

# PROCÉDURE DE TEST HIÉRARCHIQUE POUR L'ANALYSE DIFFÉRENTIELLE DE DONNÉES HI-C

Élise Jorge<sup>1,2</sup>, Pierre Neuvial<sup>3</sup>, Nathalie Vialaneix<sup>2</sup> & Sylvain Foissac<sup>1</sup>

<sup>1</sup> *GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet-Tolosan, France, {elise.jorge, sylvain.foissac}@inrae.fr*

<sup>2</sup> *Université Fédérale de Toulouse, INRAE, MIAT, 31326 Castanet-Tolosan, France, nathalie.vialaneix@inrae.fr*

<sup>3</sup> *Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse, CNRS, UPS, F-31062 Toulouse Cedex 9, France, pierre.neuvial@math.univ-toulouse.fr*

**Résumé.** Les données Hi-C fournissent une information sur l'organisation tridimensionnelle du génome à partir de mesures d'interactions entre positions génomiques le long de la chromatine. Cette structure en trois dimensions a un rôle important dans la régulation de l'expression des gènes. L'objectif de l'analyse différentielle est d'identifier, à partir de réplicats obtenus dans deux conditions biologiques différentes, des régions génomiques qui présentent des différences significatives de structure entre les deux conditions. Ici, nous proposons de nous appuyer sur une modélisation hiérarchique des données Hi-C, permettant de tenir compte de la dépendance spatiale présente dans ce type de données. En utilisant un travail précédent permettant de représenter les données par des arbres et de les utiliser pour réaliser des tests, on peut, à partir de régions définies *a priori*, identifier celles qui sont d'intérêt. On s'intéresse ici au développement d'une méthode permettant d'identifier automatiquement de telles régions.

**Mots-clés.** données Hi-C, génomique 3D, arbre binaire, classification hiérarchique, tests multiples

**Abstract.** Hi-C data provide insights into the three-dimensional organization of the genome by measuring interactions between genomic positions along the chromatin. This three-dimensional structure plays a crucial role in regulating gene expression. Differential analysis aims to identify genomic regions that display significant differences in structure between two different biological conditions. Here, we propose a hierarchical modeling approach to analyze Hi-C data, allowing to incorporate the inherent spatial dependence within the data. We build upon a recent method that represents the data as trees and performs targeted tests on predefined regions. The objective is to develop a method that automatically identifies regions of interest, streamlining the differential analysis process.

