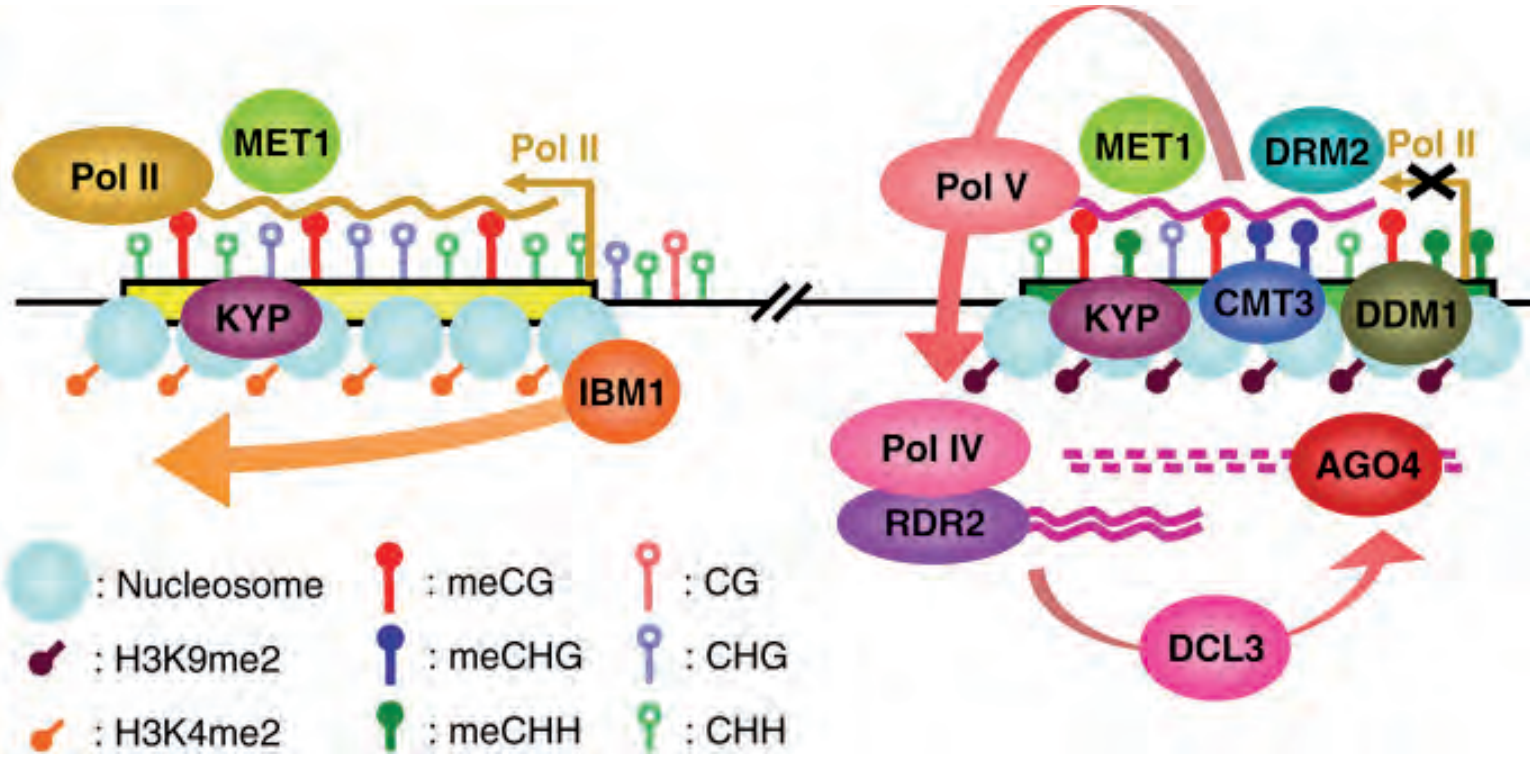
PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

# Epigenetics, an update

Jan-Peter Nap & Ad Geurts van Kessel




Report 397

PLANT RESEARCH INTERNATIONAL
WAGENINGEN UR

# Epigenetics, an update

Jan-Peter Nap[1] & Ad Geurts van Kessel[2]

Commissioned by:
The Netherlands Commission on Genetic Modification (COGEM)

[1] Wageningen University & Research centre, Wageningen
[2] Radboud University, Nijmegen

Plant Research International, part of Wageningen UR
Business Unit Bioscience
June 2011

Report 397

*Disclaimer*

The scientific contents of this report is the full responsibility of the authors only. COGEM, Supervisory Board, Wageningen University & Research centre, or Radboud University Nijmegen need not fully comply with all statements, conclusions or opinions presented in this report. Neither Wageningen University & Research centre nor Radboud University Nijmegen are responsible in any way for any damage that may be caused by the contents of this report or any use thereof.

*Figure on cover*

Epigenetic regulation of gene expression depends on the interaction of very many players (Teixeira and Colot, 2009).



**Plant Research International, part of Wageningen UR**
**Business Unit Bioscience**

Address : P.O. Box 169, 6700 AP Wageningen, The Netherlands
: Wageningen Campus, Droevendaalsesteeg 1, Wageningen, The Netherlands
Tel. : +31 317 47 70 00
Fax : +31 317 41 80 94
E-mail : info.pri@wur.nl
Internet : www.pri.wur.nl

# Table of contents

page

# Preface

In 2006, we presented an overview of the major topics and trends in epigenetic research in terms of mechanisms, examples and potential applications. The report was assigned by the Netherlands Commission on Genetic Modification (COGEM) and was aimed at providing a scientific background to informed discussions for decision and policy making with respect to genetic modification in relationship to epigenetics and its applications in the future. In view of the developments covered, it was predicted that the 'omics' angle to (epi)genetic research would accelerate the discovery and explanation of epigenetic phenomena to such an extent that the potential half-life of the 2006 report in terms of detailed explanations and models should be considered fairly limited. Therefore, it was recommended that COGEM would follow closely the developments in the field of epigenetics in order to decide whether additional measures would be necessary in the safety assessment of genetically modified organisms.

Here, we present an update of the 2006 report, including an attempt to put scientific developments into the perspective of the core interest of COGEM to allow the development of policies for the evaluation of safety and the identification of potentially adverse risks in genetically modified organisms. Epigenetic research has become an even more important part of mainstream biological research and major conceptual changes of views and models have developed in the period from 2006 on, largely fueled by technological developments that allow genome-wide analyses. The field is uncovering new levels, complexities and details in the regulation of gene expression in biological systems. This report summarizes and describes the newest data on epigenetic phenomena and mechanisms published in the extensive scientific literature. The text is based primarily on a vast number of scientific papers and reviews that was published in the latest years. Such reviews allow the interested reader to gain access to the primary literature.

