ETH zürich

ETH Library

Computational Approaches Linking Chromatin Organization and Genome Regulation

Doctoral Thesis

Author(s):
Paysan, Daniel

Publication date:
2024

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000699187

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

DISS. ETH NO. 30349

Computational Approaches Linking Chromatin Organization and Genome Regulation

A thesis submitted to attain the degree of

Doctor of Sciences
(Dr. sc. ETH Zurich)

presented by

Daniel Paysan
MSc. ETH in Statistik, ETH Zurich

born on 02.06.1996

accepted on the recommendation of

Prof. Dr. G.V. Shivashankar, examiner

Prof. Dr. Caroline Uhler, co-examiner

Prof. Dr. Catherine Jutzeler, co-examiner

2024

Summary

In cells, the often meter-long DNA is packed into nuclei that are only a few micrometer in size. The packing of the DNA results in a highly structured organization of the genome, i.e. the 3D chromatin organization. This organization varies between cell states and is closely related to the gene regulatory networks that control cellular function in cells. The connection arises from the optimized packing, which for instance is expressed by the spatial colocalization of genes for their efficient co-regulation. However, the depth of the connection between the chromatin organization and genome regulation, along with its regulating mechanisms, is not yet fully understood. It also remains unclear whether this connection is profound enough to use the chromatin organization of cells as an unified readout of functional cell states and study cell state transitions. This would have wide-ranging implications for disease diagnostics and therapeutic design.

In this thesis, we address these important questions. Specifically, we first demonstrate that deep representation learning can be used to effectively model the link between chromatin organization and gene regulation. To achieve this, we developed a computational pipeline that learns how the gene regulatory programs of cells are reflected in their chromatin organization. This enables our pipeline to accurately identify alterations in these regulatory programs simply by using cost-efficient chromatin images as a readout of their chromatin organization. Our work provides novel avenues for studying gene regulation and alterations thereof using image-based readouts of the nuclear chromatin organization and thus emphasizes the deep connection between chromatin organization and genome regulation.

While it has become evident that the functional optimization of the 3D chromatin packing gives rise to this connection, the mechanisms responsible for the optimized packing are not well understood. This particularly applies to the establishment of contacts between genomic regions from different chromosomes, which are crucial for cell function. In this thesis, we propose that the binding of long noncoding RNAs (lncRNAs) to the DNA may be a mechanism facilitating the formation and stability of these contacts, thereby enabling the colocalization and co-regulation of important gene clusters. To this end, we analyze multimodal sequencing data to identify spatial gene clusters formed by interchromosomal contacts. We show that the formation or loss of these gene clusters during aging could be explained by the changing abundance of specific lncRNAs. These

lncRNAs might act as "glues" binding the corresponding genomic regions together. Although additional validation experiments are ongoing, our findings provide convincing evidence that the transcriptional control of lncRNAs and their binding to the DNA plays an important role for the functionally optimized 3D genome organization.

Finally, we demonstrate how computational methods can be used to exploit the link between chromatin organization and genome regulation, enabling the development of cost-effective image-based chromatin biomarkers with applications in disease diagnostics and therapeutic design. We first showcase this using chromatin images of immune cells obtained from liquid biopsies of tumor patients. These images are input to a computational pipeline which we developed and which uses simple image analysis and machine learning methods. We show that this pipeline identifies chromatin biomarkers of immune cells that enable accurate tumor diagnosis. We further illustrate the potential of such chromatin biomarkers by applying a similar pipeline to characterize differences in the chromatin organization of distinct B cell populations in lymphoid tissue. Our corresponding analyses reveal chromatin condensation differences in these B cell populations, which correlate with the presence or absence of other immune cell populations in their local microenvironments. This could potentially provide novel insights into the mechanisms that control T cell infiltration in B cell lymphomas, also implying that inhibition of chromatin condensation may prevent immune cell evasion in B cell lymphomas. Thus, this study also highlights the potential of chromatin biomarkers to guide therapeutic design.