**Keywords.** Hi-C data, 3D genomics, binary tree, hierarchical classification, multiple testing

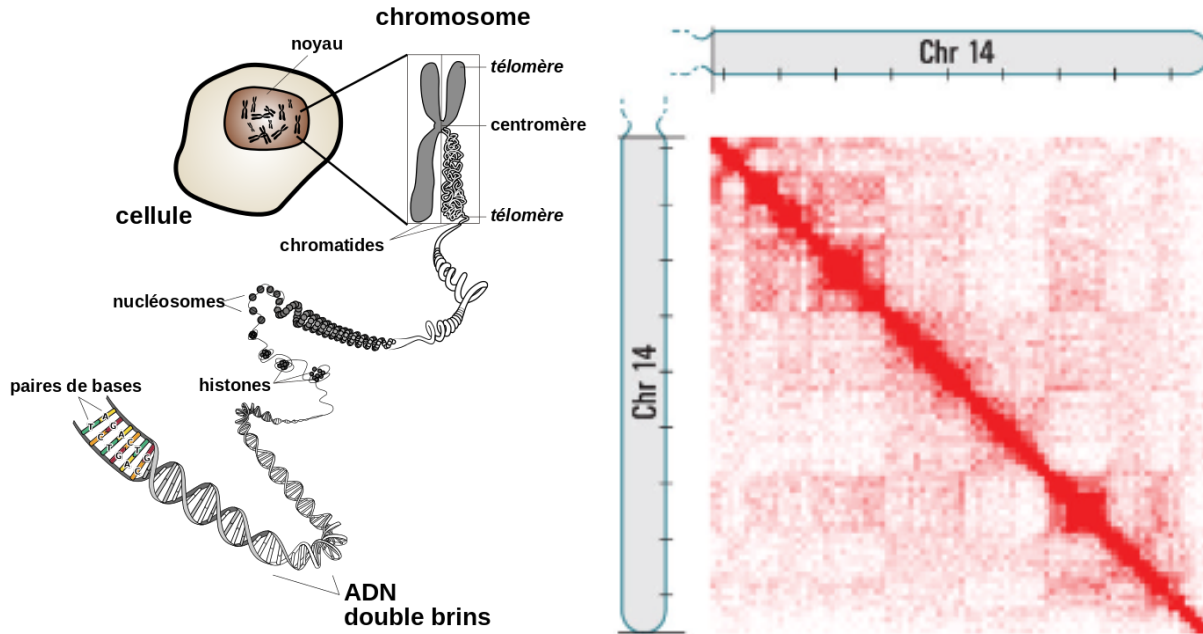


FIGURE 1 – Gauche : Schéma de la compaction de l’ADN en chromosome (“Chromosome fr” par Phrood commonswiki, Wikimedia Commons). Droite : Matrice Hi-C du chromosome 14 de [1].

# 1 Introduction

## Structure de l’ADN et données Hi-C

La chromatine est compactée au sein du chromosome selon une structure hiérarchique, comme illustré sur la figure 1 (gauche). Les données Hi-C sont des données de séquençage haut-débit qui permettent d’obtenir des informations sur l’organisation tridimensionnelle du génome dans la cellule, en mesurant la fréquence d’interactions spatiales entre régions génomiques. L’étude de ces données a permis de montrer qu’il existait, le long de la chromatine, des régions génomiques appelées TADs (*Topologically Associating Domains*) au sein desquels les interactions sont fréquentes.

L’apparition de modifications dans cette structure de compaction, par exemple la disparition d’une frontière entre deux TADs impliquant leur fusion, peut avoir un impact majeur sur l’expression des gènes dans la zone considérée. Ces modifications peuvent provoquer des pathologies neurologiques [2] ou de malformations [3].

D’une manière plus formelle, les données Hi-C se présentent sous la forme d’une matrice carrée symétrique dont l’entrée  $(i, j)$  correspond au nombre de contacts observés dans l’expérience Hi-C entre les positions génomiques  $i$  et  $j$ . La figure 1 (droite) représente une telle matrice, dans laquelle l’intensité de couleur est proportionnelle à la valeur du nombre de contacts correspondant.

## Analyse différentielle de données Hi-C

On s'intéresse ici à un problème d'analyse différentielle entre des ensembles de matrices Hi-C obtenues dans deux conditions différentes,  $\mathcal{C}_1$  et  $\mathcal{C}_2$ . L'objectif est d'identifier des régions génomiques qui présentent des différences significatives d'interactions entre ces deux conditions.

Formellement, on possède  $r = r_1 + r_2$  matrices de taille  $p \times p$  où  $M_k^{\mathcal{C}_1}$  ( $k = 1, \dots, r_1$ ) (resp.  $M_l^{\mathcal{C}_2}$  ( $l = 1, \dots, r_2$ )) correspond à la matrice obtenue pour le  $k$ -ème (resp.  $l$ -ème) réplicat de la condition  $\mathcal{C}_1$  (resp.  $\mathcal{C}_2$ ).

### État de l'art

Différentes méthodes ont été développées pour répondre à ce problème dont **diffHic** [4], **FIND** [5], **HiCcompare** [6], **multiHiCcompare** [7], **Selfish** [8] et **ACCOST** [9] mais celles-ci ne tiennent pas (ou peu) compte de la structure hiérarchique des données et produisent, de manière indépendante, une  $p$ -valeur par paire de positions (en particulier, [4, 7, 9] fondent leur approche sur une modélisation des comptages par la loi Binomiale Négative). Les détections positives sont donc fréquemment éparpillées sur l'intégralité de la matrice, sans relation avec une structure fonctionnelle du génome et donc peu interprétables.

