

La méthylation de l'ADN est caractéristique de l'inactivation épigénétique et de l'hétérochromatine dans les plantes et les mammifères, elle est généralement associée à un état réprimé de la chromatine et est largement confinée aux éléments transposables (TE) et autres séquences répétées. Les éléments transposables, en particulier les rétrotransposons, sont connus pour être activés par des stress biotiques et abiotiques (Wessler 1996) grâce à des changements épigénétiques comme la réduction de la méthylation de l'ADN. Toutefois, l'insertion de nouveaux TE dans le génome dépend d'événements non seulement au niveau d'une cellule, mais aussi au niveau de la population car il est difficile pour une variante génétique d'être fixée, sauf dans de petites populations consanguines (Husband 2004; Belyayev, et Kalendar al. 2010). Les transposons, en raison de leur capacité inhérente de se déplacer d'une position à une autre dans un génome, sont potentiellement délétères, et sont les principales cibles de la méthylation de l'ADN dirigé par l'ARN (RdDM) qui est médiée par le silencing des gènes au niveau transcriptionnel (TGS). Le RdDM est important pour l'établissement de la méthylation de l'ADN dans le SEV et la comparaison des modèles de méthylation du génome entier et de petits ARN chez *Arabidopsis* a indiqué qu'au moins un tiers des locus méthylés est associé à des groupes de siRNA (Lister, O'Malley et al. 2008). Cependant, les mécanismes permettant à la machinerie RdDM de reconnaître et d'agir spécifiquement sur les transposons et non pas sur les gènes de l'hôte restent à élucider. Il a été postulé que les transposons peuvent posséder certaines caractéristiques, comme des structures secondaires spécifiques ou de longs segments non traduits qui les rendent plus sensibles au RdDM. Cependant, plusieurs études chez *Arabidopsis* ont indiqué que les transposons ne possèdent pas de caractéristique intrinsèque au niveau de l'ADN permettant cette fonction de reconnaissance (Miura, Yonebayashi et al 2001; Kato, Takashima et al 2004; Lippman, Gendrel et al 2004; Mirouze, Reinders et al 2009). Une autre possibilité peut être leur capacité à produire des ARNs double brin (dsRNA) soit par une transcription convergente/divergente des deux brins ou à travers l'action d'une ARN-polymérase ARN dépendante (RDR) qui permettent la synthèse de dsRNA précurseurs pour produire les siRNA qui ciblent la machinerie RdDM aux séquences de transposons dans le génome.

La perte de méthylation de l'ADN dans les mutants de méthyltransférases de l'ADN ou de la protéine de remodelage de la chromatine DDM1 (diminution de 1 méthylation de l'ADN), de-réprime et mobilise plusieurs séquences de TEs (Miura, Yonebayashi et al 2001; Lippman,

Gendrel et al 2004; Mirouze, Reinders et al. 2009). Certains transposons réactivés dans ces mutants restent actifs même après réinsertion dans un fond sauvage, indiquant une transmission épigénétique de l'état d'activation du TE. Cependant d'autres transposons sont reméthylés après introduction dans un fond de type sauvage, indiquant qu'un mécanisme efficace de reméthylation peut corriger les défauts transgénérationnels de méthylation de l'ADN (Teixeira, Heredia et al. 2009). Cette reméthylation sélective et progressive est facilitée par l'ARN interférence (ARNi) et dépend de l'activité de RDR2 (ARN-polymérase ARN dépendante), une composante connue du RdDM.