In summary, in this thesis we present novel computational methods to study, describe, model, and leverage the relationship between chromatin organization and genome regulation. The results presented in this thesis improve our understanding of this connection and its regulating mechanisms as well as demonstrate its large potential for applications such as disease diagnostics and therapy evaluation.

Zusammenfassung

Die DNA von Zellen ist oft mehrere Meter lang und befindet sich dennoch in den Zellkernen, deren Durchmesser nur einige Mikrometer beträgt. Damit dies möglich ist, ist die DNA im Zellkern nicht zufällig organisiert, sondern deren Aufbau ist stark strukturiert. Die resultierende Organisation der DNA wird auch Chromatinorganisation genannt. Diese ist nicht nur abhängig vom Zustand der Zelle, sondern auch eng verbunden mit den Genregulationsnetzwerken in Zellen, die die Funktionen der Zelle massgeblich steuern. Grund für diese Verbindung ist die Optimierung der Chromatinorgansation zur effizienten Ausführung der Funktionen der Zelle. Die optimierte Chromatinorganisation zeigt sich sich zum Beispiel darin, dass Gene, die zusammen für bestimmte Funktionen der Zellen verantwortlich sind, oft nah beinander im Zellkern angeordnet sind. Diese Anordnung ermöglicht die gemeinsame und effiziente Regulation dieser Gene. Obwohl die Existenz der Verbindung zwischen der Chromatinorganisation und der Genregulation in Zellen außer Frage steht, ist noch wenig darüber bekannt wie weit diese Verbindung geht oder welche Mechanismen diese steuern. Ebenso stellt sich die Frage, ob, aufgrund dieser Verbindung, die Chromatinorganisation eine universelle Auskunft über den funktionalen Zustand von Zellen geben kann. Falls dies der Fall wäre, könnten Untersuchungen der Chromatinorganisation von Zellen zum Beispiel die kostengünstige Diagnose von Krankheiten ermöglichen oder bei der Entwicklung neuer Behandlungsmöglichkeiten solcher Krankheiten helfen.

In der vorliegenden Thesis behandeln wir diese wichtigen Fragestellungen. In diesem Rahmen, zeigen wir zuerst, dass Deep Representation Learning mittels künstlicher neuronaler Netzwerke die explizite Modellierung der Zusammenhänge zwischen der Chromatinorganisation und der Genregulation ermöglicht. Dazu entwickeln wir eine softwarebasierte Pipeline, welche mittels künstlicher Intelligenz lernt wie sich Veränderungen der Genregulation in Zellen in ihrer Chromatinorganisation wiederspiegeln. Anhand der erlernten Muster ist unsere Pipeline schlussendlich in der Lage neue und gänzlich unbekannte Veränderungen der Genregulationsnetzwerke in Zellen allein auf Basis von kostengünstigen Mikroskopbildern der Chromatinorganisation von Zellen zu identifizieren. Unsere Arbeit zeigt somit auch, dass durch die Untersuchung der Chromatinorganisation neue Erkenntnisse im Hinblick auf die Genregulation in Zellen gewonnen werden können, und dass dafür lediglich einfache Chromatinbilder benötigt werden. Damit zeigt unse-

re Arbeit wie eng die Chromatinorganisation und die Genregulationsnetwerke in Zellen miteinander verbunden sind.