La méthode *treediff* [10] permet de prendre en compte les dépendances entre positions génomiques induites par la structure 3D de la chromatine en représentant les matrices par des dendrogrammes qui sont un type particulier d'arbre binaire. À partir de cette représentation hiérarchique des données, des tests individuels position par position sont réalisés. Une méthode d'agrégation des tests individuels permet ensuite, pour une zone de la matrice donnée, d'identifier s'il existe au sein de cette zone au moins une interaction significativement différentielle entre les deux conditions. Comme nous l'expliquons ci-dessous, la limite de cette méthode est qu'elle repose sur une définition préalable des zones à tester par l'utilisateur. C'est cette limite que nous abordons dans ce travail.

## 2 La méthode *treediff*

On présente ici les principales étapes de la méthode *treediff* [10].

### 1. Étape 1 : Classification ascendante hiérarchique sous contrainte de contiguïté.

Cette première étape a pour but d'obtenir une classification des positions génomiques pour chaque matrice dans chaque condition. Pour cela, la classification hiérarchique ascendante avec contrainte de contiguïté (CAHCC) définie dans [11] est utilisée. Cette méthode consiste à appliquer une CAH à noyau, bien adaptée à des données qui sont des similarités. La contrainte de contiguïté dans la classification impose de ne regrouper que des classes et/ou des feuilles adjacentes le long du génome. On obtient, pour chaque matrice, un dendrogramme représentant une classification des données rendant compte de la dépendance spatiale dans les données Hi-C comme illustré sur la figure 2.

### 2. Étape 2 : Comparaison de dendrogrammes et tests de Student.

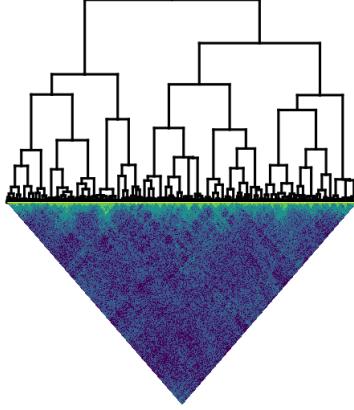


FIGURE 2 – (Demie) matrice Hi-C (en bas) et le dendrogramme associé obtenu par CAHCC (en haut).

Afin de pouvoir comparer les dendrogrammes entre eux et déterminer l'existence de différences d'interactions, des distances cophénétiques sont calculées entre toutes les paires de positions génomiques,  $i$  et  $j$ , pour chaque dendrogramme.

Ainsi,  $\forall i, j \in \{1, \dots, p\}$ ,  $k \in \{1, \dots, r_1\}$  (resp.  $l \in \{1, \dots, r_2\}$ ), on notera  $d_{(i,j)}^{\mathcal{C}_1, k}$  (resp.  $d_{(i,j)}^{\mathcal{C}_2, l}$ ) la distance cophénétique entre les positions  $i$  et  $j$  pour le  $k$ -ème (resp.  $l$ -ème) réplicat de la condition  $\mathcal{C}_1$  (resp.  $\mathcal{C}_2$ ). Ensuite, pour deux positions génomiques  $i$  et  $j$ , l'hypothèse « *Il n'existe pas de différence d'interaction entre les conditions  $\mathcal{C}_1$  et  $\mathcal{C}_2$  pour les positions  $i$  et  $j$*  » est formulée sous la forme d'une hypothèse nulle sur l'égalité des moyennes de la distance cophénétique entre  $i$  et  $j$  pour les conditions  $\mathcal{C}_1$  et  $\mathcal{C}_2$ ,  $H_0^{(i,j)} : \ll \mu_{ij}^{\mathcal{C}_1} = \mu_{ij}^{\mathcal{C}_2} \gg$ . Cette partie conduit à l'obtention d'une  $p$ -valeur  $\pi_{i,j}$  pour chaque paire de positions  $(i, j)$ .

