Corrected: Publisher Correction

Locus-specific chromatin isolation

Michiel Vermeulen[1]* *and Jérôme Déjardin*[2]*

Identifying factors bound to specific genomic loci in an unbiased manner is the holy grail of chromatin research, because it would provide an ultimate description of locus function. In this Comment, we explain why purifying a single locus represents a major technical challenge and provide recommendations for achieving this goal.

In the past four decades, fewer than 30 articles have been published that deal with the purification and characterization of specific genomic loci. This is quite a small number, particularly considering the broad scientific interest in this topic. Locus-specific chromatin purification has always been an important goal in chromatin research, because knowledge of local-chromatin composition is a prerequisite to understanding its molecular function. This paucity of publications is explained by the fact that purifying a specific chromatin locus is very challenging and potentially requires workflow adaptation for each locus. Consequently, not many successful attempts of locus-specific chromatin purification have been documented. In the past decade, several strategies have been developed that use the unique DNA sequence of a locus as a discriminating feature for locus-specific purification[1]. Some of these approaches were developed to identify a few locus-bound proteins, whereas other studies determined the 'proteome' of a locus of interest in a more comprehensive manner.

With the advent of new genome-editing technologies and improved mass spectrometry approaches and instruments, locus-specific purification appears to have been made more feasible and is therefore becoming increasingly popular. Yet, despite the recent major technological innovations, purifying a specific chromatin locus remains extremely challenging owing to its small size compared with the rest of the genome, which therefore represents a huge background signal that has to be overcome. Furthermore, maintaining the diversity and stoichiometry of the locus-bound factors during its isolation is a major feat. Moreover, irrespective of the target locus, methodology used and scientific question, it is of crucial importance to investigate the biological relevance of the identified interacting proteins. For example, the identification of a mitochondrial metabolic protein on a nuclear gene promoter could be a scientific breakthrough in the way we define that promoter, or could be a technical artefact. Below, we present some important facts and considerations that should be taken into account by those interested in performing single-locus purifications.

The abundance of a specific locus

Mass spectrometry identifies peptides from complex mixtures in the low-femtomolar range. This sets the minimum amount of material that is required for proteomic identification of a protein that binds only once to the locus of interest to half-a-billion cells. The chosen purification strategy depends on the end goal of the study. To establish the precise locus-specific protein composition, purification of the locus is required. To identify some locus-specific proteins, locus enrichment is sufficient if these proteins bind abundantly to the target locus and not much elsewhere in the genome. Importantly, the choice of method also depends on the size of the target locus. This is important because the locus size defines its relative abundance. For example, whereas a 3 kb-long gene promoter or enhancer represents 0.0001% of the ~3 GB-long human genome, telomeres comprise 300 kb and thus cover ~0.01% of the human genome. Therefore, methods that achieve 1 million-fold enrichment are required to isolate a promoter to near purity, whereas pure telomeric chromatin can be obtained using techniques that offer only 10,000-fold enrichment. Such enrichment levels may not be needed to identify locus-specific proteins that are highly enriched on the target locus compared with the rest of the genome. Nevertheless, it is clear that the amount of protein at a specific promoter is extremely low compared with the amounts bound to the rest of the genome in each cell, and this has important implications for the amount of input material that is needed and the fold enrichment that is required for successful single-locus purifications.

A chromatin locus has a complex composition

The DNA sequence and the cell type determine the local chromatin composition and function. However, structural elements such as histones are ubiquitous chromatin components, and thus their presence, although biologically crucial, does not define a specific function. By contrast, proteins that bind to DNA in a sequence-specific manner, and enzymatic complexes such as histone-modifying enzymes, ATP-dependent chromatin remodellers, DNA polymerases, RNA polymerases and

[1]*Department of Molecular Biology, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University, Nijmegen, The Netherlands.*

[2]*Institute of Human Genetics, CNRS-Université de Montpellier, Montpellier, France.*

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

https://doi.org/10.1038/s41580-020-0217-0

DNA repair enzymes, are the true determinants of the local chromatin functions. As many of these biologically relevant protein–chromatin associations are very transient and dynamic, the abundance of participating proteins can be several orders of magnitude lower than that of histones. Hence, not only can a given locus be bound by dozens of different factors that define local function, but the local dynamic range of bound proteins can also be immense. This is another major technical challenge, particularly for mass spectrometry. In fact, proteins present in minute amounts in complex mixtures tend to escape identification by mass spectrometry, which is a problem that is not unique to chromatin proteomics[2].