La méthylation de l'ADN dirigée par l'ARN est donc considérée comme un facteur important qui dicte la mise en place de la méthylation et sa transmission. Elle est souvent considérée comme étant associée à la méthylation des séquences denses en transposons aux sites CG et non-CG. Cependant, mon analyse dans le chapitre II de cette thèse montre qu'un nombre étonnamment élevé de transposons avec une dense méthylation est caractérisé par l'absence ou la quasi-absence de siRNA correspondants. J'ai fourni la preuve que dans de tels cas, la méthylation de l'ADN est probablement le résultat de propagation à partir de séquences flanquantes, ciblées elles-mêmes par les siRNAs. Cette propagation de la méthylation de l'ADN ciblée par des siRNAs pourrait être considérée à la fois dans l'hétérochromatine et l'euchromatine quoique à des degrés divers (~ 500 pb et 200 pb ~, respectivement), ce qui reflète soit un rôle de l'hétérochromatine dans la propagation de la méthylation de l'ADN, soit un effet d'inhibition de l'euchromatine, ou encore une activité plus élevée de la déméthylation de l'ADN dans l'euchromatine. Les résultats présentés dans le chapitre II permettent de fournir la preuve que le phénomène de propagation de la méthylation de l'ADN pourrait résulter d'un ensemble complexe d'interactions entre les différentes ADN méthyltransférases qui favorisent ou limitent l'ampleur de la propagation et que la contribution de la déméthylation de l'ADN n'est pas significative. Toutefois, la déméthylation active de l'ADN ne peut être exclue pour un petit sous-ensemble de séquences de TEs méthylés situés à proximité de gènes essentiels, comme récemment rapporté par Lister et al., où un sous-ensemble d'insertions de TEs à proximité de gènes est méthylé dans le sauvage et hyperméthylé dans le triple mutant rdd (Lister, O'Malley et al. 2008).

J'ai également fourni la preuve que les séquences de TE dans l'euchromatine sont préférentiellement insérées dans la chromatine ayant une conformation ouverte et sont plus abondants près de gènes. Toutefois, la distribution des séquences non méthylées par rapport aux séquences méthylées est quelque peu biaisée à l'extrémité 5' des gènes. De plus, les séquences de TE méthylées associées à des siRNA ont tendance à être moins abondantes et uniformément distribuées dans les 500 premières paires de bases en amont de l'extrémité 5' des gènes. Ces résultats suggèrent donc que les séquences méthylées de TE ont des effets délétères sur l'initiation de la transcription et que ces effets sont plus graves lorsque les séquences méthylées de TE sont associées à des siRNA. Étant donné que la méthylation de l'ADN a un effet inhibiteur sur l'activité du promoteur, il est possible que le niveau de la méthylation de l'ADN puisse expliquer en partie que les TE méthylés soient préférentiellement contre-sélectionnés à la proximité des gènes.

À la lumière de ces nouveaux résultats, il serait intéressant de déterminer les insertions de novo de séquences de TE et d'examiner comment ils affectent le contexte épigénomique des gènes environnants et influencent leur expression. Alors que les membres individuels d'une famille de transposons sont hautement conservés, leurs séquences flanquantes sont susceptibles d'être uniques. Une approche expérimentale pour déterminer les nouvelles insertions de transposons serait donc probablement basée sur l'identification des séquences contiguës d'ADN flanquant les TEs soit par séquençage en profondeur ou par l'utilisation de la technologie des biopuces pour capturer les frontières TE (Gabriel, Dapprich et al 2006;.. Wheelan, Scheifele et al 2006). Ces données pourraient être combinées avec l'ensemble des cartes de marques épigénétiques du génome comme la méthylation de l'ADN pour déterminer les délais de l'établissement de nouveaux états épigénétiques et suivre la stabilité de ces modifications épigénétiques au cours des générations.

Au chapitre III, la cartographie du génome à l'échelle d'un vaste ensemble de 11 marques d'histones et de la méthylation de l'ADN a été réalisée. Nous avons d'abord montré que même si un grand nombre de permutations théoriques de méthylation de l'ADN et des marques d'histones sont possibles, seul un nombre limité de telles combinaisons sont effectivement observées, ce qui reflète la probabilité qu'à un locus spécifique, des combinaisons de marques chromatiniennes sont impliquées dans les mécanismes de contrôle épigénétique et ne sont pas simplement des événements fortuits aléatoires. Une

