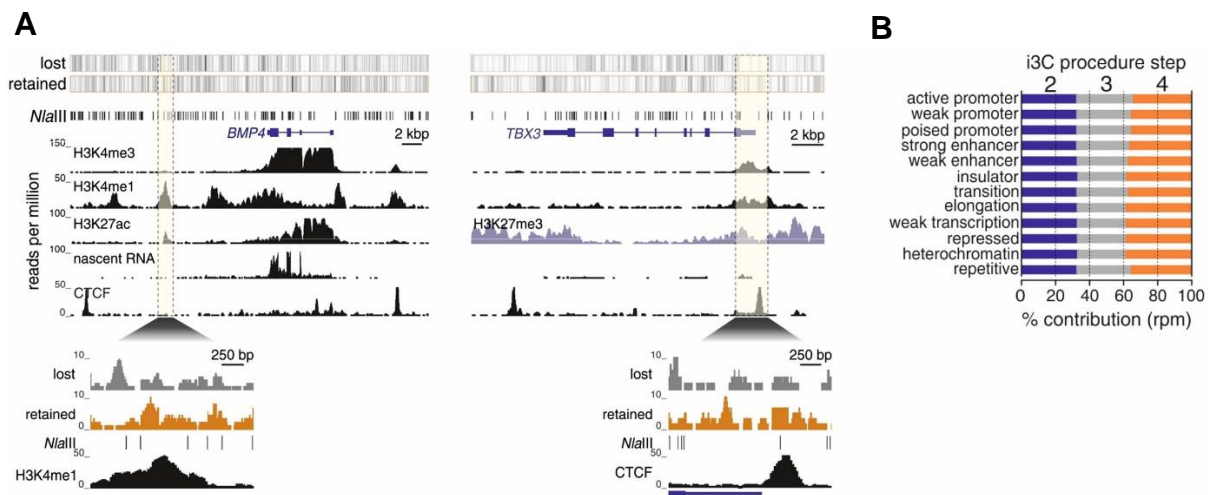
In order to investigate whether restriction of chromatin in living cells with an endonuclease triggers the DNA damage response, we compared whole-cell RNA-seq data to that from nuclei treated with more frequently cutting DNase I under same physiological conditions as in the i3C protocol (Caudron-Herger *et al.*, 2015). Fold changes in mRNA levels of a subset of DNA damage responsive genes showed no elevation (Fig. 9E), thus excluding the DNA damage response activation as a bias related to chromatin digestion.

To verify that ligation occurs *in situ* within individual nuclei under native conditions, we performed i3C on a nuclei mixture from an equal number of human endothelial (HUVEC) and mouse embryonic stem cells (mESC) and sequenced the resulting ligation hybrids. Mapping of read pairs to the human and mouse genomes, revealed

that <0.7% of the sequence hybrids match with both species (Brant *et al.*, 2016). Consistent with previously described experiments (Schönfelder *et al.*, 2010), this low value revealed that cross-ligations between mouse and human DNA fragments were rare, thereby confirming efficient single nucleus ligation.

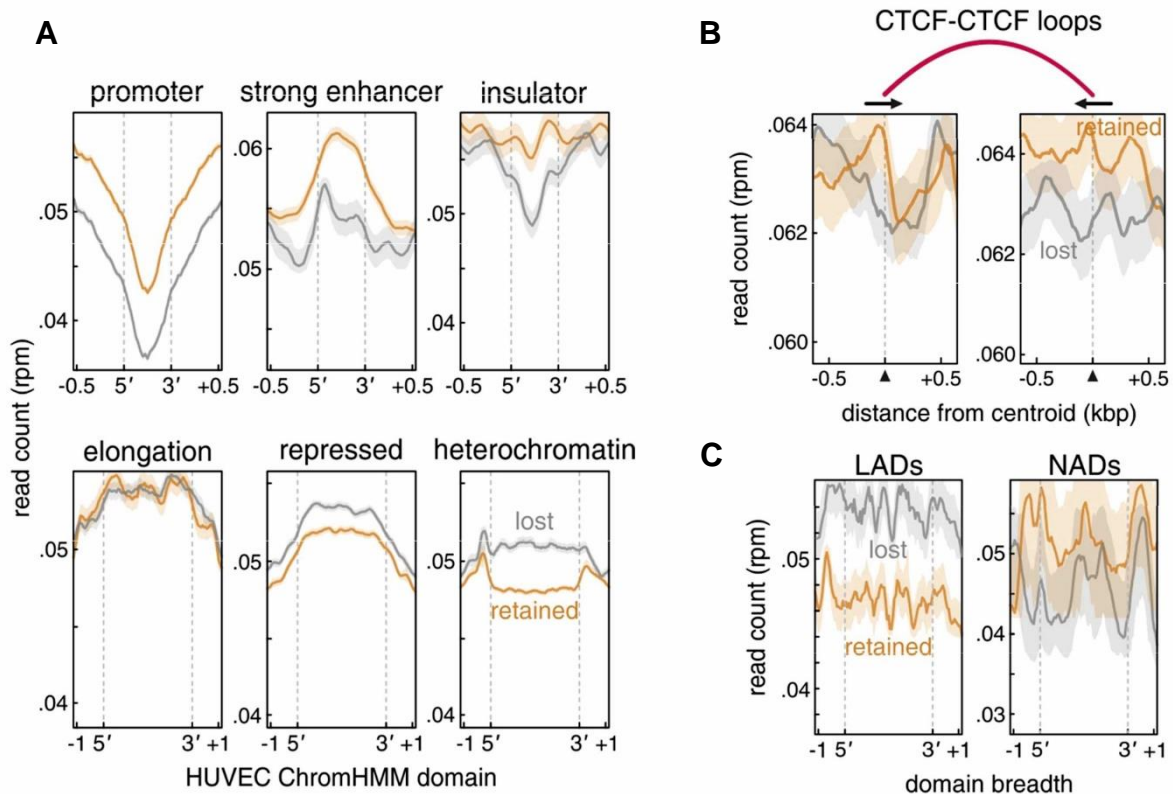
### **3.1.2 Chromatin features of the i3C template**

The i3C protocol involves loss of a substantial amount of DNA after chromatin restriction digest. To examine if cutting and subsequent removal of non-associated DNA fragments is biased towards eu- or heterochromatic regions and hence introduces an imbalanced chromatin composition between the 'lost' and 'retained' DNA fractions, we treated HUVEC nuclei with *NlaIII* and isolated DNA at steps 2 (total chromatin), 3 ('lost' chromatin), and 4 ('retained' chromatin) of the i3C procedure (Fig. 9A). Then, the DNA from each fraction was sequenced and mapped to the human reference genome. The read coverage of 'lost' and 'retained' DNA fractions was comparably divided between active and inactive loci (exemplified here via the *BMP4* and *TBX3* gene loci; Fig. 10A). Next, read pairs from the different DNA fractions were correlated to 'active' (H3K4me1/2/3, H3K27ac, H3K36me3) and 'inactive' histone marks (H3K27me3, H3K9me3) from HUVECs from the ENCODE ChIP-seq database, classified into different chromatin domains of the human genome using the Hidden-Markov-Model (HMM) (Ernst *et al.*, 2012). We found that the relative contribution of the different chromatin HMM domains did not differ between the two i3C fractions, closely mirroring HMMs in total cell chromatin (Fig. 10B).



**Figure 10 | Genomic coverage of different i3C fractions. A |** Genomic coverage of ‘lost’ and ‘retained’ fractions from steps 3 and 4 of the i3C protocol. Isolated DNA was sequenced to ~ 40 million reads and mapped to the human genome. ENCODE ChIP-seq data was aligned below. Exemplary browser views for active *BMP4* and inactive *TBX3* genes are shown. **B |** Relative contribution of HUVEC chromatin HMM features in each i3C fraction.

More detailed analysis of HMM annotation of the ‘retained’ or ‘lost’ fractions showed that the ‘retained’ fraction contains more genomic elements that are related to promoters, strong enhancers and insulators than the ‘lost’ fraction (Fig. 11A). On the other hand, the ‘lost’ fraction comprises more genomic elements from repressed and heterochromatic regions (Fig. 11A). Furthermore, we examined the read coverage around convergent CTCF sites involved in chromatin looping and lamin- (LADs) and nucleolus-associated domains (NADs) and observed an increase of reads annotated to such CTCF sites and NADs in the ‘retained’ fraction (Fig. 11B), while more regions linked to LADs were found in the ‘lost’ fraction (Fig. 11B).



**Figure 11 | Chromatin features of ‘lost’ and ‘retained’ i3C fractions.** **A** | Line plots demonstrated raw read coverage (in reads per million) of ‘retained’ (orange) and ‘lost’ (grey) i3C fractions along regions regions correlated to HUVEC ChromHMM features. **B** | Line plots demonstrated raw read coverage (in reads per million) of ‘retained’ (orange) and ‘lost’ (grey) i3C fractions around CTCF-mediated chromatin loops that were obtained from Hi-C data (Rao *et al.*, 2014). **C** | Line plots demonstrated raw read coverage (in reads per million) of ‘retained’ (orange) and ‘lost’ (grey) i3C fractions along lamin- (LADs) and nucleolus-associated domains (NADs) that were obtained from publicly available fibroblast data (Guelen *et al.*, 2008; Nemeth *et al.*, 2010).

Taken together, these results suggest that chromatin treatment with an endonuclease under physiological conditions via the i3C procedure results in read profiles of ‘lost’ and ‘retained’ nuclei fractions that show no significant biases for or against active and inactive loci; thus, the chromatin on which i3C ligation is performed adequately represents the different nuclear epigenetic landscapes.

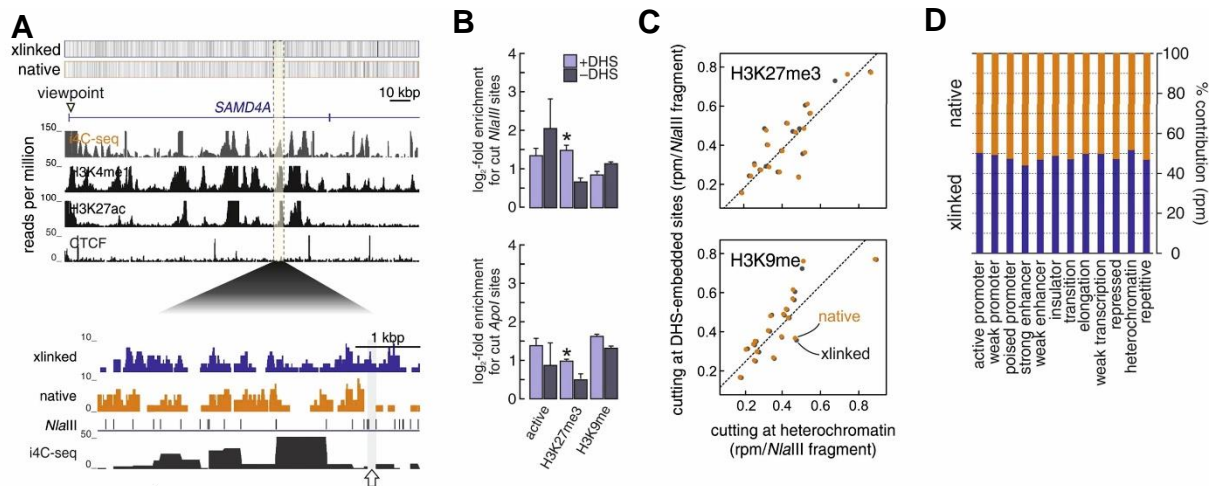
### 3.1.3 Comparison of i3C and conventional 3C templates

The conventional 3C protocol is based on cell cross-linking with formaldehyde to capture a ‘snapshot’ of its current chromatin architecture. The subsequent restriction enzyme digestion step requires prior SDS treatment of chromatin and extensive incubation times in order to expose more restrictions sites on the DNA fiber and, thus, critically improve cutting efficiency (Dekker *et al.*, 2002). In contrast, the i3C

protocol avoids chemical cell fixation and implements chromatin digestion under short and suboptimal conditions (Brant *et al.*, 2016). We therefore compared the features of the resulting cross-linked and native 3C templates. For this, nuclei from HUVECs were treated in two separate experiments with *NlaIII* as described in the conventional or the native 3C protocol (Dekker *et al.*, 2002; Brant *et al.*, 2016); then, digested DNA was isolated, sequenced, and mapped to the human genome. Exemplary coverage of DNA fragments obtained from chromatin digested under cross-linked or native conditions is illustrated for an active enhancer region of the *SAMD4A* locus (Fig. 12A). Both methods reveal an overall equal read coverage distribution along the locus, which is also congruent with the displayed i4C contacts. Yet, fragments missing from the i3C template that also lack i4C-signal, are still represented in the cross-linked template (Fig. 12A), pointing to minor differences among the conventional and the native 3C DNA template compositions.

Next, we set out to investigate if restriction enzymes cut chromatin preferably at DNase hypersensitive sites (DHS) under physiological conditions, hence resulting in contact profiles that are mostly biased towards “open” chromatin. Thereto, nuclei from HUVECs were treated either with *ApoI* or *NlaIII* according to the i3C procedure and cut DNA was used as template to measure digestion efficiency at restriction sites embedded within or outside DNase hypersensitive sites in active, Polycomb-marked (H3K27me3) or constitutively-heterochromatic (H3K9me1) regions. This line of experiments showed similar digestion efficiencies in the distinctive regions and confirmed that the chromatin restriction step in the native protocol is not restrained to “open” chromatin (Fig. 12B). To evaluate the quality of the i3C/3C template further, we compared sequenced read counts contained in DHS versus heterochromatic regions per chromosome after treating nuclei with *NlaIII* under conventional or native conditions (Fig. 12C). Our findings are depicted in a scatter plot and demonstrate how digestion efficiencies are essentially identical between the two approaches, without constraints for “open” or “closed” chromatin regions. Furthermore, read pairs from conventionally or physiologically digested chromatin were associated to ‘active’ and ‘inactive’ ENCODE HMM domains and classified into different chromatin motifs of the human genome (Hoffman *et al.*, 2013), which showed that the different chromatin HMM features contribute to a similar extent to the chromatin composition of both the i3C and 3C templates (Fig. 12D). On the basis of this data, we conclude that cutting chromatin under native or conventional conditions does not produce

different profiles as regards digestion efficiency and preferences towards “open” chromatin.

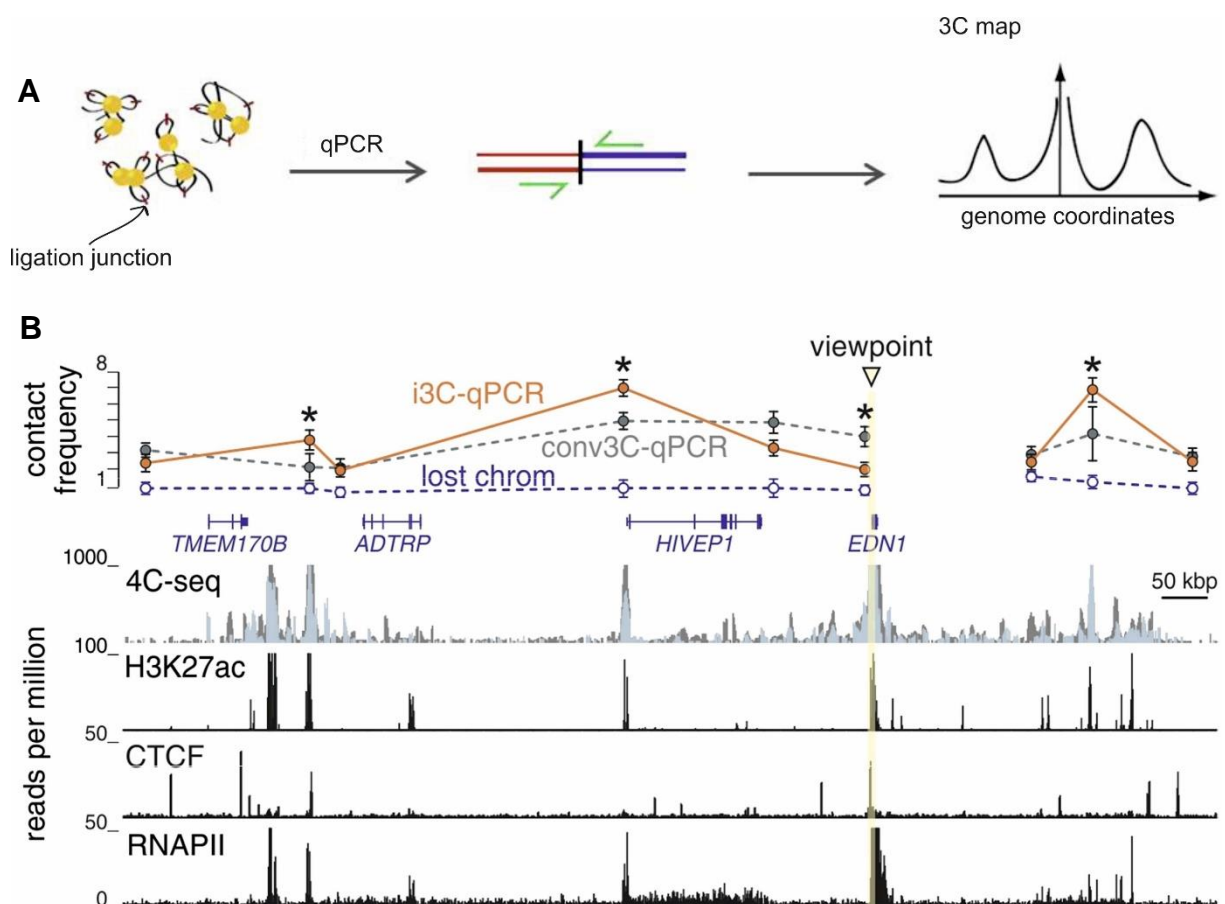


**Figure 12 | Comparison of i3C and conventional 3C DNA templates. A |** Coverage of *NlaIII* fragments digested under crosslinked (purple) or native (orange) conditions. Cut and isolated DNA from HUVECs was sequenced to ~ 40 million reads and mapped to the genome. ENCODE ChIP-seq data was aligned below the active *SAMD4A* locus. Depleted i4C signal matching depleted native read coverage is indicated by an arrow. **B |** Log<sub>2</sub>-enrichment for cut *ApoI* (bottom) or *NlaIII* (top) sites shows digestion efficiency from two replicates at DNase hypersensitive sites (grey (-) or purple (+)) and at active, polycomb-marked (H3K27me3) or heterochromatic (H3K9me1) regions. \*P<0.05; unpaired two-tailed Student’s t-test (n=4). **C |** Scatter plots compare number of reads contained in DHS versus heterochromatic regions (H3K9me1 (bottom) and H3K27me3 (top)) after cutting with *NlaIII*. Each dot represents one chromosome. Data were compared under native (orange) and cross-linked (grey) conditions. **D |** Bar plots represent relative contribution of HMM chromatin features in chromatin digested with *NlaIII* under native (orange) or cross-linked (purple) conditions. ENCODE ChIP-seq data from Hoffman *et al.*, 2013 was used.

### 3.1.4 Capturing native contacts: obtaining the first i3C interaction profile

To prove that nuclear structure is preserved during the i3C procedure and that no chemical cell-fixation is needed to capture spatial chromatin interactions, we probed for contacts made by the promoter of the constitutively expressed *EDN1* gene that were previously described using conventional 4C technology (Diermeier *et al.*, 2014). We prepared i3C and 3C templates from HUVECs using *ApoI* (Brant *et al.*, 2016; Stadhouders *et al.*, 2013) and assessed contact frequencies between an anchor fragment at the *EDN1* promoter and 9 fragments in the extended *EDN1* locus by amplification of ligation junctions with qPCR (Fig. 13A). Contact peaks of the previously-obtained conventional 4C profile were fully recapitulated in the i3C/3C

interaction graphs (Fig. 13B). However, the observed i3C contacts intensify precisely at interacting sites and decline at non-interacting control segments, while conventionally obtained 3C contacts often display inaccurate interaction levels along the locus. Since a major amount of non-associated DNA that correlates to active and inactive chromatin is lost in i3C, we performed ligation on the ‘lost’ fraction from step 3 (Fig. 9A) and checked for the possibility of a loss of contacts. As depicted by the blue dashed line in Figure 13B, no significant interactions were detected in the ‘lost’ i3C template, supporting our claim that major chromatin interactions are enriched in the i3C template.



**Figure 13 | i3C-qPCR implemented in the extended *EDN1* locus on chromosome 6. A |** Strategic overview of i3C/3C contact detection by qPCR. **B |** Contact frequency between the anchor fragment (*EDN1* TSS) and 9 surrounding segments were assessed using i3C (orange line), conventional 3C (dashed, grey line) and ligated chromatin from ‘lost’ i3C fraction (dashed, purple line) coupled to qPCR. CT values from two replicates were double-normalized to a loading control and a 3C template containing BACs that cover the examined locus. ENCODE ChIP-seq data from HUVECs were aligned below the profile. \* $P < 0.05$ ; two-tailed unpaired Student’s t-test ( $n = 2$ ).

Taken together, these results confirm that interactions between segments can be maintained and captured using the i3C protocol, even over longer-range distances

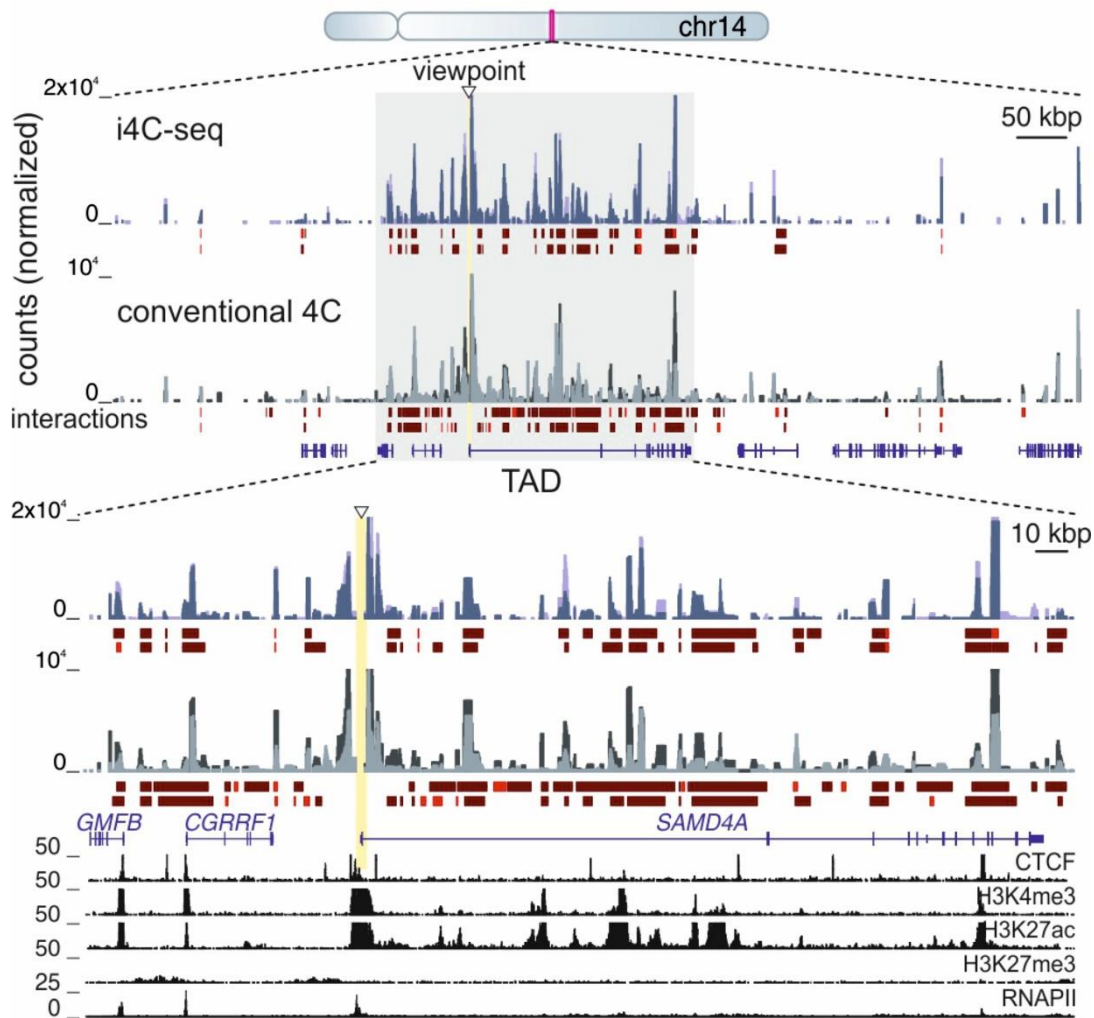


(~500 kbp in the case of the *EDN1* locus' Fig. 13B). Critically, conventional 3C profiles are accompanied with higher noise and imprecision, which the i3C method seems to be devoid of.

### **3.2 i4C technology: a native one-to-all approach**

#### **3.2.1 Implementation of i4C on the *SAMD4A* locus**

In contrast to 3C, 4C technology detects genome-wide interactions of a single locus of interest, typically called the 'viewpoint', without requiring previous knowledge of possible contacting sequences (Stadhouders *et al*, 2013). Here, we performed i4C-seq in HUVECs, which is based on re-digestion of the initial i3C template, followed by circularization via DNA ligation. Segments that contact the 'viewpoint' fragment are amplified by inverse PCR and sequenced as described (Stadhouders *et al*, 2013). We applied the approach side-by-side with conventional 4C on our model *SAMD4A* locus: This 221 kbp-long gene contains a super-enhancer core in its first intron and is transcribed in specialized 'NF $\kappa$ B'-factories upon TNF $\alpha$ -stimulation (Papantonis *et al.*, 2012), which makes it a good candidate for studying the dynamics of long-range promoter-enhancer looping. To compare interactions formed by the transcription start site (TSS) of *SAMD4A* under cross-linked and native conditions, we generated i4C- and 4C seq data from two independent replicates along a ~1-Mbp window around the 'viewpoint' (Fig. 14). 'FourSig' was used to analyse data, to correct i4C/4C reads for mapping biases, and to identify significant interactions (Williams *et al.*, 2014). Both, i4C and conventional 4C profiles display extensive similarities in their *cis*-interactions, particularly within the TAD harboring *SAMD4A* (Fig. 14). Yet, significant interactions called by 'fourSig' indicate a more frequent and non-uniform distribution of contacts in the conventional approach, which also fluctuate between replicates. On the other hand, i4C reveals only few additional contacts that are not seen with the conventional method, while replicates produced under physiological conditions closely resemble one another (Fig. 14).



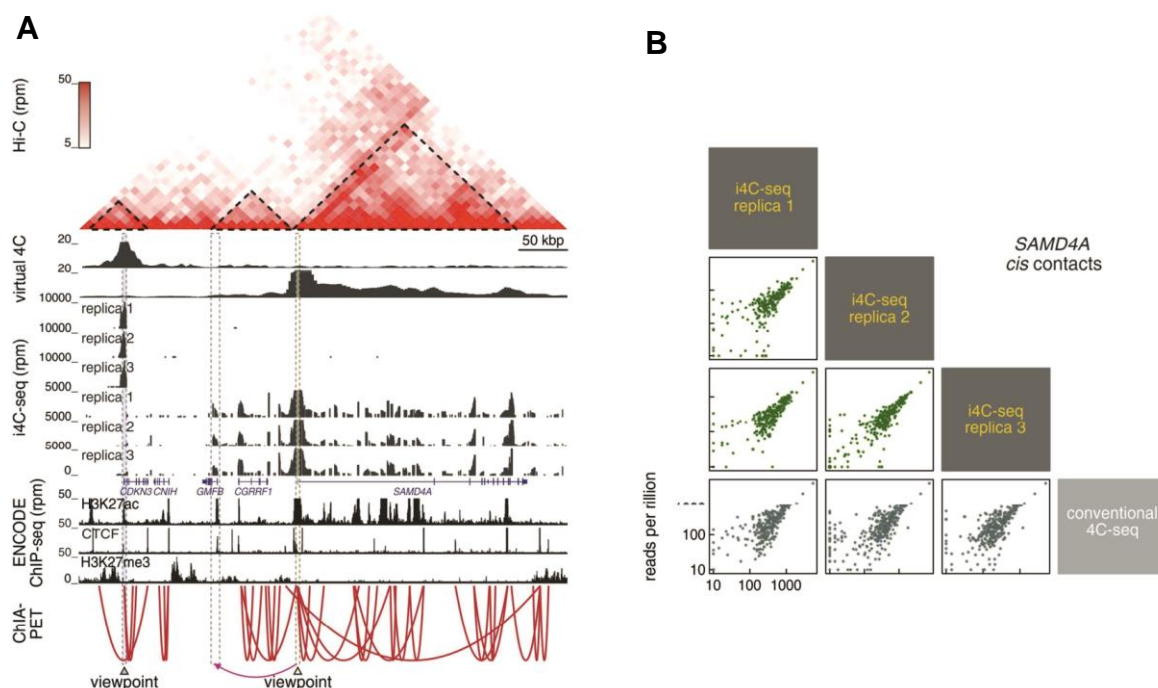
**Figure 14 | i4C – and 4C – seq were implemented on the *SAMD4A* locus.** Conventional 4C (grey shades) and i4C (purple shades) were performed in parallel in HUVECs using the *SAMD4A* TSS viewpoint and *ApoI*. Profiles from two replicates were overlaid. Strong (red) and intermediate (brown) interactions called by ‘fourSig’ were aligned below profiles. ENCODE ChIP-seq data are shown below.

Taken together, these data demonstrate comparable i4C- and 4C interaction profiles, highlighting the preservation of major structural chromatin associations under native conditions, despite the fact that a significant number of DNA fragments is lost during the i3C procedure. However, discrepancies in contact frequencies were detected between native and conventional 4C data, calling for further examination.

