

MISE AU POINT D'UN PIPELINE D'ANALYSE DES MARQUES D'HISTONES CHEZ LE POULET ACCLIMATÉ À LA CHALEUR

Gimonnet Coralie¹, David Sarah-Anne¹, Vitorino Carvalho Anaïs¹, Piégu Benoît², Hennequet-Antier Christelle¹, Brionne Aurélien¹, Bigot Yves², Collin Anne¹, Coustham Vincent^{1*}

¹ BOA, INRA, Université de Tours, 37380 Nouzilly, France

² PRC, INRA, CNRS, IFCE, Université de Tours, 37380 Nouzilly, France

vincent.coustham@inra.fr

RÉSUMÉ

Les poulets de chair sélectionnés pour leur croissance rapide sont sensibles aux variations de température. L'application d'un traitement de thermo-manipulation à la chaleur (TM) pendant l'embryogenèse permet d'accroître la tolérance à la chaleur du poulet à l'âge d'abattage (J35). Nous cherchons à montrer si les changements d'expression induits par le traitement précédemment montrés pourraient être expliqués par des reprogrammations épigénétiques mises en place pendant l'incubation et persistant au cours du développement. Pour cela, nous nous sommes intéressés à deux modifications post-traductionnelles des histones : la triméthylation des lysines 4 et 27 de l'histone 3 (H3K4me3 et H3K27me3). La distribution de ces deux marques à partir d'échantillons d'hypothalamus et de muscle à J35 chez des individus contrôles et TM a été obtenue par immunoprécipitation de la chromatine suivi d'un séquençage systématique (ChIP-seq). Afin d'analyser les données de séquençage, nous avons développé une chaîne de traitements bioinformatiques intégrant la spécificité de chacune de ces marques. Suite à l'analyse biostatistique, les pics différentiels ont été associés aux gènes co-localisés. L'étude de ces marques a révélé un impact de la TM sur l'épigénome principalement dans l'hypothalamus, suggérant l'existence d'un impact épigénétique du traitement embryonnaire persistant jusqu'à 35 jours d'âge.

ABSTRACT

Development of an analysis pipeline for histone marks of broiler acclimated to heat

Broilers selected for their fast growth are sensitive to temperature variation. Cyclic increases in egg incubation temperature (TM) were shown to improve heat tolerance of broilers at slaughter age (D35). We hypothesized that TM may induce phenotype alteration through the epigenetic reprogramming of gene expression set up during incubation and persisting throughout development. To investigate the molecular basis of TM, we studied the trimethylation on histone 3 of both lysine 4 and 27 (H3K4me3 and H3K27me3). The distribution of these two marks was studied in hypothalamus and muscle samples at D35 from control and TM broilers by chromatin immunoprecipitation, followed by high throughput sequencing (ChIP-seq). To analyze the sequencing data, we developed a bioinformatic pipeline taking into account the specificity of each mark. Differential peaks were categorized according to the genomic features where they were mapped. The study of these histone marks revealed an impact of heat treatment during incubation on the hypothalamus epigenome. This suggests that an epigenetic memory persists during heat treatment in this tissue.