analyse plus approfondie en utilisant le flou c-means partitionné, montre que l'épigénome d'*Arabidopsis* peut être divisé en quatre grands types d'état chromatinien avec des propriétés fonctionnelles distinctes. Ces états chromatiniens peuvent être distingués principalement en deux groupes, comme en témoignent les analyses par paires co-association : l'hétérochromatine constitutive (noté CS3) d'une part et trois états différents de régions euchromatiques (notés CS1, CS2 et CS4). Le premier état de l'euchromatine (CS1) correspond à des gènes transcriptionnellement actifs et est généralement enrichi en H3K4me3, H3K36me3, H3K9me3, H2Bub, H3K4me2 et H3K56Ac. CS2 représente la chromatine répressive et correspond aux régions marquées par H3K27me3 qui sont principalement associés à PRC2 (Turck, Roudier et al 2007; Zhang, Clarenz et al 2007). Cet état est également associé à H3K27me2. CS3 correspond à l'hétérochromatine classique et est presque exclusivement constituée des TEs silencieux (Lippman, Gendrel et al 2004; Bernatavichute, Zhang et al 2008), elle comprend des régions marquées par H3K9me2 et H4K20me1. Cet état est souvent co-marqué par les formes mono- ou diméthylées de H3K27 et à la méthylation de l'ADN (5mC). Toutefois, ces marques sont également trouvés sur un tiers des gènes environ. Enfin, un quatrième état de chromatine (CS4) se caractérise par l'absence de toute marque répandue et est associée à des gènes faiblement exprimés et des régions intergéniques.

L'étude des épigénomes ont également été menées chez la levure (Liu, Kaplan et al 2005.), *C. elegans* (Gerstein, Lu et al 2010; Liu, Rechtsteiner et al 2011), la drosophile (Roy, Ernst et al 2010; Kharchenko, Alekseyenko et al 2011; Riddle, Minoda et al 2011) et les cellules humaines (Wang, Zang et al 2008; honorable, Wang et al 2009; Ernst et Kellis 2010; Zhou, Goren et al . 2011). Comme chez *Arabidopsis*, elles ont toutes fait ressortir une complexité relativement faible des combinaisons des marques chromatine. Les deux états définis de la chromatine silencieuse chez *Arabidopsis* (CS2 & CS3) se retrouvent chez les métazoaires, indiquant le niveau élevé de conservation de ces états. En ce qui concerne la chromatine active, les gènes transcriptionnellement actifs correspondant à une seule combinaison (CS1) chez *Arabidopsis* présentes plusieurs combinaisons possibles dans d'autres organismes. Cette différence pourrait être due au fait que le génome d'*Arabidopsis* est relativement petit avec des gènes et des régions intergéniques en moyenne plus courtes que chez d'autres

organismes, ceci suggère que la longueur du gène pourrait avoir un rôle important dans la création des états chromatinien.

En utilisant une analyse intégrée de la distribution de ces 12 marques de la chromatine, nous avons démontré que l'épigénome d'*Arabidopsis* montre que les états de chromatine sont principalement définis au niveau des unités de transcription. Notre analyse a également confirmé que H2Bub, H3K36me3, 5mC, ainsi que H3K4me2 et H3K4me3 sont préférentiellement associés aux gènes longs, H3K27me3 d'autre part marque préférentiellement les petits gènes (Luo et Lam 2010). Contrairement à ces modifications de la chromatine, H3K27me1 ne présente aucune association préférentielle par rapport à la longueur du gène. En outre, nous avons également confirmé que H3K4me3 et H3K56ac ont tendance à marquer fortement et largement des gènes exprimés (Oh, Park et al 2008; Zhang, Bernatavichute et al 2009), et que H3K27me3 est préférentiellement associé à des gènes qui sont exprimés à des niveaux inférieurs ou de façon tissu-spécifique (Turck, Roudier et al 2007; Zhang, Clarenz et al 2007; Ah, Park et al 2008), tandis que 5mC tend à marquer des gènes exprimés modérément (Zilberman, Gehring et al 2007). De plus, nous avons noté que H2Bub, H3K36me3 et H3K9me3 ont tendance à marquer les gènes fortement exprimés, comme H3K4me3 et H3K56ac. H3K4me2 ne montre pas d'association particulière avec l'expression des gènes. Les gènes marqués par H3K27me1 ou H3K27me2 sont exprimés à des niveaux inférieurs ou d'une manière tissu-spécifique. Toutefois, H3K27me1 et H3K27me2 / 3 marquent en grande partie des ensembles de gènes avec des ontologies différentes, ce qui suggère l'existence de deux systèmes distincts de la répression des gènes associés à la méthylation de H3K27. Nous avons également observé que les niveaux d'enrichissement pour la plupart des marques de chromatine sont corrélées avec les niveaux d'expression, mais si ces corrélations reflètent l'expression de gènes dans un nombre variable de cellules, ou l'enrichissement différentielle vraie par rapport au niveau d'expression, reste à déterminer. Comme chaque type cellulaire possède un épigénome unique, qui définit son programme de régulation des gènes, il n'est finalement que la connaissance des épigénomes spécifiques de type de cellules qui permettront une compréhension complète de l'impact fonctionnel de la chromatine au niveau de la régulation sur l'activité du génome. Les technologies de séquençage de nouvelle génération sont constamment améliorées afin d'obtenir des données de séquences de haute qualité à