### 3.2.1.1 Comparison of i4C and conventional 4C data on the *SAMD4A* locus

Next, we analysed the reproducibility of i4C interaction profiles among three replicates that were generated in HUVECs using *ApoI* as first cutter and the *CDKN3* or *SAMD4A* transcription start sites as viewpoints. To prove that our native contact maps reveal interactions that can be confirmed by other 3C-techniques, we compare

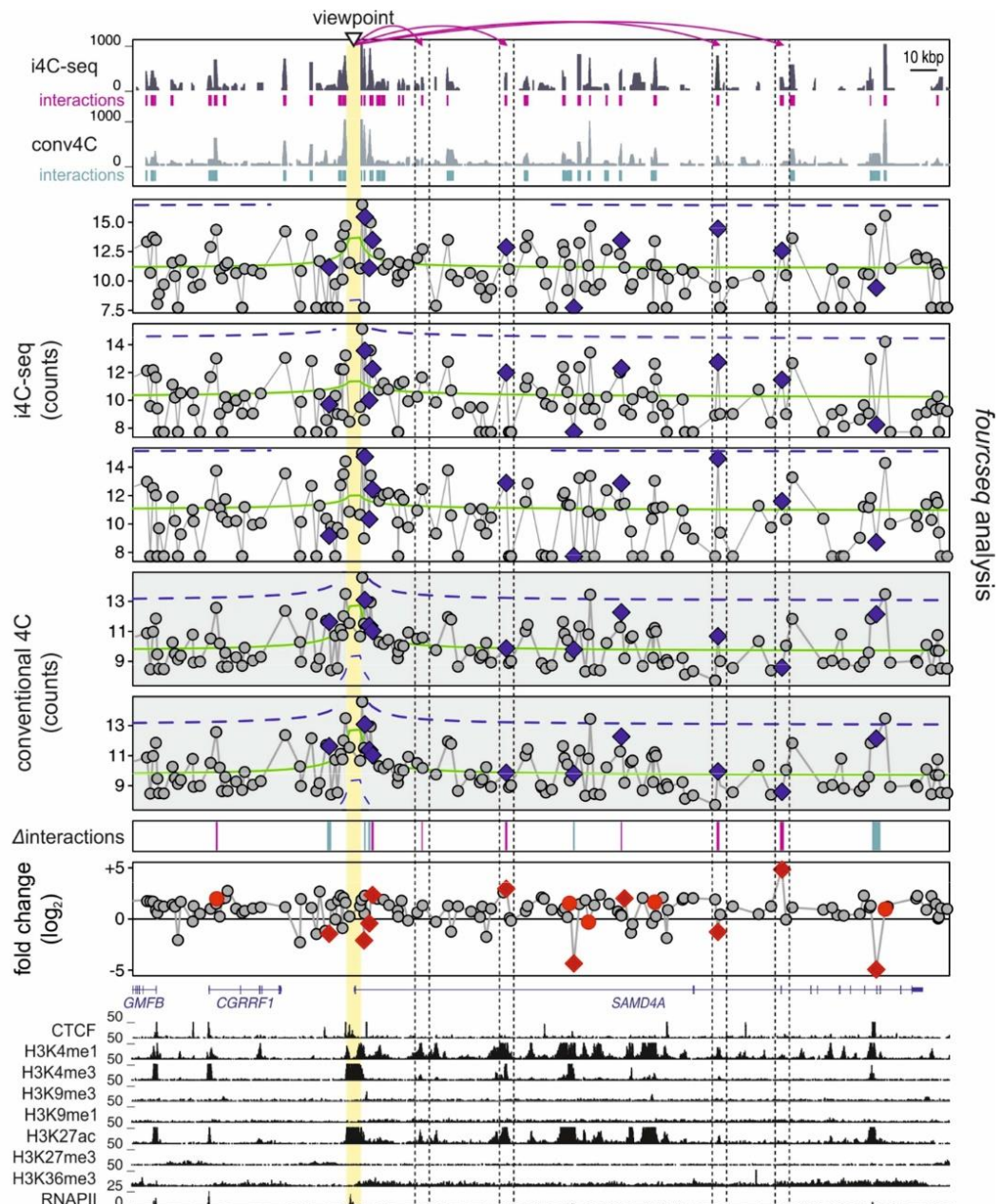
interaction profiles for this region to data from RNA polymerase II-based ChIA-PET experiments (Papantonis *et al.*, 2012) and publicly-available 5-kbp resolution Hi-C data (Rao *et al.*, 2014). In addition, ‘virtual 4C’ profiles were created for each viewpoint from that same high-resolution Hi-C data using the “ChromContact” web tool (Sato *et al.*, 2015). All i4C replicates display matching interaction profiles that align well with their respective TADs, and are in accordance with ChIA-PET, Hi-C and ‘virtual 4C’ profiles (Fig. 15A). Moreover, *cis*-interactions of the three native replicates correlated with conventional 4C data (Spearman’s correlation coefficients >0.75, see Fig. 15B). However, a promoter-promoter contact of *SAMD4A* and the upstream gene *GMFB* which was detected by i4C, was not seen in the ‘virtual 4C’ or ChIA-PET data (Fig. 15A). Based on the above, we can conclude that i4C reproducibly detects shorter- and longer-range chromatin contacts, with the vast majority of interactions residing within the relevant TADs. Yet, particular contacts were uniquely seen in i4C contact maps, and found to span TADs, questioning the strict definition of insulation at TAD boundaries.



**Figure 15 | Reproducibility of i4C interaction profiles.** **A** | Three independent i4C replicates were generated in HUVECs using *Apol* and the TSSs of *CDKN3* or *SAMD4A* as viewpoints. ENCODE ChIP-seq data were aligned below interaction profiles. Publically available Hi-C data were used to outline TADs (Rao *et al.*, 2014). Virtual 4C profiles were generated as comparison (Sato *et al.*, 2015). RNA polymerase II – driven ChIA-PET interactions were aligned below (red lines; Papantonis *et al.*, 2012). Magenta colored arrow indicates i4C contact, which was not detected by virtual 4C or ChIA-

PET. **B** | *Cis*-interactions from three i4C replicates were correlated to one conventional replicate. All Spearman's correlation coefficients were calculated  $> 0.75$ .

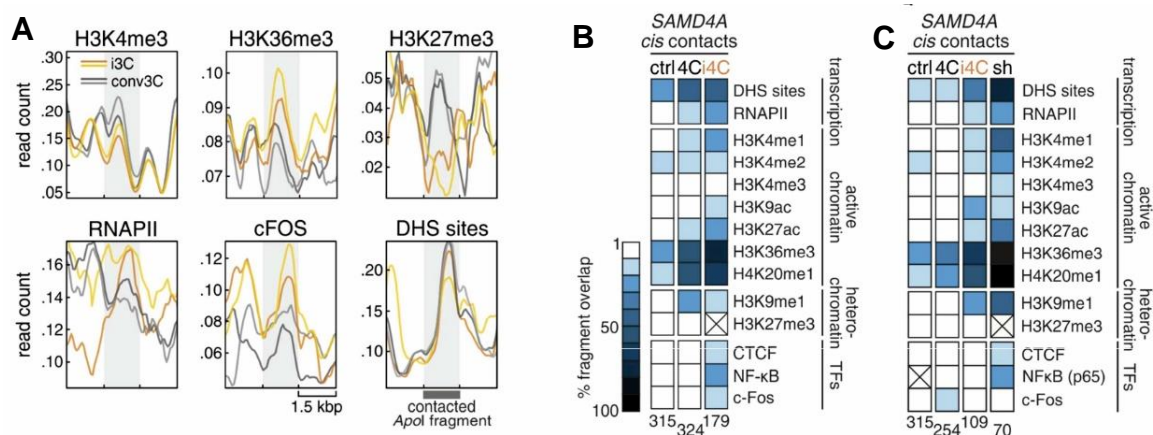
4C-seq contact frequencies decay with genomic distance from the viewpoint and are accompanied by widespread “noise”, which convolutes the identification of significant interactions. Recently, the ‘FourCSeq’ computational analysis tool was introduced to model the error-driven trend of decreasing interaction frequency with genomic distance using replicate information and to calculate strong interactions that stand out from the overall noise, hence identifying statistically-significant differences in contact frequencies between different conditions (Klein *et al.*, 2015). In order to screen the i4C- and conventional 4C data for statistically-significant changes in contact frequencies, we processed reads using FourCSeq: differential analysis of the profiles seen in figure 16, highlights strong interactions (blue squares) from three i4C- and two 4C-seq replicates produced in HUVECs using *Apol* and the *SAMD4A* TSS as a viewpoint (Fig. 16). Detected interaction differences are depicted in the color bar below the 4C profiles, with green bars representing differences unique to conventional 4C and magenta bars illustrating signal changes unique to i4C data. In agreement with the previous ‘fourSig’ analysis (Fig. 14), these results confirm changes in chromatin interaction frequencies that are detected only under physiological conditions (dashed rectangles, Fig. 16).



**Figure 16 | Differential analysis of i4C and conventional 4C data.** Data from i4C (white boxes) and 4C (grey boxes) replicates were processed using 'FourCSeq' (Klein *et al.*, 2015) in order to analyse significant interaction differences (blue squares) above the overall signal noise (grey dots). Some differences ( $\Delta$  interactions) unique for either i4C (magenta) or 4C (blue) were illustrated as bars. HUVEC ENCODE ChIP-seq data were aligned below.

In conventional 3C studies, the vast majority of pairwise *cis*-interactions between chromatin segments occur within the constraints of TADs, displaying high enrichment for regulatory elements (Jin *et al.*, 2013). These 3D interactions between *cis*-regulatory elements, such as enhancers and promoters, play a key role in transcription regulation (Noordermeer & de Laat, 2008; Vise *et al.*, 2009). Several

studies have demonstrated that these contacts are not necessarily predicted by the distance to the closest transcription start site. Indeed, *cis*-regulatory interactions are part of a complex network with promoters interacting with multiple enhancers or single enhancer contacting several promoters, as well as promoter-promoter and enhancer-enhancer associations (Sanyal *et al.*, 2012; Jin *et al.*, 2013). To investigate if characteristics of *cis*-interactions obtained from i4C data deviate from conventional 4C, we correlated *cis*-contacts from *SAMD4A* TSS i4C-/4C-seq experiments to ENCODE ChIP-seq data and classified them into different chromatin motifs of the human genome using the HMM classification (Ernst *et al.*, 2012). Approximately 65 % of all detected contacts in *cis* overlap in i4C and conventional 4C results for the investigated viewpoint, with equivalent read coverage at promoter and DNase I-hypersensitive sites (Fig. 17A). However, interactions exclusive to i4C profiles are linked to genomic regions that are enriched for active histone marks such as H3K4me1/2, H3K36me3, H3K27ac (typical enhancer mark) and H3K9ac, as well as RNA polymerase II and transcription factors (e. g. CTCF, NF- $\kappa$ B, c-Fos; Fig. 17 B&C). On the other hand, contacts that we would not expect to form, i.e. between the *SAMD4A* TSS and heterochromatic fragments, were mostly seen in conventional 4C (Fig. 17C).

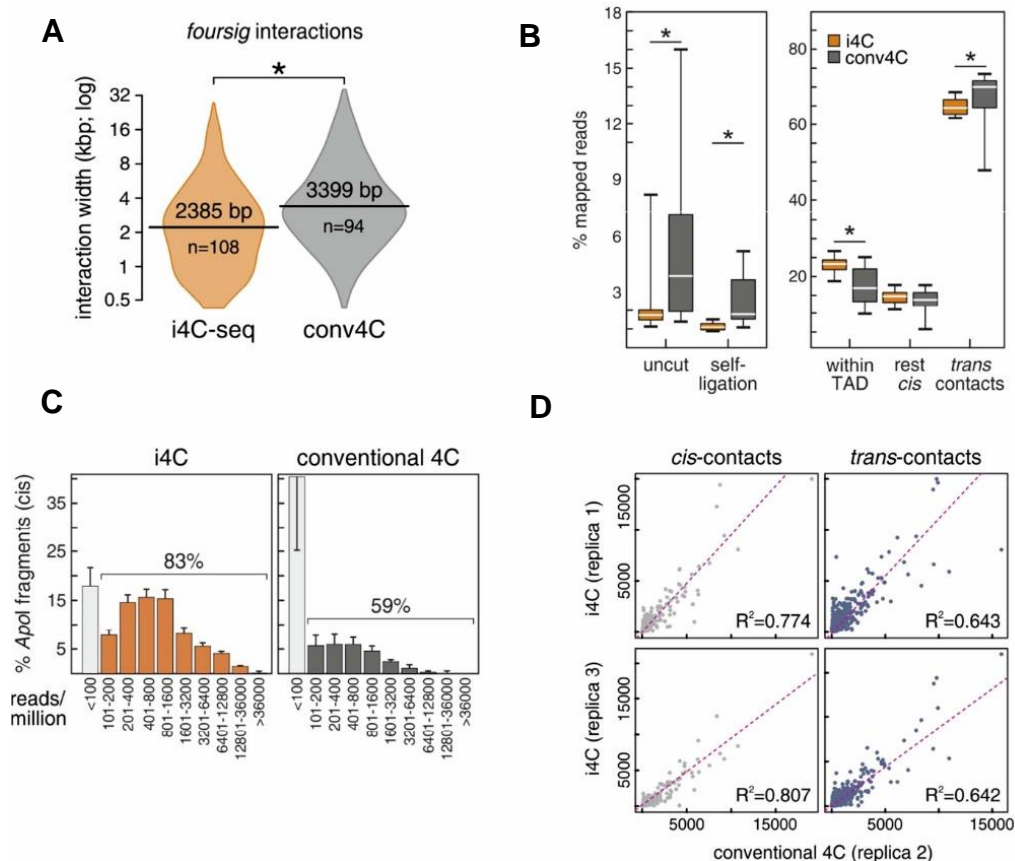


**Figure 17 | Chromatin features of *cis*-contacts obtained from i4C- and 4C-seq profiles. A |** Line plots represent raw read coverage of *cis*-contacts ( $\pm$  1.5 kbp) from two i4C-seq (orange/yellow) and two conventional 4C-seq (dark/light gray) replicates along ENCODE ChIP-/DHS-seq data. **B |** Heat map shows i4C- or 4C *cis*-contacts (Apol fragments with >100 rpm) that correlate to HUVEC ChromHMM motifs. Randomly-shuffled (ctrl) fragments serve as control. **C |** Heat map shows unique or shared (sh) i4C- or 4C *cis*-contacts (Apol fragments with >100 rpm) that correlate to HUVEC ChromHMM motifs. Randomly-shuffled (ctrl) fragments serve as control.

Although numerous parallels were observed among the native and conventional 4C interaction profiles, our results indicate that i4C determines contact frequencies more focal (Fig. 17B and Fig. 17C) and with a preference for cognate, active genomic regions.

### 3.2.1.2 Contact signal versus background noise: a closer look on i4C profiles

To test if these differences in *cis*-contact characteristics arise from variations in interaction width, we identified main interactions of *SAMD4A* TSS i4C- and conventional 4C experiments using 'fourSig' (as described in Fig. 14) and compared their distribution concerning the mean width in bp. As shown in figure 18A, signal from *cis*-contacts from i4C profiles spans shorter regions than conventionally-captured 4C signal. In parallel we also examined differences in the proportion of uncut and self-ligated fragments remaining in the two approaches. We compared the mapped reads corresponding to uncut and self-ligated fragments from *SAMD4A* TSS i4C and conventional 4C experiments (Fig. 18B) to show significantly lower numbers of uncut and self-ligation reads in i4C, and this tendency was believed to increase with milder cell fixation conditions (Werken *et al.*, 2012).



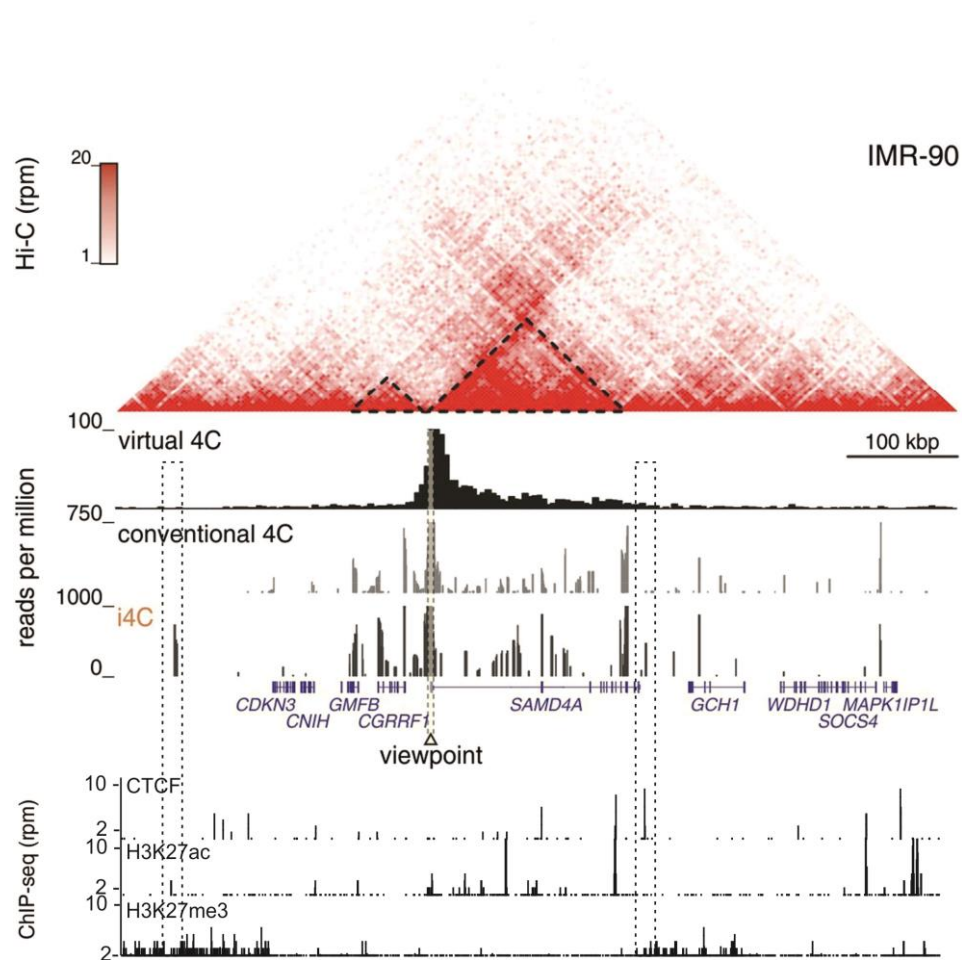
**Figure 18 | Comparison of background noise in i4C and conventional 4C. A |** Distribution of *cis*-contact width from three i4C-seq replicates (orange) and two conventional 4C-seq replicates (grey). Contacts made with the TSS of *SAMD4A* using *ApoI* were identified by ‘fourSig’. Contact number (*n*) and mean width (bp) are indicated. **B |** Percentage of mapped reads obtained from native (orange) or conventional (grey) i4C/4C experiments that correspond to uncut and self-ligated fragments (left) or to fragments mapping within, outside or in *trans* to the *SAMD4A* TAD. \* indicates significantly different mean;  $P < 0.05$ , two-tailed unpaired Student’s t-test. **C |** Percentage of all *cis*-contacts (*ApoI* fragments) from three i4C-seq replicates (orange) and two conventional 4C-seq replicates (grey) binned according to read coverage (rpm). Mapped reads below a threshold of 100 rpm are considered as background noise. **D |** Correlation of *cis* (grey) - and *trans* (purple) contact read distribution of i4C replicates to conventional 4C replicates. Spearman’s correlation coefficients ( $R^2$ ) are indicated.

Due to a substantial amount of unattached DNA being lost after chromatin digestion and prior to *in situ* ligation in the i4C protocol (Fig. 9A), contacts detected under native conditions appear as focal signal with enhanced signal-to-noise ratios. When nuclei are cross-linked and conventionally processed, resulting 4C contacts extensively display random background noise (see Fig. 18C). Thus, we evaluated the background noise level produced in i4C- and conventional 4C for the *SAMD4A* TSS viewpoint by defining a background threshold at 100 rpm, as fragments with coverage of mapped reads below this value displayed the highest deviation amongst all our replicates. Then, the percentage of mapped reads from all detected *cis*-contacts carrying >100 rpm is considered as potentially “meaningful” signal. We found that 83% of all i4C reads are above the threshold compared to <60% of conventional 4C reads (Fig. 18C), confirming elevated background noise in conventional 4C interaction profiles. Next, we investigated the correspondence of *cis*- and *trans*-contacts between i4C and/or conventional 4C replicates (Fig. 18D). Calculated Spearman’s correlation coefficients ( $R^2$ ) present high concordance for *cis*-interactions, while contacts in *trans* correlate to a lower extent. These data are in agreement with the literature, since the vast majority of detected 4C interactions occur in *cis* (Visel *et al.*, 2009; de Laat & Duboule, 2013). Yet, i4C displays more mapped reads for *cis*-contacts within TADs compared to conventional 4C. In contrast, a higher percentage of mapped *trans*-contacts is observed under conventional, cross-linked, conditions (Fig. 18D). Taken together, our results highlight major differences in i4C- and conventional 4C data outcome, with the latter being biased to insufficient chromatin digestion and ligation as well as high background noise rates.



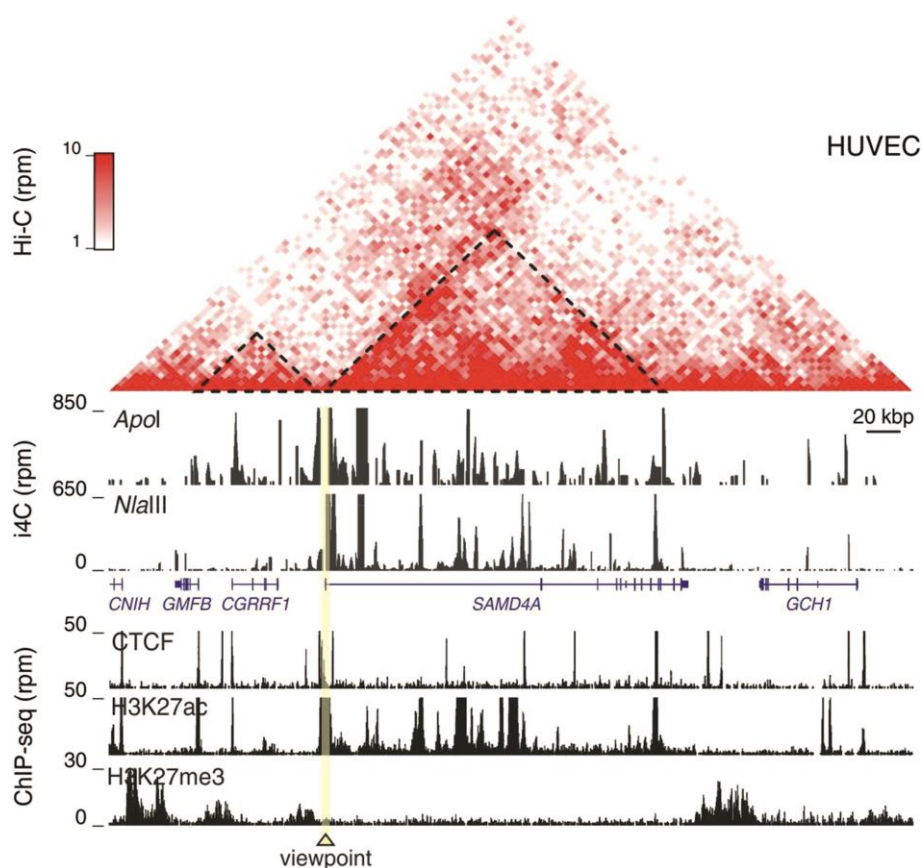
### 3.2.1.3 *SAMD4A* locus i4C using various cell types and restriction enzymes

In order to reproduce our native interaction profiles for *SAMD4A* in a different cell type, i4C-seq was performed next to conventional 4C-seq in the human primary lung fibroblast cells, IMR90, using *ApoI* and our model viewpoint (*SAMD4A* TSS). Contact profiles were compared to publicly-available Hi-C data (Rao *et al.*, 2014) and virtually generated 4C profiles for the same region. Examination of the resulting interaction profiles demonstrated many common contacts between HUVECs and IMR90s that reside within their respective TADs (Fig. 19). Furthermore, i4C and conventional 4C data are consistent with one another, with only few peaks deviating under native conditions



**Figure 19 | i4C – and 4C implemented in a different cell line.** i4C-seq and conventional 4C were performed in IMR90s using *ApoI* and the *SAMD4A* TSS viewpoint. IMR90 histone ChIP-seq data is illustrated below the profiles (Jin *et al.*, 2013). Virtual 4C and publicly available Hi-C data at 5-kb resolution are shown (Rao *et al.*, 2014). Deviating i4C signals are demarcated by dotted rectangles.

Next, we compared i4C interaction profiles generated using the 6-bp cutter *ApoI* or the 4-bp cutter *NlaIII*. The choice of restriction enzyme is critical for the resolution of chromatin interaction profiles, and may even result in the gain/loss of particular interactions depending on the enzyme recognition site relative to the interacting genomic element; 4-bp cutter enzymes allow for the study of genome organization at much higher resolution, but this denser segmentation can affect noise levels (Kalhor *et al.*, 2011). For this, i4C was conducted in HUVECs using either *ApoI* or *NlaIII* and the *SAMD4A* TSS as a viewpoint. As seen in figure 20, both profiles display a large fraction of common contacts. Nonetheless, *NlaIII* digestion may be the reason for the loss of contacts observed between *SAMD4A* and *CGRRF1/GMFB*. In conclusion, we were able to recapitulate most spatial chromatin contacts of the *SAMD4A* TSS using two different cell types and restriction enzymes.

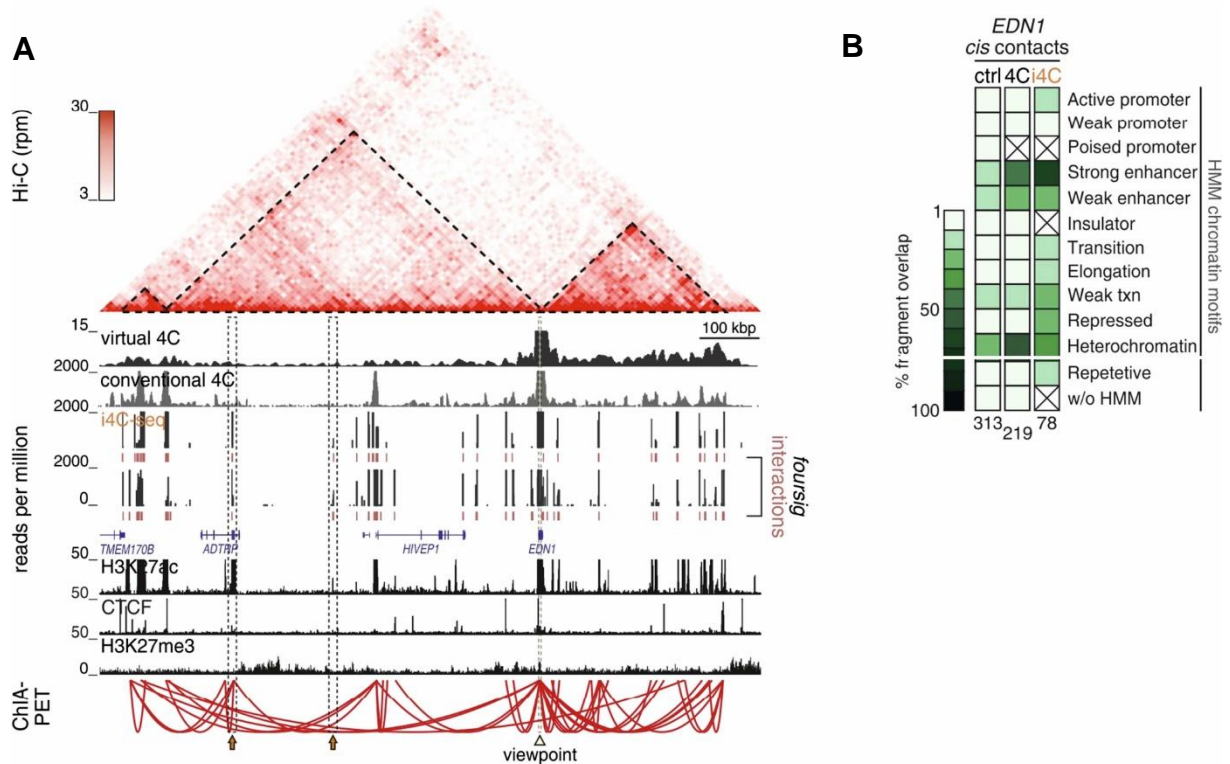


**Figure 20 | Comparison of i4C profiles using different restriction enzymes.** i4C-seq was performed in HUVECs using either a six-bp cutter *ApoI* or a 4-bp cutter *NlaIII*, and the *SAMD4A* TSS viewpoint. ENCODE ChIP-seq data is illustrated below the profiles. Publicly available Hi-C data at 5-kb resolution are shown (Rao *et al.*, 2014).

### **3.2.2 Implementation of i4C in different loci**

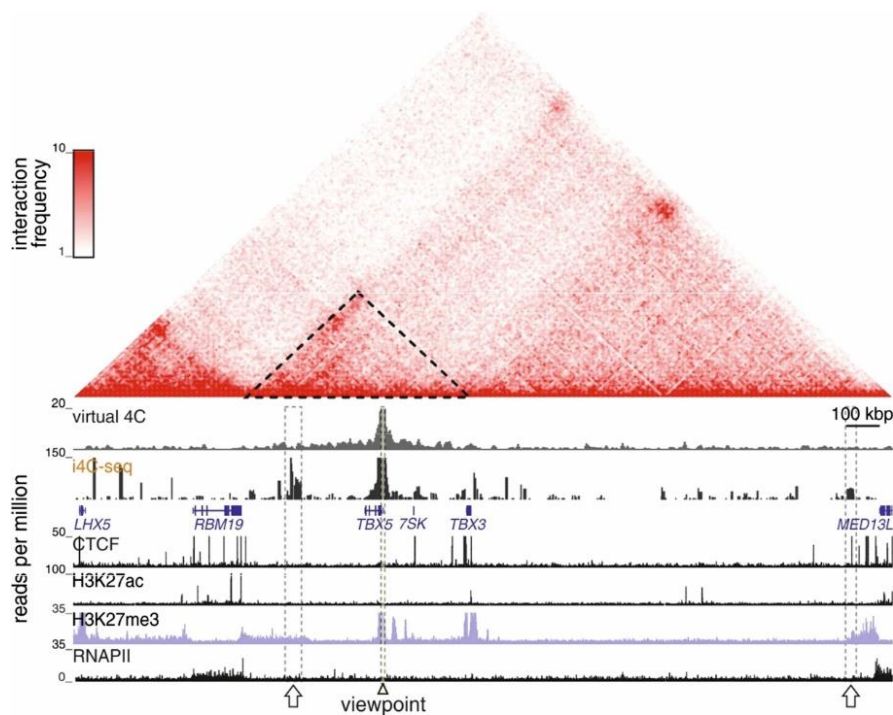
#### **3.2.2.1 i4C implementation in gene-poor and hetero-chromatinized loci**

The extended *SAMD4A* locus is densely populated by genes and *cis*-regulatory elements, making it a suitable candidate for studying chromatin interaction dynamics. To investigate if native interaction profiles can be equally informative in relatively gene-poor or hetero-chromatinized regions, i4C was applied on genes that reside within such genomic loci. First, we performed i4C alongside with conventional 4C in HUVECs using the *EDN1* TSS as a viewpoint, and compared our findings to interaction profiles from previously-obtained ChIA-PET data (Papantonis *et al.*, 2012) and virtually-generated 4C profiles (Sato *et al.*, 2015). TADs were outlined using publically available Hi-C data (Rao *et al.*, 2014). The constitutively expressed *EDN1* gene is located close to the TAD boundary of a relatively gene-poor region. Examination of the chromatin contact maps for its TSS revealed sparse but nicely matching i4C and conventional 4C profiles, while under native conditions signals are again significantly more focal (Fig. 21A). Most of the detected contacts are enriched for “active” chromatin marks relevant promoter and enhancer elements, and this is particularly so for the i4C data (Fig. 21B). However, some interactions between the *EDN1* TSS and upstream active *cis*-regulatory segments were only seen with our native method. Furthermore, both i4C and conventional 4C profiles were poorly recapitulated in virtual 4C maps, despite the good correlation with the respective TAD boundaries and ChIA-PET loops (Fig. 21A).