INTRODUCTION

Les poulets de chair ont été sélectionnés génétiquement sur des critères de performance de croissance. Cependant l'accroissement de leur masse musculaire n'a pas été accompagné par un développement des systèmes cardio-vasculaire et respiratoire (Havenstein et al., 2003), affectant ainsi leur capacité de résistance aux fluctuations de température. Pour pallier à ce manque de résistance thermique des poulets de chair, il est possible de les acclimater à la chaleur dès l'incubation en appliquant un traitement de thermo-manipulation embryonnaire à la chaleur (TM ; Figure 1). Jusqu'à l'âge d'abattage (J35), ce traitement améliore la tolérance à la chaleur des individus avec une baisse significative de la mortalité chez les mâles (Piestun et al., 2008). Une trentaine de gènes différentiellement exprimés en réponse à la TM a notamment été mise en évidence dans une étude du transcriptome du muscle *Pectoralis major* du poulet à J35 (Loyau et al., 2016). L'expression différentielle de ces gènes pourrait être liée à une modification d'une partie de l'épigénome au cours du traitement embryonnaire à la chaleur. L'épigénome est défini par l'ensemble des modifications épigénétiques impliquées dans la régulation de l'expression des gènes sans en altérer la séquence. Les modifications épigénétiques les plus étudiées sont la méthylation de l'ADN et les modifications post-traductionnelles des histones. Les histones sont des protéines liées à l'ADN. L'ajout d'une marque épigénétique (méthylation, acétylation, phosphorylation, etc) sur leur extrémité N-terminale influe sur la compaction de la chromatine (ADN + protéines) et par conséquent sur l'expression des gènes.

L'objectif de notre étude était de caractériser la distribution chez les poulets TM et contrôles (C) de deux marques d'histone (H3K4me3 et H3K27me3) dans le muscle et l'hypothalamus. Ces marques sont connues pour leur contribution dans la reprogrammation épigénétique de l'expression des gènes (Feil et Fraga, 2012). L'hypothalamus et le muscle sont deux tissus impliqués dans la thermorégulation

respectivement au niveau périphérique et central. Pour cela, la chromatine de trois échantillons TM et trois échantillons C a été immunoprécipitée pour chaque marque et chaque tissu, puis séquencée à haut débit (ChIP-seq). Cette communication décrit la mise en place des traitements bioinformatiques et biostatistiques de données ChIP-seq, dédiés à l'analyse de chaque marque d'histone étudiée.

1. MATÉRIELS ET MÉTHODES

1.1 ChIP-seq

La chromatine de muscle et d'hypothalamus de 3 échantillons TM et 3 C de poulets à J35 a été immuno-précipitée avec des anticorps dirigés contre les marques H3K4me3 (Anti-trimethyl-Histone H3 K4 Cat. #07-473, Millipore) et H3K27me3 (Anti-Trimethyl-Histone H3 K27 Cat. #07-449, Millipore). Des librairies des ADN immunoprécipités ont été réalisées (Kit NEB Next II, New England Biolabs) et séquencées par un séquenceur Illumina Hi-Seq 4000. Des échantillons C et TM de muscle et d'hypothalamus dont la chromatine n'a pas été immunoprécipitée (INPUT) ont également été séquencés afin de quantifier le bruit de fond présent dans nos échantillons et ainsi normaliser les résultats obtenus. À l'issue du séquençage, des fragments de 50pb non paires ont été obtenus.

1.2 Analyses bioinformatiques

La qualité des séquences obtenues après séquençage a été vérifiée avec FASTQC (Babraham bioinformatics). Les séquences ont été cartographiées sur le génome du poulet (*Gallus gallus* 5.0) avec Bowtie2 (Langmead and Salzberg, 2012). Les duplicats de séquences ont été supprimés avec SAMtools (Li et al., 2009). Les enrichissements de marques d'histones ont été détectés séparément entre échantillons C et TM et de façon indépendante pour chaque marque et chaque tissu. L'enrichissement en H3K4me3 a été calculé par condition en tenant compte des INPUT respectifs en utilisant l'outil PePr (Zhang et al., 2014). Les enrichissements de la marque H3K27me3 ont

été calculés par échantillon relatif à leur propre INPUT avec epic/SICER (Xu et al., 2014 ; voir Résultats et Discussion partie 2.1). Suite à ces détections, pour chaque marque et chaque tissu, l'union des pics C et TM (Figure 2) a été réalisée afin d'obtenir les pics communs mais également les pics spécifiques de chaque condition pour définir les coordonnées sur lesquelles l'analyse différentielle sera effectuée. Le comptage des lectures sur ces coordonnées a finalement été réalisé avec featureCount (Liao et al., 2014).