Although we have tried to make this report readable on its own, implying some repeat of material in the 2006 report, the 2006 report is taken as starting point. Where we refer to this report, we also refer to all references therein. We have, more explicitly than in 2006, tried to link the developments in the field of epigenetics to the issue of safety assessment of genetically modified organisms.

Jan-Peter Nap
Ad Geurts van Kessel

June 2011
Wageningen/Nijmegen

# Management samenvatting

Epigenetica wordt gedefinieerd als veranderingen in de functie van genen die in mitose (tussen cellen) of meiose (tussen generaties) overerfbaar zijn, maar die niet terug te voeren zijn op veranderingen in de onderliggende DNA sequentie. Een epigenetische toestand blijft in principe bestaan in afwezigheid van het signaal dat de toestand veroorzaakt en is ook in principe omkeerbaar. Belangrijke componenten van epigenetische effecten zijn DNA methylering, chromatine en RNA moleculen. Vergeleken met 2006 is de kennis over de epigenetische machinerie en de organisatie daarvan sterk gegroeid. Door de epigenetica is de wereld van genregulatie veel gedetailleerder en subtieler geworden en blijkt nog complexer en veel meer geïntegreerd dan eerder gedacht. Dit kan erop wijzen dat organismen een aanzienlijke redundantie in genregulatie nodig hebben om zich op de juiste manier te ontwikkelen en op een toepasselijke manier te kunnen reageren op veranderingen in hun interne of externe omgeving. In humane systemen richten toepassingen van epigenetische modificatie zich op mitotische overerving, zoals in de ontwikkeling van epigenetische geneesmiddelen tegen kanker, maar die toepassingen betreffen in het algemeen geen genetische modificatie. In planten zijn epigenetische toepassingen veelal gericht op meiotische overerving en wel gerelateerd aan genetische modificatie. Een nieuwe toepassing van epigenetica in planten betreft de overerfbare onderdrukking van genexpressie gebaseerd op DNA methylering, waarbij de onderdrukking van expressie gerealiseerd blijft ook in de afwezigheid van het onderdrukking-inducerende transgen. Het rapport evalueert diverse relaties tussen de genetische modificatie van planten en het optreden van al dan niet bedoelde epigenetische effecten in de context van de bestaande procedures voor het vaststellen van de biologische veiligheid van transgene planten in het milieu. Het concludeert dat de bestaande protocollen voor biologische veiligheidsanalyse volstaan en afdoende zijn om een mogelijk onbedoeld fenotype veroorzaakt door epigenetische effecten vast te stellen.

# Executive summary

Epigenetics is defined as changes in gene function that are mitotically (between cells) and/or meiotically (between generations) heritable and do not entail a change in DNA sequence. Epigenetic states are self-perpetuating and potentially reversible. Major players in epigenetic effects are DNA methylation, chromatin and RNA molecules. Compared to 2006, the understanding of the epigenetic machinery and its molecular organization has extended considerably. Epigenetics has made the world of gene regulation much more detailed and subtle, but also much more complex and integrated. This may indicate the presence or need for considerable redundancy in gene regulation to allow cells and organisms to develop properly and respond appropriately to changes in internal or external environments. In human systems, applications of epigenetic modification focus on mitotic inheritance, such as the use of epigenetic drugs for the treatment of cancer, but such applications generally do not involve genetic modification. In plant systems, epigenetic applications are generally associated with genetic modification. A new application of epigenetics entails heritable gene silencing based on the use of DNA methylation, in which silencing of an endogenous gene is accomplished in the absence of the transgene used to triggering the silencing. Various relationships between genetic modification of plants and the occurrence of possibly unintended epigenetic effects are here considered in the context of the existing protocols for the environmental biosafety assessment of genetically modified plants. Overall, it is concluded that the existing frameworks of biosafety assessment are in place and sufficient to detect any adverse phenotype that could be caused by epigenetic effects.

# Samenvatting

In dit rapport wordt epigenetica gedefinieerd als veranderingen in de functie van genen die in mitose of meiose overerfbaar zijn, maar die niet terug te voeren zijn op veranderingen in de onderliggende DNA sequentie. Een epigenetische toestand blijft bestaan in afwezigheid van het signaal dat de toestand veroorzaakt (signaal-onafhankelijke handhaving) en die toestand is in principe omkeerbaar. Belangrijke componenten in de epigenetica zijn: (a) DNA methylering, (b) chromatine, de combinatie van DNA en eiwit in de eukaryote kern en (c) RNA moleculen, oorspronkelijk klein, maar nu ook (heel) lang. Signaal-onafhankelijke handhaving is momenteel het best gedocumenteerd voor DNA methylering. Opgemerkt moet worden dat er onder onderzoekers nog aanzienlijke discussies zijn over de precieze definitie en reikwijdte van epigenetica, vooral met betrekking tot het concept van signaal-onafhankelijke handhaving.

Epigenetische genregulatie is belangrijk voor de juiste ontwikkeling van zowel planten als dieren. Overdracht van epigenetische informatie binnen een organisme via mitose heet mitotische epigenetische overerving. Dit komt vaker voor dan overdracht naar volgende generaties via meiose in de geslachtscellen, een proces dat meiotische of transgenerationele epigenetische overerving wordt genoemd. Het laatste type overerving is overtuigend beschreven in planten, mogelijk omdat in planten de geslachtscellen pas laat in de ontwikkeling worden gevormd. Minder duidelijk is het bestaan en relatieve belang van meiotische epigenetische overerving in zoogdier (menselijke) systemen. De omvang van het onderzoek naar epigenetische overerving groeit hard en er is veel interesse om toepassingen te ontwikkelen. Epigenetisch onderzoek en ontwikkeling hebben de laatste jaren aanzienlijke vooruitgang geboekt, vooral dankzij de ontwikkelingen in DNA en RNA sequencing waardoor dergelijke sequenties aanzienlijk sneller en goedkoper kunnen worden bepaald. Samen met analyses zoals chromatine immunoprecipitatie (ChIP) resulteert dit in indrukwekkende hoeveelheden data die bijdragen aan het ontcijferen en steeds beter begrijpen van de epigenetische informatie van cellen. De verwachting is dat de technologische ontwikkelingen verder zullen gaan en zullen bijdragen aan meer begrip en gebruik van epigenetische overerving. Verdere ontwikkelingen in onder meer DNA sequencing technologie zullen in de toekomst de analyse van het epigenoom van de individuele cel mogelijk maken.

