Article 25fa pilot End User Agreement

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication other than authorised under this licence or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library
Radboud University
Copyright Information Point
PO Box 9100
6500 HA Nijmegen

You will be contacted as soon as possible.

Purification and enrichment of specific chromatin loci

Mathilde Gauchier¹, Guido van Mierlo², Michiel Vermeulen²✉ and Jérôme Déjardin¹✉

Understanding how chromatin is regulated is essential to fully grasp genome biology, and establishing the locus-specific protein composition is a major step toward this goal. Here we explain why the isolation and analysis of a specific chromatin segment are technically challenging, independently of the method. We then describe the published strategies and discuss their advantages and limitations. We conclude by discussing why significant technology developments are required to unambiguously describe the composition of small single loci.

"Chromosomes are the largest and least well understood of cellular organelles. The challenge lies in a combination of complex biochemistry and inhomogeneous structure." This statement, made by Roger Kornberg more than ten years ago¹, reflects the longstanding interest and effort directed at understanding the regulation of the eukaryotic genome. The genome is in fact assembled into a structure termed chromatin, and chromatin composition is highly heterogeneous depending on the cell type and the cell-cycle stage. Importantly, chromatin composition is mostly locus specific and determines that locus' function. Establishing the locus-specific composition of chromatin represents a major technical challenge, and this remains an outstanding issue. Indeed, knowing the components of a system composition is a major step toward understanding how it works.

Currently, genome biology studies are based on mapping chromatin proteins and/or epigenetic marks and correlating their presence with genome activities, such as transcription, replication and repair. Combined with next-generation sequencing technologies, chromatin immunoprecipitation (ChIP) has been instrumental in advancing our knowledge of chromatin biology. However, this approach is limited by the requirement for prior knowledge of the factor to assay and is dependent on the availability and specificity of antibodies². Improvements in ChIP and in mass spectrometry (MS) techniques have also made it possible to explore the interactome of a factor in the context of chromatin and to identify novel regulators^{3–8}. However, these approaches are not suitable for describing the composition of a specific chromatin locus—by which, in the context of this work, we mean any specific genomic region, of varying size, primarily defined by its biological function: for example, a 'promoter' promotes RNA transcription of a gene or a 'telomere' protects a chromosome end.

DNA affinity pulldown is a well-known approach to identify factors binding to a specific DNA sequence⁹. Moreover, in-vitro-prepared mono- and polynucleosomes and naked DNA with specific modifications have been used for affinity pulldown in combination with MS^{10–13} to identify proteins associated with epigenetic marks. However, in vivo regulation processes, which involve specific sequences and a large and often only partially known array of specific histone and DNA modifications, are very difficult to recapitulate in vitro.

Before 2010, only a few approaches were available for isolating and analyzing specific loci as chromatin^{14–17}. With the popularization of proteomics, novel strategies based on MS were developed to comprehensively identify proteins that interact with a locus of interest^{18–23}. Although initially only relatively large parts of the genome (i.e., repetitive loci) could be quantitatively isolated to sufficient purity, a series of new approaches have been established for determining the proteome of small single-copy loci from mammalian genomes. In this critical review of currently published strategies, we first describe the biochemical challenges in obtaining a specific locus of chromatin as a pure entity for downstream MS analyses. Then, we discuss the published approaches and highlight their advantages and limitations.

The biochemical challenges

Regardless of the strategy, locus-specific chromatin purification requires a few common steps: (i) the preparation of soluble chromatin that retains as many relevant molecular interactions as possible; (ii) a specific and efficient method to capture the locus of interest; and, importantly, (iii) an enrichment factor in proportion with the relative abundance of the locus in the genome.

Chromatin preparation and stoichiometry of chromatin components. Chromatin is a heterogeneous mixture of nucleic acids and proteins, each engaged in dynamic complexes with different affinities for noncovalent interactions and with distinct stoichiometry. Moreover, the pH and ionic strength requirements to maintain these interactions and solubility in vitro vary hugely among individual chromatin components²⁴. Therefore, the ideal milieu to preserve chromatin integrity and complexity in the native state is practically unattainable in a test tube. Yet compromises can be made, and several classical native chromatin preparation protocols exist²⁴ that can be used for functional studies. However, the tradeoff necessary to maintain solubility by using chelating compounds, salts and/or detergents inevitably leads to the dissociation of important interactors or to the induction of composition rearrangements. To limit these effects, chromatin can be stabilized by covalent chemical crosslinking before extraction—usually with formaldehyde²⁵ (Box 1)—to maintain as many interactions as possible. Another key parameter is the absolute abundance of the

¹Institute of Human Genetics CNRS-Université de Montpellier UMR 9002, Montpellier, France. ²Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, Nijmegen, the Netherlands.

✉e-mail: michiel.vermeulen@science.ru.nl; jerome.dejardin@igh.cnrs.fr

Box 1 | Native chromatin preparations compared to formaldehyde crosslinking

Crosslinking is the process of joining two or more molecules by covalent bonds using chemicals or other physical means, such as exposure to UV light. Many crosslinkers are commercially available, and these vary in solubility, spacer arm length, reactivity, functional groups, reversibility and price. The chemical reaction of formaldehyde (HCHO) with amino and imino groups is commonly used to stabilize protein–DNA complexes and analyze chromatin composition. Indeed, this inexpensive reagent has critical advantages: it is cell membrane permeable, and its small size allows it to diffuse rapidly inside biological materials. Its small size also permits the crosslinking of amino acids and nucleic acids, which requires very close proximity (≤ 2 Å). It is commonly believed that formaldehyde generates nonspecific crosslinks, but we have found no experimental evidence for this in the literature. Formaldehyde is the most popular chemical used in ChIP, immunofluorescence and chromatin conformation capture approaches, and to the best of our knowledge, it does not generate artifactual signals. Similarly, the results of pioneering chromatin studies argue against nonspecific crosslinking because crosslinks could not be detected between bovine serum albumin (BSA) and DNA after incubation of DNA with highly concentrated BSA (50 mg/ml), which does not normally bind to DNA, in 1% formaldehyde for >4 days⁵⁷. This suggests that efficient crosslinking requires stable interactions. However, this also represents a limitation of formaldehyde, because true, but transient, *in vivo* interactors might not efficiently be retained on chromatin, as has been convincingly shown for some MeCP2 mutants binding to pericentromeres⁵⁸.