1.3 Analyses biostatistiques

À partir du comptage, pour chaque marque et chaque tissu, une analyse différentielle a été effectuée avec le paquet R edgeR version 3.20.8 (Robinson et al., 2010, McCarthy, 2012). Les comptages de lectures des pics détectés sont normalisés (Dillies et al., 2013) par la méthode TMM (trimmed mean of M-values). Un modèle linéaire généralisé est estimé pour chaque pic afin d'identifier les pics différentiels entre les animaux C et TM. Une correction de Benjamini-Hochberg (BH) est appliquée pour contrôler le taux de faux positifs dans le cadre des tests multiples (Benjamini, Y. and Y. Hochberg, 1995). Les pics différentiels co-localisant avec des gènes ont alors été annotés avec GenomeFeatures (<https://github.com/abrionne/GenomeFeatures>).

1.4 Visualisation des pics

Pour visualiser les pics, un script utilisant les outils BEDtools (Quinlan et al., 2010) et bigwigCompare (deepTools suite, Ramirez et al., 2014) a été écrit afin de convertir les fichiers de cartographie (.bed) en fichier BigWig (.bw) représentant les profils normalisés des enrichissements des deux marques d'histones. Le profil des pics différentiels et leurs coordonnées ont ensuite été visualisés avec IGV (Integrative Genome Viewer, Robinson et al., 2011).

2. RÉSULTATS ET DISCUSSION

Nous avons exploré différents logiciels et paramétrages afin de mieux caractériser la

distribution des marques de triméthylation de la lysine 4 et 27 de l'histone 3 (H3K4me3 et H3K27me3).

2.1 Choix de l'outil de détection de pics

Afin d'optimiser la détection de pics, nous avons d'abord sélectionné des outils de détection adaptés à chaque marque. En effet, la marque H3K4me3 est une marque connue pour être enrichie de manière très localisée sur le génome (« pics étroits ») tandis que la marque H3K27me3 est une marque présentant des profils d'enrichissement plus étalés (« pics larges » ; Figure 3, Barski et al., 2007). Ainsi, la détection des pics H3K4me3 a été réalisée avec l'outil PePr (Zhang et al., 2014) adapté aux pics étroits alors que la détection des pics H3K27me3 a été réalisée avec l'outil epic (basé sur SICER, Xu et al., 2014) adapté à la détection de pics larges. La détection des pics H3K27me3 a également été réalisée avec MACS2 (Zhang et al., 2008), outil largement utilisé, l'outil epic obtient cependant de meilleurs résultats en termes de détection de pics (David et al., 2017).

2.2 Impact de la déduplication

Alors que la taille moyenne des pics obtenus pour la marque H3K27me3 du muscle et de l'hypothalamus était conforme à la taille moyenne attendue, la taille moyenne des pics obtenus pour la marque H3K4me3 dans le muscle (3kb), était supérieure à la taille de pics attendue (1 kb, Barski et al., 2007). Nous avons donc testé l'impact de la déduplication avec l'outil SAMtools, consistant en la suppression de duplicats de PCR, sur la détection des pics avec PePr. En effet, la comparaison des visualisations des cartographies dédupliquées et non dédupliquées laissait suggérer un écrasement des pics après déduplication. Cet écrasement des pics après déduplication pourrait être lié à la très grande profondeur de séquençage utilisée pour la marque H3K4me3 jusqu'à trois fois supérieure aux recommandations ENCODE (Landt et al., 2012). Ainsi, l'étape de déduplication pourrait mener à une suppression artéfactuelle de séquences identiques liées à l'étroitesse du pic. Dans

notre étude, après avoir effectué la détection des pics H3K4me3 dans le muscle, le retrait de l'étape de déduplication a réduit la largeur des pics à la taille attendue (Tableau 1). L'étape de déduplication reste cependant recommandée pour les techniques de séquençage haut débit (Ebbert et al., 2016). Nous avons donc cherché d'autres alternatives pour réduire la taille des pics.