Wie bereits beschrieben ist die Optimierung der Chromatinorganisation eine Grundlage des direkten Zusammenhang zwischen dieser Organisation und der Genregulation. Welche Mechanismen für diese optimierte Organisation wiederum verantwortlich sind, ist jedoch noch nicht gänzlich bekannt. Dies betrifft insbesondere die Mechanismen, die dafür sorgen, dass Gene von unterschiedlichen Chromosomen in nahen Kontakt im Zellkern kommen und dadurch gemeinsam reguliert werden können. In dieser Thesis zeigen wir, dass long noncoding RNAs (lncRNAs) eine wichtige Rolle in diesem Zusammenhang spielen könnten. Unsere Ergebnisse legen nahe, dass lncRNAs sich an die DNA binden und wie eine Art Kleber fungieren. Dies führt dazu, dass unterschiedliche Segmente der DNA von verschiedenen Chromosomen zusammengehalten werden, welches wiederum zur Entstehung der gemeinsam regulierten Gencluster führt (berziehungsweise verhindert, dass sich diese Cluster auflösen). Mittels verschiedener Arten von Sequenzdaten identifizieren wir Beispiele solcher Gencluster und zeigen, dass die Entstehung oder Auflösung dieser während Alterungsprozessen durch die sich ebenfalls verändernde Verfügbarkeit von eingigen speziellen lncRNAs erklärt werden könnte. Unsere Ergebnisse legen nahe, dass diese lncRNAs in der Tat wie eine Art Kleber diese Gencluster zusammenhalten. Und obwohl die finalen Validierungsexperimente noch ausstehend sind, implizieren unsere derzeitigen Ergebnisse, dass die transkriptionelle Kontrolle von lncRNAs und die Interaktionen dieser lncRNAs mit der DNA eine wichtige Rolle in der Optimierung der Chromatinorganisation von Zellen spielen.

Im finalen Teil dieser Thesis, zeigen wir wie rechnergestüzte Methodiken, dazu verwendet werden können die Verbindung zwischen der Chromatinorganisation und der Genregulation auszunutzen. Dies ermöglicht die Entwicklung effizienter, bildbasierte Chromatinbiomarker, die zur Diagnose von Krankheiten eingesetzt werden oder auch bei der Entwicklung neuer Therapien helfen können. Dies demonstrieren wir zunächst in dem wir Mikroskopbilder der Chromatinorganisation von Immunzellen von Tumorpatienten benutzen. Diese Bilder werden von uns dann mithilfe einer selbst entwickelten rechnergestutzen Pipeline verarbeitet und mittels klassischer Bildanalyse und Methodiken des Maschinellen Lernens analysiert. Dabei identifiziert die Pipeline verschiedene Chromatinbiomarker, deren Ausprägungen in den Immunzellen eine akkurate Tumordiagnose ermöglichen.

Um das Potenzial solcher Chromatinbiomarker zu unterstreichen, benutzen wir ähnliche Methoden, um unterschiedliche Subpopulationen von B Zellen in lymphoiden Gewebe zu untersuchen. Unsere Analysen identifizieren dabei einen Zusammenhang zwischen der Kondensation des Chromatins in B Zellen und der Präsenz oder dem Fehlen von anderen Immunzellpopulationen in ihrer Umgebung. Diese Ergebnisse deuten auf einen bisher unbekannten Mechanismus der Steuerung der Infiltration von Immunzellen

in B Zell Lymphomen hin, welches auch neue Therapiemöglichkeiten offenlegt. Damit zeigt unsere Arbeit, dass die bildbasierte Chromatinbiomarker nicht nur diagnostisches Potenzial haben, sondern auch bei der Entwicklung neuer Therapien helfen können.

Zusammenfassend haben wir in dieser Arbeit neue rechnergestüzte Methodiken entwickelt und angewendet, um den Zusammenhang zwischen der Chromatinorganisation und der Genregulation in Zellen zu untersuchen, zu modellieren und schlussendlich zum Beispiel zur Diagnose von Krankheiten einzusetzen. Die Ergebnisse dieser Arbeit liefern neue Erkenntnisse bezüglich dieses Zusammenhangs und den Mechanismen, die diesen regulieren. Desweiteren zeigt unsere Arbeit das grosse Potenzial von Chromatinbiomarkern, welche diese Verbindung ausnutzen und zahlreiche medizinische Anwendungsgebiete haben.