Het begrip van de epigenetische machinerie en de kennis van de moleculaire organisatie daarvan is sinds 2006 aanzienlijk gegroeid. De kennis van epigenetische genregulatie is veel gedetailleerder geworden, maar deze is daarmee ook veel complexer gebleken. Nieuwe aspecten blijven ontdekt worden. Vooral in de RNA wereld zijn momenteel veel meer typen en moleculen bekend, waaronder lange niet-coderende RNA moleculen, dan in 2006. Ook het voorkomen en de regulerende rol van DNA methylering blijkt subtieler dan aanvankelijk gedacht: DNA methylering is niet alleen geassocieerd met gen silencing, maar in bepaalde gevallen ook met genactiviteit. Een nieuwe chemische modificatie van DNA, hydroxymethylcytosine, is ontdekt en kan nog onbekende regulerende functies hebben. De belangrijkste ontwikkeling is dat veel duidelijker is geworden dat er veel meer integratie en samenwerking is tussen alle verschillende epigenetische mechanismen dan in 2006 werd verondersteld. Deze geïntegreerde complexiteit van epigenetische mechanismen kan erop wijzen dat cellen en organismen behoefte hebben aan een aanzienlijke redundantie in genregulatie in staat te zijn zich naar behoren te ontwikkelen en op gepaste manieren te kunnen reageren op veranderingen in intracellulaire of extracellulaire veranderingen.

In zoogdier (humane) systemen richten toepassingen van epigenetische overerving zich vooral op de mitotische epigenetische overerving. Veel aandacht gaat daarbij uit naar geneesmiddelen tegen kanker die zich richten op epigenetische mechanismen. Dergelijke geneesmiddelen zijn op de markt en nieuwe zijn in ontwikkeling. Een belangrijk aandachtspunt blijft de specificiteit van de geneesmiddelen. In de toekomst zal genoom-brede epigenetische profilering gemeengoed worden voor de diagnose van kanker en voor het volgen van de voortgang van deze ziekte. Dergelijke profilering zal helpen om nieuwe toepassingen te ontwikkelen. Ook andere ziektes zullen epigenetisch geanalyseerd gaan worden. Daarbij zal rekening gehouden moeten worden met de variatie in epigenetische karakteristieken tussen individuele cellen. De ontwikkeling van technologie voor het kunnen analyseren en mogelijk modificeren van een enkele cel zal nodig zijn voor toepassingen die voldoende specifiek, doelgericht en stabiel zijn. Toepassingen in zoogdier (humane) systemen zijn doorgaans niet gebaseerd op genetische modificatie.

Toepassingen van epigenetica in planten richten zich juist op meiotische epigenetische overerving, hetzij gebruikmakend van genetische modificatie, of juist bedoeld om epigenetische variatie te genereren en te gebruiken zonder last te hebben van de huidige regelgeving rond genetische modificatie. Er bestaat een voorbeeld van directe selectie voor epigenetische variatie in koolzaad, maar het is momenteel onduidelijk hoe algemeen die strategie kan zijn. Interessant is het concept van de zogenaamde epiRILS (epigenetische recombinante inteeltlijnen), waarin de DNA methylering stabiel is veranderd en aanleiding geeft tot nieuwe en mogelijk bruikbare meiotisch overerfbare fenotypische variatie. Een belangrijke toepassing van een epigenetisch mechanisme betreft het overerfbaar uitschakelen van genexpressie met behulp van RNA-afhankelijke DNA methylering (RdDM). Momenteel is er echter slechts één voorbeeld gedocumenteerd waarin de expressie van een endogeen plantengen wordt uitgeschakeld op een manier die erfelijk is en gehandhaafd blijft nadat het transgen dat de uitschakeling heeft veroorzaakt weer afwezig is (signaal-onafhankelijke handhaving).

Een belangrijk onderwerp behandeld in dit rapport is de relatie tussen de genetische modificatie van planten en het optreden van epigenetische effecten. Epigenetische effecten in transgene planten zouden het gevolg kunnen zijn van het binnengekomen transgen dat: (1) het omringende chromatine beïnvloedt (*cis* epi-effect), of (2) de expressie van andere plantengenen beïnvloedt (*trans* epi-effect), terwijl de beïnvloeding onbedoeld is en in principe moet blijven bestaan in de afwezigheid van het transgen (signaal-onafhankelijke handhaving). Daarnaast (3) kunnen epigenetische mechanismen de expressie van het transgen zelf onbedoeld beïnvloeden en (4) zijn er toepassingen denkbaar waarin de genetische modificatie zich doelbewust richt op het bewerkstelligen van epigenetische veranderingen (epigenetische modificatie). Deze vier mogelijkheden voor het optreden van epigenetische effecten worden in dit rapport geanalyseerd in de context van de bestaande protocollen voor de analyse van de biologische veiligheid van genetisch gemodificeerde planten in het milieu.

(**1**) *Cis* epi-effecten: de aanwezigheid van het binnengebrachte transgen zou de lokale chromatine configuratie kunnen veranderen, daardoor de expressie van nabijgelegen genen en daarmee het fenotype van het organisme. De evaluatie van dergelijke onbedoelde effecten is onderdeel van de bestaande biologische veiligheidsanalyse. Er zijn bovendien geen voorbeelden in de literatuur gevonden waar het onderzoek laat zien dat een dergelijk effect, mocht het al optreden, gehandhaafd blijft in afwezigheid van het transgen, bijvoorbeeld nadat het transgen door uitkruising is verwijderd.

(**2**) *Trans* epi-effecten: het binnengebrachte transgen of diens expressieproduct zou aanleiding kunnen geven tot een epigenetische verstoring door via een epigenetisch mechanisme de expressie van een willekeurig endogeen gen te veranderen. Eventuele onbedoelde veranderingen die afhankelijk zijn van de aanwezigheid (expressie) van het transgen en die terugkeren naar de oorspronkelijke toestand in de afwezigheid van (de expressie van) dat transgen, zijn onderdeel van de bestaande biologische veiligheidsanalyse. Alleen veranderingen die blijven bestaan in de afwezigheid van (de expressie van) het transgen, betreffen daadwerkelijk onbedoelde epigenetische veranderingen. Dergelijke gevallen zijn in de literatuur niet beschreven, afgezien van één voorbeeld van gericht gebruik van RNA-afhankelijke DNA methylering (RdDM). Gezien het bestaande beoordelingskader van genetische gemodificeerde planten zullen eventuele onbedoelde epigenetische effecten tijdig ontdekt worden in de onderzoeks- en ontwikkelingsfase, inclusief veldproeven, van ieder gepland (commercieel) gebruik van genetisch gemodificeerde planten in het milieu.

(**3**) Endogene epigenetische mechanismen zouden de bedoelde expressie van het transgen kunnen beïnvloeden. Voorbeelden in de literatuur betreffen vooral de ongewenste silencing van het binnengebrachte transgen. Het verlies van het gewenste fenotype is een typisch voorbeeld van het producentenrisico. Dergelijk plantmateriaal zal niet tot een commercieel product ontwikkeld worden. Mocht een product al op de markt zijn, dan zijn de verplichte monitoring na introductie op de markt en de algemene verantwoordelijkheid van de producent voldoende om ervoor te zorgen dat het onbedoelde verlies van het gewenste fenotype ontdekt wordt en het product van de markt wordt gehaald.