**Figure 21 | i4C-seq implemented in the gene-poor *EDN1* locus. A |** Two independent i4C replicates and one conventional 4C replicate were generated in HUVECs using *ApoI* and the TSS of *EDN1* as viewpoint. Significant interactions were called by ‘fourSig’ (red bars). ENCODE ChIP-seq data were aligned below interaction profiles. Publically available Hi-C data were used to outline TADs (Rao *et al.*, 2014) and to generate virtual 4C profiles. RNA polymerase II – driven ChIA-PET interactions were aligned below (red lines; Papantonis *et al.*, 2012). Two interactions at H3K427ac-decorated sites that are unique to i4C were highlighted with dashed rectangles. **B |** Heat map shows i4C- or 4C *cis*-contacts that correlate to HUVEC ChromHMM motifs. Randomly-shuffled (ctrl) fragments serve as control.

Next, we investigated chromatin looping events that are made by the TSS of the *TBX5* gene, which lies within a heterochromatic region demarcated by the typical Polycomb-group histone mark, H3K27me3. For this, i4C was applied in HUVECs using *NlaIII* and contact maps were aligned to virtual 4C and conventional Hi-C profiles (Rao *et al.*, 2014). Our findings demonstrate interactions between *TBX5* and other H3K27me3 marked regions, including the inactive *TBX3* locus in a neighboring downstream TAD (Fig. 22). In addition, contacts with enhancer as well as CTCF-bound elements were observed. Yet, most major interactions seen in the i4C profile were not present in the conventional 3C-derived data.



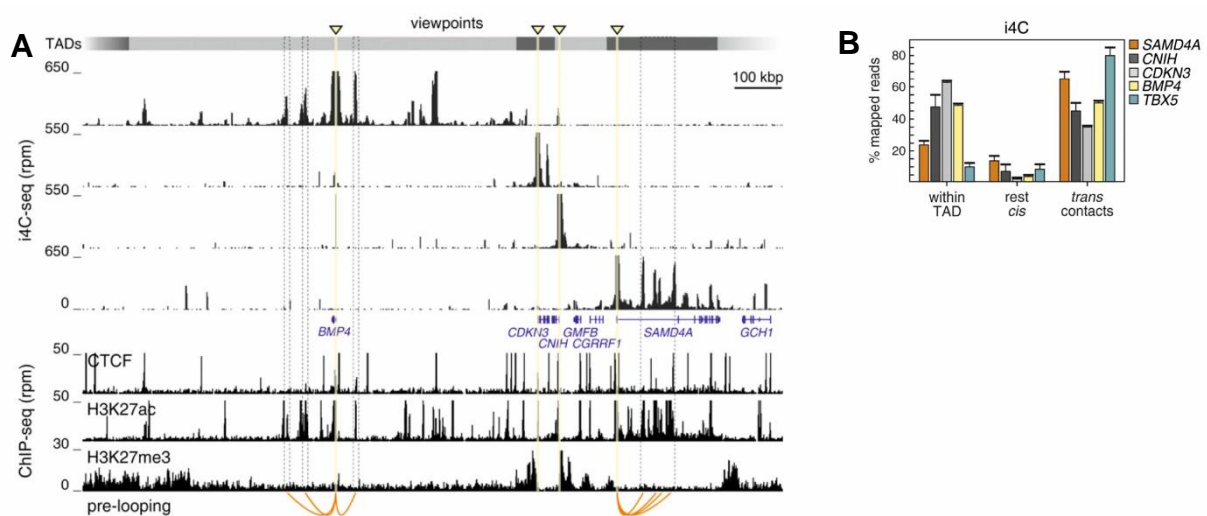
**Figure 22 | i4C-seq implemented in the heterochromatinized *TBX5* locus.** i4C-seq was performed in HUVECs using *NlaIII* and the TSS of *TBX5* as viewpoint. ENCODE ChIP-seq data were aligned below interaction profiles. Publically available Hi-C data were used to outline TADs (Rao *et al.*, 2014) and to generate virtual 4C profiles. Two strong interactions at H3K27me3-decorated sites that are unique to i4C were highlighted with dashed rectangles.

In conclusion, 3D chromatin interactions in gene-poor and heterochromatic loci can be successfully captured with our native i3C approach. For transcriptionally-active viewpoints, we could also identify contacts with other active elements that were unique to i4C and displayed lower background noise. In heterochromatic regions, i4C contacts connected other similarly-marked “inactive” fragments, hinting on possible limitations due to cross-linking in ‘closed’ chromatin when using conventional 3C.. These findings agree with former conventional chromosome conformation studies that reported higher contact frequencies in open over closed chromatin regions (Williamson *et al.*, 2014) and loss of large-scale genome organization patterns such as TADs in highly condensed mitotic chromosomes (Naumova *et al.*, 2013).

### 3.2.2.2 *Cis-regulatory interactions are mostly confined within TADs*

Higher-order genome organization studies identified TADs as molecular ties to shape and control the three-dimensional chromatin structure (Lieberman-Aiden *et al.*, 2009; Dixon *et al.*, 2012). However, regulatory influences on transcription at a small-scale are mainly driven by interactions between *cis*-regulatory elements such as promoters

and enhancers (Noordermeer & de Laat, 2008; Visel *et al.*, 2009), and complex networks of their spatial interactions are established and maintained by the binding of transcription and architectural factors (Jin *et al.*, 2013; Phillips-Cremins *et al.*, 2013). Recent studies involving depletion of such structural factors, highlighted their role in anchoring the vast majority of interactions between *cis*-regulatory elements, but almost invariably within the constraints of TAD boundaries (Sofueva *et al.*, 2013). All of our aforementioned i4C and conventional 4C data confirm that most detected *cis*-contacts align with TAD-imposed topological restrictions (Fig. 14). To further assess if this observation holds true universally, we performed i4C in HUVECs using *NlaIII* and the TSSs of the *BMP4*, *CDKN3*, *CNIH*, and *SAMD4A* genes as viewpoints, and examined interaction profiles relevant to TAD boundaries (obtained from publicly-available data; Dixon *et al.*, 2012). All viewpoints are located in the extended *SAMD4A* locus, and data analysis revealed that the majority of *cis*-contacts lie within the respective TAD for each viewpoint (Fig. 23 A&B). Only few inter-TAD contacts (such as promoter-promoter contacts between *BMP4* and *CNIH* or *CDKN3*) could be detected. Nonetheless, at least 40% of all i4C reads mapped in *trans*. Interestingly, the inactive and H3K27me<sub>3</sub>-demarcated gene *TBX5* displayed the largest proportion of putative *trans*-contacts, indicative of a higher-order genome organization based on Polycomb bodies. Moreover, promoters of the TNF $\alpha$ -responsive genes *BMP4* and *SAMD4A* also show pre-looping to enhancers prior to TNF induction (Fig. 23A), in agreement with published findings (Jin *et al.*, 2013).



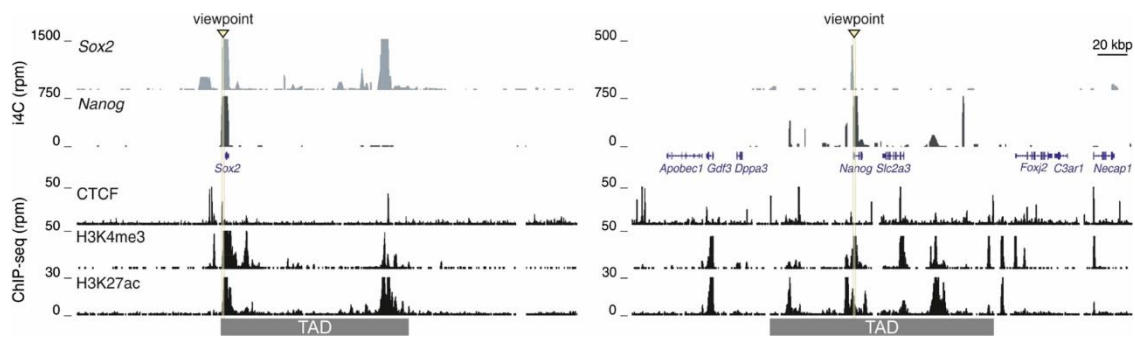
**Figure 23 | Native interactions are restricted to TAD boundaries and describe prelooping. A |** Independent i4C experiments were performed in HUVECs using *NlaIII* and the TSSs of *CDKN3*, *BMP4*, *CNIH* or *SAMD4A* as viewpoints. ENCODE ChIP-seq data were aligned below interaction

profiles. Publically available Hi-C data were used to outline TADs (Dixon *et al.*, 2012). Orange lines indicate pre-established contacts between enhancers and promoters of *BMP4* and *SAMD4A*. **B** | Percentage of reads obtained from i4C experiments for the TSSs of *CDKN3*, *BMP4*, *CNIH*, *TBX5* or *SAMD4A* mapping within, outside or in *trans* to the respective TADs.

Although approximately 50% of chromatin is lost upon restriction enzyme treatment during i4C, main conformational entities such as TADs still impose their topological effects in the natively-processed nuclei, and contribute to the spatial organization of *cis*-interactions. Yet, the detection of established *trans*-contacts (Fig. 23B) with i4C requires the development of new, more robust analysis pipelines.

### **3.2.2.3 Capturing inter-chromosomal contacts with i4C**

Typical 4C results reveal the bulk of detected signal close to the respective viewpoint, with higher numbers of intra-chromosomal captures compared to inter-chromosomal ones (Simonis *et al.*, 2006). Previous 4C studies in mouse pluripotent stem-cells showed that clustering of pluripotency factor-bound genomic elements bring together distant loci from different chromosomes (de Wit *et al.*, 2013). To investigate whether i4C technology efficiently captures *trans*-chromatin interactions, we reproduced previously published contacts between the promoter regions of the genes encoding the pluripotency factors *Sox2* and *Nanog* (de Wit *et al.*, 2013). Native i4C was applied to mouse embryonic stem cells (mESCs) using *Apo1* and the CTCF peak proximal to the *Sox2* promoter, or the *Nanog* TSS as viewpoints. The CTCF-bound site next to the *Sox2* promoter connects mostly other CTCF-associated elements in *cis*, while the *Nanog* TSS is engaged in contacts with surrounding enhancer and CTCF elements (Fig. 24). Like before, both i4C interaction profiles correlate well with their respective TAD boundaries. Notably, the investigated viewpoints, located on chromosome 3 and 6, form strong reciprocal *trans*-interactions (Fig. 24). Consistent with our earlier documented results of elevated levels of mapped i4C and 4C reads in *trans* (Fig. 23B), we now also detect inter-chromosomal contacts without a need for cross-linking.

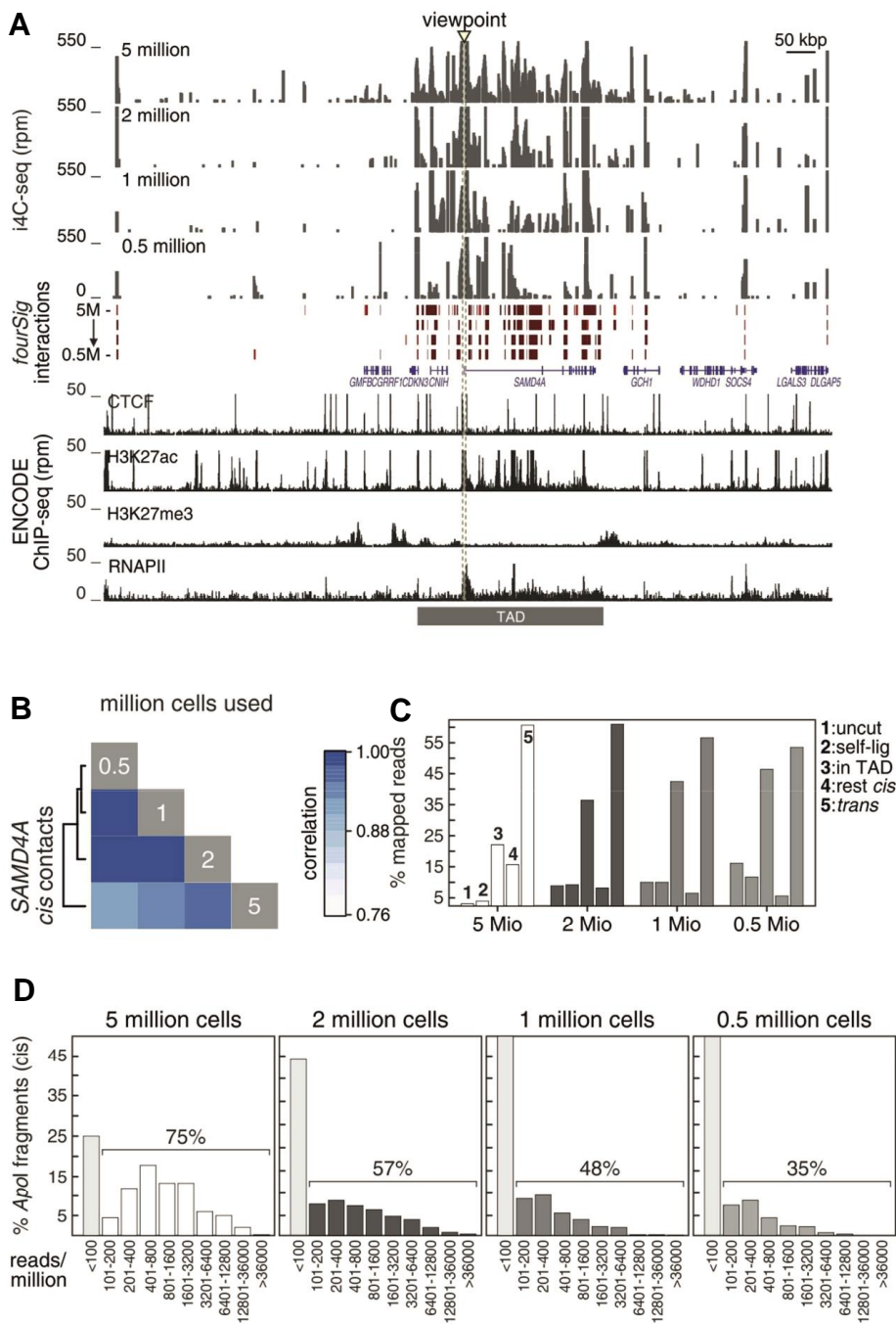


**Figure 24 | *trans*-interactions captured by i4C in the *Sox2* and *Nanog* loci in mESCs.** i4C experiments were performed in mESCs using *Apol* and the TSS of *Nanog* as well as the CTCF peak proximal to the promoter of *Sox2* as viewpoints. ENCODE ChIP-seq data were aligned below interaction profiles. Publically available Hi-C data were used to outline TADs (de Wit *et al.*, 2013).

### 3.2.3 Enrichment of i4C signal allows reduction of input material

It is advisable to typically perform 3C on a pool of approximately ten million cells, which yields averaged contact frequencies (Stadhouders *et al.*, 2013). Most probably due to the lack of cell cross-linking and the removal of large amounts of DNA, the i4C protocol yields notably enhanced signal-to-noise ratios without a need for increasing the initially used amount of cell material (as is required for “native ChIP” for example; Kasinathan *et al.*, 2014; Fig. 18). Indeed, our findings encouraged us to apply our native approach to low counts of cell nuclei. i4C was performed on 5-, 2-, 1- and 0.5 million HUVEC nuclei using *Apol* and the well-studied *SAMD4A* TSS viewpoint. The obtained contact i3C profiles contained significant and reproducible interactions (as called by FourSig) with increasingly lower cell counts (Fig. 25A). Spearman’s correlation coefficients of all *SAMD4A* *cis*-contacts between distinct experiments were calculated and confirmed good accordance between i4C data from 0.5 to 2 million nuclei, while differences in contact frequencies moderately increased when profiles from 0.5 and 5 million cells were compared (Fig. 25B). Examination of the read coverage on every *Apol* fragment revealed increased uncut and self-ligation events upon cell number reduction (Fig. 25C), hence the elevated background signal in lower cell count experiments (Fig. 25D).





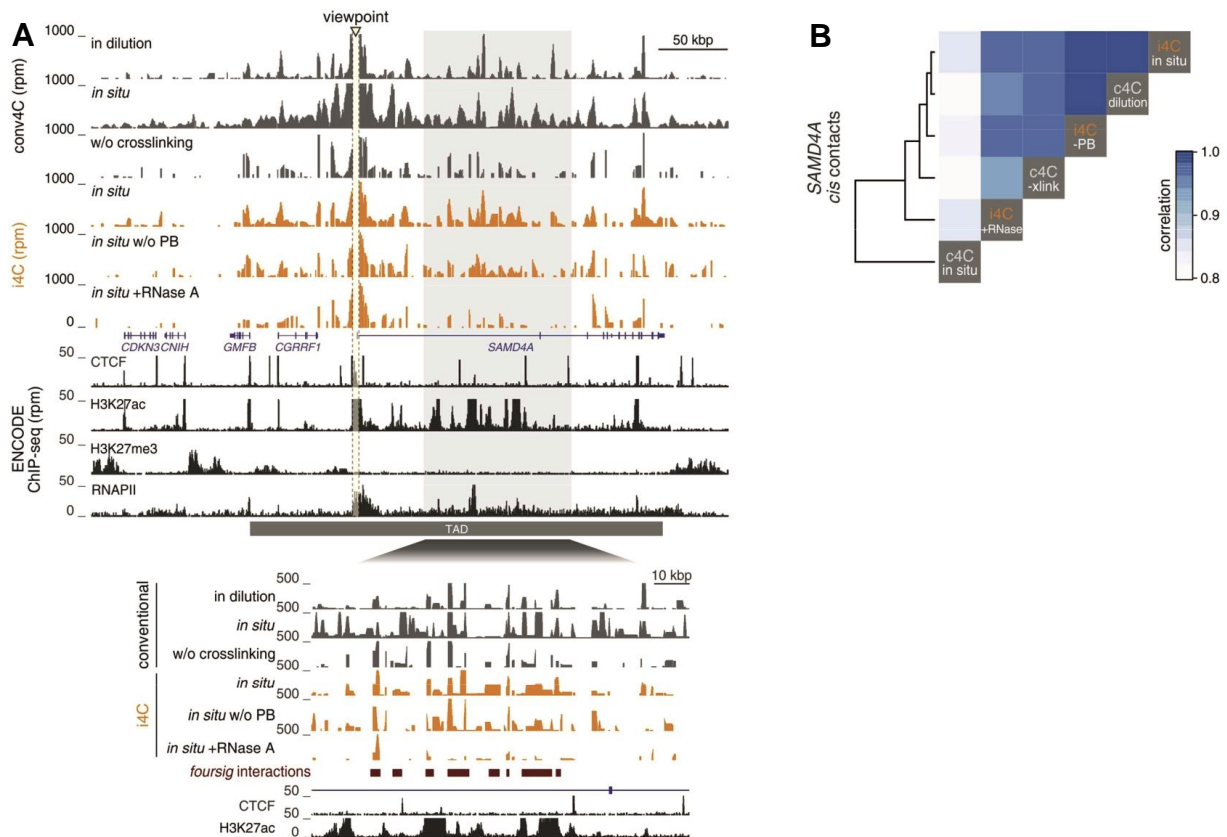
**Figure 25 | i4C-seq implemented using decreasing numbers of isolated nuclei. A |** i4C-seq was performed in 5, 2, 1 and 0.5 million HUVECs nuclei using *ApoI* and the *SAMD4A* TSS viewpoint. ENCODE ChIP-seq data is illustrated below the profiles. Publically available Hi-C data outline respective TAD (Rao *et al.*,2014). Significant interactions were called with ‘fourSig’ (red bars). **B |** Cis-interactions from i4C experiments were correlated to each other. All Spearman’s correlation coefficients were calculated  $> 0.75$ . **C |** Percentage of reads obtained from i4C experiments mapping to uncut or self-ligated fragments as well as within, outside or in *trans* to the respective TADs. Low cell counts display increased uncut/self-ligated fragments. **D |** Percentage of all cis-contacts (*ApoI* fragments) from different i4C-seq experiments binned according to read coverage (rpm). Mapped

reads below a threshold of 100 rpm are considered as background noise. Low cell counts display higher background noise.

The aforementioned data demonstrate that adequately reproducible *cis*-interaction profiles can be achieved with as few as  $10^6$  cells, albeit at the expense of diminished signal-to-noise ratios, and, thus, i4C may serve as the tool of choice for studying genome organization when starting material is limited (e.g., patient tissues, model organism *C. elegans*).

### **3.2.4 It is all about controls: different conditions produce different outcomes**

Recently, an '*in situ* Hi-C' protocol was introduced, which is based on the original Hi-C protocol (Lieberman-Aiden *et al.*, 2009), but relies on restriction digest and ligation steps performed in intact nuclei so as to reduce random ligation frequencies (Rao *et al.*, 2014). In addition, '*in situ* Hi-C' was applied to uncross-linked lymphoblastoid cells after embedding them in agar. This returned sparse interaction heat maps, yet converging to those obtained from cross-linked cells (Rao *et al.*, 2014). To compare chromatin interaction profiles from our native i4C protocol with the above described (as well as with other) experimental conditions, we generated side-by-side i4C contact profiles in HUVECs using *Apo1* and our model *SAMD4A* locus. Figure 26 shows interaction profiles upon variable treatments that include typical 4C, 4C following the '*in situ* Hi-C' procedure with or without cell cross-linking, i4C using our physiological buffer or commercially-available buffers, and i4C on nuclei pretreated with 30  $\mu$ g RNase A. As expected, conventional 4C and i4C display strong similarities, replacing 'PB' with commercial buffers only moderately affects the *SAMD4A* contact profile, while the '*in situ*' version of 4C exhibited strong background signal throughout the locus (Fig. 26B). However, 4C according to the '*in situ* Hi-C' protocol in the absence of cross-linking yields markedly de-enriched profiles with diminished contact frequencies between the *SAMD4A* TSS and the dense enhancer elements in its first intron – and these looping events essentially vanish once nuclei are pre-treated with RNase A, hinting to a stabilizing function of RNA in spatial chromatin organization (Fig. 26A and B).



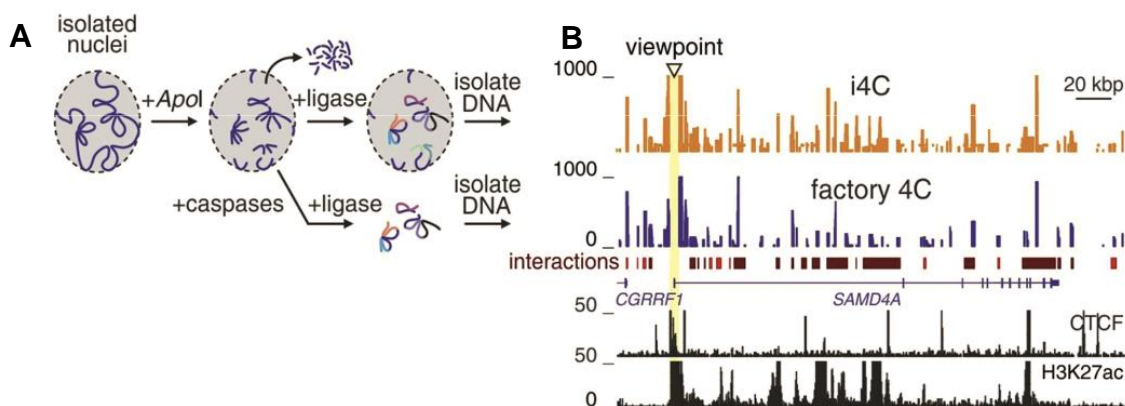
**Figure 26 | i4C-seq performed under different conditions. A |** i4C-seq (orange) and conventional 4C (grey) were performed in HUVECs using *Apo1* and the *SAMD4A* TSS viewpoint under variable conditions. From top to bottom: typical 4C-seq, 4C-seq following the *in situ* Hi-C protocol (Rao *et al.*, 2014) with or without cross-linking, the typical i4C-seq protocol with PB or commercially available buffers and i4C where nuclei were pretreated with 30  $\mu$ g RNase A. ENCODE ChIP-seq data is illustrated below the profiles. Publically available Hi-C data outline respective TAD (Rao *et al.*, 2014). Significant interactions were called with ‘fourSig’ (red bars). **B |** *Cis*-interactions from different i4C/4C experiments were correlated to each other. All Spearman’s correlation coefficients were calculated > 0.75.

This line of investigations revealed i4C and conventional 4C interaction profiles that correlate with another, albeit no agreement was observed when these contact maps were compared to experimental settings comprising steps from the ‘*in situ* Hi-C’ protocol with or without cross-linking. Indeed, in our hands the small-volume ligation step described by Rao *et al.* resulted in notably higher rates of background ligation events. Moreover, most chromatin interactions did not persist upon harsh treatments involving SDS and long restriction digestion incubation times when cell fixation was avoided. Striking loss of contacts and aggregation of new chromatin clusters that did not correlate with functional histone marks were also observed when RNase A was

used to degrade RNA prior to i4C implementation, highlighting a stabilizing role of RNA in chromatin morphology (Nickerson *et al.*, 1989).

### 3.2.5 Transcription factories: factory-i4C captures chromatin interactions

Next, we investigated interactions between loci that are associated with transcription factories by combing a transcription factory isolation protocol with i4C (Melnik *et al.*, 2012; Brant *et al.*, 2016). In brief, nuclei were isolated from HUVECs under mild conditions in 'PB' and digested with *Apo1* as described before (Fig. 27A). Cut and non-associated DNA fragments were removed and the resulting nuclear fractions were treated with caspase group-III proteins to release transcription factories into the supernatant as previously described. Cut chromatin still-associated with isolated transcription factory fragments were ligated under dilute conditions to enrich for ligation events between spatially proximal genomic segments, and further processed to obtain i4C contact profiles for the *SAMD4A* TSS. Previous studies described native chromatin interaction captures on DNA templates that were linked to isolated transcription factory complexes by performing ligation on such complexes embedded in an agarose gel (Melnik *et al.*, 2012).



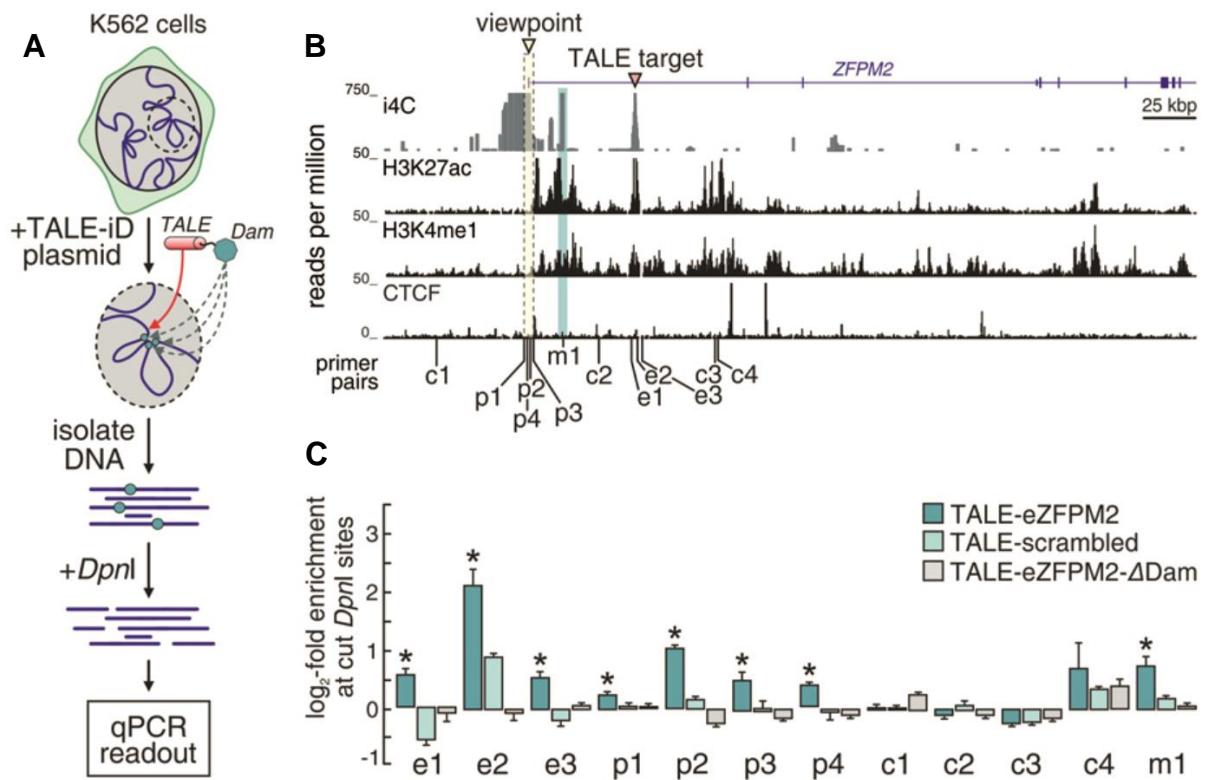
**Figure 27 | Factory-i4C implemented in HUVECs.** **A** | Transcription factories were isolated in HUVECs using *Apo1* and Caspases group III according Melnik *et al.*, 2011 and sequences were ligated under diluted conditions. **B** | Factory i4C profiles (purple) were compared to i4C profiles (orange) for the *SAMD4A* TSS viewpoint. ENCODE ChIP-seq data are aligned below.

The earlier-documented contacts between the *SAMD4A* TSS and enhancer elements and CTCF sites were preserved in this factory i4C dataset, thus confirming that transcription of our model gene is associated to these specific nucleoplasmic bodies, known as transcription factories. However, the *in vivo* effects of caspase treatment remain unknown and factory-i4C needs to be evaluated for genomic loci located in

both active and inactive regions to define its bias towards contacts captured only when using active genes as viewpoints.