2.3 Mise au point de la taille de fenêtre de détection de pics

Par défaut, la taille de fenêtre de détection de pics est un paramètre calculé automatiquement par l'outil PePr. Elle a été définie pour la marque H3K4me3 à 240 pb dans l'hypothalamus et 1000 pb dans le muscle, ce qui pourrait expliquer les différences de tailles de pics entre ces deux tissus. Afin de comprendre l'impact de la taille de fenêtre de détection de pics sur la taille moyenne des pics détectés, nous avons testé différentes tailles de fenêtre. Une taille de fenêtre de détection des pics fixée à 240 pb produit une taille moyenne de pics d'environ 1kb, ce qui est cohérent avec ce qui était attendu et l'observation des profils des pics (Figure 4). Nous avons donc fixé la taille de fenêtre de détection de pics pour la marque H3K4me3 dans le muscle à 240 pb dans l'outil PePr pour la suite des analyses.

2.4 Analyses différentielles

Après union des pics C et TM, le comptage des lectures sur les coordonnées de l'union des pics a été réalisé afin de permettre

d'effectuer une analyse différentielle entre pics C et TM. L'analyse différentielle des marques H3K4me3 et H3K27me3 entre échantillons C et TM a révélé un impact de la TM principalement sur les marques H3K4me3 au niveau de l'hypothalamus avec plus de 700 pics différentiels (p-valeur ajustée par BH < 0,05) contre environ 150 pics différentiels pour la marque H3K27me3. Au contraire, les marques H3K4me3 et H3K27me3 semblent peu impactées dans le muscle avec une cinquantaine de pics différentiels pour les deux marques. L'analyse fonctionnelle des gènes sur lesquels ont été trouvés des pics différentiels est en cours afin de mettre en évidence des voies biologiques d'intérêt. Ces résultats suggèrent qu'une mémoire épigénétique de la TM pourrait persister dans l'hypothalamus. D'autres mécanismes peuvent être à l'origine de la réponse physiologique et métabolique dans le muscle.

CONCLUSION

L'exploration des outils et de leurs paramètres a permis de mettre au point un pipeline d'analyse des marques d'histone H3K4me3 et H3K27me3 chez le poulet en cours de développement avec le gestionnaire Nextflow. L'utilisation de ce pipeline pour l'analyse différentielle entre poulet C et TM suggère qu'il pourrait exister une mémoire épigénétique du traitement embryonnaire dans l'hypothalamus. D'autres mécanismes peuvent être impliqués dans la réponse à la TM du muscle.

Remerciements

Cette étude a été financée par un projet INRA PHASE CI STRESSEPIMARK et de l'ANR JCJC QuailHeatE (ANR-15-CE02-0009-01). Les auteurs remercient l'équipe MOQA pour l'aide technique et l'UE PEAT pour l'élevage des animaux.

RÉFÉRENCES BIBLIOGRAPHIQUES

Barski, Artem, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, et Keji Zhao. 2007. *Cell* 129 (4): 823-37.
Benjamini, Y. and Y. Hochberg (1995). *Journal of the Royal Statistical Society Series B-Methodological* 57(1): 289-300.