partir d'un génome isolé d'une seule cellule. Des études épigénétiques de molécules unique par séquençage en temps réel (Pushkarev, Neff et al 2009; Korfach, Bjornson et al 2010) sont susceptibles de révolutionner la recherche en épigénétique dans l'avenir et d'éliminer de nombreux problèmes d'interprétation de l'épigénome, qui sont actuellement les produits d'une population mixte de cellules.

Enfin, dans le dernier chapitre des résultats de cette thèse (chapitre IV), une modification d'histone spécifiquement choisie (H2Bub) a été utilisée pour fournir une vision dynamique de l'épigénome et ses effets sur la transcription de la chromatine sous-jacente en réponse à la lumière. Compte tenu de la nature dynamique de la marque H2Bub sur la régulation des gènes chez la levure et les mammifères, et les connaissances fragmentaires sur les coactivateurs transcriptionnels agissant par l'intermédiaire de l'ubiquitination de H2B, j'ai choisi de me concentrer spécifiquement sur la marque H2Bub. Plus généralement, le rôle de la mono-ubiquitinylation des histones est peu connue chez les plantes, en dehors de la description récente de l'ubiquitination H2A associée à la machinerie reliée aux Polycombs et le rôle des dépôts de H2Bub dans le contrôle de l'expression du gène FLC chez *A. thaliana* (Cao Dai et al 2008; Bratzel, López-Torrejón et al 2010).

H2Bub marque plus de 10.000 gènes chez *A. thaliana*. Comme chez la levure et les mammifères, nous avons observé que H2Bub est limité aux régions transcrites et surtout associé à des gènes actifs (Roudier, Ahmed et al. 2011). Par ailleurs, la marque H2Bub chez *Arabidopsis* corrèle avec plusieurs autres marques d'histone associé à la transcription, à savoir la triméthylation de l'histone H3 sur les lysines 4, 9 et 36. Les plantes mutantes pour la H2B ubiquitine-ligase 1 (HUB1), qui est un homologue de la levure Bre1 (Fleury, Himanen et al. 2007), montrent une perte complète de la marque H2Bub sur tout le génome. Ainsi, le mutant *hub1-3* ne présentant pas de marque H2Bub visible, il fournit un outil idéal pour explorer les rôles globaux de cette marque.

Les premiers travaux en laboratoire ont révélé que les plantes mutantes d'*Arabidopsis hub1-3* présentent des défauts dans le processus de dé-étiolement, suggérant un rôle pour H2Bub dans la transition obscurité-lumière. Par conséquent, dans le chapitre IV de cette thèse, le j'ai utilisé le mutant *hub1-3* comme un outil pour corrélérer les analyses de transcriptome et épigénomiques afin d'évaluer l'impact de la dynamique de la marque H2Bub sur les