### **3.3 TALE-iD: an orthogonal approach to verify native chromatin looping events**

Contacts between loci that were revealed with 3C-derived techniques are typically validated using FISH approaches (Williamson *et al.* 2014). However, these methods also include a chemical cell cross-linking step, rendering them unsuitable to verify i4C interactions. This encouraged us to develop an alternative, orthogonal technique that can detect chromatin interactions under native conditions called 'TALE-iD' (Brant *et al.*, 2016). The basic idea relies on the fusion of a custom transcription activator like-effector (TALE) DNA-binding domain that specifically binds to a genomic segment of interest (e. g. promoter or enhancer; Mendenhall *et al.*, 2013), with an active bacterial adenosine methylase (Dam; Vogel *et al.* 2006). Target specificity of TALE proteins is based on a highly conserved DNA binding domain, consisting of 13 to 28 amino-acid tandem repeats. Each repeat is identical except for repeat-variable di-residues at position (RVDs) 12 and 13, which specifically recognize one base pair of the target DNA sequence with high precision (Boch *et al.*, 2009; Bogdanove and Voytas, 2011). The methylation-based technique DamID identifies genomic regions interacting with proteins of interest (e.g. transcription factors, lamin B1 proteins, and histones) in a non-crosslinked manner (Vogel *et al.*, 2006, Kind *et al.* 2013). A combination of highly-specific DNA sequence targeting by TALEs with DamID resulted in our novel method TALE-iD. Following its introduction in living cells, TALE-Dam constructs bind the targeted genomic element and the associated Dam methylase adds methyl-groups to adenosine residues in spatially-proximal consensus GATC sequences. Digestion with the restriction enzyme *DpnI*, which only recognizes methylated GATC sites, allows the examination of digestion efficiencies as an indirect measure of chromatin interaction frequency (Fig. 28A).



**Figure 28 | TALE-iD verifies native looping interactions at the human *ZFP M2* locus.** **A** | Strategic TALE-iD overview. K562 cells were nucleofected with constructs that are fused to a bacterial DNA methylase (Dam) and target a strong enhancer in the first intron of *ZFP M2* or a scrambled non-binding control region. DNA was isolated post 48 hours and digested with *DpnI*. qPCR using primer pairs flanking *DpnI* sites was performed to evaluate contact frequency. **B** | i4C was performed in K562 cells using *ApoI* and the TSS of *ZFP M2*. ENCODE ChIP-seq data are shown below. **C** | qPCR readout at different *DpnI* sites. *DpnI* sites at the *ZFP M2* promoter (p1–p4) and enhancer (e1–e3; positions in panel B) were targeted in qPCRs after restriction digest. Bar plots show log<sub>2</sub>-fold enrichment of cut sites (1/DDCt) over background *DpnI* cutting levels in un-transfected K562 cells. Regions c1–c4 serve as controls, region m1 (an enhancer shown to interact with the TSS by i4C) is also methylated as part of a multi-loop structure. The same *DpnI* sites were also tested in transfections involving a construct encoding either a non-targeting (“scrambled”) TALE domain or the targeting domain fused to an inactive Dam protein (“DDam”). \*P < 0.05; two-tailed unpaired Student’s t-test (n = 3). The bars and error bars denote mean ± SEM.

First, we performed i4C in human erythroleukemia cells (K562) using *ApoI* and the TSS of the TNF $\alpha$ -responsive and 486 kbp-long *ZFP M2* gene (Wada *et al.*, 2009). Prominent interactions between the promoter region and enhancer elements located in the first intron of the gene were observed (Fig. 28B). To verify these i4C interactions, we transfected K562 cells with a TALE-iD construct targeting one enhancer site marked by a signal peak in the *ZFP M2* i4C profile. As a control, we examined the same locus in cells that were transfected with TALE-iD constructs that

either carry a not-targeting DNA-binding domain ('TALE-scrambled') or that bind the targeted enhancer but only carry an inactive Dam protein ('TALE-eZFPM2- $\Delta$ Dam'). qPCR readouts at different *DpnI* sites along the promoter (P1-P4), the targeted enhancer (E1-E3) and negative control regions (C1-C4) confirmed the interaction between the promoter and its contacting enhancer (Fig. 28C). An additional interaction between the promoter and an enhancer located further upstream was also revealed using TALE-iD, which hints to a multi-loop structure in the locus (Fig. 28 B&C). On the other hand, 'scrambled' and ' $\Delta$ Dam' experiments returned diminished or no contact signals. Collectively, these data demonstrate that i4C-captured contacts can be orthogonally validated via this novel and native TALE-iD approach. Moreover, TALE-iD may serve as an effective tool to analyze three-dimensional chromatin interactions under native conditions and with high precision. Nevertheless, the amount of transfected fusion protein needs to be extensively titrated for each individual assay and data outcome requires several normalization controls.

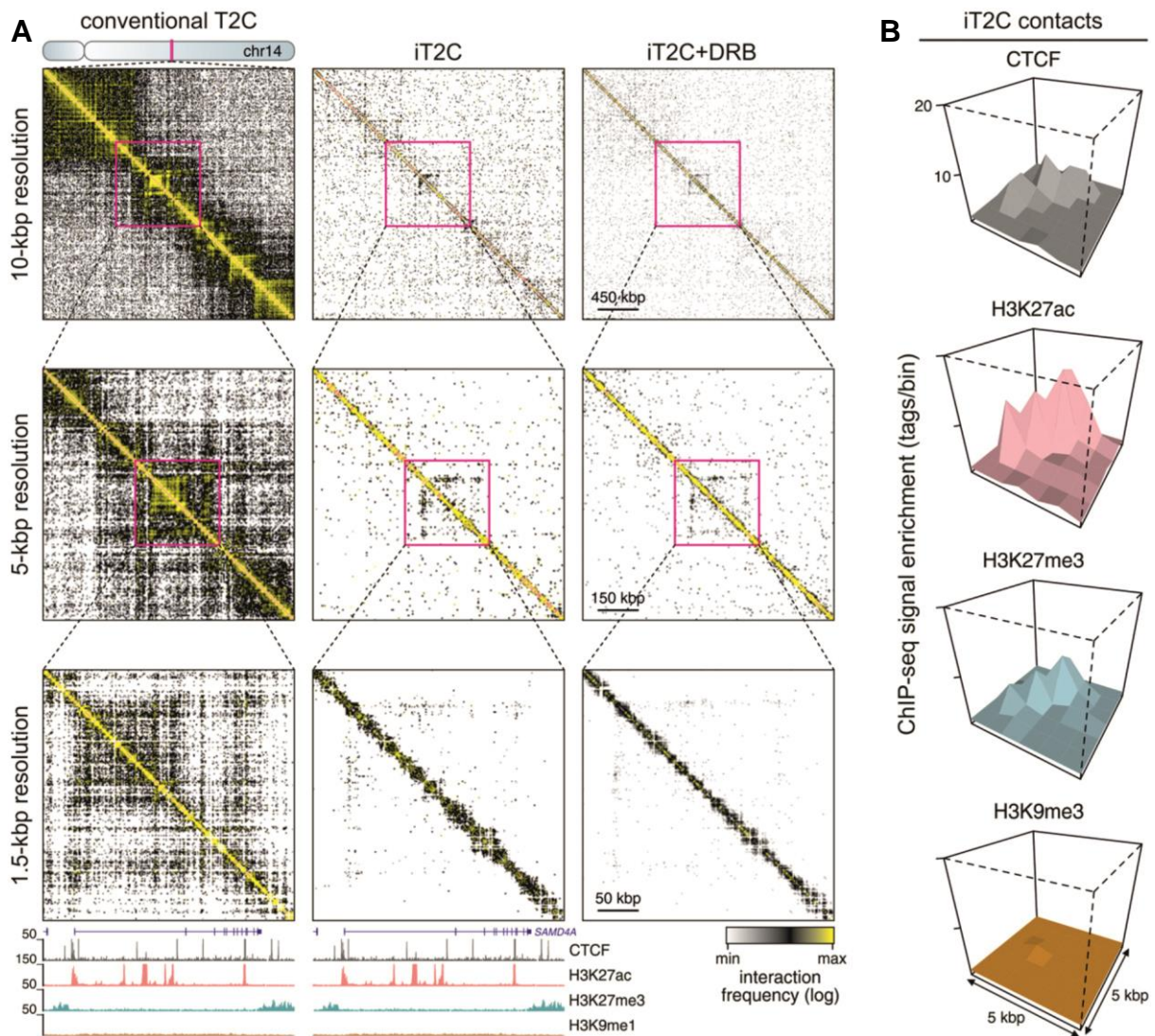
### **3.4 iT2C: an intrinsic Targeted Chromatin Capture method**

Analysis of chromatin interaction frequencies with 3C or 4C technologies has the advantage of high resolution of contacts within the investigated locus (Dekker *et al.*, 2002; Stadhouders *et al.*, 2013), however, they are both restricted to low throughput as regards the "viewpoint" fragment and by the requirement of some previous knowledge/assumptions on the locus' interactome. On the other hand, Hi-C allows investigation of interactions between all individual loci across the genome (Lieberman-Aiden *et al.* 2009); but, to obtain sufficient coverage massive sequencing effort is necessary, rendering Hi-C too costly for most research groups. To overcome the aforementioned limitations a novel targeted chromatin capture (T2C) approach was developed, which enables analysis of all interactions within a specific region of interest at single restriction-fragment resolution (Kolovos *et al.*, 2014). In brief, 3C ligation products are digested with a secondary restriction enzyme and oligonucleotide probes spanning the examined locus are used to subselect those interactions that emanate from the locus and sequence them in a high throughput fashion (for more details see Methods 2.2.4.4).

### **3.4.1 Implementation of iT2C on a 2.8-Mbp locus of chromosome 14**

On the basis of the protocol for conventional T2C, we developed a native version called iT2C (Brant *et al.*, 2016). We performed iT2C in HUVECs using probes targeting every *ApoI* fragment in a 2.8-Mbp locus containing our model *SAMD4A* gene, and compared local iT2C interaction maps to conventional T2C profiles (Fig. 29A). At low and higher resolution, regular T2C heat maps display genome organization in TADs/subTADs, which are consistent with Hi-C maps that were generated for this locus (Fig. 29A; Rao *et al.*, 2014). In contrast, natively performed experiments revealed that TAD structures are outlined by interaction signal at lower resolutions, while individual chromatin loops are visualized against ultra-low background signal at higher ones (Fig. 29A). To assess whether our native contacts constitute random ligation signals, we analyzed which epigenetic factors (e.g., CTCF and histone marks) are enriched at the basis of these captured iT2C interactions by alignment of the iT2C results to ENCODE ChIP-seq data. The corresponding PE-scan graphs (de Wit *et al.*, 2013) demonstrated that iT2C contacts arise by the interaction of chromatin segments marked by CTCF, H3K27ac or H3K27me3 (Fig. 29B). As a control, contact enrichment for the repressive histone mark H3K9me3 was not observed (Fig. 29B). Next, we studied effects of inhibited transcriptional elongation on spatial chromatin structure by pre-treating HUVECs with DRB, a chemical agent known to hinder transcription elongation by RNA polymerase II via the inhibition of CDK9 (Yamaguchi *et al.*, 1999). In agreement with previous reports, transcriptional disruption only partially alters native genomic organization (Fig. 29A), in support of the notion of an overarching and transcription-induction-independent chromatin framework (Dixon *et al.*, 2015).



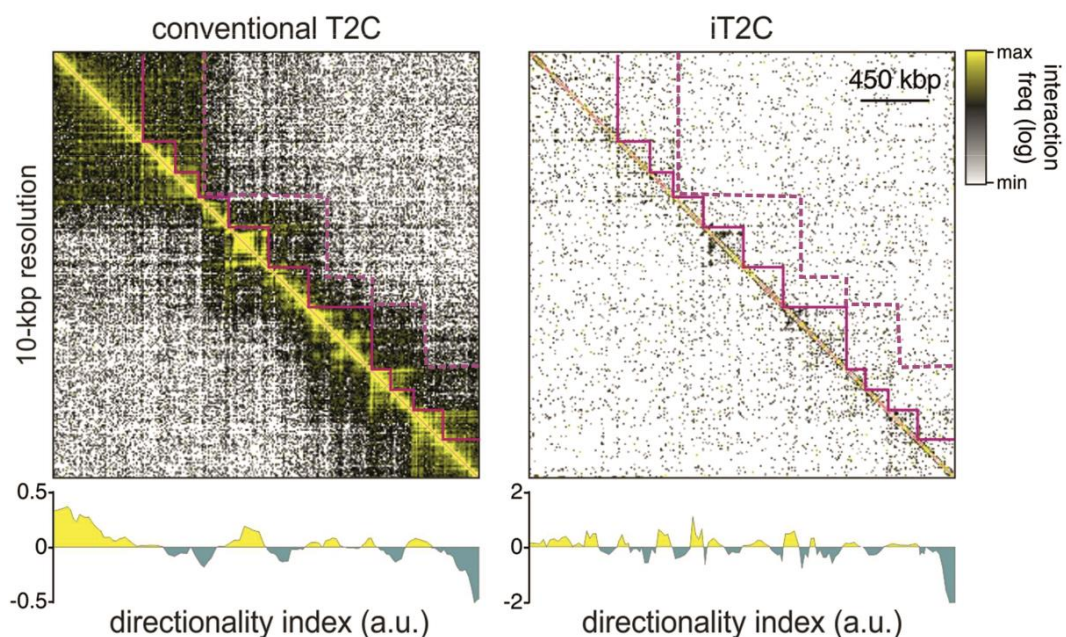


**Figure 29 | Expanded view on chromatin organization on chromosome 14 by iT2C and conventional T2C. A |** Interaction maps from conventional T2C (left) and iT2C (middle) in the 2.8 Mbp locus surrounding SAMD4A on chromosome 14 at increasing resolution. ENCODE ChIP-seq data from HUVECs are aligned below 250 kbp window around SAMD4A. iT2C was performed in HUVECs pretreated with the transcription inhibitor DRB (right). **B |** PE-Scan graphs were plotted as described in de Wit *et al.*, 2013 and show the enrichment of iT2C interactions +/- 5 kbp close to CTCF (grey), H3K27ac (pink) and H3K27me3 (blue) sites, while H3K9me3 (brown) that is absent from this region serves as a control.

Taken together, iT2C data show major differences when native and conventional T2C interaction maps are compared, with native profiles being devoid of background noise from bystander ligations. Although both methods result in equivalent higher-order chromatin structures suggestive of TAD-based organization, individual chromatin loops are uncovered more robustly with iT2C. Moreover, high contact frequencies close to the diagonal in iT2C profiles are reminiscent of short and dense contact domains observed via 'Micro-C' experiments in yeast (Hsieh *et al.*, 2015).

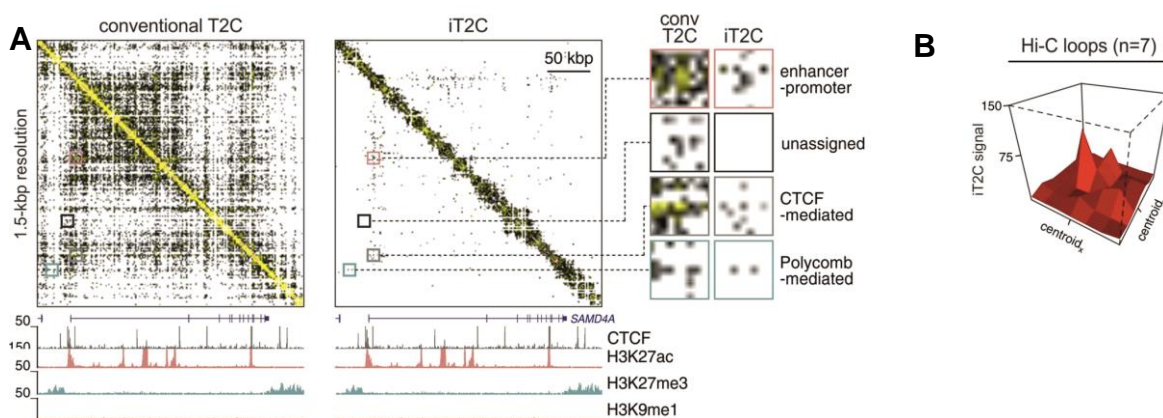
### 3.4.2 Characteristics of iT2C and conventional T2C interaction maps

Although iT2C and conventional T2C contact heat maps display an overall consistent larger-scale genome organization in the 2.8-Mbp locus around *SAMD4A*, native profiles sensitively reveal more sub-TADs. To investigate characteristics that define TAD identification in both approaches, we calculated the ‘directionality index’ (DI), which is a typical tool for TAD boundaries identification (Dixon *et al.* 2012). In brief, DI quantifies the bias of upstream versus downstream contact frequencies in T2C heat maps, and the points of local change (“local insulation”) indicate TAD boundaries. Interestingly, once calculated on iT2C data, DI revealed numerous sub-TADs, while conventional T2C identifies bigger sized TADs (Fig. 30). A closer inspection on both heat maps, however, demonstrates that sub-TADs found under native conditions are also visualized in conventional profiles (magenta colored line). Vice-versa, large TADs seen in conventional data are outlined in native interaction profiles (dashed magenta colored line). These results reflect that DI calculations are differentially affected by the underlying data and new bioinformatics tools for iT2C are needed.



**Figure 30 | Comparison of TAD boundaries in iT2C and conventional T2C maps.** Interaction data from conventional T2C (left) and iT2C (right) experiments in the 2.8 Mbp locus surrounding *SAMD4A* on chromosome 14 were used to calculate the directionality index, which is an indicator for TAD boundaries (Dixon *et al.*, 2012). Solid magenta colored line illustrates sub-TADs, while dashed magenta line represents TADs.

An additional feature of iT2C experiments is the enhancement of individual chromatin loops that are not obvious in signal-crowded T2C maps. To exemplify this, we examined iT2C contact signals around our model *SAMD4A* gene: these were linked to specific enhancer-promoter contacts, to CTCF- or Polycomb-mediated regions. Comparison of iT2C signal around these contacts to that in conventional T2C maps revealed that conventional T2C contacts suffered from increased, potentially artefactual ligation events that impair discrimination of the loop signal (Fig. 31A). Moreover, unassigned signals generated under cross-linked conditions are not at all seen in iT2C interaction profiles. Then, to establish whether known chromatin loops are recapitulated with the native method, we correlated iT2C data to convergent CTCF-CTCF interactions previously identified in that region using publicly-available Hi-C data (Rao *et al.*, 2014). Our findings show that all seven previously-reported loops were captured by iT2C (Fig. 31B).

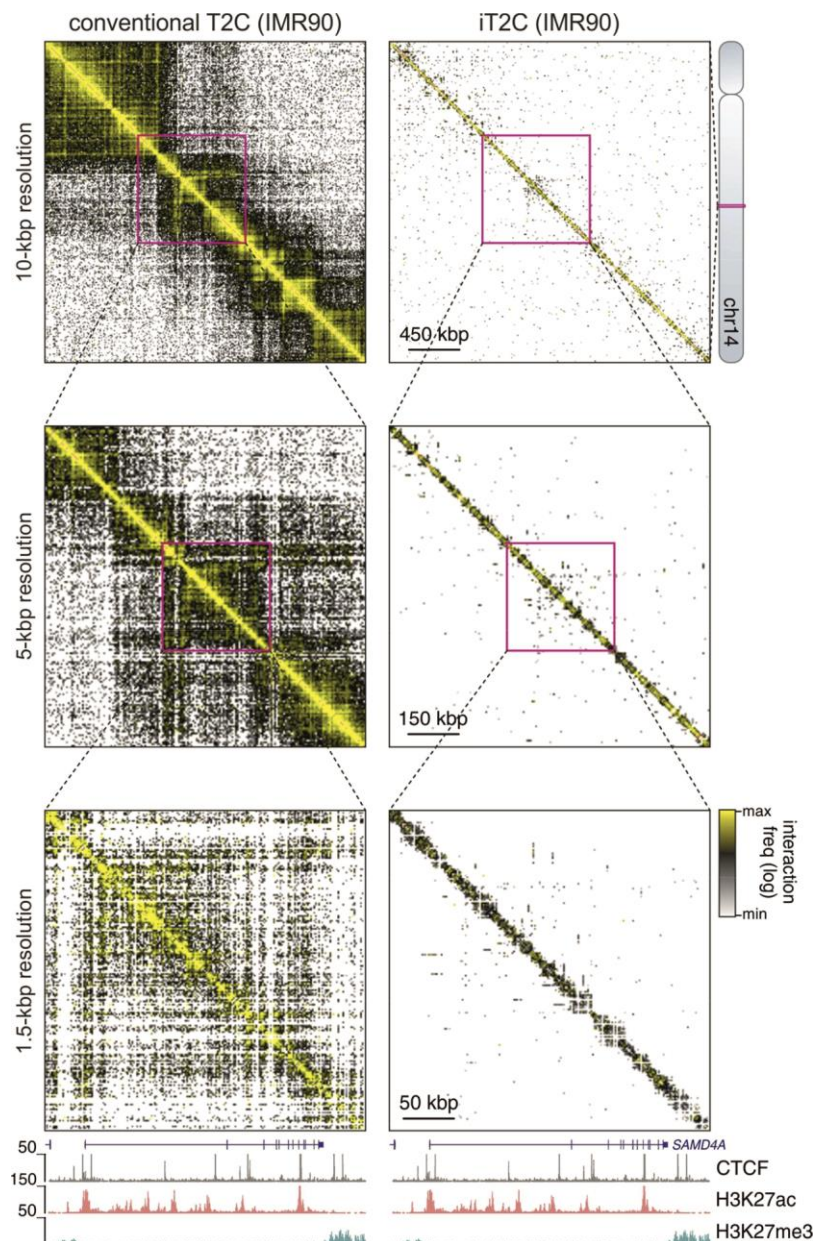


**Figure 31 | Comparison of interaction sites captured by iT2C or conventional T2C. A |** Interaction maps from conventional T2C (left) and iT2C (right) in the 2.8 Mbp locus surrounding *SAMD4A* on chromosome 14 at increasing resolution. ENCODE ChIP-seq data from HUVECs are aligned below. Exemplary interactions are magnified. **B |** PE-Scan graph was plotted as described in de Wit *et al.*, 2013 and shows the enrichment of iT2C interactions at CTCF-CTCF chromatin loops (n=7) that were previously called by Hi-C (Rao *et al.*, 2014).

These data show that the native iT2C protocol results in sparse interaction profiles at the sub-chromosomal level; crucial structural attributes like TADs, sub-TADs, and discrete chromatin looping events are delineated using iT2C. When compared to conventional T2C, marked differences are observed in signal enrichments and TAD boundary definition using the DI, allowing for a finer appreciation of chromatin folding.

### **3.4.3 Implementation of iT2C in a different cell line**

In order to recapitulate our local contact maps from native and conventional capture experiments in an additional cell type, iT2C was performed side-by-side to standard T2C in the human primary lung fibroblast cell line, IMR90, using *ApoI* and the same 2.8-Mbp locus around *SAMD4A*. Contact profiles were correlated to ENCODE ChIP-seq data and visualized at different resolutions. Analysis of the resulting interaction profiles demonstrate many common contacts between HUVECs and IMR90s that are associated to characteristic genomic elements within the TAD containing *SAMD4A* (Fig. 32). At a lower resolution, TADs and subTADs that were also reported in HUVEC are outlined. Furthermore, and again consistent to data obtained in HUVECs, iT2C and conventional T2C contact maps show major variations in signal enrichment. However, a common higher-order chromatin organization was still observed. These data lead to the conclusion that different cell types display a correlated over-arching genome architecture, which is in agreement with studies of cell types derived from stem cell differentiation (Dixon *et al.*, 2015).



**Figure 32 | Recapitulation of iT2C and conventional T2C heat maps in IMR90s.** Interaction maps from conventional T2C (left) and iT2C (right) in the 2.8 Mbp locus surrounding SAMD4A on chromosome 14 at increasing resolution. ENCODE ChIP-seq data from IMR90s are aligned below 250 kbp window around SAMD4A.

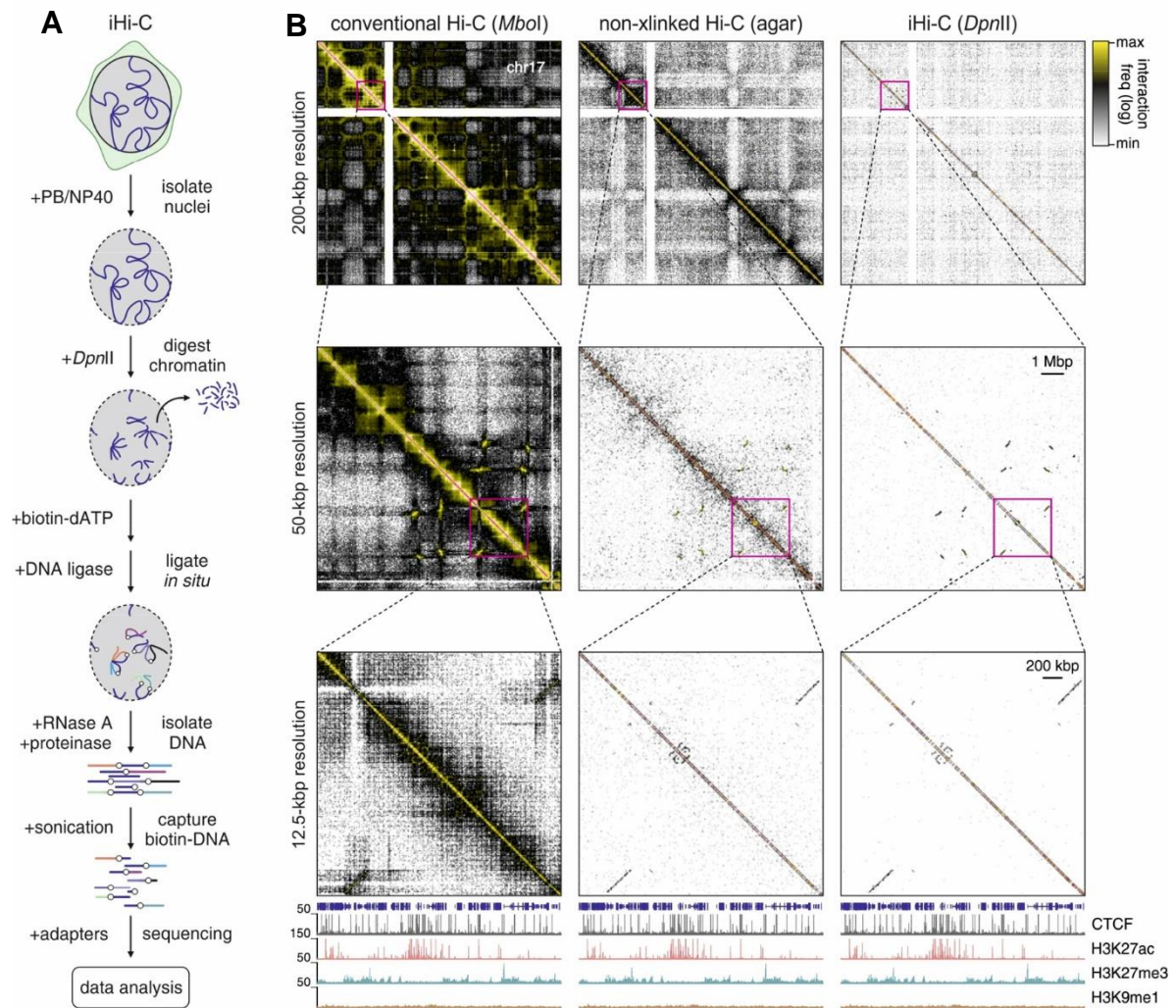
### 3.5 Proof of principle iHi-C: generation of intrinsic genome-wide contact maps

Contrary to the above described techniques, Hi-C represents an untargeted method allowing the analysis of ‘all versus all’ interactions, thereby suitable for obtaining a whole-genome pairwise interaction map (Lieberman-Aiden *et al.*, 2009). Previous Hi-C studies have uncovered a partitioning of the genome into active and inactive A/B compartments (Lieberman-Aiden *et al.*, 2009), and identified TADs as structural units that shape nuclear architecture during interphase (Dixon *et al.*, 2012; Naumova *et al.*,

2013). Recently, cost-intensive sequencing efforts were also made in order to increase resolution of contact maps and uncover new aspects of genome folding, like the formation of loops almost exclusively between thousands of convergent CTCF-bound sites (Rao *et al.*, 2014).

### **3.5.1 Implementation of iHi-C**

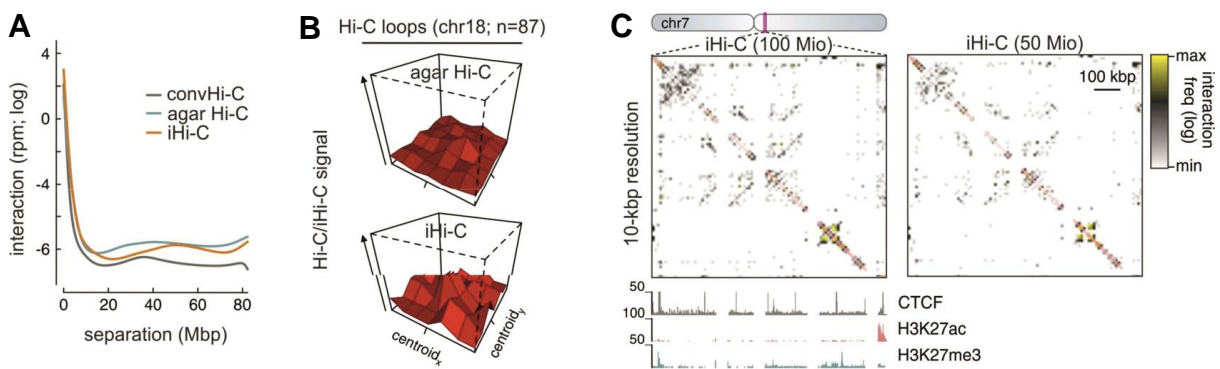
Adapted from the *in situ* Hi-C protocol (Rao *et al.*, 2014) but implemented on the basis of our i3C method, we developed a native version called ‘iHi-C’ to explore chromatin contacts at the genome-wide level, while still avoiding biases that involve chemical cross-linking and harsh detergent treatments (Brant *et al.*, 2016). For this purpose we treated nuclei from HUVECs with the 4-bp cutter *DpnII*, incorporated biotinylated-dATP into ligation junctions to mark them, and subjected the resulting templates to paired-end sequencing to ~300 million reads (Fig. 33A). Following data analysis, data from conventional Hi-C performed *in situ* on crosslinked cells or on non-crosslinked but agar-embedded cells were compared and contrasted. Figure 33B gives exemplary maps from human chromosome 17 at increasingly higher resolutions, and aligned to ENCODE ChIP-seq data. We can deduce that large A/B compartments and the established TAD “triangles” are seen in conventional maps, but are reduced to a compendium of loops in the native iHi-C. Notably, these maps display less signal density compared to our iT2C reports, which might indicate a need for “deeper” sequencing, but at the same time hint less of the compartmentalization of the chromosomal landscape, and more for a model that combines spatially-unstructured and structured domains.