Formaldehyde also readily generates protein–protein crosslinks, which results in the recovery of proteins indirectly bound to specific regions through other proteins²⁵. This forms the rationale for chromosome conformation capture approaches. This feature of formaldehyde might also lead to the nonspecific crosslinking of abundant proteins to specific loci. Therefore, any ‘unexpected’ protein association to a given locus requires validation by orthogonal methods, and could explain the presence of some noncanonical proteins in locus-specific purifications. Moreover, some proteins might not be crosslinked by formaldehyde, and thus will be lost during preparation. This issue might be addressed by the use of other crosslinking agents, which can help stabilize certain transcription factors on chromatin⁵⁹. Furthermore, formaldehyde can affect epitope recognition by antibodies that are used for ChIP, and this may influence the yield. Finally, enzymatic treatments of chromatin after crosslinking are often inhibited or less efficient, preventing functional analyses. For these reasons, several research groups in the chromatin field have developed native chromatin purification approaches. Native conditions allow very efficient immunoprecipitation, a major advantage for ChIP experiments, which can then be quantitatively analyzed without PCR amplification steps⁶⁰. However, native chromatin preparation requires enzymatic digestion, and consequently some chromatin fractions are poorly solubilized, thus preventing their analysis. Moreover, native preparations are not suitable for proteins that are not tightly bound to chromatin and are susceptible to generating more non-specific chromatin rearrangements.

chromatin locus to be purified because this will determine the amount of required starting material. The current detection limit for peptide identification by standard MS in a practical setting is in the low femtomole (10^{-15} mol) range (Box 2). This represents a minimum of 600 million molecules. Assuming a (theoretical) purification yield of 100%, the isolation of a specific single-copy locus requires half a billion diploid cells as input material to identify the most abundant components at that locus (often structural components, such as histones). In practice, however, the amount of starting material should be carefully considered in light of the end goal of locus-specific proteome characterization because regulatory components (such as enzymes) are usually several orders of magnitude less abundant than structural components. Therefore, their identification might require greater amounts of starting material. For example, at pericentromeric regions, the SUV39H histone methyltransferase is two orders of magnitude less abundant than nucleosomes²⁶. In nonsynchronized cells, the abundance of the telomerase enzyme, which is crucial for telomere maintenance, is several orders of magnitude lower on telomeres than are the subunits of the most abundant protein complex associated with telomeres, the shelterin complex. Consequently, substantially more starting material is required to identify telomerase in telomere chromatin preparations compared to shelterin subunits.

Locus-specific capture. From a biochemical point of view, locus-specific purification represents a tremendous technical challenge partly because a specific locus often shares components (such as histones) with other loci. This strongly limits the efficiency of classical separation approaches that rely on differential biophysical separation criteria (for example, ionic strength, size or solubility). Moreover, as a small and unique locus represents only a minute fraction of the entire genome, the chosen strategy must not only discriminate, but also sufficiently enrich, that locus from a gigantic amount of similar entities that tend to copurify. The challenge

is comparable to that of finding a specific straw, not a needle, in a haystack. Few targets are uniquely bound by specific protein factors and are abundant enough to be isolated using a single immunoaffinity for a protein specific to the target^{27,28}. However, in most cases, the uniqueness of a chromatin locus is embodied by its DNA sequence and its distinctive combination of bound factors. Therefore, sequence-specific hybridization capture has been the strategy of choice. Alternatively, a target locus can be modified by adding exogenous sequences bound by unique adapters, as we will also detail later.

The enrichment factor: the key signal-to-noise ratio parameter. In ChIP, chromatin components are mapped to specific genomic regions. As mapping is the endpoint in this analysis, background signal is unimportant as long as the enrichment factor is sufficient to discriminate between specific binding to a unique region and background binding to a non-relevant region of the same size. This is usually assessed by quantitative PCR. In a ChIP-sequencing experiment, only a minority of the tens of millions of sequencing reads are informative and contribute to the genuine signals (‘peaks’), without major impact on mapping. Locus-specific isolation methods have an entirely different objective. Indeed, the enrichment factor of the target locus should be calculated relative to the rest of the genome and not relative to another non-targeted region of similar size, because in these approaches, all non-targeted regions (i.e., the rest of the genome) and not just one control region contribute to the background (Figs. 1 and 2a). This is crucial for the success of the experiment and for high-confidence MS identification of locally bound factors.

The relative abundance of the target (i.e., the target size relative to the genome size) drastically affects the purification requirements. For instance, mouse pericentromeric repeats (comprising ~3% of the genome) can be captured to levels above background with strategies that enrich this locus at least 17-fold (i.e., 51% of the purified