(**4**) Het realiseren van epigenetische veranderingen als doel van de genetische modificatie (epigenetische modificatie) kan aantrekkelijk lijken, maar tot nu toe is er maar één duidelijk voorbeeld gepubliceerd op basis van RNA-afhankelijke DNA methylering (RdDM). Deze toepassing en de gevolgen ervan voor wet- en regelgeving in de Europese Unie (EU) worden momenteel onderzocht door een EU werkgroep 'Nieuwe veredelingstechnologieën'. Toepassing van chromatine elementen, doelgerichte RNA interferentie (RNAi) en andere toepassingen richten zich ook op de epigenetische component van genregulatie. Dergelijke toepassingen zijn voor zover bekend geheel afhankelijk van de aanwezigheid van de gebruikte transgene elementen en vallen daardoor in het kader van de huidige biologische veiligheidsbeoordeling van genetisch gemodificeerde planten.

Alles overziende kan op basis van de huidige kennis en literatuur geconcludeerd worden dat de kans erg klein is dat een binnengebracht transgen aanleiding geeft tot een epigenetische verandering die gehandhaafd blijft in een cel - of organisme- zonder dat transgen en leidt tot onbedoelde effecten, mede ook gezien de achtergrond ('ruis') aan epigenetische verandering en variabiliteit in cellen en organismen die al door ontwikkeling, omgeving en andere bronnen wordt veroorzaakt. Mocht het ingebrachte transgen toch onverwacht aanleiding geven tot een onbedoeld epigenetisch effect, dan is het bestaande beoordelings-kader voor de biologische veiligheid van genetisch gemodificeerde planten voldoende om dat tijdig te ontdekken. Dergelijke eventuele onbedoelde epigenetische effecten zullen niet gemist worden in de bestaande protocollen voor het vaststellen van de biologische veiligheid van genetisch gemodificeerde planten in het milieu.

De voortgaande ontwikkeling van epigenetische technologie, vooral gericht op het hele genoom, zal verder bijdragen aan kennis over de epigenetische status van cellen of organismen en mogelijkheden om deze doelgericht aan te passen om gewenste fenotypes te krijgen. Het onderzoeksveld ontwikkelt zich nog steeds in een razend tempo. Veel zal beter begrepen moeten worden voordat het zo ver is en voordat het zinvol wordt te overwegen of de toevoeging van epigenetische profilering aan de bestaande beoordelingskaders meerwaarde heeft voor de evaluatie van de biologische veiligheid van genetisch gemodificeerde planten.

# Summary

In this report, epigenetics is defined as the changes in gene function that are mitotically and/or meiotically heritable and do not entail a change in DNA sequence. Epigenetic states are typically maintained in the absence of the signal that initiated the state (self-perpetuation) and are potentially reversible. Major players in epigenetics are: (a) DNA methylation, (b) chromatin, the complex of DNA and proteins in the eukaryotic nucleus and (c) RNA molecules, initially involving only small RNA, but now also (very) large molecules. Self-perpetuation is currently best documented for DNA methylation. It should be noted, that among researchers there is currently still considerable debate on the precise scope and definition of epigenetics, notably with respect to the concept of self-perpetuation.

In both mammals and plants, epigenetic gene regulation is important for proper development. The transmission of epigenetic information within an organism (mitotic epigenetic inheritance) occurs more frequent than the transmission of such information through generations via the germ-line, known as meiotic or transgenerational epigenetic inheritance. The latter is now firmly established in plants, likely due to the late formation of germ cells in plant systems. The occurrence and relative importance of meiotic epigenetic inheritance in mammalian (human) systems is less well established. Research into epigenetic inheritance is booming and the interest to develop applications of such inheritance is on the rise. Epigenetic research and development have progressed tremendously in recent years, notably because of the developments in DNA and RNA sequencing in terms of higher throughput and lower costs. Combined with analyses such as chromatin immunoprecipitation (ChIP), this is resulting in impressive amounts of data that are helping to unlock the information present in the epigenetic complement of cells. It is anticipated that the technological developments will continue to advance and will contribute to more understanding and use of epigenetic inheritance. The future is likely to see further developments in DNA sequencing and single-cell technology. This way, the analysis of the epigenome of the single cell will come in reach.

The understanding of the epigenetic machinery and knowledge of its molecular organization have extended considerably compared to the situation in 2006. The knowledge of epigenetic gene regulation has become much more detailed, but also indicated that such regulation is much more complex. New players continue to be identified. Notably the RNA world now includes many more types and molecules, including long non-coding RNA molecules. The regulatory roles and occurrence of DNA methylation have become more subtle: it is not only associated with gene silencing, but in other cases also with gene activity. A new chemical modification of DNA, hydroxymethylcytosine, was discovered and may have additional regulatory roles. Above all, it has become more clear that there is much more integration and cross-talk between all different epigenetic regulatory mechanisms than had been anticipated in 2006. The integrated complexity of epigenetic mechanisms may indicate the presence or need for considerable redundancy in gene regulation to allow cells and organisms to develop properly and respond appropriately to changes in internal or external environments.

In mammalian (human) systems, applications of epigenetic inheritance primarily focus on mitotic epigenetic inheritance. Currently, a lot of attention is devoted to epigenetic drugs for the treatment of cancer. Such drugs are on the market and new are in development, but their specificity is still a major challenge. In the future, genome-wide epigenetic profiling for the diagnosis of cancer types and monitoring of disease progression is likely to become standard practice. Such profiling may be of help to develop additional applications in the future. Disorders other than cancer are likely to be approached with similar epigenetic technology. To account for cell-to-cell variation in epigenetic characteristics, future development of single-cell analyses and modification may be required for applications becoming sufficiently specific, targeted and stable. The applications in mammalian systems generally do not involve genetic modification.

In contrast, the applications of epigenetic inheritance in plant science generally aim at meiotic epigenetic inheritance, either using genetic modification, or in developing strategies that generate and use epigenetic variation without the burden of regulation currently associated with genetic modification. An example of direct selection for epigenetic variation in oilseed rape awaits confirmation in other crops. Interesting is the concept of so-called epiRILS (epigenetic recombinant inbred lines), in which DNA methylation is affected to generate additional and potentially useful phenotypic variation that is meiotically inherited. An important application of an epigenetic mechanism entails heritable gene silencing based on the use of RNA-directed DNA methylation (RdDM). Currently, there is only one example of RdDM documented in which heritable, self-perpetuating silencing of an endogenous gene is accomplished in the absence of the transgene used to trigger the silencing (self-perpetuation).