**Figure 33 | Proof-of-principle: iHi-C generated in HUVECs. A** | Strategic overview of the iHi-C protocol. Prior to ligation, cohesive fragment-ends were filled-in with biotin-dATP, allowing pull-down of all ligation hybrids. **B** | Comparison of interaction maps for conventional Hi-C (HUVECs, *HindIII*, sequenced to 500 million reads), non-crosslinked and agar embedded Hi-C from Rao et al., 2014 (lymphoblasts, *Mbol*, 100 million reads) and iHi-C (HUVECs, *DpnII*, 150 million reads) for chromosome 17 at increasing resolution. ENCODE ChIP seq data were aligned below a 2 Mbp window.

Now, to assess whether interaction frequencies differ with increasing distance, we compared the three approaches. In all three more interactions are captured at the shorter-ranges, while longer-range ones display lower frequencies – and, on average, frequency decay profiles were consistent among conventional *in situ* Hi-C and iHi-C (Fig. 34A). Next, we evaluated the distribution of iHi-C interactions relative to CTCF-bound sites by correlating native and *in situ* Hi-C signals to known CTCF-CTCF loops on chromosome 18. PE-Scan graphs in figure 34B demonstrate that iHi-C recapitulates CTCF-mediated loops more robustly than *in situ* Hi-C on agar-embedded cells, revealing that the latter is biased for random ligation events. Finally,

we tested if the loss of non-associated DNA fragments and the improved signal-to-noise ratio of iHi-C allow for lower sequencing depth without contact detection impairment. We generated iHi-C interaction maps for a 0.85-Mbp region on chromosome 7 at 10 kb resolution using 100 or 50 randomly-selected million reads *in silico*. The resulting contact maps are strikingly similar, and highlight how effective contact detection is over near-zero background noise via our native protocol (Fig. 34C).



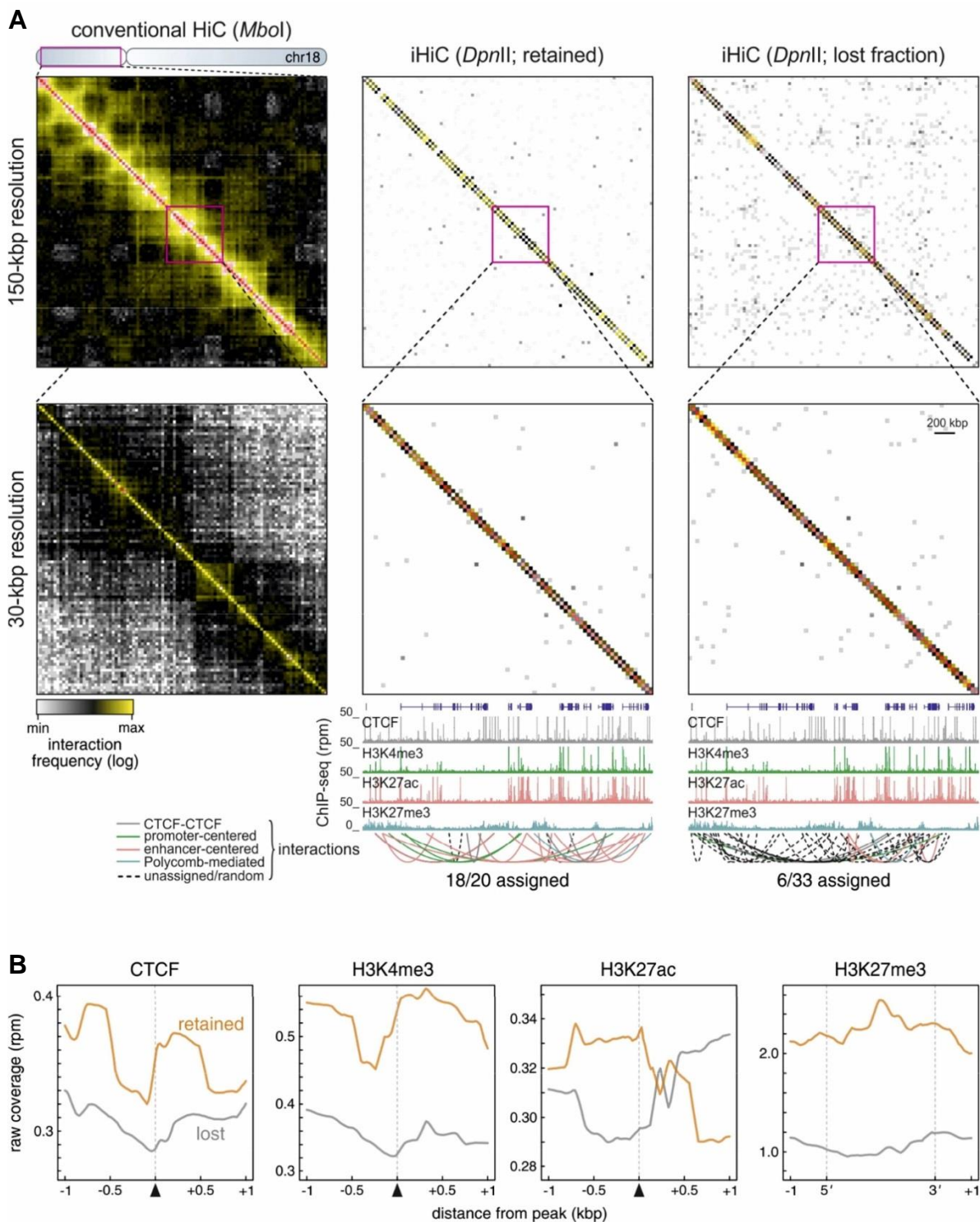
**Figure 34 | A closer look on iHi-C features.** **A** | Line plot shows the combined  $\log_{10}$ -interaction (in rpm) at increasing distance (Mbp) in iHi-C (orange), conventional Hi-C (black) and non-crosslinked agar Hi-C (blue). **B** | PE-Scan graphs were plotted as described in de Wit *et al.*, 2013 and show the enrichment of iHi-C and agar Hi-C interactions at CTCF-CTCF chromatin loops ( $n=87$ ) that were previously called by Hi-C (Rao *et al.*, 2014). **C** | iHi-C interaction maps from HUVECs for 0.85 Mbp region on chromosome 7 using either 100 million reads or 50 million reads. ENCODE ChIP seq data were aligned below.

Taken together, we provide proof-of-principle that iHi-C outlines discrete chromatin interactions in a more accurate fashion than the previously introduced uncross-linked, agar-embedded *in situ* Hi-C. Nonetheless, iHi-C profiles lack much signal compared to conventional Hi-C, especially at the large-scale A/B compartment level. These data beg the question if any major contacts are lost during the iHi-C procedure. If so, this might be either due to loss of interacting genomic segments during cutting/pelleting at the very beginning of the protocol, or a result of insufficient biotinylation of cohesive DNA ends in PB, which can lead to reduced ligation efficiency. The latter can be probably disregarded, given the broad coverage of diverse interactions along the contact maps, even at lower sequencing depths; the former is queried below.



### **3.5.2 What is lost? iHi-C generated in 'retained' and 'lost' fractions**

The key steps of all i3C-derived variants involve circumvention of chemical cell cross-linking and thus removal of substantial amounts of DNA after chromatin digestion. We have already demonstrated that the 'retained' and 'lost' fractions contain equal chromatin composition (Fig. 11). Although i3C and i4C experiments reproduced similar contact profiles to conventional methods, iT2C and especially iHi-C reveal strikingly sparse contact maps. Therefore, we investigated if major pairwise contacts are lost during the native iHi-C procedure by ligating biotinylated DNA fragments that are close in space in both the 'retained' and the 'lost' nuclear HUVECs fraction according to the standard iHi-C protocol (see 33A) or under dilute conditions, respectively. After sequencing to ~200 million reads, we generated heat maps at different resolutions for a region of chromosome 18 and aligned them to ENCODE ChIP-seq data (Fig. 35A). When compared to conventional Hi-C profiles, our data reveal that more contacts associated to genomic elements such as promoters, enhancers, Polycomb-, and CTCF-bound sites, were retrieved in the 'retained' iHi-C fraction. In contrast, only few meaningful contacts could be assigned using the 'lost' iHi-C data, accompanied by seemingly random ligation events. Next, we examined the quality of detected signals by comparing read coverage from 'retained' and 'lost' iHi-C profiles around CTCF-, promoter-, enhancer-, and Polycomb-centered peaks (Fig. 35B). Our results support the notion that DNA segments lost after chromatin digestion in the i3C protocol are not engaged in "meaningful" chromatin interactions. Nonetheless, the molecular biases leading to varying signal enrichment in conventional and native Hi-C experiments require further examinations and novel analytical tools.

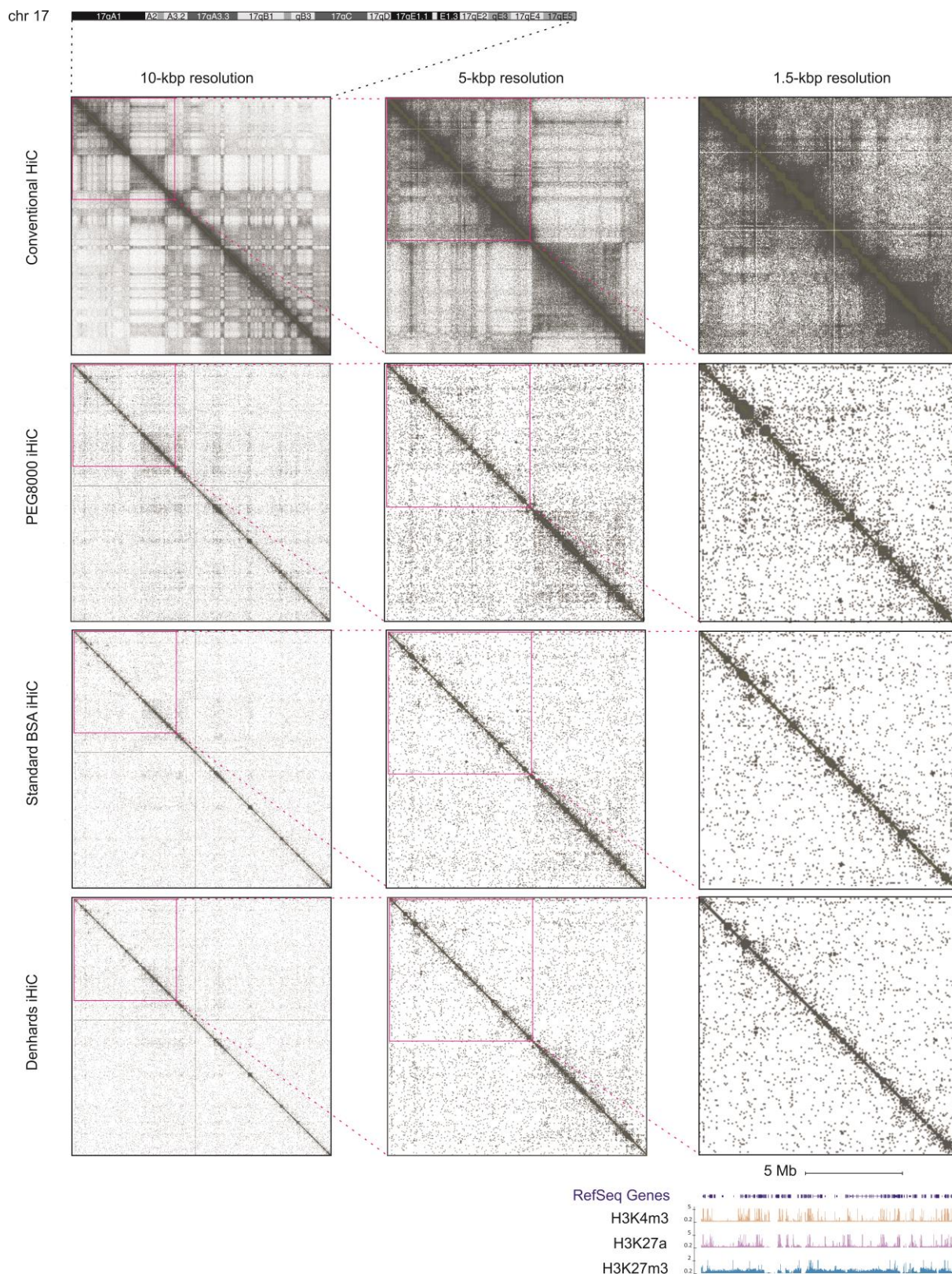


**Figure 35 | iHi-C generated in the 'lost' or 'retained' fractions of HUVECs. A |** Comparison of interaction maps for conventional Hi-C (HUVECs, *HindIII*, sequenced to 500 million reads), iHi-C on the 'retained' fraction (HUVECs, *DpnII*, 200 million reads) or iHi-C on the 'lost' fraction (HUVECs, *DpnII*, 200 million reads) for chromosome 18 at increasing resolution. ENCODE ChIP seq data were aligned below a 2 Mbp window. **B |** Line plots showing the differences in raw read coverage (in rpm) of 'retained' (*orange*) and 'lost' iHi-C data (*grey*) around CTCF, H3K4me3, H3K27ac, and H3K27me3 peaks. *Below*: Curved lines (*below*) connect interacting bins and are colour-coded as indicated.

### **3.5.3 Improvement of ligation efficiency leads to better contact enrichment**

Having established that i3C-based variants reliably capture chromatin interactions, we sought to further address the outstanding concern over signal sparsity in our genome-wide contact maps (compared to conventional heat maps; see Fig. 33). To investigate if we could enrich our datasets for “meaningful” ligations, we performed pilot iHi-C experiments in mESCs under altered ligation steps. We took advantage of the phenomenon of macromolecular crowding, whereby addition of crowding agents to our PB solution would reduce the volume of available solvent for other molecules, thus increasing their effective concentration. Under such conditions macromolecules such as proteins were shown to become stabilized and enzymatic kinetics increased (Norris & Malys, 2011). We applied iHi-C in mESCs as described in 2.2.4.5, but instead of supplementing PB with 100 µg of bovine serum albumin (BSA) during the ligation procedure, we added either polymer polyethylene-glycole (PEG8000) to a final concentration of 20% or Denhardt’s (contains polymers such as ficoll, polyvidone and serum albumin) diluted to 5X concentration. After paired-end sequencing, reads were mapped to the mouse reference genome and pairwise interaction heat maps were compared to those from conventional Hi-C data (Geeven *et al.*, 2015). The addition of PEG8000 stands out, as it remarkably intensifies interaction signals and reproduces TAD outlines, as well as strong chromatin loops that were observed with conventional Hi-C at lower and higher resolutions (exemplified by maps for chromosome 17 in Fig. 36). On the other hand, addition of BSA or 5X Denhardt’s to the iHi-C ligation reactions did not significantly improved the basic iHi-C results (Fig. 36).

In summary, although the PCR amplification steps in i4C and i3C-seq procedures sufficed for obtaining enriched contact profiles, iHi-C involves sensitive biotinylation and ligation product pull-down steps that can critically affect signal intensity. Still, enhanced ligation efficiencies and richer chromatin interaction maps were obtained upon addition of PEG to PB, demonstrating that our i3C protocol can be applied in a genome-wide manner to analyze chromatin conformation at various resolutions.



**Figure 36 | Improvement of iHi-C signal enrichment in mESCs.** Comparison of interaction maps for conventional Hi-C using 100  $\mu$ g BSA (mESCs, *DpnII*, Geeven *et al.* 2015), iHi-C using 20 % PEG8000 (mESCs, *DpnII*, sequenced to 200 million reads), standard iHi-C using 100  $\mu$ g BSA (mESCs, *DpnII*, sequenced to 200 million reads) or iHi-C using 5 x Denhardt's (mESCs, *DpnII*, sequenced to 200 million reads) on chromosome 17. ENCODE ChIP seq data were aligned below a 10 Mbp window.

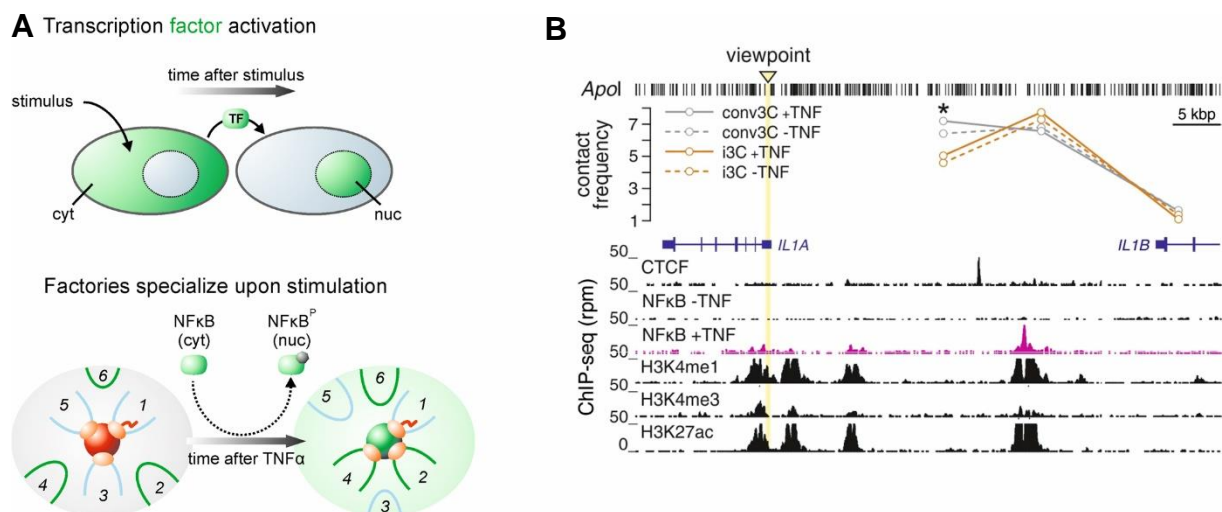
### **3.6 Dynamics of chromatin interactions upon cytokine signalling**

The tumor necrosis factor alpha ( $\text{TNF}\alpha$ ) is a strong cytokine that orchestrates the inflammatory response through nuclear factor kappa B ( $\text{NF}\kappa\text{B}$ ) signaling (Smale, 2010). The transcription factor  $\text{NF}\kappa\text{B}$  commonly consists of a p65/p50 heterodimer that is bound by inhibitor of nuclear factor kappa B ( $\text{I}\kappa\text{B}$ ) proteins in the cytoplasm, blocking its nuclear localization. Upon stimulation of cells with cytokines such as  $\text{TNF}\alpha$ , subunits of the  $\text{I}\kappa\text{B}$  kinase complex (IKK) are activated leading to phosphorylation of  $\text{I}\kappa\text{B}$  proteins and subsequent ubiquitination followed by proteosomal degradation of  $\text{I}\kappa\text{B}$ . This then translocates  $\text{NF}\kappa\text{B}$  into the nucleus where it activates promoters of several hundred responsive genes (Ashall *et al.*, 2009; Smale, 2010).

Current knowledge sees “activating” chromatin looping events involving particular *cis*-regulatory elements such as active promoters, enhancers and CTCF-binding sites (Sanyal *et al.*, 2012). On the other hand, inactive promoters and segments enriched for repressive chromatin marks (e.g., H3K27me3) interact preferentially with similarly classified *cis*-elements to form compact three-dimensional structures (Francis *et al.*, 2004). Especially promoter-enhancer contacts are embedded in complex networks to regulate gene expression. Previous studies have shown that genes sharing enhancers, for instance  $\text{NF-}\kappa\text{B}$ -responsive genes, are coordinately induced by  $\text{TNF}\alpha$  signaling and co-transcribed in “specialized” transcription factories (i.e., in factories rich in  $\text{NF-}\kappa\text{B}$  (Fig. 37A); Papantonis *et al.*, 2012). Moreover,  $\text{TNF}\alpha$  induction was used to dynamically regulate transcription of long human genes by RNA polymerase II in a temporal manner. This indicates a strong relationship between dynamic long-range chromatin interactions and transcription factor dependent gene expression regulation (Wada *et al.*, 2009). Although many genes co-associate only after  $\text{TNF}\alpha$  stimulation, numerous pre-established promoter-enhancer contacts have been recorded already before cytokine induction in human cells (Jin *et al.* 2013). Hence, distinct transcription programs might be controlled by pre-existing chromatin looping structures in order to facilitate connections between enhancers and their target genes. The mechanisms that link transcriptional regulation with chromatin architecture are still elusive and were investigated during this work.

### 3.6.1 Promoter-enhancer crosstalk after TNF $\alpha$ stimulation using 3C or i3C

The promoter of the TNF $\alpha$  responsive *IL1A* gene interacts with a downstream enhancer prior to binding by the transcription factor NF- $\kappa$ B at that enhancer site (Jin *et al.*, 2013). To investigate if pre-existing contacts are an artefact of cell cross-linking and random bystander-ligation events, we explored this well-studied locus using conventional 3C-qPCR in parallel with native i3C-qPCR. We prepared i3C and 3C templates from HUVECs that were untreated or stimulated with TNF $\alpha$  for 60 minutes, and assessed interaction frequencies between the promoter, a distal NF- $\kappa$ B-bound enhancer and intervening regions not enriched for “active” histone marks. As shown in Figure 37B, DNA pre-looping between the *IL1A* promoter and the downstream enhancer was captured independently of cytokine induction both by conventional as well as by native 3C. However, under cross-linked conditions significant contact frequencies were also detected at the intervening control regions, thus blurring the interaction profile for the *IL1A* promoter (Fig. 37B).



**Figure 37 | Detection of pre-looping in a TNF- $\alpha$  responsive locus in HUVECs.** **A** | Schematic illustration of specialized factories. Cytoplasmic NF $\kappa$ B translocates into the nucleus upon stimulation with TNF- $\alpha$ . **B** | Co-transcription of TNF- $\alpha$  responsive genes at specialized ‘NF $\kappa$ B factories’ is induced after binding of nuclear NF $\kappa$ B (Papantonis & Cook, 2013). **B** | i3C (orange) and conventional 3C (grey) coupled to qPCR were performed in the locus around the TSS of *IL1A* prior (solid lines) and post 60 minute TNF- $\alpha$  stimulation (dashed lines) of HUVECs. Data were aligned to ENCODE ChIP-seq and p65 ChIP-seq data (Papantonis *et al.*, 2012).  $P < 0.05$ ; two-tailed unpaired Student’s t-test ( $n = 2$ ).

In agreement with previously-published data (Jin *et al.*, 2013), we recapitulated a pre-existing promoter-enhancer contact that forms independent of TNF $\alpha$  stimulation and was detected using either conventional or native capture methods, highlighting that

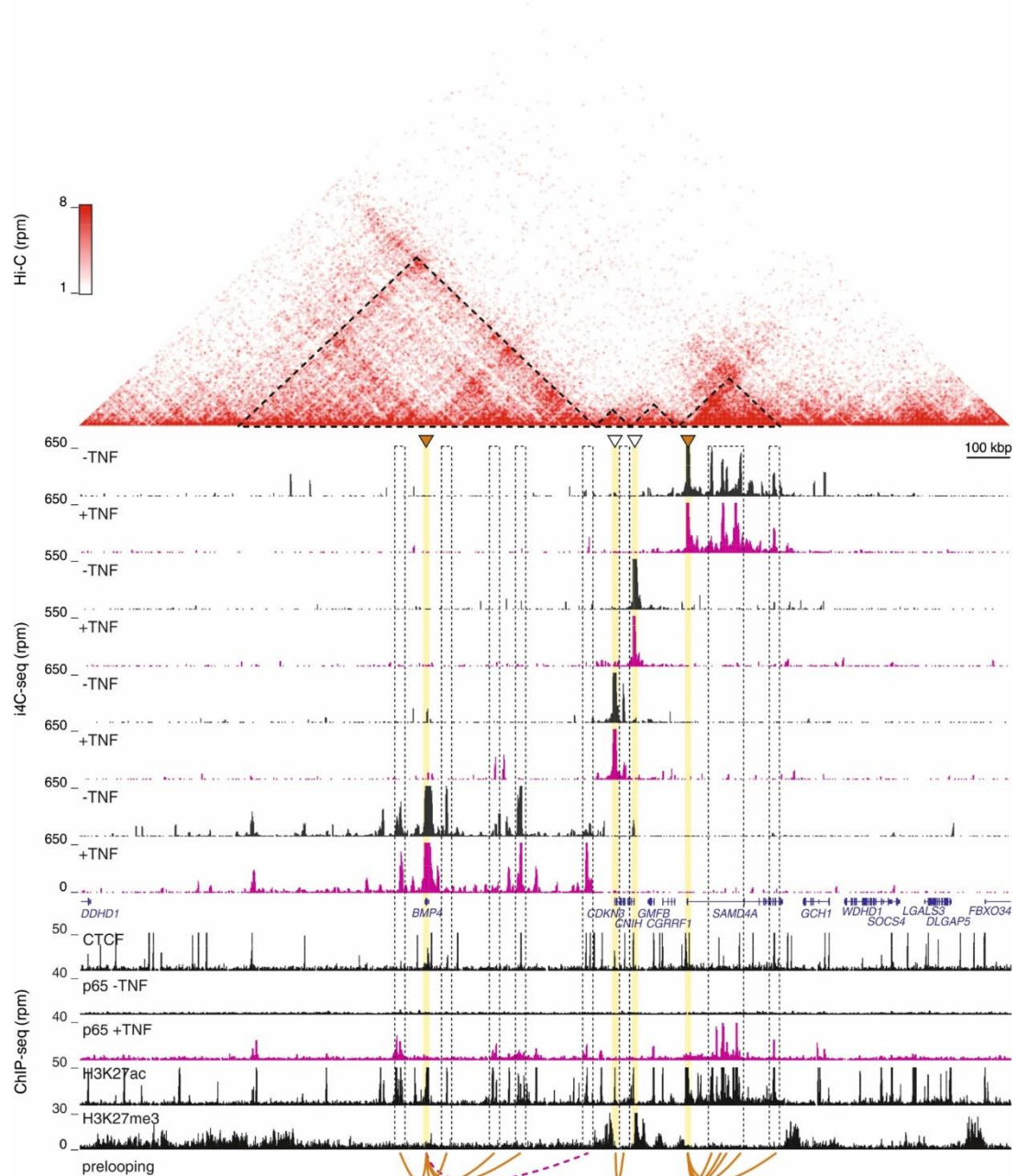
pre-established interactions between *cis*-regulatory elements are not a result of artefactual cross-linking. Consistent with previous findings (Fig. 14), i3C displayed improved signal-to-noise ratios compared to the conventional technique, allowing for a more accurate contact determination.

### **3.6.2 Tracking chromatin interaction dynamics on chromosome 14 using i4C**

Since i4C experiments yielded enhanced signal-to-noise levels, we hypothesized that focal i4C contacts would allow for a perhaps more detailed characterization of promoter-enhancer contact dynamics before and after TNF $\alpha$  stimulation (Brant *et al.*, 2016). Hence, we applied i4C and conventional 4C on HUVECs that were untreated or stimulated by a 60-minute pulse of TNF- $\alpha$ . We explored interactions made by the TSSs of the TNF $\alpha$  responsive genes: *SAMD4A*, *BMP4* as well as with the non-responsive *CDKN3* and *CNIH* genes on human chromosome 14, each previously shown to reside within its own TAD (Fig. 23). ChIP-seq data of the NF- $\kappa$ B heterodimer p65 component were aligned to the i4C/4C interaction profiles to enable tracking of p65-centered spatial contacts along the locus (Papantonis *et al.*, 2012). Our findings revealed that the *SAMD4A* and *BMP4* promoters interact with enhancers decorated by the expected H3K27ac histone mark within their respective TADs already prior to TNF $\alpha$  treatment (Fig. 38). Interestingly, i4C profiles show that prelooped contact signal intensities remain either unchanged or increase after incubation with TNF $\alpha$ , and these enhancers notably became bound by p65 within the 60-min stimulation window (Fig. 38). This suggests that the 3D landscape, once established in a cell-type specific manner, is mostly stable and “primed” to facilitate the access of transcription factors to regulatory elements for prompt gene regulation in response to extra-cellular stimuli like TNF $\alpha$ . On the other hand, some enhancers that did not become occupied by p65, either re-engaged with or dissociated from the promoters of *BMP4* and *SAMD4A* upon TNF $\alpha$  treatment. Furthermore, the non-responsive *CDKN3* and *CNIH* genes that are flanked by repressive H3K27me3 marks, responded to TNF $\alpha$  stimulation by forming few promoter-enhancer contacts (Fig. 38). Our data illustrate spatial chromatin interactions that change dynamically upon extracellular signaling, confirming that not all promoter-enhancer contacts are prelooped and stable over time.

The majority of pre-established interactions with *cis*-regulatory elements that were detected in i4C profiles, were also seen when conventional 4C was applied (Fig. 39).