Box 2 | The MS challenges

Although the theoretical detection limit for peptide identification of a digested purified protein is in the zeptomole (10^{-21}) to attomole (10^{-18}) range^{61,62}, the current detection limit for peptide identification from a complex peptide mixture by standard MS in a practical setting is in the low femtomole (10^{-15} mol) range. As this requires a minimum of 600 million molecules, it is critical to achieve high enrichment of a single-locus target while avoiding sample loss. Although adaptations of biochemical procedures will allow improving enrichment and purity, the inclusion of one or several recently optimized MS workflows might be required to obtain a comprehensive single-locus proteome. First, the data acquisition mode of modern mass spectrometers can be adjusted between 'conventional' data-dependent acquisition (DDA) and data-independent acquisition (DIA). Although DDA is extremely powerful, it relies on 'random' selection of peptides. In practice this means that peptides with the highest intensity are selected first, resulting in the overrepresentation of high-intensity peptides. Conversely, in DIA all peptides in a defined mass-to-charge window are selected and fragmented, thus allowing the identification of all peptides in a complex mixture⁶³. The superiority of DIA for measuring low-abundance proteins in a complex mixture has been recently demonstrated⁶⁴. This acquisition mode should be especially powerful for the analysis of samples that are dominated by highly abundant structural proteins, such as histones, in the case of single-locus purification. Second, new methods are currently

being developed for the online fractionation of peptide mixtures using ion separation in inert gas under the influence of an electrical field. These approaches, collectively referred to as ion mobility spectrometry (IMS), exist in several versions depending on the IMS instrumentation. IMS can be broadly divided in two groups: drift-tube ion mobility spectrometry (DTIMS) and high-field asymmetric-waveform ion-mobility spectrometry (FAIMS). These methods are used to reduce peptide coelution and cofragmentation, and they can significantly improve the accuracy and comprehensiveness of multiplexed proteomic analyses⁶⁵. Recently, it was shown that FAIMS implementation in the MS workflow results in higher sensitivity and increases the quantifiable peptides from a complex mixture compared to experiments without FAIMS^{66,67}. Finally, as single-locus purification experiments will yield only limited amount of material for liquid chromatography–mass spectrometry (LC-MS) analysis, even when using billions of input cells, it will be critical to reduce sample loss before and during MS measurements. This could be achieved by performing (part of) the experiment and/or sample preparation on a microfluidics device because this makes it possible to identify protein complexes from minute amounts of input material⁶⁸. During MS measurements, sample loss could be further minimized by spiking in labeled peptides that serve as carrier peptides. This approach has recently proven its value toward developing single-cell proteomics workflows that also deal with low amounts of input material⁶⁹.

material will be pericentromeric). Enriching 3% of the genome by 33-fold would lead to near-homogeneity of the purified preparation (i.e., 99% of the purified material would be pericentromeric). On the other hand, for a single-copy 3-kb promoter element (0.0001% of the human genome), a 1-million-fold enrichment would be required to reach near homogeneity (Fig. 2b).

Although an adequate enrichment factor depends on the size of the target site relative to the genome size and on the capture efficiency, noise depends on the capture specificity and on the stringency of the procedures used to eliminate off-target and non-specifically associated factors. An easy way to estimate target-locus enrichment is to assess its abundance relative to that of another non-target locus of similar size by quantitative PCR. In the starting material, this ratio should be equal to 1, while the value in the purified fraction represents a good estimate for enrichment.

In most cases, single-step affinity purification generates at best a 1,000-fold enrichment for unique proteins in solution. Therefore, if not combined with additional purification steps, the 1,000-fold enrichment of a human 3-kb locus will yield a proteome constituting 99.9% background proteins and 0.1% locus-specific factors. Nevertheless, even with such enrichment values, it is still possible to identify single-locus-specific factors, if they are significantly bound to the target and not elsewhere, but not to determine the comprehensive locus-specific composition.

Locus-specific isolation strategies

Direct hybridization of capture probes to chromatinized DNA. DNA sequence specificity can be exploited for chromatin isolation. Hybridization-based capture approaches are probably the most straightforward methods because they do not require genetic engineering. On the other hand, they always involve chemical crosslinking to maintain protein–DNA interactions because hybridization requires denaturation of the target dsDNA or other means to make the target sequence single stranded (such as exonuclease digestion).

Proteomics of isolated chromatin segments (PiCh) was the first reported method that combines locus purification and MS (Fig. 3a).

PiCh does not require either prior knowledge about the proteins that bind to the locus of interest or genetic engineering of target cells. PiCh is based on the hybridization of synthetic nucleic acid probes ('locked' nucleic acids as well as modified RNA) that are complementary to the locus of interest. These probes bind to DNA with very high specificity and stability and harbor the biotin analog desthiobiotin, which can be captured on streptavidin-conjugated beads and eluted with biotin. This allows the specific release of probe–chromatin hybrids in solution, while a significant amount of non-target chromatin remains bound to the streptavidin matrix. This considerably reduces the noise and significantly contributes to the enrichment value. After chromatin decrosslinking, proteins are identified by MS. PiCh was initially used to purify telomeric chromatin (0.01% of the human genome). As expected, the most abundant proteins were the six shelterin proteins, validating the approach. Moreover, PiCh revealed that the composition of telomeric chromatin is more complex than anticipated, and allowed the identification of novel factors and consequent characterization of novel regulation processes^{18,29–34}. PiCh has been combined with SILAC for quantitative analyses, for instance to measure changes in pericentromeric and telomeric regulation by heterochromatin enzymes^{35,36}. In this setup, PiCh allowed the identification and quantification of post-translational modifications of histones. At both targets, the ensemble of histone post-translational modifications (the code) was consistent with the proteome data (readers and writers/erasers), indicating that the method is suitable for studying chromatin regulation and for testing the histone code hypothesis. An important limitation of this method is the difficulty of obtaining PiCh probes from oligonucleotide companies. Moreover, PiCh capture probes only work on the 3' end of the target fragment, a constraint that has major implications for the preparation and fragmentation of soluble chromatin³⁷. In its current version, PiCh provides an enrichment factor of up to 10,000-fold. Therefore, it is not appropriate for studying the composition of single-copy small loci (~3 kb) from mammalian genomes. Indeed, as discussed before, only 1% of PiCh-purified proteins would be target specific in such a case.

Fig. 1 | The potential information gained from locus-specific research depends on the target sample enrichment. Indeed, establishing the local composition and identifying novel factors that uniquely bind to the locus of interest are two entirely different questions that require vastly different enrichment factors.