Purification strategies

Specific loci can be distinguished only on the basis of their DNA sequence. We think that this feature is the most relevant for locus isolation. Indeed, sequence-specific hybridization capture by oligonucleotides and sequence-specific binding by adaptor proteins (for example, RNA-guided nuclease-dead-Cas9 harbouring affinity purification tags) have been used as locus isolation strategies[3–8]. These techniques have allowed the successful characterization of the composition of large (repetitive) loci and also the identification of some small, single-locus-specific proteins. Nevertheless, technologies that allow determining the composition of small unique loci in a comprehensive manner are still lacking. Methods that involve single affinity purification steps are inherently unable to provide the required enrichment factor (1 million-fold) to ensure that each isolated protein originates from the purified locus instead of being a background protein.

Regardless of their relative abundance, different genomic elements have specific functions, which are defined by their interacting proteins: telomere repeats are bound by shelterin proteins, centromeres are bound by more than a dozen dedicated CENP nucleosome-associated factors, promoters are bound by transcription factors and transcription start sites are bound by the RNA polymerase II pre-initiation complex and by the RNA polymerase II holoenzyme, origins of DNA replication are bound by replication factors and so on. Multi-step purification strategies need to be developed to isolate a unique genome segment to sufficient purity to enable to retrieve primarily these factors. For any target locus, including those mentioned above, if most of the proteins expected for the locus are not recovered and identified, results should be interpreted with extreme caution, particularly unanticipated protein associations. A chromatin locus may be bound by non-canonical proteins that confer unexpected biological functions, but the validation of any novel single-locus isolation strategy requires it to identify most of the known and expected proteins as an obligate starting point.

Practical recommendations

A brief checklist for those interested in performing single-locus proteomics:

- Ensure you have enough starting material: genomics analyses can be performed using low amounts of material, because they always involve a faithful amplification step. By contrast, performing proteomics analyses of complex mixtures still requires relatively high amounts of material, because peptides cannot be amplified. The required amount of the target locus should therefore be carefully estimated, and sufficient amounts of starting material should be prepared accordingly.
- Use the right methodology: depending on the relative abundance of the target locus, methods providing sufficient enrichment should be developed. Whereas relatively abundant targets can be isolated with single-step isolation protocols, small, unique loci within large genomes will require multi-step isolation strategies.
- Critically assess the results: if the purpose of your study is to define the 'proteome' of the target locus, the results should first and foremost recapitulate most of the known binders at the target. Identifying only a few known proteins indicates that the purification of the target locus has not maintained biologically crucial proteins or that locus enrichment was insufficient. If this is the case, the identified proteins might still be relevant and should be independently validated, but such experiments cannot be interpreted as successfully describing local protein composition.

1. Gauchier, M. et al. Purification and enrichment of locus-specific proteomes: challenges and possibilities. *Nat. Methods* (in the press) (2020).
2. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867 (2002).
3. Déjardin, J. & Kingston, R. E. Purification of proteins associated with specific genomic loci. *Cell* **136**, 175–186 (2009).
4. Pourfarzad, F. et al. Locus-specific proteomics by TChP: targeted chromatin purification. *Cell Rep.* **4**, 589–600 (2013).
5. Ide, S. & Dejardin, J. End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. *Nat. Commun.* **6**, 6674 (2015).
6. Buxton, K. E. et al. Elucidating protein-DNA interactions in human alphoid chromatin via hybridization capture and mass spectrometry. *J. Proteome Res.* **16**, 3433–3442 (2017).
7. Liu, X. et al. In situ capture of chromatin interactions by biotinylated dCas9. *Cell* **170**, 1028–1043 (2017).
8. Tsui, C. et al. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc. Natl Acad. Sci. USA* **115**, 201718844 (2018).

Acknowledgements

Work in the laboratory of J. Dejardin is supported by grants from ARC Equipe Labélisée 2016, the ERC CoG METACHROM, from INCA, from the Fondation Schlumberger pour l'Education et la Recherche and from Merck (MSD). The laboratory of M. Vermeulen is part of the Oncode Institute, which is partly funded by the Dutch Cancer Society.

Competing interests

The authors declare no competing interests.