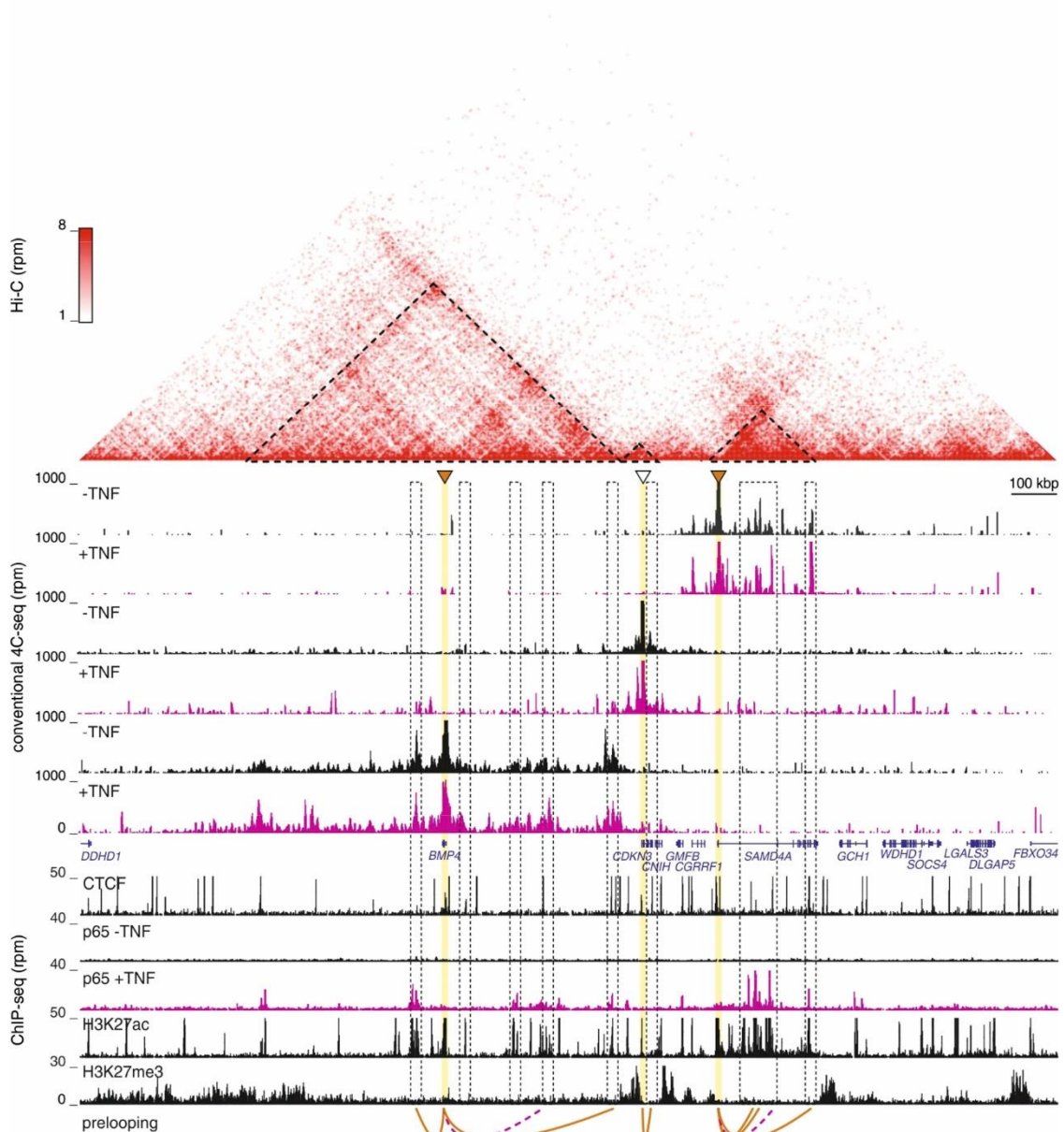
However, the dynamics of discrete contacts were often masked by the elevated background levels and the less focal signal peaks. Moreover, changes in contact frequencies were severely dampened upon TNF $\alpha$  treatment.



**Figure 38 | Native interactions are restricted to TAD boundaries and describe prelooping.** Independent i4C experiments were performed in HUVECs using *NlaIII* and the TSSs of *CDKN3*, *BMP4*, *CNIH* or *SAMD4A* as viewpoints prior and after a 60 minute TNF- $\alpha$  pulse. ENCODE ChIP-seq data p65 ChIP-seq data (Papantonis et al., 2012) were aligned below interaction profiles. Publically available Hi-C data were used to outline TADs (Rao et al., 2014). Orange lines indicate pre-



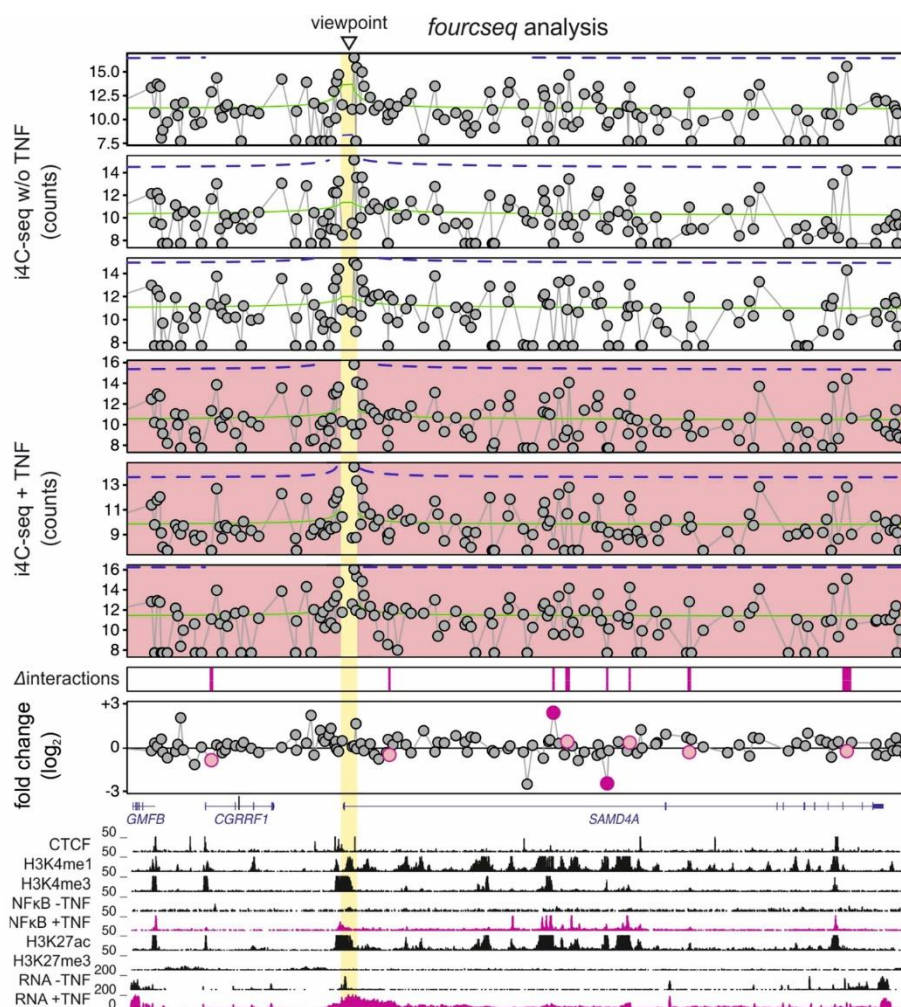
established contacts between enhancers and promoters of *BMP4*, *CDKN3* and *SAMD4A*. Dotted rectangles and magenta colored line indicate interaction dynamics.



**Figure 39 | Conventional 4C interactions upon TNF- $\alpha$  stimulation.** Independent 4C experiments were performed in HUVECs using *NlaIII* and the TSSs of *CDKN3*, *BMP4* or *SAMD4A* as viewpoints prior and after a 60 minute TNF- $\alpha$  pulse. ENCODE ChIP-seq data and p65 ChIP-seq data (Papantonis et al., 2012) were aligned below interaction profiles. Publically available Hi-C data were used to outline TADs (Rao et al., 2014). Orange lines indicate pre-established contacts between enhancers and promoters of *BMP4*, *CDKN3* and *SAMD4A*. Dotted rectangles and magenta colored line indicate interaction dynamics.

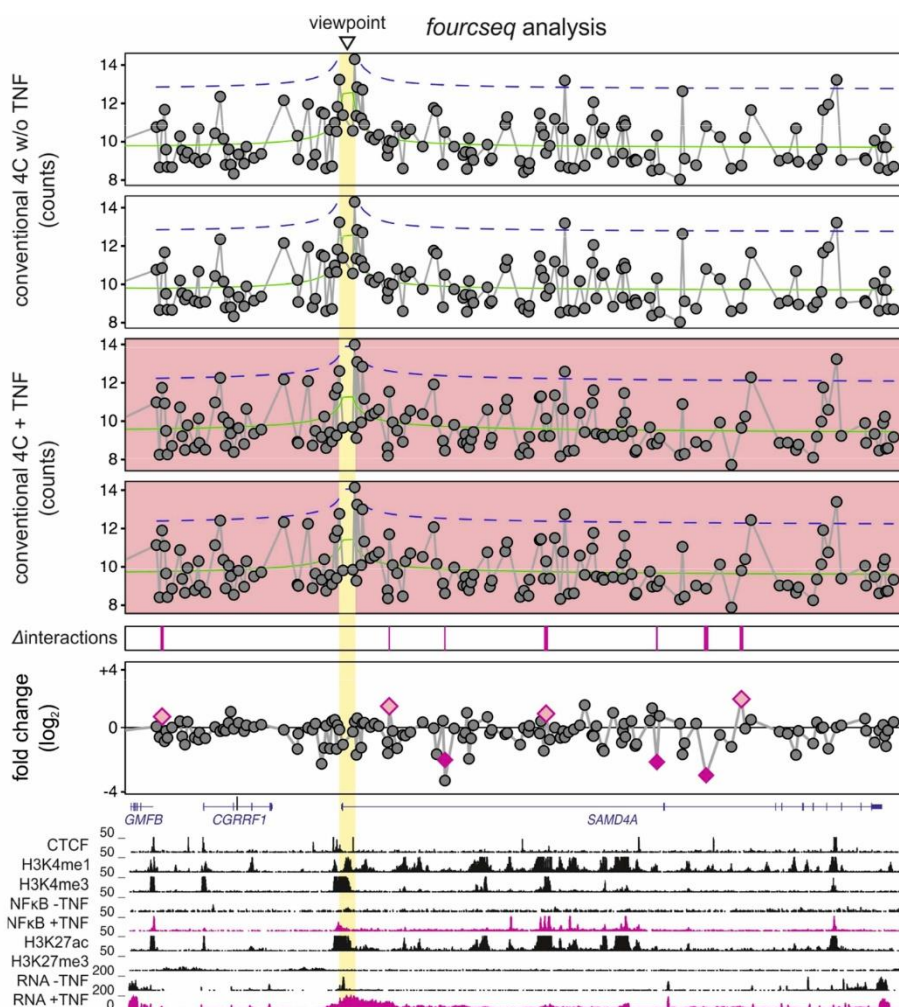
In order to identify significant changes in contact frequencies related to spatial chromatin rearrangements upon TNF $\alpha$  stimulation, we processed *cis*-interactions

detected by i4C or by conventional 4C using the differential analysis tool 'FourCSeq' (Klein *et al.*, 2015). For this, we compared replicated data obtained for the *SAMD4A* TSS viewpoint before and after cytokine induction, because most interaction changes were observed within the TAD surrounding this TNF $\alpha$  responsive gene (Figs. 38 & 39). The differential analysis of profiles in figures 40 and 41, highlight strong interaction changes that supersede the overall noise (pink/magenta dots – unique to i4C; pink/magenta diamonds – unique to conventional 4C, Figs. 40 and 41). Consistent with the visual inspection of i4C/4C contact maps shown above (Figs. 38 and 39), these analyses show significant differences after TNF $\alpha$  stimulation at sites carrying enhancer marks (H3K27ac) and differentially-bound by p65 (Fig. 40). On the other hand, numerous contact changes recorded after differential analysis of conventional 4C data did not correlate with enhancer marks or p65-bound sites (Fig. 41).



**Figure 40 | Differential analysis of i4C data upon TNF- $\alpha$  stimulation.** Data from i4C replicates using *Apo1* and the *SAMD4A* TSS as viewpoint, before (white boxes) and after 60 min TNF- $\alpha$

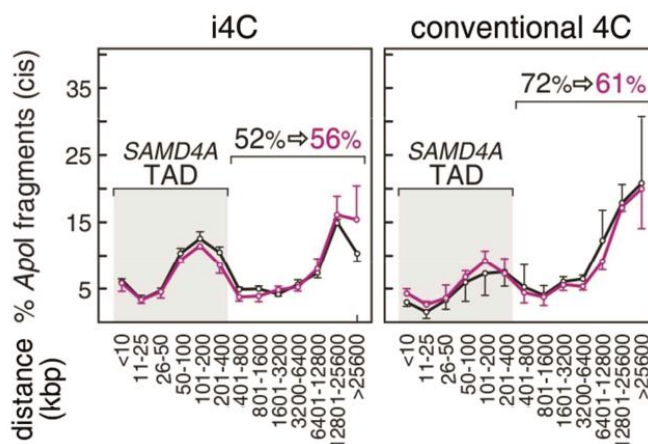
stimulation of HUVECs (pink boxes), were processed using 'FourCSeq' (Klein *et al.*, 2015) in order to analyse significant interaction differences. Some differences ( $\Delta$  interactions) were illustrated as magenta colored bars. HUVEC ENCODE ChIP-seq data, p65 ChIP-seq data and RNA-seq data (Papantonis *et al.*, 2012) were aligned below.



**Figure 41 | Differential analysis of conventional 4C data upon TNF- $\alpha$  stimulation.** Data from 4C replicates using *Apo1* and the *SAMD4A* TSS as viewpoint, before (white boxes) and after 60 min TNF- $\alpha$  stimulation of HUVECs (pink boxes), were processed using 'FourCSeq' (Klein *et al.*, 2015) in order to analyse significant interaction differences. Some differences ( $\Delta$  interactions) were illustrated as magenta colored bars. HUVEC ENCODE ChIP-seq data, p65 ChIP-seq data and RNA-seq data (Papantonis *et al.*, 2012) were aligned below.

Next, we investigated if the partial remodeling of chromatin interactions upon cytokine stimulation resulted in the formation of new interactions that extend beyond TAD-boundary restrictions. As a measure for this, we calculated the percentage of *cis*-contact signal by the *SAMD4A* TSS before and after TNF $\alpha$  induction, to its separation

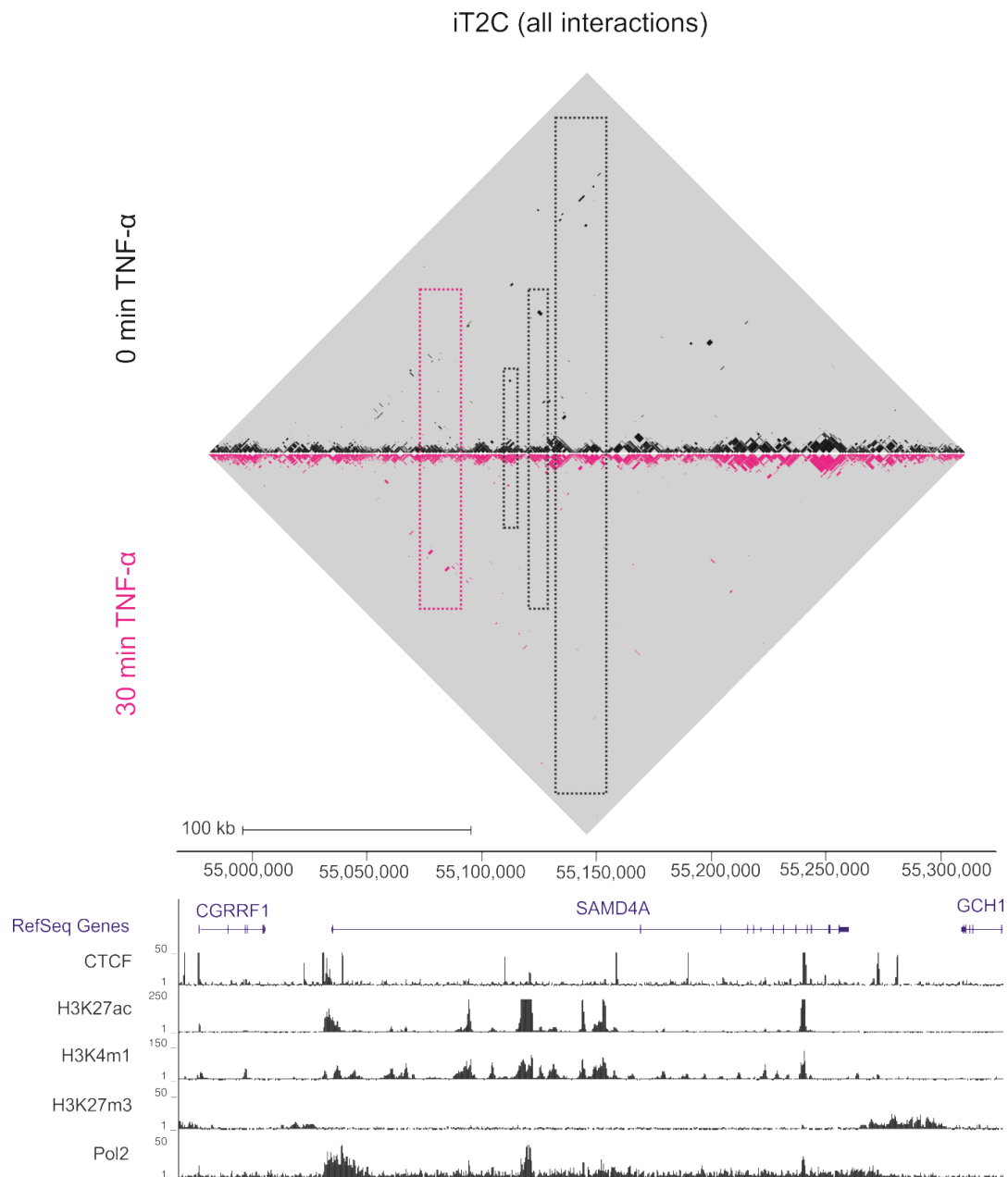
from the viewpoint. Data obtained from i4C and conventional 4C experiments were compared, and the results revealed that the majority of chromatin interactions that were detected under native or conventional conditions, remain restricted within TAD boundaries (Fig. 42). Thus, chromatin dynamics occur in bulk within TADs, with the conventional 4C data showing a ~10% increase in longer range contact signal.



**Figure 42 | TNF $\alpha$  stimulation partially remodels prelooping within the *SAMD4A* TAD.** Percentage of cis-contacts (*Apol* fragments) made by the TSS of *SAMD4A* mapped within and outside its TAD. Results were obtained before (black) and after (magenta) TNF $\alpha$  stimulation using three i4C replicates and two conventional 4C replicates.

### 3.6.3 Analysis of chromatin dynamics on chromosome 14 using iT2C

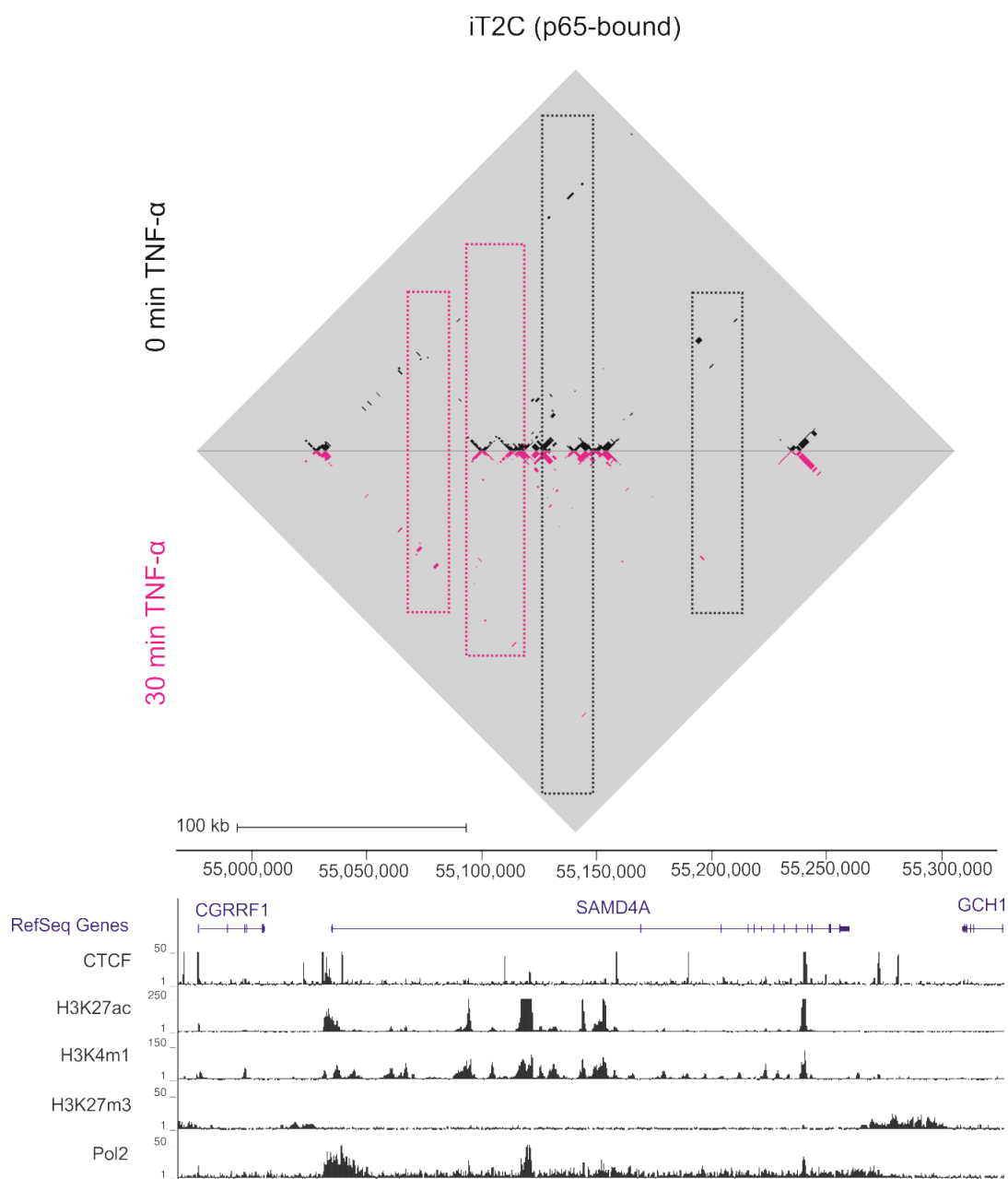
Seeking to exclude that the aforementioned changes in contact frequencies resulted from biases in the PCR amplification during the i4C/4C procedures, we examined chromatin interaction dynamics after cytokine induction in a ‘many-to-many’ fashion. We applied iT2C on our 2.8-Mbp model locus on chromosome 14 in IMR90s before and after a 30-minute TNF $\alpha$  pulse, and iT2C heat maps were aligned to ENCODE ChIP-seq data. First, captured chromatin interactions agree with the known TAD boundaries in the locus, and confirmed that chromatin contacts reside predominantly within domain boundaries both before and after TNF $\alpha$  stimulation (Fig. 43). Second, we observed pre-looping between the *SAMD4A* promoter and downstream enhancer sites, as well as between enhancer-enhancer and CTCF-bound sites that remained stable upon cytokine treatment (Fig. 43). However, a number of contacts recorded in unstimulated cells were lost upon TNF $\alpha$  induction (highlighted by grey rectangles; Fig. 43), while few promoter-enhancer contacts formed *de novo* after stimulation (highlighted by magenta rectangles; Fig. 43).



**Figure 43 | iT2C interactions in the SAMD4A locus upon TNF- $\alpha$  stimulation.** iT2C was performed in IMR90s using Apol and oligo probes capturing interactions in a 2.8 Mbp locus around SAMD4A before (black) and after 30 minutes TNF- $\alpha$  treatment (magenta). Pre-looping interactions unique to unstimulated cells are depicted with black, dashed rectangles. Interaction changes linked to TNF- $\alpha$  signaling are depicted with magenta, dashed rectangles. ENCODE CHIP-seq data were aligned below.

In order to investigate the relation of these chromatin dynamics with the binding of the transcription factor NF- $\kappa$ B, we focused only on iT2C interactions emanating from p65-bound sites in this 2.8-Mbp locus. Notably, the few p65-based contacts detected in unstimulated cells were lost upon TNF $\alpha$  induction, exemplified by the *SAMD4A* promoter which dissociates from the pre-looped enhancer site far downstream its gene body (grey rectangles; Fig. 44). As p65 is not bound at these genomic sites in

unstimulated cells, this observation suggests that NF- $\kappa$ B binding itself promotes the dissociation. On the other hand, p65-associated promoter-enhancer contacts were observed within the super-enhancer in the first *SAMD4A* intron after stimulation (magenta rectangles; Fig. 44), whereas pre-stimulation of enhancer-enhancer contacts further downstream were lost on cytokine induction (second grey rectangle; Fig. 44).

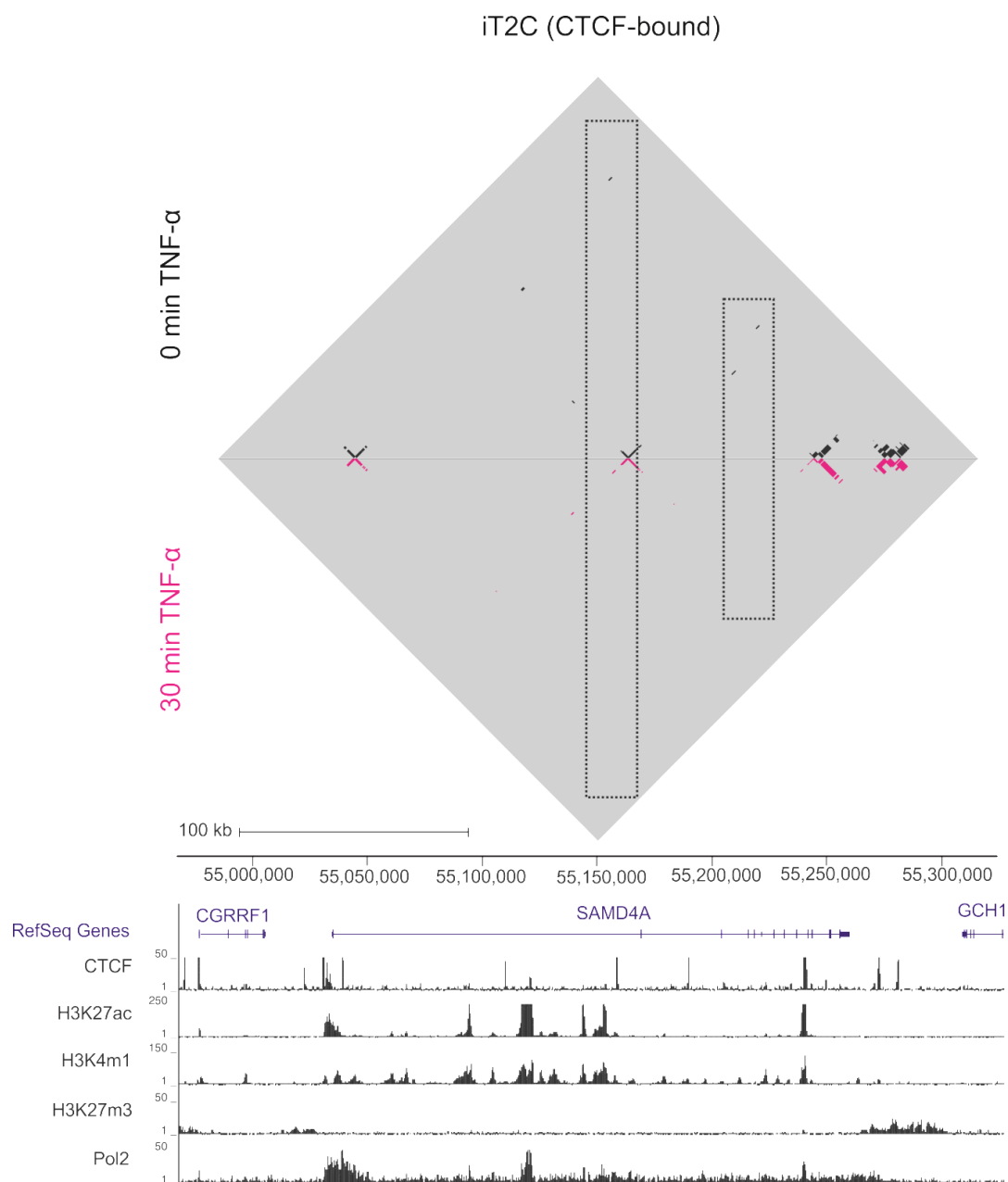


**Figure 44 | iT2C interactions in the *SAMD4A* locus that are bound by p65 upon TNF- $\alpha$  stimulation.** iT2C was performed in IMR90s using Apol and oligo probes capturing interactions in a 2.8 Mbp locus around *SAMD4A* before (black) and after 30 minutes TNF- $\alpha$  treatment (magenta). Contacts associated with p65-binding are displayed. Pre-looping interactions unique to unstimulated

cells are depicted with black, dashed rectangles. Interaction changes linked to TNF- $\alpha$  signaling are depicted with magenta, dashed rectangles. ENCODE ChIP-seq data were aligned below.

Collectively, these data show that many chromatin interactions are established before p65 binds and are preserved after TNF $\alpha$  stimulation. Still, several promoter-enhancer contacts are reshuffled. While the *SAMD4A* promoter tends to interact frequently with the strong enhancer elements in its first intron upon TNF $\alpha$  induction, contacts with enhancers further downstream are lost. Furthermore, most of the interactions amongst enhancers and CTCF-bound sites are not as stable as might be assumed. However, since many enhancer sites overlap CTCF-bound sites, the precise involvement of such elements in the TNF $\alpha$  response was distinguished by focusing only on iT2C interactions emanating from standalone CTCF sites (Fig. 45). Strong CTCF-based contacts correlated well with CTCF ChIP-seq data. Interestingly, interactions involving the last enhancer element and the promoter region were detected in unstimulated cells and were lost after TNF- $\alpha$  stimulation. Similar findings were reported for CTCF-mediated contacts between the middle and last enhancer segments. It is important to note that the same interaction dynamics were previously observed for p65-bound contacts (Fig. 44).

In summary, we confirmed that both p65-bound and -unbound chromatin interactions are dynamically remodeled after only 30 min of cytokine signaling. Furthermore, pre-established interactions associated with CTCF sites were lost after TNF $\alpha$  treatment, confirming that iT2C efficiently captured chromatin interaction changes.



**Figure 45 | iT2C interactions in the SAMD4A locus that are bound by CTCF upon TNF- $\alpha$  stimulation.** iT2C was performed in IMR90s using Apol and oligo probes capturing interactions in a 2.8 Mbp locus around SAMD4A before (black) and after 30 minutes TNF- $\alpha$  treatment (magenta). Contacts associated with CTCF-binding are displayed. Pre-looping interactions unique to unstimulated cells are depicted with black, dashed rectangles. Interaction changes linked to TNF- $\alpha$  signaling are depicted with magenta, dashed rectangles. ENCODE ChIP-seq data were aligned below.