A main issue considered in this report is the relationship between genetic modification of plants and the occurrence of epigenetic effects. Epigenetic effects in transgenic plants could be the result of the incoming transgene that: (1) affects the surrounding chromatin (*cis* epi-effect), or (2) affects the expression of endogenous genes (*trans* epi-effect), whereas the effect considered should be unintended and persist in the absence of the transgene. In addition, (3) endogenous epigenetic mechanisms could affect the expression of the transgene unintentionally. Also, (4) applications should be considered in which epigenetic changes are the deliberate target of the genetic modification (epigenetic engineering). These four possibilities for epigenetic effects are here evaluated in the context of the existing protocols for the environmental biosafety assessment of genetically modified plants.

(**1**) *Cis* epi-effects: the presence of the incoming transgene could change the local chromatin characteristics, which could have a notable influence on the expression of neighboring genes and the phenotype of the resulting organism. Such unintended effects are part of the current biosafety assessment protocols. There are no examples in the literature where it is shown that such effects, if any, occur and persist upon removal of the transgene, for example by outcrossing.

(**2**) *Trans* epi-effects: the incoming transgene or its expression product could act as an 'epigenetic disruptor' and affect the expression of any endogenous gene by an epigenetic mechanism. As long as the observed and unintended changes depend on the presence (expression) of the transgene as inducer and revert to the original situation in the absence of the transgene (expression), such changes, if any, are covered by the current biosafety assessment and monitoring protocols. Only in case the unintended changes persist after the transgene has left the genome, they should be considered epigenetic. Such effects have not been described in the literature, apart from one example of the targeted application of RdDM. In view of the current biosafety assessment protocol for genetically modified plants, such unintended effects are likely to be detected in the research and assessment phase, including field trials, of any anticipated commercial release of a genetic modification event in plants.

(**3**) Endogenous epigenetic mechanisms could affect the intended expression of the transgene. Examples include the inadvertent gene silencing of the transgene introduced. Such effects are well known in the literature. The loss of the intended phenotype is a typical example of the 'producer's risk'. Generally, such material will not be developed into a commercial product for release. In case the plant material is already on the market, the mandatory post-market monitoring and general stewardship of the product by the producer will be sufficient to detect (and withdraw) such events.

(**4**) Epigenetic changes as deliberate target of genetic modification (epigenetic engineering) may seem attractive, but as yet only one clear example using RdDM, discussed above, has been put forward in the scientific literature. This RdDM technology and its ramifications is currently subject of assessment in an EU Working Group 'New Breeding Technologies' for future consideration in EU procedures. Although the use of chromatin boundaries, targeted RNA interference (RNAi) and other examples of applications do target the epigenetic machinery of a cell, such approaches are essentially genetic in nature and will be subject to current assessment protocols.

Overall it is concluded that on the basis of current knowledge and literature, the likelihood that an incoming transgene generates an epigenetic modification that is maintained in a cell/organism without that transgene and gives rise to unintended effects, is very small, also considered against the background of epigenetic change and variability already generated by development, environmental stress and/or other sources of epigenetic variation. In the unlikely case that the incoming transgene results in an unintended phenotype due to epigenetic mechanisms, the existing frameworks of biosafety assessment are in place and sufficient to detect such an adverse phenotype. It is considered sufficiently unlikely that such putative epigenetic changes will not be noticed in the biosafety procedures now in place.

The continued development of epigenetic technology, notably at the whole genome level, will help to understand and in the future possibly modify the epigenomic status of genomes to obtain phenotypes of interest. The field is evolving at a very rapid pace and a great deal still needs to be learned prior to be able to consider any added value of comprehensive epigenetic evaluations into future biosafety assessments.

# List of abbreviations

| | |
|---|---|
| 3C | chromosome confirmation capture |
| 5hmC | 5-hydroxymethylcytosine |
| 5mC | 5-methylcytosine |
| Ac | acetylation |
| bp | base pair(s) |
| ChIP | chromatin immunoprecipitation |
| COGEM | Netherlands Commission on Genetic Modification |
| DAM | DNA adenine methyltransferase |
| DIP | DNA immunoprecipitation |
| dsRNA | double stranded RNA |
| DNA | deoxyribonucleic acid |
| DNMT | DNA methyltransferase |
| EMBL | European Molecular Biology Laboratory |
| epiRIL | epigenetic recombinant inbred line |
| ES | embryonic stem cells |
| EU | European Union |
| FISH | fluorescent in situ hybridization |
| GA | Genome Analyzer (Illumina) |
| GM | genetic modification or genetically modified |
| H3K9, | histone 3, lysine 9 (example) |
| HAT | histone acetyltransferase |
| HDAC | histone deacetylase |
| HMG | high mobility group |
| HMT | histone methyl transferase |
| HPLC | high performance liquid chromatography |
| iPSC | induced pluripotent stem cells |
| kb | kilo base(s) |
| LNA | locked nucleic acid |
| lncRNA | long non-coding RNA |
| MBD | methyl DNA binding |
| mC | methylcytosine |
| Me | methylation |
| miRNA | microRNA |
| mRNA | messenger RNA |
| MS | mass spectrometry or methylation-sensitive (depending on context) |
| NChIP | native ChIP |
| NGS | next (now) generation sequencing |
| ncRNA | non coding RNA |
| nt | nucleotide(s) |
| PCR | polymerase chain reaction |
| PGM | Personal Genomics Machine |
| PTGS | post transcriptional gene silencing |
| RdDM | RNA-dependent DNA methylation |
| RNA | ribonucleic acid |
| RNAi | RNA interference |

| | |
|---|---|
| siRNA | small interfering RNA; repeat-associated = rasi; transacting = tasi |
| SMRT | Single Molecule Real Time |
| SNP | Single Nucleotide Polymorphism |
| sRNA | small RNA |
| TE | transposable element |
| VN | vegetative nucleus |
| ZF | zinc finger |

# 1. Introduction

## 1.1 Definition: what is epigenetics?

In the context of this report, epigenetics is defined as 'changes in gene function that are mitotically and/or meiotically heritable and do not entail a change in DNA sequence' (Russo et al., 1996; Nap and Geurts van Kessel, 2006). Epigenetic states are typically maintained in the absence of the signal that initiated the state (self-perpetuation) and are potentially reversible. Major players in epigenetics are DNA methylation and chromatin, the physical association of DNA with numerous proteins that is supposed to carry and convey information. Yet, it is important to realize that there is still quite some debate among experts about this definition and its consequences, notably which phenomena should be included and which not in the context of epigenetics. For example, 'what is epigenetics?' was a major issue in many discussions at the 9th EMBL conference on Transcription and Chromatin (Jose Muiño, Wageningen, personal communication, 2010). The term 'epigenetics' clearly means different things to different people. The British geneticist Adrian Bird has commented: 'Epigenetics is a useful word if you don't know what's going on - if you do, you use something else.' (Anon., 2010).