### 3. Étape 3 : Agrégation de Simes.

La dernière partie consiste à utiliser les  $p$ -valeurs individuelles pour réaliser un test de l'hypothèse  $H_0^C = \bigcap_{i,j \in C, i < j} H_0^{(i,j)}$ , où  $C$  est un intervalle de positions génomiques contigües. Pour ce faire, les  $\pi_{i,j}$  sont agrégées par la méthode de Simes [12]. De manière plus précise, si  $n = \frac{|C|(|C|-1)}{2}$  est le nombre de paires de positions génomiques dans  $C$ , l'agrégation de Simes consiste à calculer une  $p$ -valeur corrigée :

$$\pi_{\text{Simes}}^C = \min \left\{ n \times \frac{\pi_{(k)}}{k}; k = 1, \dots, n \right\}, \quad (1)$$

où  $\pi_{(k)}$  est la  $k$ -ème plus grande  $p$ -valeur parmi les  $(\pi_{i,j})_{i,j \in C, i < j}$ .

Sous des hypothèses de dépendance positive de type PRDS (Positive Regression Dependency on a Subset) [13] sur les  $p$ -valeurs individuelles, l'erreur de type I est contrôlée au niveau  $\alpha$ , càd :  $\mathbb{P}_{H_0^C}(\pi_{\text{Simes}}^C \leq \alpha) \leq \alpha$ . L'hypothèse PRDS est classique en tests multiples et considérée comme réaliste en génomique [14].

En résumé, la méthode *treediff* permet d'inférer l'existence d'au moins une interaction différentielle au sein d'une zone prédéfinie de la matrice. Cette méthode requiert donc une connaissance préalable des données afin de choisir des zones à tester. De plus, lorsque plusieurs zones sont testées, une correction de tests multiples est nécessaire.

### 3 Une méthode « *data-driven* » d'analyse différentielle de données Hi-C

L'objectif de cette section est de présenter comment, à partir de la méthode *treediff*, nous proposons une méthode qui utilise les données pour déterminer quelles zones de la matrice tester tout en conservant le contrôle de l'erreur. Pour cela, nous avons considéré plusieurs méthodes de tests multiples sur des données structurées que nous présentons dans la section 4.1 avant de décrire notre proposition dans la section 4.2.

#### 3.1 Tests multiples dans un cadre hiérarchique

La plupart des méthodes de contrôle de tests multiples exploitant une structure hiérarchique ont été développées dans un cadre de sélection de variables ou d'un autre type de test portant sur des variables  $(X^j)_{j=1,\dots,p}$  dans lequel on connaît une structure hiérarchique entre les  $X^j$  (généralement un dendrogramme obtenu à partir d'une classification ascendante hiérarchique des  $X^j$ ). Dans ce cadre, on connaît la  $p$ -valeur  $\phi_j$  associé à chaque  $X^j$  indépendamment et contrôlant l'hypothèse  $H_0^j$  portant sur  $X^j$ . On s'intéresse alors à trouver les groupes (classes) de variables,  $C$ , rejetant l'hypothèse nulle  $H_0^C = \cap_{j \in C} H_0^j$ . Ce cadre diffère donc un peu de celui que nous avons décrit précédemment puisqu'il ne considère que les hypothèses nulles liées directement aux variables, et pas l'ensemble des hypothèses nulles liées aux interactions entre ces variables.

Nous définissons deux types d'erreurs classiques pour le cadre des tests multiples : le FWER (*Family Wise Error Rate*) [15] et le FDR (*False Discovery Rate*) [16]. Soit  $\mathcal{H}$  un ensemble d'hypothèses et  $\mathcal{H}_0 \subseteq \mathcal{H}$  le sous-ensemble des hypothèses vraies. Soit  $R$  l'ensemble des hypothèses rejetées par la procédure de test et  $|R \cap \mathcal{H}_0|$  le nombre de faux positifs. Le FDR est défini par  $\mathbb{E}(\frac{|R \cap \mathcal{H}_0|}{|R| \vee 1})$  qui est l'espérance de la proportion de faux positifs parmi les hypothèses rejetées. Le FWER correspond à  $\mathbb{P}(|R \cap \mathcal{H}_0| \geq 1)$ , la probabilité de faire au moins une erreur de type I.