A similar strategy, hybridization capture of chromatin-associated proteins for proteomics (HyCCAPP), was later developed to characterize the composition of repetitive and single-copy loci in yeast. Depending on the target, HyCCAPP provides an enrichment of 80- to 800-fold relative to another non-target locus, making it suitable for characterizing repeated (or large) loci^{38,39}. HyCCAPP uses DNA probes that are easier to synthesize than PICh probes, a critical advantage when targeting large loci with complex sequences. However, DNA probes bind with lower stability, and this partly explains the lower target enrichment compared with PICh, although the comparison is difficult because the targeted loci are different. HyCCAPP was used to purify a single-copy ~1.4-kb locus from the yeast genome (~0.01% of the genome)²³. With an enrichment value of 1,000-fold, only 10% of the material (i.e., 0.01% × 1,000) in the purified fraction should be locus specific. Such a percentage might be sufficient to discover specific interactors if they are abundant at the target (signal) and not elsewhere (noise), but insufficient to enrich for proteins that are also abundant elsewhere (such as histones). Thus, in its current form, HyCCAPP is unlikely to be suitable to describe single-locus-specific protein composition and the local histone code regulation.

New, more efficient hybridization-based locus-specific chromatin purification approaches might be developed by taking advantage of DNA nanotechnologies that could help to improve the kinetics of assembly and stabilization of the hybridized probes. Multiplexed

HyCCAPP has been used to target several loci from the same starting material and to effect the orderly release of each target for downstream analysis⁴⁰. We are currently developing novel PICh capture probe designs and optimizing multistep enrichment protocols. Our efforts, in combination with improved MS instruments and pipelines, might allow the development of a robust method to establish the composition of single-copy targets from mammalian genomes.

Purification based on nuclease-dead Cas9. CRISPR-Cas9 technology enables precise targeting of regions of interest in the genome, and this has been exploited for identifying locus-specific proteins. In these approaches, hybridization to the target DNA sequence is also used for locus-specific isolation, but the probe, made of a 20-nucleotide-long guide RNA, is associated with a nuclease-dead Cas9 (dCas9) that can be functionalized with affinity tags (Fig. 3b). In the initial attempts using single-antibody-based affinity purification after binding of the guided dCas9 to the target, the genome-scale specificity and biological relevance of the interacting proteins were not evaluated^{22,41}. In fact, a single-locus guided dCas9 can bind to tens to hundreds of off-target sequences⁴², and this could lead to the copurification of relatively large amounts of irrelevant regions. This issue was recently addressed by using a biotinylatable dCas9 protein to purify specific chromatin loci in human cells⁴³. The method is named CAPTURE (CRISPR affinity purification in situ of regulatory elements) and involves the simultaneous expression in

Fig. 2 | Signal-to-noise ratio issues. **a**, Example of the purification of a telomere locus from the human genome. Telomere-binding proteins have distinct stoichiometry for telomere DNA and represent minute amounts of the total proteome. To establish the telomere composition, telomere-binding proteins must represent the major components after purification. **b**, Enrichment fold required for locus characterization according to locus relative abundance.

a given cell type of a specific guide RNA along with a dCas9 harboring a target peptide that can be biotinylated *in vivo* by the bacterial biotin ligase BirA (also coexpressed in these cells; see Fig. 3b). As a proof of principle, CAPTURE was first used to purify telomeres, and it allowed the identification of several telomere-specific binding proteins. This suggests that CAPTURE can sufficiently enrich telomeric chromatin. However, three major shelterin proteins, TRF1, TIN2 and TPP1, were not retrieved. As dCas9 is a large protein known to perturb chromatin regulation upon binding, it is possible that dCas9 binding along the telomere sequence might exclude some relevant proteins from telomeres *in vivo*. CAPTURE was then used to purify small single-copy loci from the human β -globin cluster, and the amount of captured off-target material was evaluated by ChIP-seq analysis. The specific enrichment of each single-copy targeted locus was relatively high, in the range of 10^3 , while that of the predicted off-targets was lower (10^0 to 10^2). Using label-free approaches, appropriate negative controls (including a cell line in which the target was removed by knockout) and multiple replicates, CAPTURE allowed the retrieval of several locus-specific proteins, demonstrating its usefulness.

More recently the Tjian laboratory developed a method called CLASP (Cas9-locus-associated proteome)⁴⁴. Unlike with *in situ* targeting methods, in CLASP affinity capture is performed *in vitro* using solubilized crosslinked chromatin on an immobilized dCas9 that binds to specific guide RNAs. This approach does not require genetic engineering, and it circumvents the issue of possible changes in the local regulation and endogenous composition upon dCas9 tethering *in vivo*. As a proof of principle, CLASP was first employed to retrieve the telomere proteome from human cells. CLASP successfully recovered several shelterin proteins; however, TRF1, TIN2 and POT1 were not retrieved, suggesting a bias in the method. Then, CLASP was used to purify the repetitive histone locus from *Drosophila* cells. The characterization of two novel histone gene regulators revealed that they were bound to the histone mRNA,

implying that CLASP might also target ribonucleoprotein complexes. Classical transcription factors and molecules involved in the main cell machineries (such as the replisome and spliceosome) were absent, suggesting that significant DNA-based processes can be missed with CLASP.