- David S-A, Piégu B, Hennequet-Antier C, Pannetier M, Aguirre-Lavin T, Crochet S, Bordeau T, et al. 2017. *Biological Procedures Online*
- Dillies M.-A., et al. 2013. *Briefings in Bioinformatics* 14(6): 671-683.
- Ebbert, Mark T. W., Mark E. Wadsworth, Lyndsay A. Staley, Kaitlyn L. Hoyt, Brandon Pickett, Justin Miller, John Duce, John S. K. Kauwe, et Perry G. Ridge. 2016. *BMC Bioinformatics* 17 (Suppl 7).
- Feil, Robert, et Mario F. Fraga. 2012. *Nature Reviews Genetics* 13 (2): 97-109.
- Havenstein, G. B., P. R. Ferket, et M. A. Qureshi. 2003. *Poultry Science* 82 (10): 1500-1508.
- Landt, Stephen G., Georgi K. Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, Serafim Batzoglou, Bradley E. Bernstein, et al. 2012. *Genome Research* 22 (9): 1813-31.
- Langmead, Ben, et Steven L. Salzberg. 2012. *Nature Methods* 9 (4): 357-59.
- Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, et 1000 Genome Project Data Processing Subgroup. 2009. *Bioinformatics (Oxford, England)* 25 (16): 2078-79.
- Liao, Yang, Gordon K. Smyth, et Wei Shi. 2014. *Bioinformatics (Oxford, England)* 30 (7): 923-30.
- Loyau T, Hennequet-Antier C, Coustham V, Berri C, Leduc M, Crochet S, Sannier M, et al. 2016. *BMC Genomics* 17: 329.
- McCarthy, DJ, Chen, Y, Smyth, GK. 2012. *Nucleic Acids Research* 40, 4288-4297
- Piestun, Y., D. Shinder, M. Ruzal, O. Halevy, J. Brake, et S. Yahav. 2008. *Poultry Science* 87 (8): 1516-25.
- Quinlan, Aaron R., et Ira M. Hall. 2010. *Bioinformatics (Oxford, England)* 26 (6): 841-42.
- Ramírez, Fidel, Dündar F, Diehl S, Björn A. Grüning, et Manke T. 2014. *Nucleic Acids Research* 42 (Web Server issue): W187-191.
- Robinson, Mark D., Davis J. McCarthy, et Gordon K. Smyth. 2010. *Bioinformatics (Oxford, England)* 26 (1): 139-40.
- Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, et Jill P. Mesirov. 2011. *Nature Biotechnology* 29 (1): 24-26.
- Xu, Shiliyang, Sean Grullon, Kai Ge, et Weiqun Peng. 2014. *Methods in molecular biology (Clifton, N.J.)* 1150: 97-111.
- Zhang, Yong, Tao Liu, Clifford A Meyer, Eeckhoutte J, David S Johnson, Bradley E Bernstein, Chad Nusbaum, et al. 2008. *Genome Biology* 9 (9): R137.
- Zhang, Yanxiao, Yu-Hsuan Lin, D. Johnson T, S. Rozek L, et A. Sartor M. 2014. *Bioinformatics* 30 (18): 2568-75.