Une des premières méthodes permettant d'aborder la question du contrôle de l'erreur dans un contexte de données structurées hiérarchiquement est celle décrite dans [17] qui propose une procédure de test d'hypothèses structurées en un unique arbre et assurant un contrôle du FDR. Toutefois, comme discuté dans [14], dans le cadre de tests portant sur un cluster d'hypothèses, le contrôle du FDR ne donne pas une garantie suffisante sur les résultats.

Une approche permettant de contrôler le FWER dans un cadre similaire a été proposée dans [18]. Cette méthode s'applique à des données organisées hiérarchiquement en arbre et consiste à parcourir l'arbre de haut en bas afin d'identifier les plus petits clusters rejetant l'hypothèse nulle tout en assurant un contrôle global du FWER sur tout le parcours de l'arbre. Cette méthode a été proposée dans le cadre de la sélection de variables, en particulier dans le cas où celles-ci sont fortement corrélées.

Depuis, d'autres approches similaires ont été proposées, comme celle de [19], qui est une méthode de test pour des données organisées en DAG (*Directed Acyclic Graph*). Celle-ci est

appliquée à un graphe d'ontologie de gènes. Un dendrogramme pouvant être vu comme un type de DAG, la structure de données utilisée dans cette méthode est compatible avec le cas d'application de *treediff*. Cependant, une des hypothèses utilisées par les auteurs est que, pour trois clusters  $A$ ,  $B$  et  $C$  tels que  $A = B \cup C$ , alors si  $H_0^A$  et  $H_0^B$  sont vraies,  $H_0^C$  est vraie aussi. Ceci n'est pas vérifié dans notre cadre car si aucune interaction différentielle n'est trouvée dans le cluster  $A$  et le cluster  $B$ , il pourrait quand même y avoir une interaction différentielle entre une position  $i$  de  $A$  et une position  $j$  de  $B$  donc l'hypothèse  $H_0^C$  ne serait pas vérifiée. Cette méthode n'est donc pas adaptée à notre cadre d'étude.

Enfin, une procédure de test pour des données ordonnées en temps ou en espace est proposée dans [20]. La méthode consiste à représenter les données sur un graphe puis à utiliser une procédure de test le long de ce graphe qui permet un contrôle du FWER. Bien que les hypothèses que l'on souhaite tester soient organisées spatialement, le fait que cette méthode ne propose pas d'utiliser de structure hiérarchique *a priori* sur les données rend difficile son adaptation au contexte des données Hi-C.

Nous avons donc choisi de nous appuyer sur le cadre formel de [18], bien adapté au contexte des données Hi-C, et de l'étendre au cas particulier où l'hypothèse nulle correspond à une intersection d'hypothèses nulles sur des interactions.

### 3.2 Contrôle de tests multiples pour l'analyse de données Hi-C

Pour adapter l'approche de [18] au cadre décrit dans la section 2, nous avons besoin d'une structure hiérarchique unique pour appuyer le parcours des tests à effectuer. Aussi, en complément des CAHCC réalisées sur chacune des matrices Hi-C individuelles, nous proposons de construire une CAHCC « consensuelle » basée sur la matrice  $M = \sum_{k=1}^{r_1} M_k^{C_1} + \sum_{k=1}^{r_2} M_k^{C_2}$ .