Although epigenetics clearly refers to the inheritance of variation above and/or beyond ('epi') changes in DNA sequence, the term is becoming shorthand for a variety of regulatory systems involving DNA methylation, chromatin structure, histone modification, nucleosome positioning and noncoding RNA (Riddihough and Zahn, 2010). Notably, the regulatory role of post-translational modifications of histones and their correlation with transcriptional states has promoted a wider use of the term epigenetics in the literature (Bonasio et al., 2010). Epigenetics is than defined as 'the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states', and includes any molecular signature found on chromosomes, especially histone marks (Bird, 2007). In contrast, a more narrow operational definition defines epigenetics as 'a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence' (Berger et al., 2009). In the 2010 special issue on epigenetics of Science, the introductory paper uses the more traditional definition 'the inheritance of variation (-genetics) above and beyond (epi-) changes in the DNA sequence' (Bonasio et al., 2010). The authors outline that for any system considered to be 'epigenetic', it should be heritable, self-perpetuating and reversible. To distinguish various epigenetic states and/or phenomena, numerous modifiers are added, such as *cis* and *trans*, mitotic (or somatic) and meiotic (or transgenerational), as well as vertical and horizontal. In this report we will include these additional modifiers to distinguish various epigenetic states and/or phenomena. We will explain these modifiers below.

- ***Cis* and *trans* epigenetics**
  *Cis* and *trans* refer to the physical association of epigenetic signals: *cis*-epigenetic signals are physically associated with the chromosome (or gene) on which they act (e.g. DNA methylation), whereas *trans*-epigenetic signals are not, or not necessarily so (e.g. miRNA). It may not always be easy to discriminate between these two signals. In current thinking about epigenetics, attention for the importance and role of cytosolic factors, such as small RNA molecules that are transmitted by partitioning of the cytosol during cell division (therefore *trans* epigenetic signals) is on the rise, in addition to the studies of the molecular signals that are physically associated with DNA (*cis*-epigenetic signals) and inherited via chromosome segregation during cell division (Bonasio et al., 2010).

- **Mitotic and meiotic epigenetics**
  Mitotic (or somatic) epigenetics refers to signals propagated through mitosis, whereas meiotic (or trans-generational) epigenetics refers to signals propagated through meiosis.

- **Vertical and horizontal epigenetics**
  Transmission of epigenetic signals to daughter cells, such as in mitotic as well as in meiotic epigenetics, is called vertical epigenetics, whereas epigenetic signals can also be transmitted between sibling cells in the same organisms, e.g. through the transport of miRNA. The latter phenomenon is referred to as horizontal epigenetics (Bonasio et al., 2010).

Epigenetic factors beyond DNA - and not covered in our previous nor in this review - are the prion proteins. These could be considered 'epigenetic' in the extreme (Halfmann and Lindquist, 2010).

Epigenetics as defined above generates its own terminology, most of which is analogous to the DNA sequence-based terminology. The 'epigenome' of an organism is the genome with all its epigenetic modifications. Each individual cell may, therefore, have its own distinguishable epigenome and an organism should be thought of as a combination of numerous different epigenomes. Whereas in genetics the term allele is used to distinguish between different forms of a gene that differ in their DNA sequence (located on equivalent positions on a chromosome), the term epi-allele refers to genes (alleles) that do not differ in their DNA sequence, but carry different epigenetic marks. The term (or concept) 'epi-gene' is not used.

Overall, there is some tendency to label any 'non-genetic' phenomenon as 'epigenetic' by default. In a 2010 Science video (http://videolab.sciencemag.org/Featured/ 650920373001/1), several well-known experts in the field provide their view on the current definition of epigenetics, nicely showing the disagreements and potential confusions in the terminology used today. In the video, Peter Fraser (Cambridge, United Kingdom) gets the last word: 'it is like the 'X-files' of chromatin'. Given the discussions in the field as summarized above, we feel that it is not the purpose of this report to take position, nor declare some given viewpoints as better than others. Therefore, we here continue to define epigenetics as before as: 'changes in gene function that are mitotically and/or meiotically heritable and do not entail a change in DNA sequence' and we will include the role of DNA methylation, histone modification, small RNA and chromosome structure. In classical genetics, the notion that an epigenetic state established in the parent, either stochastically in response to the environment or by a predetermined signal, can be inherited by the offspring, has some Lamarckian flavor. Therefore, it continues to be received with resistance, suspicion and interest (Nap and Geurts van Kessel, 2006; Youngson and Whitelaw, 2008; Daxinger and Whitelaw, 2010). After all, the inheritance of acquired characteristics was also suggested by Darwin, who proposed the existence of 'gemmules' as somatic particles that entered the germ line to contribute to the characteristics of the next generation (Martienssen, 2010). Regulatory mobile small RNA may represent such gemmules (Martienssen, 2010) or RNA granules (Eichler et al., 2010). Accumulated evidence continues to indicate that the detailed mechanisms of epigenetic inheritance are as complex as, if not considerably more complex than, the mechanisms implied in the genetic code. This growing awareness tends to have epigenetics presented in the more popular press as a revolutionary new science that presents 'the antidote to the idea that we are hard-wired by our genes' (Bird, 2007).

In contrast, there is the belief, notably among geneticists, that 'in the end' it will all boil down to the DNA sequence, genes and environment. Chromatin marks and other epigenetic modifications of DNA are the consequences of sequence-specific interactions of proteins (and RNA) that recruit modifying enzymes to appropriate targets. All these marks and modifications are the result of sequence-specific regulatory interactions and are, therefore, dependent on the genomic sequence (Ptashne et al., 2010). In other cases, the explanation of epigenetics can only reflect a lack of detailed genetic (or DNA) knowledge. For example, a presumed epimutation involved in susceptibility to colorectal cancer could be attributed to DNA micro-deletions that interfere with a normal stop of transcription (Venkatachalam et al., 2010). A conceptual escape is known as epistasis (Nap and Geurts van Kessel, 2006). Epistasis, or epistatic interaction, is a concept from Mendelian genetics basically to explain non-Mendelian inheritance. In epistasis, the action of one gene (or gene product, i.e. protein) is modified by one or more other genes (or gene products), that segregate independently. Complex epistatic interactions may show non-Mendelian inheritance, for example related to linkage disequilibrium, and as a result they are interpreted as epigenetic. As outlined before, the future may unveil an explanation for many, if not all, epigenetic phenomena in terms of epistatic (genetic) interactions (Nap and Geurts van Kessel, 2006).