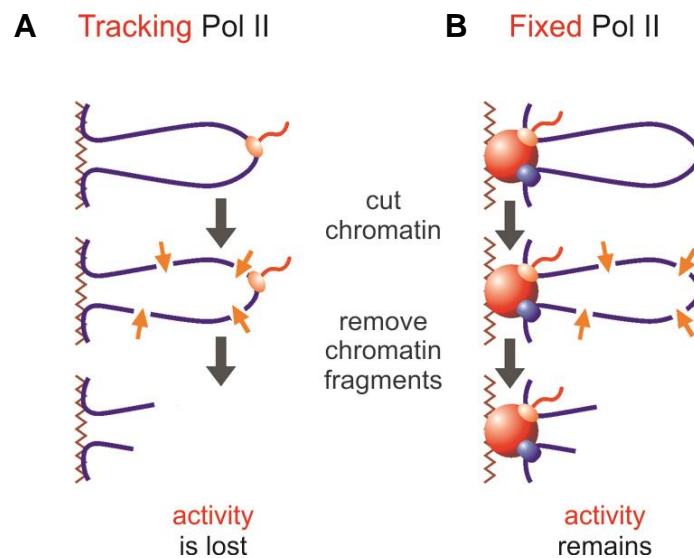
## 4. Discussion

The molecular composition and three-dimensional organization of the chromatin fiber inside the nucleus is achieved via complex mechanisms. The efforts of large consortia, like the ENCODE Project, have helped map the positions of DNA-binding proteins and histone modifications along the linear chromatin fiber, thus allowing for the characterization of regulatory and functional elements in the human and other model-organism genomes (ENCODE Project Consortium, 2012). Over the past two decades scientists have invested tremendous efforts for interpreting such data and for dissecting how chromatin architecture controls gene expression via the spatial interactions of various genomic elements; hence the development of a series of chromosome conformation capture-based approaches to probe chromatin folding at increasing resolution and throughput. Typically, these methods involve the cross-linking of cells with formaldehyde to capture loci that are in close physical proximity, followed by the pairwise determination of contact frequency between any two loci (Denker & de Laat, 2016). However, chemical cross-linking may introduce biases that compromise contact data evaluation, since *in vivo* effects of fixation efficiency and specificity remain obscure. Furthermore, conventional 3C technologies are not devoid of random ligation events between “bystander”, non-directly interacting fragments, which convolute the identification of meaningful chromatin interactions (Dekker *et al.*, 2006). To overcome the limitations that arise from harsh manipulation of chromatin networks when using cross-linking agents, non-physiological buffers and detergents, we developed an intrinsic 3C method (i3C), which enables examination of the native organization of mammalian genomes (Brant *et al.*, 2016).

### 4.1 Features of i3C-based techniques

First evidence for the association of transcription sites with the nucleoskeleton was shown in a decisive experiment that implemented cell permeabilization in a modified ‘physiological buffer’ (PB) with subsequent nuclease treatment and electro-elution to remove most of the chromatin from the nucleus. Strikingly, active RNA polymerases remained attached to sub-nuclear structures and generated nascent RNA, which was revealed via nuclear ‘run-on’ experiments (Fig. 46; Jackson & Cook, 1985). These findings did not only substantiate the “transcription factory” model, which involves immobilized RNA polymerases at transcription factories that transcribe DNA by

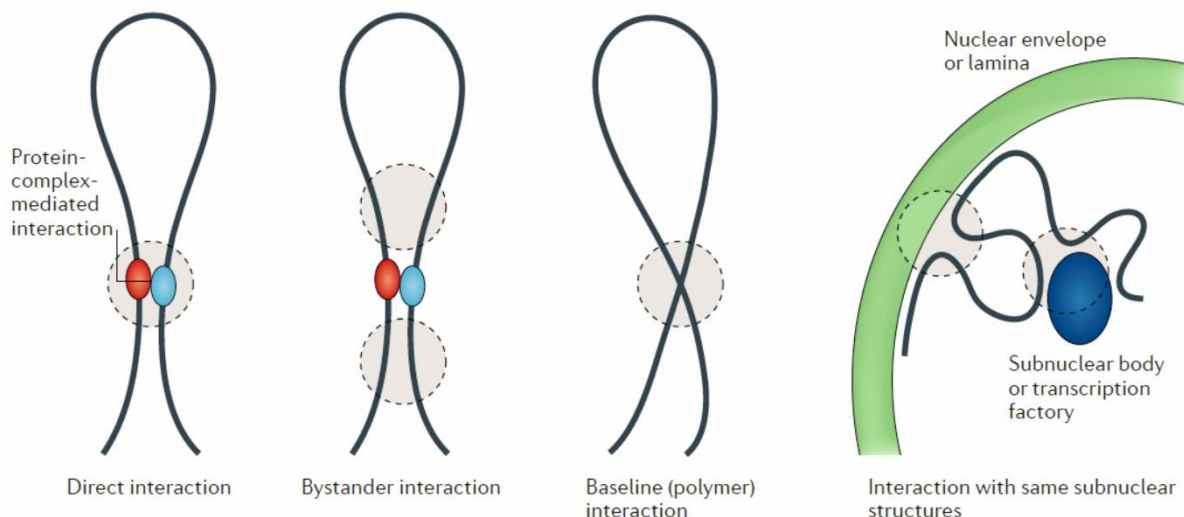
reeling it in rather than tracking along the chromatin fiber, but also highlight the relevance of a nuclear core cage comprised of structural entities (e.g., transcription factories, Polycomb bodies, LADs, nucleoli) that contribute to cellular function.



**Figure 46 | Two models of transcription and the evidence that active RNA polymerase are immobilized at transcription factories. A |** The model describes RNA polymerases (orange oval) that move along chromatin to transcribe DNA. After chromatin treatment with nucleases and electro-elution of cut fragments from the nucleus, essentially all activity is lost. **B |** The model shows active RNA polymerases (orange oval) that are fixed on transcription factories and transcribe DNA by reeling in the long molecule and extruding the nascent transcript. After chromatin treatment with nucleases and electro-elution of cut fragments from the nucleus, essentially all activity is retained (adapted from Papantonis & Cook, 2013).

In the case of i3C, we exploited the attributes of the ‘physiological buffer’ that retains most transcriptional activity and thus preserves nuclear chromatin structure (Jackson *et al.*, 1981). This permits the study of native interactions by retaining DNA molecules in close spatial proximity without the need for chemical cross-linking. Hence, nuclear isolation and *in situ* chromatin digestion, as well as *in situ* ligation, can successfully be performed in PB under non-disruptive conditions for chromatin and nuclear architecture. Consistent with the pioneering studies by Jackson & Cook, i3C involves elimination of a substantial amount of chromatin fragments that are not attached to subnuclear entities after endonuclease restriction, thus enriching for structures bound to the nucleoskeleton and to nuclear bodies (e.g., transcription factories, Polycomb bodies, LADs). Perhaps surprisingly, i3C results highlight the preservation of inherent spatial proximity, despite the loss of almost 50% of the cell’s DNA content.

Despite major differences in chromatin treatment, including cross-linking, harsh detergent utilization, extensive incubation under non-physiological temperatures, and mechanical rotation during the implementation of conventional 3C, both native and cross-linked templates recorded similar chromatin interactions, reproducing, for example, the whole range of previously reported contacts formed by the *EDN1* TSS (Diermeier *et al.*, 2014). On the other hand, loss of non-attached chromatin fragments during i3C yielded higher signal-to-noise ratios, and i3C contacts were enriched more focally at relevant interaction sites (e.g. enhancer elements, CTCF sites, Polycomb histone marks), compared to conventional 3C data. These findings are consistent with former observations of bystander ligation products resulting from random ligation events within the cross-linked and rigid chromatin network of conventionally-treated nuclei (Fig. 47; Dekker *et al.*, 2013).

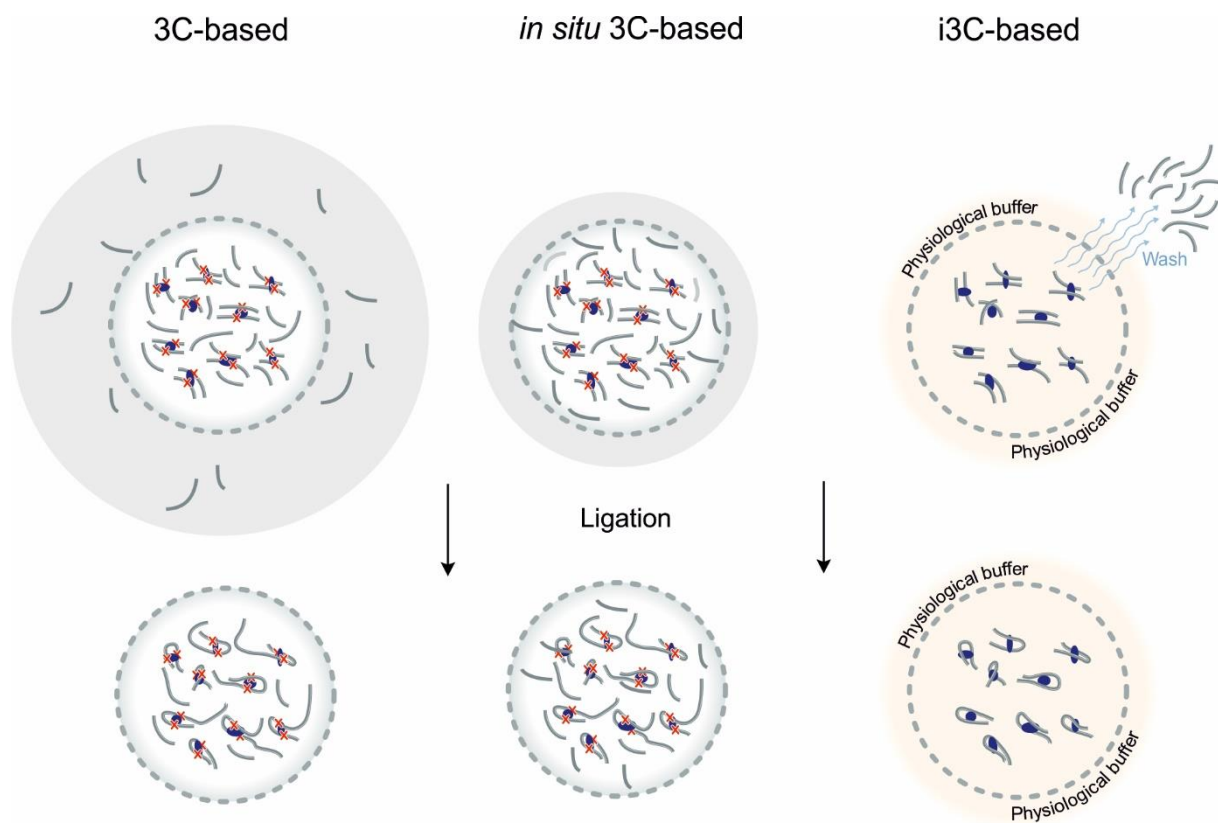


**Figure 47 | Looping events of the chromatin fiber.** 3C detects random contacts resulting from bystander ligation and cross-linking of the baseline polymer fiber. Moreover, efficiency of cross-linking at heterochromatic sub-nuclear structures such as LADs remains obscure (Dekker *et al.*, 2013).

Hence, conventional 3C-based assays report contact frequencies between loci that are harshly cross-linked, and it can be difficult to distinguish putatively functional from non-functional interactions. In contrast, nuclei handled in PB harbor flexible chromatin fibers; hence i3C provides insights into 3D chromatin looping, which might be missed when using standard approaches.

Another key characteristic of intrinsic i3C-based assays is the ligation within intact nuclei, in a small reaction volume, whereas standard C-technologies include a ligation step under highly diluted conditions to achieve ligation of diffusing cross-

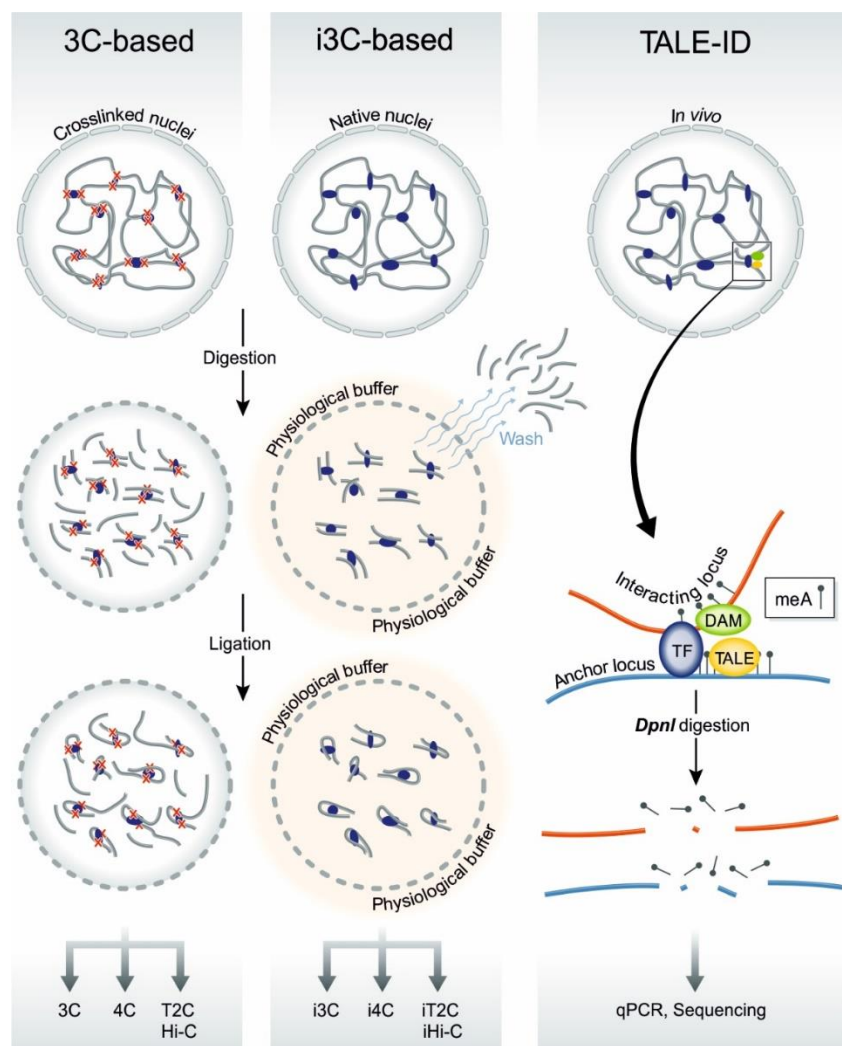
linked segments (Fig. 48). In 2014, Rao *et al.* developed an alternative Hi-C approach, the '*in situ* Hi-C', which aimed at diminishing background signal by performing ligation within intact nuclei in a small volume (Fig. 48, Rao *et al.*, 2014). Moreover, experiments from the same study included interaction maps from nuclei that were not cross-linked but embedded in agar plugs in order to maintain chromatin architecture during the ensuing digestion and ligation procedures. In theory, this *in situ* ligation step of chromatin according to Rao *et al.* will still involve random fragments retained in the sample, thus, actually increasing background noise in the data. This was confirmed in our results and is in agreement with standard 3C protocols (Dekker *et al.*, 2002). Furthermore, avoiding cell cross-linking but implementing harsh manipulation of chromatin, as described by Rao *et al.*, destabilizes the chromatin network and results in interaction loss (Fig. 26).



**Figure 48 | Comparison of ligation steps during chromosome conformation capture approaches.** Diluted ligation according to Dekker *et al.*, (left) and *in situ* ligation according to Rao *et al.*, (middle) of cross-linked nuclei are shown. Intrinsic ligation of non-crosslinked nuclei in PB (right). It is important to note that all methods perform *in situ* ligation (adapted from Rowley & Corces, 2016).

Long-range chromatin interactions that were identified by 3C-derived technologies require verification with an independent method such as RNA/DNA FISH. However,

discrepancies regarding contact detection between microscopy studies and 5C assays in the *HoxD* locus were recently reported (Williamson *et al.*, 2014), which might stem from differential formaldehyde cross-linking efficiency. Since FISH also requires harsh cell cross-linking, we developed a novel orthogonal approach called 'TALE-iD' (Fig. 49), which allows detection of native chromatin interactions by combining locus-specific targeting via TAL-effector binding domains with the activity of an adenosine methylase (Dam; Brant *et al.*, 2016). A similar approach was introduced after TALE-iD, which enables capture of chromatin interactions by biotinylated dCas9, supporting the role of targeted *in situ* chromatin capture assays (Liu *et al.*, 2017). However, RNA-guided Cas9 nucleases display unspecific off-targeting and a DNA recognition pattern (Hsu *et al.* 2013) that is less precise than that of TALEs. Thus, TALE fusions with Dam may serve as an effective tool to analyze three-dimensional chromatin interactions under native conditions, and with high specificity.



**Figure 49 | Comparison of methods to study chromatin structure.** Conventional 3C-based (left) and intrinsic i3C-based approaches (middle) are schematically illustrated. A strategic overview of TALE-iD is shown (right; adapted from Rowley & Corces, 2016).

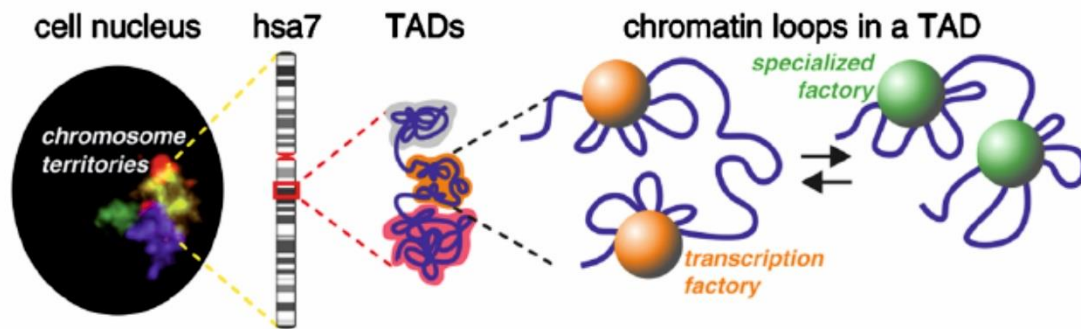
Moreover, the robust and time-saving i3C protocol can be performed within <1.5 days with limited hands-on time, while the conventional 3C procedure requires ~3 days before a 3C template is obtained. Also, we were able to perform i3C on cells of different types using fairly low cell amounts, and achieved high contact resolution (pending on the choice of restriction endonuclease). These key features render our intrinsic approach applicable for a variety of research questions.

#### **4.2 Role of i3C-based assays in to date genome organization knowledge**

In the past two decades, 3C-derived techniques substantially advanced our understanding of the genome's structure-to-function relationship. Basic 3C studies have contributed to major findings for gene regulation, like long-range interactions between the locus control region (LCR) and distal  $\beta$ -globin genes in a tissue specific manner (Tolhuis *et al.*, 2002). However, this approach is prone to inaccuracies related to varying amplification efficiencies of the designed primer pairs and was soon followed by more systematic and high-throughput methods. Using the locus-specific 4C approach, genome organization during cell lineage commitment was, for instance, characterized by identifying a specialized and widespread long-range interaction network that is engaged with key pluripotency factors (Apostolou *et al.*, 2013; Wei *et al.* 2013; de Wit *et al.*, 2013). The ensuing 5C and Hi-C assays allowed looping investigation in a genome-wide manner and contributed to the discovery of key structural entities such as topological associated domains (TADs), as well as the segregation of the genome into active and inactive A/B compartments (Dostie *et al.*, 2006; Lieberman-Aiden *et al.*, 2009; Dixon *et al.*, 2012). Moreover, ChIA-PET unraveled chromatin interactions that are bound by specific proteins such as transcription factors, highlighting the strong link between chromatin looping and transcriptional regulation (Fullwood *et al.*, 2009; Li *et al.*, 2012). More recently, contact profiles obtained from single cells using conventional genome-wide capture methods revealed extensive cell-to-cell heterogeneity of three-dimensional chromatin structure, emphasizing the importance of future studies on individual cells (Nagano *et al.*, 2013; Flyamer *et al.* 2017). Hence, conventional 3C-based techniques

constitute a core of tools to examine genome organization, but still begging the question, what can we learn from our intrinsic i3C assay?

To address this question, we mainly applied i3C-derived approaches on the *SAMD4A* model locus on human chromosome 14. Previous studies by our group have characterized *SAMD4A* as a suitable gene for studying long-range interactions upon cytokine signaling (Papantonis *et al.*, 2012). Our results revealed significantly more focal i4C interaction profiles that enable enhanced precision in the determination of interacting loci, especially when an interacting region is densely populated by *cis*-regulatory elements (e.g., the *SAMD4A* super-enhancer). These observations were supported by the identification of significant interactions with ‘fourSig’ analysis (Williams *et al.*, 2014), which additionally highlighted the fidelity in replicating of i4C contact maps (while conventional 4C replicates displayed stronger variations among each other). Moreover, we detected intrachromosomal interactions that reside within the constraints of TAD boundaries and are associated preferentially with regulatory elements such as promoters, enhancers, or CTCF-bound sites, suggesting that these typically long-range *cis*-contacts are functional and not simply predicted by the closest distance to each other. Indeed, previous studies have shown that three-dimensional interactions between *cis*-regulatory elements are part of a complex network, which plays a key role in transcription regulation (Noordermeer & de Laat, 2008; Vise *et al.*, 2009). Notably, we recorded native *cis*-contacts that are associated with the transcription factor NF $\kappa$ B following TNF $\alpha$  signaling, the residence time of which on chromatin is only few milliseconds, and its identification typically requires extensive cross-linking (Nowak *et al.*, 2005). Our intrinsic assay also identified pre-established chromatin looping between TNF-responsive promoters and their cognate enhancers, presumably at the surface of subnuclear structures such as “transcription factories” (Fig. 50). These findings suggest that immobilization of paused and active RNA polymerases on transcription factories does shape the 3D genome (Cook, 2010; Melnik *et al.*, 2011; Ferrai *et al.*, 2010). Contrary to previous reports (Jin *et al.*, 2013), intrinsic, prelooped chromatin interactions were partially remodeled following cytokine signaling, rendering i3C assays a suitable instrument to study temporally-resolved contact dynamics.



**Figure 50 | Hierarchical layers of genome organization.** Chromosomes occupy distinct territories, which can intermingle in the interphase nucleus. TADs harbor intra-domain chromatin loops that are associated with transcription factories (orange sphere). Upon extracellular signaling or cell differentiation, chromatin loops are partially remodeled in a specialized factory (adapted from Brant & Papantonis, 2015).

In another variation tested, the interactions of the *SAMD4A* TSS were fully lost upon treatment with the RNase A nuclease to remove RNA from the “nuclear matrix”. The term describes a structural framework that hosts many cellular functions such as DNA replication, RNA synthesis, and splicing (Berezney & Coffey, 1974). Apparently, the disruption of nuclear RNA leads to a breakdown of the nuclear matrix thus disrupting chromatin architecture via the formation of chromatin clumps (Nickerson et al., 1989). These results of ours highlight a stabilizing role of RNA in 3D chromatin organization.

Taken together, our investigations using i3C-based techniques underline the preservation of a replicable, core nuclear scaffold that retains structural (e.g., TADs, LADs, Polycomb-based) and functional features (e.g., transcriptional activity) of mammalian genome organization. However, the principles that drive chromatin folding still remain elusive. To address these concerns, we focused on two main observations that were made using i3C, such as chromatin-interaction dynamics and insulation at TAD boundaries upon extracellular signaling.

#### **4.3 Static and dynamic chromatin: what forces shape the genome?**

The NF $\kappa$ B signaling pathway plays a key role in inflammatory responses, as well as in organismal aging and cancer (Tilstra *et al.*, 2011). Inflammatory and oxidative stress stimulates the translocation of the transcription factor into the nucleus, where it regulates the expression of genes encoding cytokines, growth factors, and factors



associated with apoptosis, senescence, and inflammation (Adler *et al.*, 2007). To understand how transcriptional regulation of these processes is linked to 3D genome structure, we treated primary human cells with short pulses of tumor necrosis factor  $\alpha$ , and utilized i3C- and 3C-based approaches to study chromatin dynamics. In agreement with previous studies, we were able to show that TNF $\alpha$ -responsive enhancers also natively interact with their target promoters prior to extracellular signaling (Jin *et al.*, 2013). These observations suggest that an overarching chromatin architecture is established in every cell type, which regulates transcriptional activation by selectively controlling the communication of enhancers with their target genes. Enhancer sites engaged in prelooping become bound by NF- $\kappa$ B, pointing out that distant enhancers loop to promoters before gene induction to facilitate transcriptional activation by key TNF $\alpha$ -responsive transcription factors.

Nonetheless, we observed that some promoter-enhancer interactions were reshuffled in response to TNF $\alpha$  signaling. Observed signal differences were more obvious when intrinsic i4C- and iT2C assays were applied, indicating that cross-linking of chromatin can impair precise discrimination of fine-tuned chromatin dynamics. Interestingly, many prefolded promoter-enhancer contacts also involve CTCF-bound elements, hinting towards a role of architectural proteins to shape the basic 3D chromatin network (Phillips-Cremins *et al.*, 2013).

Other key mediators of pre-established genome topology are paused RNA polymerases that tether promoter-enhancer elements to the respective transcription factory prior to gene induction (Ghavi-Helm *et al.*, 2014). Once released from the transcription start site, active RNA polymerases transcribe long TNF $\alpha$ -responsive genes with a speed of  $\sim$ 3 kbp/min, and produce a wave of nascent transcripts deployed in a temporal-dependent manner (Wada *et al.*, 2009). This means that one complete transcription cycle of *SAMD4A* takes  $\sim$ 75 minutes. Consistent with these findings, we reported that some pre-existing looping-interactions changed according to time-point at which they were investigated. A 30-min signaling pulse resulted in contact loss mainly between the *SAMD4A* promoter and enhancer elements close to the gene terminus, because at that time point the RNA polymerase is engaged in the first third of the gene's length (Figs 43-45). On the other hand, a 60-min TNF $\alpha$  pulse revealed chromatin dynamics mostly between the promoter and enhancer elements within the first intron, while interactions between the promoter and the enhancer sites

further downstream remained stable (Figs 38&39). These results confirm that the RNA polymerase functions as a major tie that determines the basic chromatin architecture by pre-looping genomic elements to target gene promoters in order to ensure efficient activation of transcription, while also driving chromatin dynamics upon transcriptional induction in a temporal-specific manner. Nonetheless, some pre-looped chromatin contacts do not change upon extracellular signaling, suggesting either parallel associations of more than one RNA polymerases on one target gene (Ghavi-Helm *et al.*, 2014), or just that these responses are not synchronized and biallelic, but i3C/3C methods always provide us with the average interactions profiles of both alleles in the cell population (and not just the active ones).

#### ***4.4 Topologically-associated domains: are they key mediators of chromatin folding?***

The newly-developed ChromEMT method allowed visualization of the ultrastructure of chromatin in the nucleus by a combination of electron microscopy tomography with DNA labeling (Ou *et al.*, 2017). This tool uncovered that the chromatin fiber is a flexible and mostly disordered chain in the range of 5 to 24 nm in diameter, and differentially arranged into subnuclear structures. These novel findings contrast conventional 3C studies and highlight that cells fixation creates a rigid chromatin network, which increases the density of chromatin fibers and produces compact chromatin structures that might not accurately represent the profiles of chromatin folding. Moreover, recent reports have shown that individual chromatin clusters vary from cell to cell, and only by averaging many cell contact maps, typical features of chromatin organization such as TADs appear (Flyamer *et al.* 2017). These findings lead to the question: do TADs really exist *in vivo*?

Locus-specific assays, like conventional 4C and intrinsic i4C, produce robust chromatin interaction profiles for selected sites and uncover structural and functional genomic loop formations (e.g., TADs). Nonetheless, the obtained contact profiles represent chromatin interactions from an average of cells. To overcome limitations for quantitative examination of contact frequencies, we applied our intrinsic chromosome conformation capture approach to PCR-free and genome-wide technologies such as Hi-C and the targeted T2C variant (Lieberman-Aiden *et al.*, 2009; Kolovos *et al.*, 2014). iT2C and iHi-C contact maps outline conformations that resemble TADs and sub-TADs from conventional assays. However, native contact maps are markedly

depleted for interaction signal within these domains when compared to genome-wide profiles obtained from cross-linked cells. Moreover, individual chromatin loops also span across TAD boundaries. Still, intrinsic contacts correlate well with genomic elements such as promoter, enhancer, and CTCF-bound sites, whereas the majority of conventionally-detected signals seem to be of a non-specific nature. One explanation for these findings is that only a few strong interactions exist in the interior of domains, while most other associations are natively weak and random and cell cross-linking might artificially intensify these – hence reveal the globule-like structures or TADs in the cell population data. The borders of TADs are frequently associated with binding of CTCF, which functions as an insulator between neighboring domains. Recently, CTCF-depletion studies revealed that disruption of TAD insulation does not really impact higher-order compartmentalization of chromatin, thus suggesting that mechanisms other than TAD assembly drive large-scale folding (Nora et al., 2017). Furthermore, a flexible and disordered chromatin fiber structure (like that observed using ChromEMT), fits well with our genome-wide native chromatin interaction data. Then, intrinsic 3C-based approaches provide strong hints on the true nature of genome organization in the nucleus, which now opens additional questions that will need to be investigated in further experiments.

#### ***4.5 Future perspectives***

Pioneering studies unraveled a tight link between the 3D chromosome architecture in the nucleus and genome function. However, an outstanding question is, what are the structural principles that determine folding and function of the genome in interphase- and in mitotic cells? The work presented here provides alternative and time-saving tools that allow investigation of chromosome conformation architecture under native conditions, without chemical and mechanical manipulations of the chromatin network.

While native methods generate focal and replicable contact signals between genomic loci, conventionally-obtained interaction maps display elevated levels of background ligations. Moreover, i3C data showed that the vast majority of intra-chromosomal interactions are indeed insulated by TAD boundaries, but weakly compared to when conventional methods are applied to identify TAD restrictions. Recent ChromEMT and Hi-C studies on individual mammalian nuclei, directly question the relevance of TAD assembly in higher-order genome organization (Flyamer et al. 2017; Ou et al., 2017). Future work involving iHi-C experiments in single nuclei will help unravel

genome conformation under native conditions, and examine significance of individual loops previously detected by population-based iT2C and iHi-C assays. Furthermore, the ligation step of iHi-C in single nuclei will avoid the previous need for biotinylation of cohesive DNA fragments, and hence reduce any biases arising from the pull-down of ligation products. Instead, incorporation of specific barcodes using the activity of the Tn5 transposase (Picelli *et al.*, 2014) will allow generation of contact maps in individual nuclei. Moreover, a combination of these aforementioned experiments with targeted protein degradation in cells (see Nora *et al.*, 2017), will shed light on the role of CTCF and other architectural proteins (such as cohesin or HMGB2) in genome folding.