Le parcours du dendrogramme se fait dans le sens inverse de l'agrégation des classes (c'est-à-dire, de la racine du dendrogramme, jusqu'aux feuilles, en parcourant chaque nœud du dendrogramme, correspondant à une classe, avant de parcourir ses nœuds enfants). Pour chaque classe  $C$  rencontrée dans le parcours du dendrogramme, on dispose de la  $p$ -valeur obtenue par la méthode *treediff*,  $\pi^C$  : cette  $p$ -valeur contrôle la probabilité de rejet à tort de  $C$  sous l'hypothèse  $H_0^C$ . Afin d'assurer un contrôle du FWER sur l'ensemble de l'arbre, on réalise alors, pour chaque classe, deux ajustements de cette  $p$ -valeur, un premier relatif à la taille de la classe considérée et un second relatif à la hiérarchie :

1. on appelle *p-valeur ajustée de C* la quantité :

$$\pi_{adj}^C = \frac{p}{|C|} \pi^C \quad (2)$$

Ainsi, plus une classe est petite (et donc plus elle est « lointaine » dans le parcours de la hiérarchie), plus elle est pénalisée ;

2. on appelle *p-valeur hiérarchiquement ajustée* la quantité :

$$\pi_{adj,h}^C = \max_{D, D \supset C} \pi_{adj}^D \quad (3)$$

Ainsi, les classes rejetées au seuil  $\alpha$  par cette procédure sont des classes dont tous les parents ont été rejetés. Réciproquement, cet ajustement hiérarchique permet de s’assurer que si une classe n’est pas rejetée, ses descendantes ne le sont pas non plus. La procédure est illustrée dans la figure 3. Enfin, nous avons prouvé que cette approche permet bien un

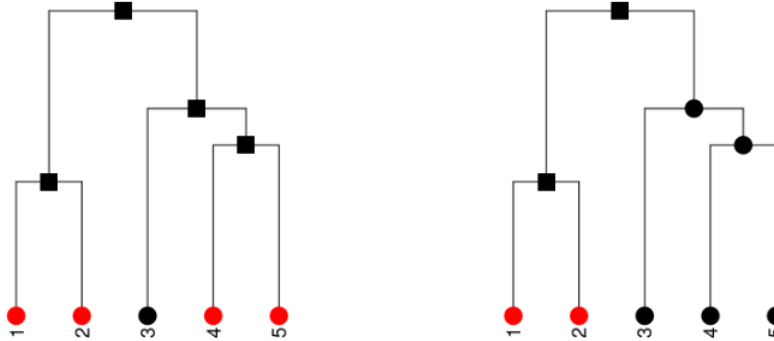


FIGURE 3 – Illustration de la procédure globale de parcours du dendrogramme consensus. Les classes rejetées sont représentées par un carré et les classes non-rejetées sont représentées par un cercle. Les feuilles (qui, dans le cas de l’approche *treediff*, ne peuvent être testées) sont représentées par un cercle. Les feuilles qui correspondent aux positions génomiques identifiées par la procédure sont représentées en rouge.

contrôle du FWER global, comme dans le cadre décrit par [18] (résultat non montré ici).

## 4 Application, résultats et perspectives

Nous avons implémenté la méthode décrite en section 4.2, et l’avons testée sur des données Hi-C issues de lignées cellulaires murines [21] pour deux conditions biologiques qui correspondent à des stades de différenciation cellulaire différents de cellules neuronales.

L’analyse des premiers résultats nous permet de faire deux remarques principales : en premier lieu, on observe ce qui est appelé le *spatial specificity paradox* et qui n’est pas spécifique à notre contexte applicatif. Ce paradoxe postule que les plus grosses classes rejetées de la hiérarchie ne sont pas les plus informatives. En second lieu, on observe que le fait que les tests sont, dans le cas des données Hi-C, des tests d’interactions modifie l’interprétation des résultats.

Les dendrogrammes de la figure 4 nous permettent d’illustrer ces deux remarques.