réponses transcriptionnelles. Mon analyse a indiqué que H2Bub peut ne pas être nécessaire pour déterminer les états ON/OFF d'activation des gènes, mais peut plutôt contribuer à la dynamique d'expression et de manière sélective réguler finement le niveau d'induction ou de répression de l'expression génique en réponse à des signaux lumineux. J'ai démontré que la marque H2Bub est très dynamique sur des centaines de gènes et que de changements significatifs dans les niveaux d'enrichissement se produisent sur de nombreux gènes après seulement une heure d'illumination. Seulement quelques gènes montrent une perte d'enrichissement de H2Bub après 6 heures de lumière, et nous avons observé que la régulation négative du gène n'était généralement pas associée à une perte significative de la marque. Ceci est en contraste évident avec le gain régulier de cette marque associée à l'induction de l'expression. Ces observations pourraient s'expliquer par une grande stabilité de cette marque au niveau des gènes réprimés, H2Bub pourrait alors avoir un rôle dans la répression de l'expression soit au niveau transcriptionnel, où H2Bub inhiberait la transcription, ou à un niveau post-transcriptionnelle, où HUB1 serait impliquée directement ou indirectement dans la dégradation des ARNm (Cano, Miranda-Saavedra et al. 2010). Ce résultat suggère donc d'autres expériences, comme la réalisation d'une cinétique afin de déterminer le temps moyen pour l'enlèvement de H2Bub après l'arrêt du stimulus lumineux, en fonction de l'activité des gènes. Enfin, nous avons observé que la marque H2Bub est plus prononcée sur les gènes les plus longs, alors que les gènes de moins de 2 kb n'ont pas montré de différences significatives dans l'expression entre le sauvage et hub1-3, alors que les gènes de plus de 2 kb ont montré une induction significativement plus élevée chez le sauvage par rapport à hub1-3. Ce résultat concorde avec le rôle prévu de la marque H2Bub en facilitant l'élongation de la transcription, plus les gènes sont longs, plus ils sont particulièrement sensibles à la cinétique d'éviction des nucléosomes et/ou l'ouverture de la chromatine et la processivité de l'ARN polymérase.

Dans cette thèse, j'ai fourni une vue spatiale et temporelle de l'épigénome d'*A. thaliana* en utilisant des contributions importantes de bioinformatique. Toutes les méthodes expérimentales pour la cartographie épigénomique génèrent d'énormes quantités de données qui nécessitent des méthodes efficaces pour des niveaux de base de traitement des données et de contrôle de qualité. Alors que le développement des outils informatiques et des ressources pour l'analyse de données à l'échelle de l'épigénome s'accélère

rapidement, différents défis existent qui découlent de l'analyse des données épigénomiques. Ils comprennent, mais ne sont pas limités à, des considérations de méthodes basées sur l'hybridation comme l'image et des artefacts de numérisation, correction de fond, la normalisation tableau, et pour la séquence des analyses fondées sur l'alignement appliquées à un génome de référence avec précision, rapide utilisation de la mémoire, et la flexibilité du programmes d'alignement. D'autres considérations pour les méthodes basées sur le séquençage serait de permettre aux inadéquations de plusieurs correspondances dans l'alignement, et l'exigence de plus ou jumelé lit-end. Une grande partie des algorithmes ont été écrites pour trouver les pics ou les régions à destination des ensembles de données de de CHIP-on-chip ou CHIP-seq, mais le problème de trouver les pointes n'a pas encore été entièrement résolu. Par exemple, la plupart des algorithmes disponibles de recherche de pointe ne fonctionnent pas particulièrement bien avec les modifications des histones qui couvrent des domaines étendus tels que H3K27me3 et ils semblent manquer un nombre important de régions faiblement liées pour de telles modifications. Enfin, étant donné que les données épigénomique croient à un rythme sans précédent, le développement de navigateurs de Génome puissant et convivial est la nécessité de l'heure pour faciliter une compréhension et une interprétation significative de l'information biologique à partir des masses de données.