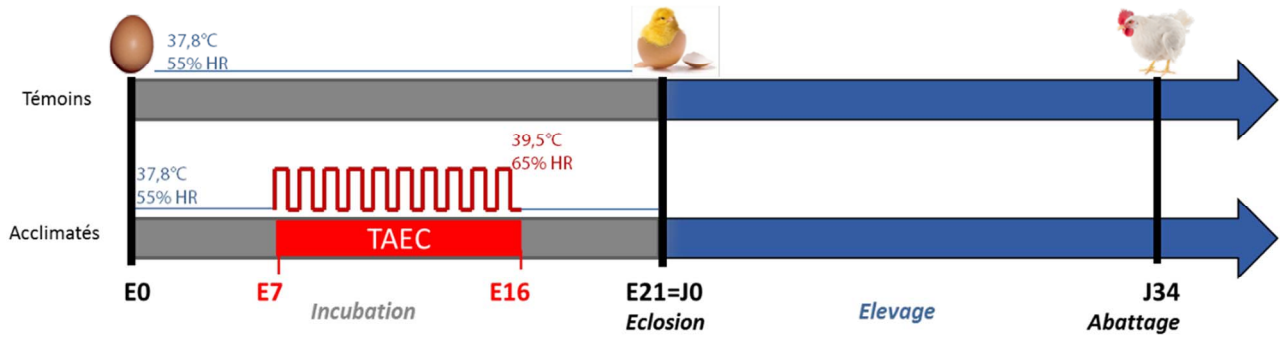
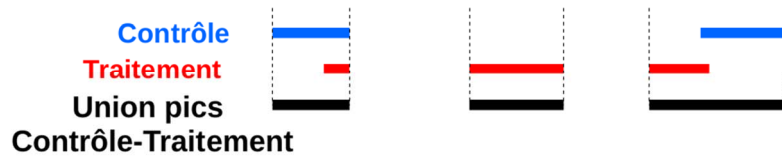
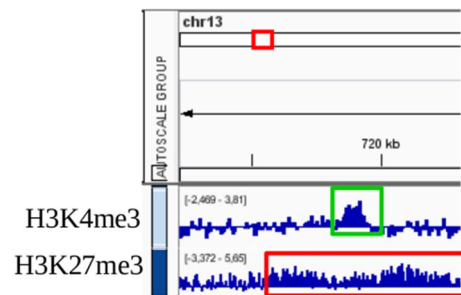
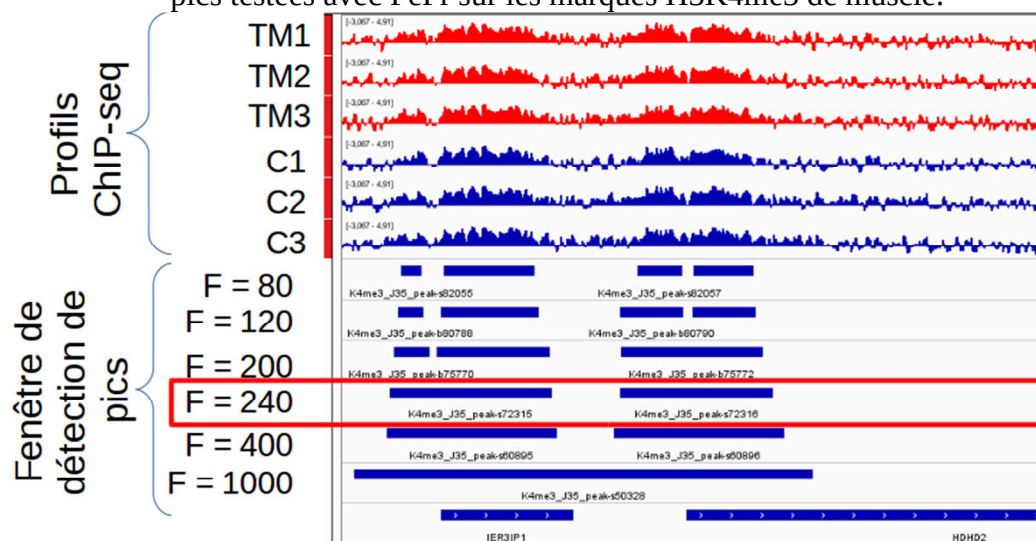
Figure 1. Protocole expérimental d'acclimation embryonnaire à la chaleur.**Figure 2.** Union entre les pics contrôles et les pics thermo-manipulés.**Figure 3.** Enrichissement des différentes marques. Les pics sont encadrés en vert pour H3K4me3 (pic étroit) et rouge pour H3K27me3 (pic large).**Figure 4.** Visualisation des profils ChIP-seq des échantillons contrôles (C) et thermo-manipulés (TM) et des coordonnées des pics détectés selon les différentes tailles de fenêtre de détection de pics testées avec PePr sur les marques H3K4me3 de muscle.

Tableau 1. Tailles moyennes de pics H3K4me3 détectés avec PePr sur des cartographies dédupliquées ou non-dédupliquées.

	Taille moyenne des pics
Théorie (Barski et al., 2007)	1kb
H3K4me3 – Hypothalamus Dédupliquées (N = 3)	1,15kb
H3K4me3 – Muscle Dédupliquées (N=3)	3,07kb
H3K4me3 – Muscle Non-dédupliquées (N=3)	1,01kb