Dans le dendrogramme de gauche, la classe qui contient toutes les positions est rejetée et ses deux classes descendantes ne le sont pas. C’est donc la région génomique correspondant à l’ensemble des positions qui est identifiée. Or, l’information que nous donne la procédure est qu’au sein des classes  $\{1, 2, 3, 4, 5, 6, 7\}$  et  $\{8, 9, 10\}$ , il n’existe pas d’interaction différentielle. Le rejet de la classe contenant toutes les positions signifie donc qu’il existe probablement au moins une interaction différentielle entre une position de la classe  $\{1, 2, 3, 4, 5, 6, 7\}$  et une

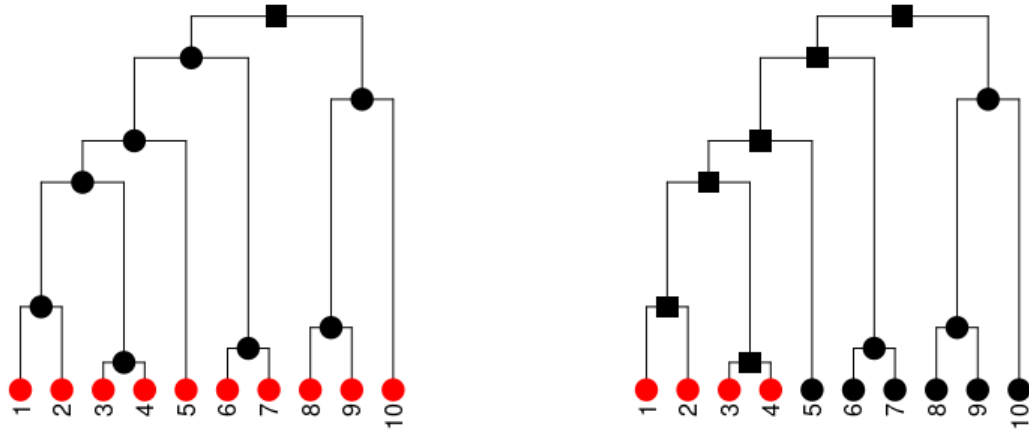


FIGURE 4 – Deux dendrogrammes illustrant des spécificités observées dans les résultats. Les classes rejetées sont représentées par un carré et les classes non-rejetées sont représentées par un cercle. Les feuilles ne pouvant être testées, elles sont représentées par un cercle. Les feuilles qui correspondent aux positions génomiques identifiées par la procédure sont représentées en rouge.

position de la classe  $\{8, 9, 10\}$ . Cet exemple illustre que, pour une région identifiée de taille significative, on ne possède pas d'information sur le nombre d'interactions différentielles ni sur leur localisation précise.

Sur le dendrogramme de droite, la classe contenant toutes les positions est également rejetée. Néanmoins, comme la classe  $\{1, 2, 3, 4, 5\}$  est, cette fois-ci, aussi rejetée, ce n'est pas la classe globale qui est identifiée par la méthode. Ainsi, bien que des interactions différentielles puissent exister entre les classes  $\{1, 2, 3, 4, 5, 6, 7\}$  et  $\{8, 9, 10\}$ , cette information est ignorée dans la suite de l'exploration de la hiérarchie. Sur ce même dendrogramme, on peut également pointer une limite de la méthode liée au fait de descendre au maximum le long de la hiérarchie : les classes  $\{1, 2\}$  et  $\{3, 4\}$  sont identifiées par la méthode mais de manière disjointe. En effet, on déduit de la procédure de test qu'il existe au moins une interaction différentielle dans  $\{1, 2\}$  et de même dans  $\{3, 4\}$ . Cependant, bien que la classe  $\{1, 2, 3, 4\}$  ait été rejetée par la méthode, on ne sait pas s'il existe des interactions différentielles entre les classes  $\{1, 2\}$  et  $\{3, 4\}$ .

Ces exemples illustrent certaines limites de la méthode et démontrent la nécessité d'un travail supplémentaire d'analyse et de représentation des résultats afin de tirer un maximum d'information de la procédure de test et d'obtenir des régions génomiques d'intérêt d'un point de vue biologique. Les perspectives envisagées en ce sens ainsi qu'une analyse plus complète des résultats seront discutées lors de la présentation.