## 1.2 Epigenetics, a brief outline

Epigenetics here defined as 'changes in gene function that are mitotically and/or meiotically heritable and do not entail a change in DNA sequence' is thought to comprise a variety of molecular mechanisms that are all interrelated. We here give a brief overview of the main mechanisms. Generally, four different epigenetic mechanisms are distinguished in the literature (Nap and Geurts van Kessel, 2006):

- DNA methylation (and de-methylation)
- Protein (notably histone) modification
- RNA-based mechanisms
- Higher-order chromatin-based mechanisms

Of these, data currently available indicate that only DNA methylation satisfies all criteria for 'true' epigenetics in terms of heritability, reversibility and self-propagation (Bonasio et al., 2010). In order to understand and appreciate how these mechanisms interrelate and can regulate gene expression, as described in the following sections of this report, it is necessary to comprehend how DNA is organized and transcription is regulated in the nucleus of a cell (Figure 1). The underlying 'textbook knowledge' (Allis, 2007) is here summarized as primer for the more detailed considerations below. The terms that will reappear in the subsequent sections and/or are considered important are here given in bold.

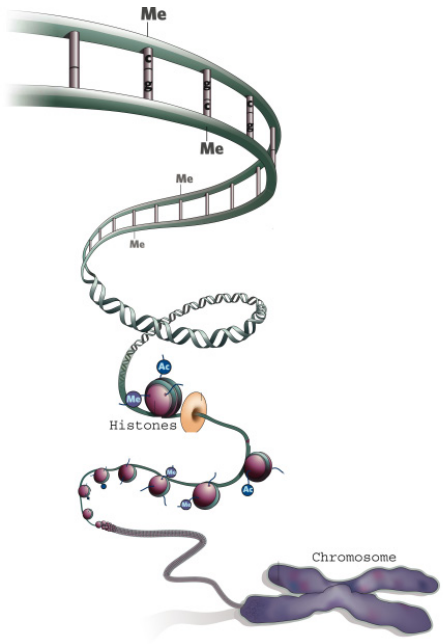



*Figure 1. Epigenetic marks that affect transcription include DNA cytosine methylation (Me) and covalent histone acetylation (Ac) or methylation (Venkatachalam et al. 2010).*

The central dogma of biological information flow is summarized as 'DNA makes RNA makes protein' (Allis, 2007). Proteins do the work in the biological unit called a cell, which contains a nucleus with DNA. The DNA in that nucleus harbors genes, although the precise definition of 'what is a gene', is still open to debate (Pearson, 2006; Falk, 2010). Active genes generate a code in RNA in a process called **transcription**. RNA leaves the nucleus and moves to the cytoplasm. There, the RNA type known as mRNA is used as template for the generation of protein in a process called **translation**. However, there are many exceptions to this 'dogmatic' route. Some viruses invert the route and make DNA out of RNA. Only some DNA gives rise to RNA and even that not all the time or in all cells. Also, not all RNA gives rise to protein.

The need for epigenetic regulation of gene function stems from the observation that every cell in multi-cellular organisms arises from a single-cell precursor and carries the same DNA, yet the organism is composed of different cell types with different phenotypes and functions. Epigenetic modifications of DNA and changes in chromatin structure play well-appreciated roles in cellular differentiation and development. In addition, such alterations have emerged as important factors in the development and progression of disease. Many diseases, or otherwise unwanted phenomena, can be traced back to proteins and/or RNA molecules that are in the wrong place and/or the wrong time, or are malfunctioning for other reasons. Most of these errors relate to the DNA used at the wrong time and/or place. In research and application, it is often attempted to employ the biological information flow, either to understand a biological phenomenon to address a given organism's state (e.g. in case of pharmaceuticals), or to adjust the biological information flow in an organism to suit the purposes of the engineer or physician. In all cases, the challenge for bio(techno)logical research is to understand - and subsequently use or modify - the normal course of events, as well as the exceptions to the rules.

The DNA molecule is a linear chain of nucleotides organized in a double helix structure that in a nucleus is tightly folded around proteins (Figure 2). The combination of protein and DNA is called chromatin. The major DNA-protein complex is called the **nucleosome** and the main protein component of the nucleosome is called **histone**. The nucleosome core consists of two of each of four different histone proteins, H2A, H2B, H3 and H4. These histones make up the central core particle of the nucleosome and act as spools around which DNA winds. The DNA is wrapped around this protein complex in about 2 turns, comprising about 147 base pairs. Another, larger histone, H1 or 'linker histone', binds the nucleosome as well as the entry and exit sites of the DNA, thus locking the DNA into place, allowing the formation of higher order structure. H1 is thought to act as a clamp with a stabilizing function. Together with the DNA linking two nucleosomes, there are about 200 bp (base pairs) of DNA per nucleosome. The precise distribution of the nucleosomes over the DNA strand, known as **nucleosome positioning**, has regulatory functions as well.

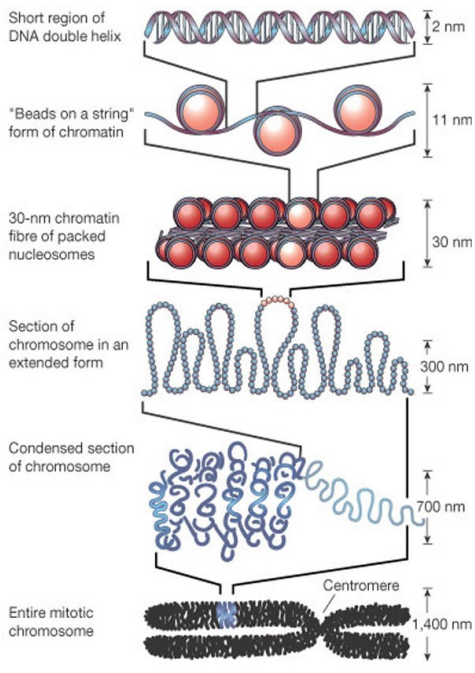



*Figure 2. Different levels of DNA condensation (Felsenfeld and Groudine, 2003).*

The nucleosome core is composed of two H2A-H2B dimers and two H3-H4 dimers, forming two nearly symmetrical halves. The core histones are relatively similar in structure. They are highly conserved through evolution, all featuring 'helix-turn' motifs which allow easy dimerization. They share the feature of long 'tails' on one end of the amino acid structure. Histones allow for different types of physical-chemical interactions with DNA and, via their tails, with the

world outside DNA. In addition, there are special types of histone proteins, or **histone variants** that can have regulatory functions in particular cases. Various other proteins are associated with the nucleosome, including enzymes and scaffold proteins. For example, so-called high mobility group (HMG) proteins help the nucleosomes to form higher order chromosome structures such as 10-nm-fibers. A chain of nucleosomes can be arranged in a 30-nm fiber, a compacted structure thought to have a zigzag ribbon structure or to have no regular structure. Beyond the 30-nm fiber the structure of chromatin is poorly understood, but it is suggested that the 30-nm fiber is arranged into loops along a central protein scaffold to form transcriptionally competent **euchromatin**. Further compaction involving the positively charged histone N-termini interacting with the negatively charged phosphate groups of the DNA backbone results in transcriptionally inactive, so-called 'closed' chromatin, or **heterochromatin**. Notably euchromatin is thought to adopt additional higher-order chromatin structures that play a role in transcriptional gene regulation.