Purification based on the binding of sequence-specific proteins. If the locus of interest is abundant enough and is significantly bound by proteins not found elsewhere in the genome, these proteins can be crosslinked to chromatin, immunoprecipitated and analyzed by MS-based approaches (Fig. 3c). For instance, a telomere proteome was established using anti-shelterin antibodies^{27,28}. When the locus of interest is not uniquely bound by an endogenous protein, sequence-specific adaptor proteins can be designed and expressed to bind to the region of interest. Transcription-activator-like effector (TALE) domains can be designed to bind to specific DNA sequences⁴⁵. When tagged, the bound TALE proteins can be immunoprecipitated to enrich the associated chromatin segment (Fig. 3d). This strategy was used to enrich for the *GAL1* gene under native conditions in yeast^{19,46} and for telomeric chromatin in fixed cultured cells⁴⁷. However, in this last study, shelterin proteins were absent, suggesting that TALE binding might compete with the binding of biologically relevant telomere proteins *in vivo*. Another strategy involves significant genetic engineering of the target cells (Fig. 3e). The locus of interest is modified to harbor specific DNA sequences that can be tightly bound by adaptor proteins, such as Tet or LexA. After insertion of these specific sequences at the locus of interest and coexpression of the adaptor protein, the locus is enriched by immunoprecipitation. By introducing an orthogonal enrichment step that involved the specific excision of the target from the bulk chromatin in yeast, Kornberg's group was able to enrich the *PHO5* gene promoter by 100,000-fold⁴⁸. However, the use of native conditions might have led to loss of important interactors, as confirmed by MS analysis, in which only histones and few regulatory proteins were

Fig. 3 | Different approaches for locus-specific chromatin characterization. **a**, Oligonucleotide-based capture (PICCh and HyCCAPP). **b**, dCas9-based capture (CAPTURE, CLASP, CRISPR ChAP-MS and enCHIP). **c**, Immunoprecipitation of factors restricted to the locus of interest (QTIP). **d**, TALE-based capture. **e**, Capture methods that require editing the target sequence (genomic engineering). **f**, Proximity biotinylation (CasID, GloPro, C-BERST and CAPLOCUS); note that this does not address locus composition but rather is intended to discover locus-specific factors. **g**, Genomic strategy developed in yeast (Epi-Decoder).

recovered^{21,48}. The LexA affinity strategy was also used to investigate the proteins and histone post-translational modifications associated with the single-copy *GAL1* promoter region in yeast¹⁹ under conditions of transcription silencing or induction. With this method, the

1,000-bp target was enriched by about fivefold relative to a non-targeted region (*ACT1*), suggesting that *GAL1* represented only 0.04% of the purified material. The chicken β -globin HS4 insulator was targeted using the same approach⁴⁹.

Fig. 4 | Major limitations of the CRISPR-dCas9- and TALE-based targeting systems. **a**, dCas9 binds to nonspecific targets with a significant probability. **b**, As most of the expressed dCas9 is not bound to the target, when dCas9 is fused with BirA or APEX, the promiscuous biotinylation of neighboring factors generates a very high background. **c**, Chromatin conformation can modulate Cas9 binding efficiency. **d**, Simple-repeat multitargeting by dCas9 can compete with resident factors and change the endogenous composition.

The Tet system was used to enrich a single-copy transgene (the human fetal- γ -globin gene) in mice²⁰. To this aim, the Tet adaptor was fused with different tags, including a biotag that can be biotinylated by BirA. The method allowed the identification of 14 unique proteins, including regulatory proteins involved in γ -globin gene repression.

Proximity-labeling approaches. The BirA protein targets a lysine in a specific peptide sequence for biotinylation. However, after mutation into BirA*, its biotin ligase activity becomes promiscuous and it can biotinylate any lysine in the surrounding proteins. These biotinylated factors can then be captured with streptavidin and identified by MS. This feature was first used to develop a method named BioID to identify proteins in the vicinity of a given protein factor⁵⁰. Briefly, the protein of interest is fused with BirA* and expressed in cell lines that are cultured in the presence of biotin. Successful experiments using BioID largely depend on the distribution of BirA-tagged proteins during the biotinylation step, which can last up to several hours. The first chromatin-based method using BirA* was named CasID⁵¹. In CasID (Fig. 3f), BirA* is fused to dCas9, and then dCas9 is guided by specific guide RNAs to telomeres and pericentromeric regions of mouse cells. The great advantage of using promiscuous BirA* biotinylation is that there is no need for specific strategies to maintain relevant protein interactions in the chromatin preparations *in vitro*. Indeed, biotinylation around dCas9 occurs *in vivo*, and after cell lysis, biotinylated proteins are recovered in a single high-stringency affinity enrichment step. CasID successfully

recovered some shelterin proteins at telomeres⁵¹, but as with other dCas9-based methods described above, abundant telomeric factors, such as TRF1, RAP1 and POT1, were missing. Another approach, in which BirA* was fused to TRF1, was used successfully to analyze the proteome of TRF1 neighboring factors and retrieved both expected and novel telomere-binding proteins⁵².

Other proximity-labeling-based approaches have been developed more recently: dCas9–APEX2 biotinylation at genomic elements by restricted spatial tagging (C-BERST), CAPLOCUS and genomic locus proteomics (GLOPro). All three methods are based on dCas9 fused to the engineered ascorbate peroxidase APEX2, a protein that generates highly reactive biotin-phenoxyl radicals using biotin-phenol and hydrogen peroxide^{53,54,55}. These radical entities are short-lived and can interact with tyrosine residues in proteins present in a restricted 20-nm radius. APEX2 biotinylation is a very rapid and efficient process, and incubation times are significantly shorter (about a minute) than with BirA*-based methods. C-BERST and CAPLOCUS were used to establish the proteome of telomeres as a proof of principle. C-BERST was also successfully employed to identify alpha-satellite pericentromeric and centromeric proteins⁵⁴. GLOPro was used to identify proteins bound to single-copy promoter elements of the telomerase *TERT* and of *MYC*⁵⁵. Many factors not known to have canonical functions in transcription were identified, and some were validated by ChIP. The authors noted that expected binding proteins, such as canonical promoter factors (for example, general transcription factors), or even RNA polymerase II subunits, were not enriched in the analysis because they had