## Remerciements

Ce travail est soutenu par le groupe de travail ChrocoNET financé par le métaprogramme INRAE DIGIT-BIO. La thèse d'Élise Jorge est financée par INRAE.

## Bibliographie

- [1] Erez Lieberman-Aiden, Nynke L. Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner, Richard Sandstrom, Bradley Bernstein, M.A. Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander, and Job Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009.
- [2] Malte Spielmann, Darío G Lupiáñez, and Stefan Mundlos. Structural variation in the 3d genome. *Nature Reviews Genetics*, 19(7):453–467, 2018.
- [3] Darío G Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, Hülya Kayserili, John M Opitz, Renata Laxova, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5):1012–1025, 2015.
- [4] Aaron T.L. Lun and Gordon K. Smyth. diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics*, 16:258, 2015.
- [5] Mohamed Nadhir Djekidel, Yang Chen, and Michael Q. Zhang. FIND: difFerential chromatin INteractions Detection using a spatial Poisson process. *Genome Research*, 28:412–422, 2018.
- [6] John C. Stansfield, Kellen G. Cresswell, Vladimir I. Vladimirov, and Mikhail G. Dozmorov. HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *BMC Bioinformatics*, 19:279, 2018.
- [7] John C. Stansfield, Kellen G. Cresswell, and Mikhail G. Dozmorov. multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments. *Bioinformatics*, 2019. Forthcoming.
- [8] Abbas Roayaei Ardakany, Ferhat Ay, and Stefano Lonardi. Selfish: discovery of differential chromatin interactions via a self-similarity measure. *Bioinformatics*, 35(14):i145–i153, 2019.
- [9] Kate B. Cook, Borislav H. Hristov, Karine G. Le Roch, Jean-Philippe Vert, and William Stafford Noble. Measuring significant changes in chromatin conformation with ACCOST. *Nucleic Acids Research*, 48(5):2303–2311, 2020.

- [10] Pierre Neuvial, Nathanaël Randriamihamison, Marie Chavent, Sylvain Foissac, and Nathalie Vialaneix. A two-sample tree-based test for hierarchically organized genomic signals. *Journal of the Royal Statistical Society, Series C*, 2024. Forthcoming.
- [11] Nathanaël Randriamihamison, Nathalie Vialaneix, and Pierre Neuvial. Applicability and interpretability of Ward’s hierarchical agglomerative clustering with or without contiguity constraints. *Journal of Classification*, 38:363–389, 2021.
- [12] R. John Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.
- [13] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188, 2001.
- [14] Jelle J. Goeman and Aldo Solari. Multiple testing for exploratory research. *Statistical Science*, 26(4):584–597, November 2011. Publisher: Institute of Mathematical Statistics.
- [15] Carlo Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3–62, 1936.
- [16] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.
- [17] Daniel Yekutieli. Hierarchical false discovery rate – Controlling methodology. *Journal of the American Statistical Association*, 103(481):309–316, 2008.
- [18] Nicolai Meinshausen. Hierarchical testing of variable importance. *Biometrika*, 95(2):265–278, 2008.
- [19] Rosa J. Meijer and Jelle J. Goeman. A multiple testing method for hypotheses structured in a directed acyclic graph. *Biometrical Journal*, 57(1):123–143, 2015. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.201300253>.
- [20] Rosa J. Meijer, Thijmen J. P. Krebs, and Jelle J. Goeman. A region-based multiple testing method for hypotheses ordered in space or time. *Statistical Applications in Genetics and Molecular Biology*, 14(1):1–19, 2015.
- [21] Boyan Bonev, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L Papadopoulos, Yaniv Lubling, Xiaole Xu, Xiaodan Lv, Jean-Philippe Hugnot, Amos Tanay, and Giacomo Cavalli. Multiscale 3D genome rewiring during mouse neural development. *Cell*, 171(3):557–572.e24, 2017.