## References:

- Belyayev, A., R. Kalendar, et al. (2010). "Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat." *Mobile DNA* **1**(1): 6.
- Bernatavichute, Y. V., X. Zhang, et al. (2008). "Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*." *PloS One* **3**(9): e3156.
- Bratzel, F., G. López-Torrejón, et al. (2010). "Keeping cell identity in *Arabidopsis* requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination." *Current Biology: CB* **20**(20): 1853-1859.
- Cano, F., D. Miranda-Saavedra, et al. (2010). "RNA-binding E3 ubiquitin ligases: novel players in nucleic acid regulation." *Biochemical Society Transactions* **38**(6): 1621-1626.
- Cao, Y., Y. Dai, et al. (2008). "Histone H2B monoubiquitination in the chromatin of FLOWERING LOCUS C regulates flowering time in *Arabidopsis*." *The Plant Cell* **20**(10): 2586-2602.
- Ernst, J. and M. Kellis (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome." *Nature Biotechnology* **28**(8): 817-825.
- Fleury, D., K. Himanen, et al. (2007). "The *Arabidopsis thaliana* homolog of yeast BRE1 has a function in cell cycle regulation during early leaf and root growth." *The Plant Cell* **19**(2): 417-432.
- Gabriel, A., J. Dapprich, et al. (2006). "Global Mapping of Transposon Location." *PLoS Genet* **2**(12): e212.
- Gerstein, M. B., Z. J. Lu, et al. (2010). "Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project." *Science (New York, N.Y.)* **330**(6012): 1775-1787.
- Hon, G., W. Wang, et al. (2009). "Discovery and annotation of functional chromatin signatures in the human genome." *PLoS Computational Biology* **5**(11): e1000566.
- Husband, B. C. (2004). "Chromosomal variation in plant evolution." *American Journal of Botany* **91**(4): 621 - 625.
- Kato, M., K. Takashima, et al. (2004). "Epigenetic control of CACTA transposon mobility in *Arabidopsis thaliana*." *Genetics* **168**(2): 961-969.
- Kharchenko, P. V., A. A. Alekseyenko, et al. (2011). "Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*." *Nature* **471**(7339): 480-5.
- Korlach, J., K. P. Bjornson, et al. (2010). "Real-time DNA sequencing from single polymerase molecules." *Methods in Enzymology* **472**: 431-455.
- Lippman, Z., A. V. Gendrel, et al. (2004). "Role of transposable elements in heterochromatin and epigenetic control." *Nature* **430**(6998): 471-6.
- Lister, R., R. C. O'Malley, et al. (2008). "Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*." *Cell* **133**(3): 523-36.
- Liu, C. L., T. Kaplan, et al. (2005). "Single-nucleosome mapping of histone modifications in *S. cerevisiae*." *PLoS Biology* **3**(10): e328.
- Liu, T., A. Rechtsteiner, et al. (2011). "Broad chromosomal domains of histone modification patterns in *C. elegans*." *Genome Res* **21**(2): 227-36.

- Luo, C. and E. Lam (2010). "ANCORP: a high-resolution approach that generates distinct chromatin state models from multiple genome-wide datasets." The Plant Journal: For Cell and Molecular Biology **63**(2): 339-351.
- Mirouze, M., J. Reinders, et al. (2009). "Selective epigenetic control of retrotransposition in Arabidopsis." Nature **461**(7262): 427-430.
- Miura, A., S. Yonebayashi, et al. (2001). "Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis." Nature **411**(6834): 212-214.
- Oh, S., S. Park, et al. (2008). "Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis." PLoS Genetics **4**(8): e1000077.
- Pushkarev, D., N. F. Neff, et al. (2009). "Single-molecule sequencing of an individual human genome." Nature Biotechnology **27**(9): 847-850.
- Riddle, N. C., A. Minoda, et al. (2011). "Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin." Genome Research **21**(2): 147-163.
- Roudier, F., I. Ahmed, et al. (2011). "Integrative epigenomic mapping defines four main chromatin states in Arabidopsis." The EMBO Journal **30**(10): 1928-1938.
- Roy, S., J. Ernst, et al. (2010). "Identification of functional elements and regulatory circuits by Drosophila modENCODE." Science (New York, N.Y.) **330**(6012): 1787-1797.
- Teixeira, F. K., F. Heredia, et al. (2009). "A role for RNAi in the selective correction of DNA methylation defects." Science **323**(5921): 1600-4.
- Turck, F., F. o. Roudier, et al. (2007). "Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27." PLoS Genetics **3**(6): e86.
- Wang, Z., C. Zang, et al. (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.
- Wessler, S. R. (1996). "Turned on by stress. Plant retrotransposons." Current Biology: CB **6**(8): 959-961.
- Wheelan, S. J., L. Z. Scheifele, et al. (2006). "Transposon insertion site profiling chip (TIP-chip)." Proceedings of the National Academy of Sciences **103**(47): 17632 -17637.
- Zhang, X., Y. V. Bernatavichute, et al. (2009). "Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana." Genome Biology **10**(6): R62.
- Zhang, X., O. Clarenz, et al. (2007). "Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis." PLoS Biol **5**(5): e129.
- Zhou, V. W., A. Goren, et al. (2011). "Charting histone modifications and the functional organization of mammalian genomes." Nat Rev Genet **12**(1): 7-18.
- Zilberman, D., M. Gehring, et al. (2007). "Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription." Nature Genetics **39**(1): 61-69.