Table 1 | List of locus-specific purification methods

Principle	Name	Input (cell numbers)	Species	Target	Crosslinked / native	References	Comments
DNA-based hybridization	PICh	$\sim 10^9$	Mouse, human, <i>Drosophila</i>	Telomeres, rDNA, major satellites, minor satellites (unp.), retrotransposon (unp.)	Crosslinked	18,29,35–37	Does not affect the chromatin composition. Requires crosslinking. Restricted to repeats.
	HyCCAPP	3.3×10^9 to $>10^{11}$	Yeast, human	α satellites, rDNA, X element, <i>GAL1-10</i> , <i>ENO2</i> etc.	Crosslinked	23,38,40	Does not affect the composition of the chromatin. Requires crosslinking.
Immunopurification TRF	QTIP	5×10^7 to 10^8	Human	Telomeres	Crosslinked	27,28	Restricted to telomeres.
Proximity biotinylation	TRF1-BioID	6×10^7	Human	Telomeres	Native	52	Requires establishment of stable cell lines. Restricted to telomeres.
Genetic engineering: LexA	RS-TAP*	10^{11} to 1.5×10^{12}	Yeast	rDNA, <i>PHO5</i>	Native	21	First single-locus purification effort with the goal of analyzing chromatin structure.
	ChAP-MS	10^{11}	Yeast	rDNA, <i>GAL1</i>	Crosslinked	19	Requires establishment of stable cell lines.
	i-ChIP	5×10^7	Mouse	β -globin HS4 insulator	Crosslinked	49	Requires establishment of stable cell lines.
	i-ChIP	5×10^7	Chicken	<i>Pax5</i>	Crosslinked	49	Requires establishment of stable cell lines.
Genetic engineering: TetO	TChP	3×10^9	Mouse	γ -globin	Crosslinked	20	Requires establishment of stable mouse lines.
TALE system: immunoprecipitation	TAL-ChAP-MS	10^{11}	Yeast	<i>GAL1</i>	Native	46	Requires establishment of stable cell lines.
	enChIP	2×10^7	Mouse	Telomeres	Crosslinked	47	Requires establishment of stable cell lines.
CRISPR-dCas system: immunopurification	enChIP	6×10^7	Human	<i>IRF1</i>	Crosslinked	47	Requires establishment of stable cell lines.
	CRISPR-ChAP-MS	1×10^{11}	Yeast	<i>GAL1</i>	Crosslinked	22	Requires establishment of stable cell lines.
	CAPTURE	0.25×10^9 to 1×10^9	Human	β -globin cluster, telomeres	Crosslinked	43	Requires establishment of stable cell lines.
	CLASP (in vitro)	2×10^9	<i>Drosophila</i> , human	Histone cluster, telomeres	Crosslinked	44	Recombinant dCas9 used as a hook in vitro; does not require establishment of stable cell lines.
CRISPR-dCas system: proximity biotinylation	CasID	4×10^7	Mouse	Telomeres, major and minor satellites	Native	51	Requires establishment of stable cell lines.
	CRISPR/ Cas9 APEX GloPro	n.d.	Human	<i>hTERT</i> , <i>cMYC</i>	Native	55	Requires establishment of stable cell lines.
	C-BERST	6×10^7	Human	Telomere, α satellites	Native	54	Requires establishment of stable cell lines.
	CAPLOCUS	4×10^7	Human	Telomere	Native	53	Requires transient transfections.
TAG-ChIP-Barcode-Seq	Epi-Decoder	1.5×10^9	Yeast	Barcoded reporter gene at the <i>HO</i> locus	Crosslinked	56	Independent of MS but relies on high-throughput sequencing. Restricted to yeast.

unp., unpublished; n.d., not determined.

comparable abundance in the negative controls. This suggests that non-guided dCas9 or dCas9 associated with a negative-control guide RNA significantly biotinylates biologically relevant proteins

that are discarded in the analyses. One caveat with all proximity-labeling methods is that most of the expressed dCas9 molecules (hundreds, if not thousands, per cell) are not associated with the

Box 3 | Epi-Decoder: a genomic strategy to quantify proteins bound to a small single locus in yeast

An elegant genomic method named Epi-Decoder (Fig. 3g), based on systematic protein tagging, transgene barcoding and ChIP analysis, was recently developed in yeast to identify factors bound to a single-copy reporter gene³⁶. Unlike proteomics and biochemistry, genomics does not require very large amounts of starting material (genomic libraries can be amplified easily by PCR, while peptides cannot). Results clearly indicate that Epi-Decoder allows the establishment of local composition. The transgene is mostly bound by RNA polymerase II subunits and proteins of the preinitiation complex, while the proteins at the transcriptional termination site are different. The method also hints at the relative protein abundances at the locus. Epi-Decoder is currently limited only by the number of proteins that can be tagged. If applicable to mammalian systems, this alternative strategy might currently represent the most convincing approach for establishing single-copy locus composition.

target locus, and some are even significantly bound to nonspecific targets (Fig. 4a). This is particularly problematic when a small single-copy locus is targeted, because only one dCas9 molecule per genome is specifically bound at any given time. This implies that the vast majority of the biotinylated proteome is nonspecific and contributes to noise (Fig. 4b). Moreover, as dCas9 is a rather large molecule, its binding efficiency might be greatly influenced by the chromatin structure (Fig. 4c). Finally, dCas9 binding might compete out the binding of biologically relevant interactors (Fig. 4d). Appropriate controls and multiple replicates might be sufficient to mitigate these issues. Nevertheless, dCas9 proximity labeling-based methods might be useful to identify proteins that bind only to the single-copy target and not elsewhere.