Still, the principles that govern higher-order chromatin organization remain obscure. Are RNA polymerase major mediators of chromatin clustering at discrete nucleoplasmic foci? Consistent with already published data, we observed that pre-established chromatin looping partially persists after transcriptional activation, but distinct chromatin dynamics could be recorded via i3C in a temporal-specific manner. This suggests that RNA polymerases function as molecular ties shaping both basic and dynamic chromatin folding. This hypothesis can be tested with iHi-C experiments coupled to immunoprecipitation of RNA polymerase II, p65, or CTCF before and after cytokine signaling. Moreover, CRISPR-based knock-down of enhancer elements in our model *SAMD4A* locus will allow examination of the relationship between transcriptional elongation and chromatin looping. Finally, multiple rosette-like structures of chromatin loops are associated with subnuclear entities – thus, i3C templates should contain several fragments ligated together. However, 3C assays to date only investigate pairwise interactions, and may this miss significant structural and functional chromatin information. Future work should incorporate “multi-i3C” experiments that take advantage of latest generation “long-read” sequencing to find multi-contact interaction hubs in the genome.

Finally, using fusions of specific TALE DNA-binding proteins to fluorophores can be introduced into living cells in physiological buffer and combined with live-cell imaging to investigate chromatin dynamics upon extracellular signaling in real time *in vivo*.

In summary, native technologies serve rich scope in getting closer to the true organization of genomes in living cells.

## References

- Adler AS, Sinha S, Kawahara TL, Zhang JY, Segal E, Chang HY. Motif module map reveals enforcement of aging by continual NF-kappaB activity. *Genes Dev.* 2007 Dec 15;21(24):3244-57. Epub 2007 Nov 30.
- Apostolou E, Ferrari F, Walsh RM, Bar-Nur O, Stadtfeld M, Cheloufi S, Stuart HT, Polo JM, Ohsumi TK, Borowsky ML, Kharchenko PV, Park PJ, Hochedlinger K. Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell.* 2013 Jun 6;12(6):699-712.
- Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, Ryan S, Spiller DG, Unitt JF, Broomhead DS, Kell DB, Rand DA, Sée V, White MR. Pulsatile stimulation determines timing and specificity of NF-kappaB dependent transcription. *Science.* 2009 Apr 10;324(5924):242-6.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell.* 2007 May 18;129(4):823-37.
- Beneke S, Meyer K, Holtz A, Hüttner K, Bürkle A. Chromatin composition is changed by poly(ADP-ribosylation) during chromatin immunoprecipitation. *PLoS One.* 2012;7(3):e32914.
- Berezney R, Coffey DS. Identification of a nuclear protein matrix. *Biochem Biophys Res Commun.* 1974 Oct 23;60(4):1410-7.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science.* 2009 Dec 11;326(5959):1509-12.
- Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. *Science.* 2011 Sep 30;333(6051):1843-6.
- Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Müller S, Eils R, Cremer C, Speicher MR, Cremer T. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* 2005May;3(5):e157.
- Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet.* 2001 Feb 1;10(3):211-9.
- Brackley CA, Brown JM, Waithe D, Babbs C, Davies J, Hughes JR, Buckle VJ, Marenduzzo D. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol.* 2016 Mar 31;17:59.

- Brant L, Georgomanolis T, Nikolic M, Brackley CA, Kolovos P, van Ijcken W, Grosveld FG, Marenduzzo D, Papantonis A. Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. *Mol Syst Biol*. 2016 Dec 9;12(12):891.
- Brant L, Papantonis A. Contribution of 3D Chromatin Architecture to the Maintenance of Pluripotency. *Curr. Stem Cell Reports*, Sep. 2015, Volume 1, Issue 3, pp 170–175
- Brown KE, Guest SS, Smale ST, Hahm K, Merckenschlager M, Fisher AG. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*. 1997 Dec 12;91(6):845-54.
- Caudron-Herger M, Cook PR, Rippe K, Papantonis A. Dissecting the nascent human transcriptome by analysing the RNA content of transcription factories. *Nucleic Acids Res*. 2015 Aug 18;43(14):e95.
- Cook PR, Brazell IA. Supercoils in human DNA. *J Cell Sci*. 1975 Nov;19(2):261-79.
- Cook PR. A general method for preparing intact nuclear DNA. *EMBO J*. 1984 Aug;3(8):1837-42.
- Cook PR. A model for all genomes: the role of transcription factories. *J Mol Biol*. 2010 Jan 8;395(1):1-10.
- Cremer T, Cremer C, Schneider T, Baumann H, Hens L, Kirsch-Volders M. Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum Genet*. 1982;62(3):201-9.
- Cremer T, Cremer M, Dietzel S, Müller S, Solovei I, Fakan S. Chromosome territories--a functional nuclear landscape. *Curr Opin Cell Biol*. 2006 Jun;18(3):307-16.
- Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*. 1999 Jun 14;145(6):1119-31.
- Cullen KE, Kladde MP, Seyfred MA. Interaction between transcription regulatory regions of prolactin chromatin. *Science*. 1993 Jul 9;261(5118):203-6.
- de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*. 2013 Oct 24;502(7472):499-506.
- de Wit E, Bouwman BA, Zhu Y, Klous P, Splinter E, Verstegen MJ, Krijger PH, Festuccia N, Nora EP, Welling M, Heard E, Geijsen N, Poot RA, Chambers I, de Laat W. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*. 2013 Sep 12;501(7466):227-31.
- Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet*. 2013 Jun;14(6):390-403.

- Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002 Feb 15;295(5558):1306-11.
- Dekker J. The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods*. 2006 Jan;3(1):17-21.
- Denker A, de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev*. 2016 Jun 15;30(12):1357-82.
- Diermeier S, Kolovos P, Heizinger L, Schwartz U, Georgomanolis T, Zirkel A, Wedemann G, Grosveld F, Knoch TA, Merkl R, Cook PR, Längst G, Papantonis A. TNF $\alpha$  signalling primes chromatin for NF- $\kappa$ B binding and induces rapid and widespread nucleosome repositioning. *Genome Biol*. 2014 Dec 3;15(12):536.
- Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, Diao Y, Liang J, Zhao H, Lobanenkov VV, Ecker JR, Thomson JA, Ren B. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015 Feb 19;518(7539):331-6.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012 Apr 11;485(7398):376-80.
- Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*. 2006 Oct;16(10):1299-309.
- Ea V, Baudement MO, Lesne A, Forné T. Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. *Genes (Basel)*. 2015 Jul 27;6(3):734-50.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6;489(7414):57-74.
- Entrevan M, Schuettengruber B, Cavalli G. Regulation of Genome Architecture and Function by Polycomb Proteins. *Trends Cell Biol*. 2016 Jul;26(7):511-525.
- Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*. 2012 Feb 28;9(3):215-6.
- Ferrai C, Xie SQ, Luraghi P, Munari D, Ramirez F, Branco MR, Pombo A, Crippa MP. Poised transcription factories prime silent uPA gene prior to activation. *PLoS Biol*. 2010 Jan;8(1):e1000270.
- Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, Razin SV, Mirny LA, Tachibana-Konwalski K. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. 2017 Apr 6;544(7648):110-114.
- Fox CH, Johnson FB, Whiting J, Roller PP. Formaldehyde fixation. *J Histochem Cytochem*. 1985 Aug;33(8):845-53.

- Francis NJ, Kingston RE, Woodcock CL. Chromatin compaction by a polycomb group protein complex. *Science*. 2004 Nov 26;306(5701):1574-7.
- Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, Barbieri M, Moore BL, Kraemer DC, Aitken S, Xie SQ, Morris KJ, Itoh M, Kawaji H, Jaeger I, Hayashizaki Y, Carninci P, Forrest AR; FANTOM Consortium, Semple CA, Dostie J, Pombo A, Nicodemi M. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol*. 2015 Dec 23;11(12):852.
- Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, et al., An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*. 2009 Nov 5;462(7269):58-64.
- Gavrilov AA, Gushchanskaya ES, Strelkova O, Zhironkina O, Kireev II, Iarovaia OV, Razin SV. Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res*. 2013 Apr 1;41(6):3563-75.
- Geeven G, Zhu Y, Kim BJ, Bartholdy BA, Yang SM, Macfarlan TS, Gifford WD, Pfaff SL, Verstegen MJ, Pinto H, Vermunt MW, Creighton MP, Wijchers PJ, Stamatoyannopoulos JA, Skoultchi AI, de Laat W. Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. *Genome Biol*. 2015 Dec 23;16:289.
- Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*. 2014 Aug 7;512(7512):96-100.
- Guelen L, Pagie L, Brassat E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 2008 Jun 12;453(7197):948-51.
- Guillot PV, Xie SQ, Hollinshead M, Pombo A. Fixation-induced redistribution of hyperphosphorylated RNA polymerase II in the nucleus of human cells. *Exp Cell Res*. 2004 May 1;295(2):460-8.
- Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes JA, Birney E, Hardison RC, Dunham I, Kellis M, Noble WS. Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res*. 2013 Jan;41(2):827-41.
- Hsieh TH, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*. 2015 Jul 2;162(1):108-19.
- Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013 Sep;31(9):827-32.



Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* 2013 Sep;31(9):827-32.

Jackson DA, Cook PR. Transcription occurs at a nucleoskeleton. *EMBO J.* 1985 Apr;4(4):919-25.

Jackson DA, McCready SJ, Cook PR. RNA is synthesized at the nuclear cage. *Nature.* 1981 Aug 6;292(5823):552-5.

Jackson DA, Yuan J, Cook PR. A gentle method for preparing cyto- and nucleoskeletons and associated chromatin. *J Cell Sci.* 1988 Jul;90 ( Pt 3):365-78.

Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature.* 2013 Nov 14;503(7475):290-4.

Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol.* 2011 Dec 25;30(1):90-8.

Kasinathan S, Orsi GA, Zentner GE, Ahmad K, Henikoff S. High-resolution mapping of transcription factor binding sites on native chromatin. *Nat Methods.* 2014 Feb;11(2):203-9.

Kind J, Pagie L, Ortazokoyun H, Boyle S, de Vries SS, Janssen H, Amendola M, Nolen LD, Bickmore WA, van Steensel B. Single-cell dynamics of genome-nuclear lamina interactions. *Cell.* 2013 Mar 28;153(1):178-92.

Klein FA, Pakozdi T, Anders S, Ghavi-Helm Y, Furlong EE, Huber W. FourCSeq: analysis of 4C sequencing data. *Bioinformatics.* 2015 Oct 1;31(19):3085-91.

Kolovos P, van de Werken HJ, Kepper N, Zuin J, Brouwer RW, Kockx CE, Wendt KS, van IJcken WF, Grosveld F, Knoch TA. Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. *Epigenetics Chromatin.* 2014 Jun 16;7:10.

Kumarasinghe MP, Constantine SR, Hemamali RL. Methanol as an alternative fixative for cytological smears. *Malays J Pathol.* 1997 Dec;19(2):137-40.

Lander ES et al.;International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature.* 2001 Feb 15;409(6822):860-921.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009 Oct 9;326(5950):289-93.

- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010 Mar 1;26(5):589-95.
- Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, Cao H, Ni M, Liu Y, Gu Z, Dickerson KE, Xie S, Hon GC, Xuan Z, Zhang MQ, Shao Z, Xu J. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*. 2017 Aug 24;170(5):1028-1043.e19.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-(\Delta\Delta C(T))}$  Method. *Methods*. 2001 Dec;25(4):402-8.
- Martin P, McGovern A, Orozco G, Duffus K, Yarwood A, Schoenfelder S, Cooper NJ, Barton A, Wallace C, Fraser P, Worthington J, Eyre S. Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat Commun*. 2015 Nov 30;6:10069.
- Meaburn KJ, Misteli T. Cell biology: chromosome territories. *Nature*. 2007 Jan 25;445(7126):379-781.
- Melnik S, Deng B, Papantonis A, Baboo S, Carr IM, Cook PR. The proteomes of transcription factories containing RNA polymerases I, II or III. *Nat Methods*. 2011 Sep 25;8(11):963-8.
- Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE. Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol*. 2013 Dec;31(12):1133-6
- Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, Dekker J. Organization of the mitotic chromosome. *Science*. 2013 Nov 22;342(6161):948-53.
- Németh A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Péterfia B, Solovei I, Cremer T, Dopazo J, Längst G. Initial genomics of the human nucleolus. *PLoS Genet*. 2010 Mar 26;6(3):e1000889.
- Nickerson JA, Krochmalnic G, Wan KM, Penman S. Chromatin architecture and nuclear RNA. *Proc Natl Acad Sci U S A*. 1989 Jan;86(1):177-81.
- Noordermeer D, de Laat W. Joining the loops: beta-globin gene regulation. *IUBMB Life*. 2008 Dec;60(12):824-33.
- Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. 2017 May 18;169(5):930-944.e22.
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012 Apr 11;485(7398):381-5.

- Norris MG, Malys N. What is the true enzyme kinetics in the biological system? An investigation of macromolecular crowding effect upon enzyme kinetics of glucose-6-phosphate dehydrogenase. *Biochem Biophys Res Commun*. 2011 Feb 18;405(3):388-92.
- Nowak DE, Tian B, Brasier AR. Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques*. 2005 Nov;39(5):715-25.
- Orlando V, Strutt H, Paro R. Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods*. 1997 Feb;11(2):205-14.
- Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH, O'Shea CC. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*. 2017 Jul 28;357(6349).
- Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet*. 2003 Oct;35(2):190-4.
- Papantonis A, Cook PR. Transcription factories: genome organization and gene regulation. *Chem Rev*. 2013 Nov 13;113(11):8683-705.
- Papantonis A, Kohro T, Baboo S, Larkin JD, Deng B, Short P, Tsutsumi S, Taylor S, Kanki Y, Kobayashi M, Li G, Poh HM, Ruan X, Aburatani H, Ruan Y, Kodama T, Wada Y, Cook PR. TNF $\alpha$  signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J*. 2012 Nov 28;31(23):4404-14.
- Pederson T. The nucleolus. *Cold Spring Harb Perspect Biol*. 2011 Mar 1;3(3). pii: a000638.
- Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J, Corces VG. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. 2013 Jun 6;153(6):1281-95.
- Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res*. 2014 Dec;24(12):2033-40.
- Pombo A, Dillon N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol*. 2015 Apr;16(4):245-57.
- Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26: 841–2
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014 Dec 18;159(7):1665-80.

- Ripper D, Schwarz H, Stierhof YD. Cryo-section immunolabelling of difficult to preserve specimens: advantages of cryofixation, freeze-substitution and rehydration. *Biol Cell*. 2008 Feb;100(2):109-23.
- Rowley MJ, Corces VG. Capturing native interactions: intrinsic methods to study chromatin conformation. *Mol Syst Biol*. 2016 Dec 9;12(12):897.
- Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012 Sep 6;489(7414):109-13.
- Sato T, Suyama M. ChromContact: A web tool for analyzing spatial contact of chromosomes from Hi-C data. *BMC Genomics*. 2015 Dec 15;16:1060.
- Schermelleh L, Heintzmann R, Leonhardt H. A guide to super-resolution fluorescence microscopy. *J Cell Biol*. 2010 Jul 26;190(2):165-75.
- Schmiedeberg L, Skene P, Deaton A, Bird A. A temporal threshold for formaldehyde crosslinking and fixation. *PLoS One*. 2009;4(2):e4636.
- Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre BM, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, Dimitrova E, Dimond A, Edelman LB, Elderkin S, Tabbada K, Darbo E, Andrews S, Herman B, Higgs A, LeProust E, Osborne CS, Mitchell JA, Luscombe NM, Fraser P. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res*. 2015 Apr;25(4):582-97.
- Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS, Eskiw CH, Luo Y, Wei CL, Ruan Y, Bieker JJ, Fraser P. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*. 2010 Jan;42(1):53-61.
- Servant N, Lajoie BR, Nora EP, Giorgetti L, Chen CJ, Heard E, Dekker J, Barillot E. HiTC: exploration of high-throughput 'C' experiments. *Bioinformatics*. 2012 Nov 1;28(21):2843-4.
- Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*. 2012 Feb 3;148(3):458-72.
- Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*. 2006 Nov;38(11):1348-54.
- Smale ST. Selective transcription in response to an inflammatory stimulus. *Cell*. 2010 Mar 19;140(6):833-44.
- Sofueva S, Yaffe E, Chan WC, Georgopoulou D, Vietri Rudan M, Mira-Bontenbal H, Pollard SM, Schroth GP, Tanay A, Hadjur S. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J*. 2013 Dec 11;32(24):3119-29.

- Solomon MJ, Varshavsky A. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci U S A*. 1985 Oct;82(19):6470-4.
- Spector DL, Lamond AI. Nuclear speckles. *Cold Spring Harb Perspect Biol*. 2011 Feb 1;3(2).
- Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, Palstra RJ, Wendt KS, Grosveld F, van Ijcken W, Soler E. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc*. 2013 Mar;8(3):509-24.
- Teytelman L, Thurtle DM, Rine J, van Oudenaarden A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A*. 2013 Nov 12;110(46):18602-7.
- Tilstra JS, Clouston CL, Niedernhofer LJ, Robbins PD. NF- $\kappa$ B in Aging and Disease. *Aging Dis*. 2011 Dec;2(6):449-65.
- Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*. 2002 Dec;10(6):1453-65.
- van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, Verstegen MJ, de Wit E, Tanay A, de Laat W. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods*. 2012 Oct;9(10):969-72.
- van Steensel B, Henikoff S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol*. 2000 Apr;18(4):424-8.
- Visel A, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA. Functional autonomy of distant-acting human enhancers. *Genomics*. 2009 Jun;93(6):509-13.
- Vogel MJ, Peric-Hupkes D, van Steensel B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc*. 2007;2(6):1467-78.
- Wada Y, Ohta Y, Xu M, Tsutsumi S, Minami T, Inoue K, Komura D, Kitakami J, Oshida N, Papantonis A, Izumi A, Kobayashi M, Meguro H, Kanki Y, Mimura I, Yamamoto K, Mataka C, Hamakubo T, Shirahige K, Aburatani H, Kimura H, Kodama T, Cook PR, Ihara S. A wave of nascent transcription on activated human genes. *Proc Natl Acad Sci U S A*. 2009 Oct 27;106(43):18357-61.
- Wei Z, Gao F, Kim S, Yang H, Lyu J, An W, Wang K, Lu W. Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell*. 2013 Jul 3;13(1):36-47.
- Wijgerde M, Grosveld F, Fraser P. Transcription complex stability and chromatin dynamics in vivo. *Nature*. 1995 Sep 21;377(6546):209-13.

Williams RL Jr, Starmer J, Mugford JW, Calabrese JM, Mieczkowski P, Yee D, Magnuson T. fourSig: a method for determining chromosomal interactions in 4C-Seq data. *Nucleic Acids Res.* 2014 Apr;42(8):e68.

Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, Dostie J, Bickmore WA. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev.* 2014 Dec 15;28(24):2778-91.

Yamaguchi Y, Wada T, Watanabe D, Takagi T, Hasegawa J, Handa H. Structure and function of the human transcription elongation factor DSIF. *J Biol Chem.* 1999 Mar 19;274(12):8085-92.

Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IJcken WF, Grosveld FG, Ren B, Wendt KS. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A.* 2014 Jan 21;111(3):996-1001.

## Figure Index

Figure 2  Chromatin looping.	13
Figure 2   Overview of 3C-derived methods.	14
Figure 3   Examples of 3C, 4C, T2C and HiC interaction profiles.	16
Figure 4   Hierarchical genome organization.	19
Figure 5   Cross-linking biases in 3C technologies.	21
Figure 6   Transcriptional activity is preserved in PB.	23
Figure 7   TALE-iD plasmid.	37
Figure 8   Strategy for 4C primer design.	45
Figure 9   Features of the i3C approach.	50
Figure 10   Genomic coverage of different i3C fractions.	52
Figure 11   Chromatin features of 'lost' and 'retained' i3C fractions.	53
Figure 12   Comparison of i3C and conventional 3C DNA templates.	55
Figure 13   i3C-qPCR implemented in the extended <i>EDN1</i> locus on chr 6.	56
Figure 14   i4C – and 4C – seq were implemented on the <i>SAMD4A</i> locus.	58
Figure 15   Reproducibility of i4C interaction profiles.	59
Figure 16   Differential analysis of i4C and conventional 4C data.	61
Figure 17   Chromatin features of cis-contacts obtained from i4C- and 4C-seq profiles.	62
Figure 18   Comparison of background noise in i4C and conventional 4C.	63
Figure 19   i4C – and 4C implemented in a different cell line.	65
Figure 20   Comparison of i4C profiles using different restriction enzymes.	66
Figure 21   i4C-seq implemented in the gene-poor <i>EDN1</i> locus.	68
Figure 22   i4C-seq implemented in the heterochromatinized <i>TBX5</i> locus.	69
Figure 23   Native interactions are restricted to TAD boundaries and describe prelooping.	70
Figure 24   <i>trans</i> -interactions captured by i4C in the Sox2 and Nanog loci.	72
Figure 25   i4C-seq implemented using decreasing numbers of isolated nuclei.	73
Figure 26   i4C-seq performed under different conditions.	75
Figure 27   Factory-i4C implemented in HUVECs.	76
Figure 28   TALE-iD verifies native looping interactions at the <i>ZFPM2</i> locus.	78
Figure 29   Expanded view on chromatin organization on chromosome 14 by iT2C and conventional T2C.	81

Figure 30   Comparison of TAD boundaries in iT2C and conv. T2C maps.	82
Figure 31   Comparison of interaction sites captured by iT2C or conv. T2C.	83
Figure 32   Recapitulation of iT2C and conventional T2C heat maps in IMR90s.	85
Figure 33   Proof-of-principle: iHiC generated in HUVECs.	87
Figure 34   A closer look on iHiC features.	88
Figure 35   iHiC generated in the 'lost' or 'retained' fractions of HUVECs.	90
Figure 36   Improvement of iHiC signal enrichment in mESCs.	92
Figure 37   Detection of pre-looping in a TNF- $\alpha$ responsive locus in HUVECs.	94
Figure 38   Native interactions are restricted to TAD boundaries and describe prelooping.	96
Figure 39   Conventional 4C interactions upon TNF- $\alpha$ stimulation.	97
Figure 40   Differential analysis of i4C data upon TNF- $\alpha$ stimulation.	98
Figure 41   Differential analysis of conventional 4C data upon TNF- $\alpha$ stimulation.	99
Figure 42   TNF $\alpha$ stimulation partially remodels prelooping within the <i>SAMD4A</i> TAD.	100
Figure 43   iT2C interactions in the <i>SAMD4A</i> locus upon TNF- $\alpha$ stimulation.	101
Figure 44   iT2C interactions in the <i>SAMD4A</i> locus that are bound by p65 upon TNF- $\alpha$ stimulation.	102
Figure 45   iT2C interactions in the <i>SAMD4A</i> locus that are bound by CTCF upon TNF- $\alpha$ stimulation.	104
Figure 46   Two models of transcription and the evidence that active RNA polymerase are immobilized at transcription factories.	106
Figure 47   Looping events of the chromatin fiber.	107
Figure 48   Comparison of ligation steps during chromosome conformation capture approaches.	108
Figure 49   Comparison of methods to study chromatin structure.	109
Figure 50   Hierarchical layers of genome organization.	112



## Table Index

Table 1: Chemicals	25
Table 2: Buffer and solutions	27
Table 3: Enzymes	29
Table 4: Kits and systems	30
Table 5: General equipment	30
Table 6: General reagents used for molecular biology approaches	31
Table 7: Antibiotics used for cloning experiments	32
Table 8: 3C-qPCR primers	32
Table 9: 4C primers	33
Table 10: TALE-iD cloning primers	34
Table 11: Bacterial strains used for cloning experiments	34
Table 12: List of bacterial artificial chromosomes	35
Table 13: List of plasmids used for TALE-iD cloning	35
Table 14: List of mammalian cell lines and corresponding culture media	37
Table 15: qRT-PCR cycling conditions	41

## List of Abbreviations

µg	microgram
µl	microliter
2D	two-dimensional
3'	three prime-end of DNA sequences
3C	chromosome conformation capture
3D	three-dimensional
4C	circularized chromosome conformation capture
5'	five prime-end of DNA sequences
5C	carbon copy chromosome conformation capture
bp	base pair
BrdU	5-bromo-2`desoxyuridine
ChIP	chromatin immunoprecipitation
CT	chromosome territory
CTCF	CCCTC-factor
Dam	DNA-methyltransferase
DHS	DNase I hypersensitive site
DNA	deoxy-ribonucleic acid
e. g.	for example
for	forward
h	hour
Hi-C	higher-order chromosome conformation capture
HMM	Hidden-Markov-Model
i.p.	intraperitoneal
LAD	lamina associated domain
LCR	locus control region
M	molar
Min	minute
NAD	nucleolus associated domain
NFκB	nuclear factor kappa B
ng	nanogram
P	p-value
P65	unit of nuclear factor kappa B

PB	physiological buffer
PCR	polymerase chain reaction
Pol II	RNA polymerase II
rev	reverse
RNA	ribonucleic acid
rpm	rounds per minute or reads per million
RT	reverse transcriptase/room temperature
SD	standard deviation
SDS	sodium-dodecyl-sulfate
sec	second
Seq	sequencing
T2C	targeted conformation capture
TAD	topological associated domain
TALE	transcription activator-like effector
TNF $\alpha$	tumor necrosis factor alpha
TSS	transcription start site
U	unit
UV	ultra-violet
V	volt
wt	wild-type
$\Delta$	deleted / mutated allele

## Acknowledgement

First of all, I would like to thank Akis Papantonis for giving me the opportunity to work on this interesting project and the excellent supervision during my PhD. He is an outstanding group leader who inspired me with his patience, continuous encouragement and support during difficult times. His imperturbable passion for science motivates me to follow my career path with the same dedication.

I would like to thank Prof. Dr. Björn Schumacher, Prof. Dr. Peter Nürnberg and Prof. Dr. Siegfried Roth for agreeing to build my thesis committee.

I would like to thank Theodore Georgomanolis, Milos Nikolic, Petros Kolovos and Davide Marenduzzo for supporting my work with extensive bioinformatic analysis.

Moreover, I would like to thank all lab members and the Rada-Iglesias group for helpful discussions, continuous support and the pleasant time during and outside lab work.

I would like to thank our lunch group for making every day enjoyable. Especially I would like to thank Sara who always encouraged me in every aspect of my life and became a good friend.

Ich danke meinen Freunden Elena, Federica, Lara, Isabella und Ayse für die mentale Unterstützung in all den Jahren. Mit eurer Geduld und Motivation konnte ich alle Hürden während der Zeit als Doktorandin meistern. Auch wenn ich meistens im Labor versackt bin und selten Zeit hatte, wusste ich, dass ihr immer für mich da seid.

Desweiteren danke ich meinem kleinen Bruder für seine unbeschwehrliche und positive Art, welche mich immer wieder aufs Neue erstaunt und inspiriert. Ich danke auch meiner großen und lauten Familie für die Ablenkung vom stressigen Alltag. Ihr habt mich in guten sowie schlechten Zeiten aufgefangen und immer an mich geglaubt.

Mein ganz besonderer Dank gilt Maxim für seine Geduld mit meinen Launen, wenn mal ein Versuch nicht funktioniert hat. Dennoch hast du mich motiviert und aufgemuntert. Deine Unterstützung und Liebe haben mir geholfen diese Arbeit fertigzustellen.

Дорогие мама и папа,

без вас я бы никогда не стала тем, кем я являюсь сегодня. Вы приехали в Германию, не побоявшись трудностей, чтобы подарить нам с Альфредом лучшую жизнь. Все, чего я добилась в жизни, стало возможно только благодаря вам. Я горжусь, что у меня такие родители, и надеюсь, вы тоже можете мною гордиться. Спасибо вам за все. Я вас люблю.

## Eidesstattliche Erklärung

Ich versichere hiermit, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Prof. Dr. Björn Schumacher betreut worden.

Köln im September, 2017

Lilija Brant

Nachfolgende Teilpublikationen liegen vor:

Brant L, Georgomanolis T, Nikolic M, Brackley CA, Kolovos P, van Ijcken W, Grosveld FG, Marenduzzo D, Papantonis A. Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. *Mol Syst Biol.* 2016 Dec 9;12(12):891.

# Curriculum vitae

## Persönliche Daten

---

Name: Lilija Brant  
Adresse: Furtwänglerstraße 8  
40724 Hilden  
Email: lilija.brant@uk-koeln.de  
Geburtsort: Malowodnoje, Kasachstan  
Geburtstag: 19.05.1988  
Nationalität: Deutsch

## Akademischer Werdegang

---

09/2013 bis 09/2017 Promotion am Zentrum für Molekulare Medizin Köln (ZMMK)  
(unter der Leitung von Dr. Argyris Papantonis)

Mitglied in der Graduiertenschule "Graduate School for Biological Sciences" der Universität zu Köln

10/2010 – 10/2012 Master Program Biological Science  
Universität zu Köln

Masterarbeit am Institut für Mikrobiologie  
Ludwig-Maximilians-Universität München  
(Prof. Dr. Marc Bramkamp)  
Abschluss: Master of Science

10/2007 – 10/2010 Bachelor Biologie  
Universität zu Köln

Bachelorarbeit am Institut für Biochemie  
Universität zu Köln  
(Prof. Dr. Waffenschmidt)  
Abschluss: Bachelor of Science

## Publikationen

---

### 2017

Zirkel A\*, Nikolic M\*, Sofiadis K\*, Mallm JP, **Brant L**, Becker C, Altmueller J, Franzen J, Koker M, Gusmao EG, Costa IG, Ullrich RT, Wagner W, Nuernberg P, Rippe K, Papantonis A. Topological demarcation by HMGB2 is disrupted early upon senescence entry across cell types and induces CTCF clustering. (submitted).

Debès C\*, Grönke S\*, Karalay Ö\*, Tain L, Nakamura S, Hahn O, Zirkel A, Sofiadis K, **Brant L**, Wollnik B, Kubacki T, Schermer B, Müller RU, Benzing T, Papantonis A, Antebi A, Partridge L, Beyer A. Ageing-associated changes in transcription and splicing efficiency influence metazoan longevity. (submitted).

## 2016

**Brant L**, Georgomanolis T, Nikolic M, Brackley CA, Kolovos P, van Ijcken W, Grosveld FG, Marenduzzo D, Papantonis A. Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. *Mol Syst Biol*. 2016 Dec 9;12(12):891.

Kolovos P, Georgomanolis T, Koeflerle A, Larkin JD, **Brant L**, Nikolic M, Gusmao EG, Zirkel A, Knoch TA, van Ijcken WF, Cook PR, Costa IG, Grosveld FG, Papantonis A. Binding of nuclear factor  $\kappa$ B to noncanonical consensus sites reveals its multimodal role during the early inflammatory response. *Genome Res*. 2016 Nov;26(11):1478-1489.

Melnik S, Caudron-Herger M, **Brant L**, Carr IM, Rippe K, Cook PR, Papantonis A. Isolation of the protein and RNA content of active sites of transcription from mammalian cells. *Nat Protoc*. 2016 Mar;11(3):553-65.

## 2015

**Brant L**, Papantonis A. Contribution of 3D Chromatin Architecture to the Maintenance of Pluripotency. *Curr. Stem Cell Reports*, Sep. 2015, Volume 1, Issue 3, pp 170–175