Concluding remarks

In the genome, each locus is defined by a unique combination of proteins that support local functions. For instance, telomeres are bound by shelterin proteins for chromosomal capping, centromeres by CENP proteins that specify kinetochore position in mitosis, transcribed genes by RNA polymerases and cofactors, and replication origins by replication initiation factors. In many of the studies described in this article (Table 1), particularly when focused on the analysis of single-copy small loci, proteomic analysis often retrieved noncanonical and unexpected protein factors with high confidence, while expected proteins that normally define the target biological function were absent. In our view, this represents a matter of concern and a clear indication of the limitations of the current methods: although such approaches can be used to discover some locus-specific proteins, the major signal-to-noise ratio problem will need to be addressed before they can fully describe the specific composition of single-copy loci. We think that this is a biochemical challenge that should be prioritized in developing future locus-specific capture methods. A second problem is single-locus abundance, which, by definition, is only two molecules per diploid genome. Without peptide amplification methods, this implies working with huge amounts of starting material. This challenge will probably be solved by the continuous improvements to MS technology (Box 2) or through the use of entirely different strategies, as exemplified by the recent development of Epi-Decoder (Box 3). Epi-Decoder relies on generating yeast library strains in which genes of interest carry a universal affinity tag as well as a unique, gene-specific DNA barcode. All tagged strains are subsequently subjected to a single chromatin immunoprecipitation using a tag-specific antibody, after which gene-specific barcodes are sequenced

and quantified. Quantification of the barcodes directly reflects the abundance of each assayed protein on a given target. Epi-Decoder was used to interrogate the composition of a transcription start site of a reporter gene using 4,000 tagged yeast strains and retrieved, as expected for this type of locus, RNA polymerase II subunits and canonical transcription initiation factors. This 'genomics' method is a very promising alternative for single-locus affinity purification, although it has thus far only been applied to yeast.

Received: 12 November 2019; Accepted: 29 January 2020;
Published online: 9 March 2020

References

- Kornberg, R. D. & Lorch, Y. Chromatin rules. *Nat. Struct. Mol. Biol.* **14**, 986–988 (2007).
- Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Engelen, E. et al. Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat. Commun.* **6**, 7155 (2015).
- Ji, X. et al. Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl Acad. Sci. USA* **112**, 3841–3846 (2015).
- Mohammed, H. et al. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat. Protoc.* **11**, 316–326 (2016).
- Rafiee, M. R., Girardot, C., Sigismondo, G. & Krijgsvel, J. Expanding the circuitry of pluripotency by selective isolation of chromatin-associated proteins. *Mol. Cell* **64**, 624–635 (2016).
- van den Berg, D. L. et al. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**, 369–381 (2010).
- Soldi, M. & Bonaldi, T. The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components. *Mol. Cell. Proteom.* **12**, 764–780 (2013).
- Kadonaga, J. T. & Tjian, R. Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl Acad. Sci. USA* **83**, 5889–5893 (1986).
- Mittler, G., Butter, F. & Mann, M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res.* **19**, 284–293 (2009).
- Bartke, T. et al. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470–484 (2010).
- Vermeulen, M. et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967–980 (2010).
- Spruijt, C. G. et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
- Boffa, L. C., Carpaneto, E. M. & Allfrey, V. G. Isolation of active genes containing CAG repeats by DNA strand invasion by a peptide nucleic acid. *Proc. Natl Acad. Sci. USA* **92**, 1901–1905 (1995).
- Higashinakagawa, T., Wahn, H. & Reeder, R. H. Isolation of ribosomal gene chromatin. *Dev. Biol.* **55**, 375–386 (1977).
- Workman, J. L. & Langmore, J. P. Nucleoprotein hybridization: a method for isolating specific genes as high molecular weight chromatin. *Biochemistry* **24**, 7486–7497 (1985).
- Zhang, X. Y. & Hörz, W. Analysis of highly purified satellite DNA containing chromatin from the mouse. *Nucleic Acids Res.* **10**, 1481–1494 (1982).
- Déjardin, J. & Kingston, R. E. Purification of proteins associated with specific genomic loci. *Cell* **136**, 175–186 (2009).
- Byrum, S. D., Raman, A., Taverna, S. D. & Tackett, A. J. ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. *Cell Rep.* **2**, 198–205 (2012).
- Pourfarzad, F. et al. Locus-specific proteomics by TChP: targeted chromatin purification. *Cell Rep.* **4**, 589–600 (2013).
- Hamperl, S. et al. Compositional and structural analysis of selected chromosomal domains from *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **42**, e2 (2014).
- Waldrup, Z. J. et al. A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics* **9**, 1207–1211 (2014).
- Guillen-Ahlens, H. et al. HyCCAPP as a tool to characterize promoter DNA-protein interactions in *Saccharomyces cerevisiae*. *Genomics* **107**, 267–273 (2016).
- van Holde Kensal, E. *Chromatin* (Springer, 1989).
- Hoffman, E. A., Frey, B. L., Smith, L. M. & Auble, D. T. *J. Biol. Chem.* **290**, 26404–26411 (2015).
- Müller-Ott, K. et al. Specificity, propagation, and memory of pericentric heterochromatin. *Mol. Syst. Biol.* **10**, 746 (2014).

27. Grolimund, L. et al. A quantitative telomeric chromatin isolation protocol identifies different telomeric states. *Nat. Commun.* **4**, 2848 (2013).
28. Majerská, J., Redon, S. & Lingner, J. Quantitative telomeric chromatin isolation protocol for human cells. *Methods* **114**, 28–38 (2017).
29. Marzec, P. et al. Nuclear-receptor-mediated telomere insertion leads to genome instability in ALT cancers. *Cell* **160**, 913–927 (2015).
30. Feng, X. et al. The telomere-associated homeobox-containing protein TAH1/HMBOX1 participates in telomere maintenance in ALT cells. *J. Cell Sci.* **126**, 3982–3989 (2013).
31. Kappei, D. et al. HOT1 is a mammalian direct telomere repeat-binding protein contributing to telomerase recruitment. *EMBO J.* **32**, 1681–1701 (2013).
32. Conomos, D., Reddel, R. R. & Pickett, H. A. NuRD-ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. *Nat. Struct. Mol. Biol.* **21**, 760–770 (2014).
33. Gong, Y., Handa, N., Kowalczykowski, S. C. & de Lange, T. PHF11 promotes DSB resection, ATR signaling, and HR. *Genes Dev.* **31**, 46–58 (2017).
34. Li, J. S. et al. TZAP: A telomere-associated protein involved in telomere length control. *Science* **355**, 638–641 (2017).
35. Saksouk, N. et al. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol. Cell* **56**, 580–594 (2014).
36. Gauchier, M. et al. SETDB1-dependent heterochromatin stimulates alternative lengthening of telomeres. *Sci. Adv.* **5**, eaav3673 (2019).
37. Ide, S. & Dejardin, J. End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. *Nat. Commun.* **6**, 6674 (2015).
38. Kennedy-Darling, J. et al. Discovery of chromatin-associated proteins via sequence-specific capture and mass spectrometric protein identification in *Saccharomyces cerevisiae*. *J. Proteome Res.* **13**, 3810–3825 (2014).
39. Buxton, K. E. et al. Elucidating protein-DNA interactions in human aliphoid chromatin via hybridization capture and mass spectrometry. *J. Proteome Res.* **16**, 3433–3442 (2017).
40. Dai, Y. et al. Multiplexed sequence-specific capture of chromatin and mass spectrometric discovery of associated proteins. *Anal. Chem.* **89**(15), 7841–7846 (2017).
41. Fujita, T. & Fujii, H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem. Biophys. Res. Commun.* **439**, 132–136 (2013).
42. Kescu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683 (2014).
43. Liu, X. et al. In situ capture of chromatin interactions by biotinylated dCas9. *Cell* **170**, 1028–1043 (2017).
44. Tsui, C. et al. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc. Natl Acad. Sci. USA* **115**, E2734–E2741 (2018).
45. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2013).
46. Byrum, S. D., Taverna, S. D. & Tackett, A. J. Purification of a specific native genomic locus for proteomic analysis. *Nucleic Acids Res.* **41**, e195 (2013).
47. Fujita, T. et al. Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Sci. Rep.* **3**, 3171 (2013).
48. Griesenbeck, J., Boeger, H., Strattan, J. S. & Kornberg, R. D. Affinity purification of specific chromatin segments from chromosomal loci in yeast. *Mol. Cell Biol.* **23**, 9275–9282 (2003).
49. Fujita, T. & Fujii, H. Direct identification of insulator components by insertional chromatin immunoprecipitation. *PLoS One* **6**, e26109 (2011).
50. Roux, K. J., Kim, D. I. & Burke, B. BioID: a screen for protein-protein interactions. *Curr. Protoc. Protein Sci.* **74**, 19.23.1–19.23.14 (2013).
51. Schmidtman, E., Anton, T., Rombaut, P., Herzog, F. & Leonhardt, H. Determination of local chromatin composition by CasID. *Nucleus* **7**, 476–484 (2016).
52. Garcia-Exposito, L. et al. Proteomic profiling reveals a specific role for translesion DNA polymerase η in the alternative lengthening of telomeres. *Cell Rep.* **17**, 1858–1871 (2016).
53. Qiu, W. et al. Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. *Nucleic Acids Res.* **47**, e52 (2019).
54. Gao, X. D. et al. C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat. Methods* **15**, 433–436 (2018).
55. Myers, S. A. et al. Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat. Methods* **15**, 437–439 (2018).
56. Korthout, T. et al. Decoding the chromatin proteome of a single genomic locus by DNA sequencing. *PLoS Biol.* **16**, e2005542 (2018).
57. Solomon, M. J. & Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc. Natl Acad. Sci. USA* **82**, 6470–6474 (1985).
58. Schmiedeberg, L., Skene, P., Deaton, A. & Bird, A. A temporal threshold for formaldehyde crosslinking and fixation. *PLoS One* **4**, e4636 (2009).
59. Festuccia, N. et al. Transcription factor activity and nucleosome organization in mitosis. *Genome Res.* **29**, 250–260 (2019).
60. O'Neill, L. P. & Turner, B. M. Immunoprecipitation of native chromatin: NChIP. *Methods* **31**, 76–82 (2003).
61. Belov, M. E., Gorshkov, M. V., Udseth, H. R., Anderson, G. A. & Smith, R. D. Zeptomole-sensitivity electrospray ionization–Fourier transform ion cyclotron resonance mass spectrometry of proteins. *Anal. Chem.* **72**, 2271–2279 (2000).
62. Keller, B. O. & Li, L. Detection of 25,000 molecules of Substance P by MALDI-TOF mass spectrometry and investigations into the fundamental limits of detection in MALDI. *J. Am. Soc. Mass Spectrom.* **12**, 1055–1063 (2001).
63. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45 (2004).
64. Searle, B. C. et al. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat. Commun.* **9**, 5128 (2018).
65. Pfammatter, S., Bonnell, E. & Thibault, P. Improvement of quantitative measurements in multiplex proteomics using high-field asymmetric waveform spectrometry. *J. Proteome Res.* **15**, 4653–4665 (2016).
66. Baker, E. S. et al. Ion mobility spectrometry-mass spectrometry performance using electrodynamic ion funnels and elevated drift gas pressures. *J. Am. Soc. Mass Spectrom.* **18**, 1176–1187 (2007).
67. Pfammatter, S. et al. A novel differential ion mobility device expands the depth of proteome coverage and the sensitivity of multiplex proteomic measurements. *Mol. Cell. Proteom.* **17**, 2051–2067 (2018).
68. Furlan, C. et al. Miniaturised interaction proteomics on a microfluidic platform with ultra-low input requirements. *Nat. Commun.* **10**, 1525 (2019).
69. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).

Acknowledgements

Work in the Dejardin lab is supported by grants from ARC équipe labélisée 2016, from the ERC CoG METACHROM, from INCA, from the Fondation Schlumberger pour l'éducation et la recherche and from Merck (MSD Gnostic). M.G. is supported by the University of Montpellier and by ARC. The Vermeulen lab is part of the Oncode Institute, which is partly funded by the Dutch Cancer Society.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to M.V. or J.D.

Peer review information Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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

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

© Springer Nature America, Inc. 2020