In the closed chromatin environment of heterochromatin, genes cannot be transcribed as the transcription factors face steric hindrance to trigger mRNA synthesis: the genes are silent or silenced. Various modifications of DNA or nucleosomes (see below) are thought to open up the chromatin to allow transcription or, vice versa, prevent transcription. In this context, the importance of the occurrence of **DNA methylation**, the presence of a covalent methyl group at cytosine residues, as well as **histone modifications**, various posttranslational modifications of histone proteins, must be understood. For example, aberrant DNA methylation is associated with a wide variety of human diseases and is the focus of active investigation.

In general, genes that are actively transcribed tend to have less methylated DNA, less bound histones and less associated proteins, while transcriptionally inactive genes tend to be highly associated with histones. This tight association presents a fundamental challenge to DNA template processes, such as transcription, replication and repair, which must occur in the context of chromatin. Transcription of a gene by RNA polymerase is thought to involve a complex wading of the polymerase protein through nucleosome complexes by continuous assembly and disassembly. In the cytoplasm, where the transcribed mRNA should be translated into protein, a lot of processes are involved with the appropriate (or inappropriate) translation of mRNA into protein. Next to regulated degradation, a role of growing importance is played by different types of small RNA molecules that themselves will not give rise to protein, but prevent or delay translation. The role and regulation of such **RNA-based mechanisms** in gene regulation represents a very active area of current research. Results indicate that the impact and complexity of the role of RNA in gene regulation appears considerably larger than previously anticipated.

The phenomena described above all relate to the structural features of the packaged form of (genomic) DNA 'on top' ('**epi**') of the nucleotide sequence itself, with the notable exception of RNA-based mechanisms, most of which involve cytosolic factors that are transmitted by partitioning of the cytosol during cell division. To establish true epigenetics, these 'epi' features have now to be combined with the genetics of cell division for somatic cellular propagation (mitosis) or sexual reproduction (meiosis). In either cell division, most chromatin structure is condensed further when the microscopically visible chromosomes are formed prior to DNA replication. There is supposed to be no gene expression (transcription) during cell division, although in yeast, heterochromatin may generate transcripts during DNA replication. Yet, cells have or must have a memory of which genes were active prior to cell division. Cells are able to re-install that situation after cell division when the chromosomes de-condense into interphase and resume gene expression. Such memory should be considered the 'epitome of epigenetics'.

Several mechanisms have been proposed to explain such epigenetic memory in mitosis. It is thought to involve re-installed DNA methylation, copying histone modification marks, and possibly also by nucleosome (re)positioning and through the small RNA complement of cells. Meiotic epigenetic inheritance crosses (possibly multiple) generations and is for that reason looked upon with some unease by many, as it is often considered to contradict the fundamentals of Darwinism. In plants, this is easier to imagine: germ cells develop late in development and can accumulate epigenetic changes over a longer time period than mammalian germ cells. Understanding how such re-installment of epigenetic marks works and can be used or modified is the ultimate research goal of epigenetics research. Using that understanding for subsequent application establishes the field of what could be considered **epigenetic engineering**.

## 1.3 Motivation: why is epigenetics important?

Understanding epigenetics is at the core of understanding and, possibly, the modification of (ab)normal development and gene expression. Moreover, epigenetic phenomena are important factors in the development and progression of disease. The apparently non-genetic (or non-Mendelian) 'memory' involved in proper differentiation and development can go astray in diseases such as cancer. Human cancer in its wide variety of occurrence is one of hotter areas of research with respect to mitotic epigenetics and the use of epigenetic drugs, whereas in plants, meiotic epigenetic inheritance gets a lot of attention. In various diseases other than cancer, epigenetic phenomena attract attention either for diagnostics or therapy (Portela and Esteller, 2010). Especially because epigenetic phenomena are generally considered to be essentially reversible, they may be more suitable for targeting and cure than irreversible changes in DNA.

In addition to its role in development, epigenetic research has in recent years become intricately connected with research into the influence of the environment, as well as research into complex traits and complex diseases. Fueled in part by the technological developments of whole genome sequencing and screening, the heritability of complex traits has become an important topic of research. The apparent lack of genetic elements that fully explain the heritability of such traits, a phenomenon referred to as 'missing heritability' (Eichler et al., 2010), as well as the failure to identify the genetic causes for some complex traits (diseases), stimulates the suggestion of possible epigenetic mechanisms for the missing heritability or failed inferences. Better understanding of the epigenetic regulation of gene expression will have applications in human and veterinary medicine as well as in agriculture, involving cloning, cell reprogramming, epigenetic engineering, epigenetic medication and/or epigenetic epidemiology. The growing importance of epigenetics as a separate field of science and investigation is also witnessed by the recent appearance of numerous dedicated scientific journals, such as 'Epigenetics' (Landes Bioscience), 'Clinical Epigenetics' (Springer) and 'Epigenetics and Chromatin' (BioMedCentral).

In the context of biosafety assessment of genetically modified organisms, the issue to be considered is the relationship between genetic modification and epigenetic effects. Can an incoming transgene be an epigenetic disruptor, causing unintended *cis* or *trans* effects that are undesirable from a safety perspective? Or, conversely, can epigenetic modification result in unpredicted unintended effects on the newly introduced transgene itself? And if so, what are the consequences of such effects for regulatory procedures? Can such effects be spotted early on? Are current procedures sufficient or are new procedures necessary? In order to help regulators answer such complex questions, first an overview of epigenetic technology, epigenetic mechanisms and current applications of such epigenetic mechanisms is presented.

# 2. Epigenetic technology

In recent years, various technologies used for epigenetic analysis or epigenetic research have progressed tremendously in terms of throughput, resolution and costs (Schones and Zhao, 2008). Most of the advances are due to the combination of existing technologies, notably mapping chromatin modifications with the help of chromatin immunoprecipitation (ChIP), with recent developments in DNA/RNA sequencing, often called 'next generation sequencing' (NGS), but in view of technological developments perhaps more aptly referred to as 'now-generation' sequencing, involving massively parallel sequencing. In this chapter, a short overview of the technologies currently used and/or on the horizon is given for the study of DNA methylation, chromatin modification and the other components that make up epigenomes as defined in the previous chapter.

## 2.1 High throughput DNA sequencing

A clear revolution has taken place within the last years in high-throughput DNA sequencing. Reduced costs (albeit sometimes relatively) and markedly large volumes of sequence data characterize these new methods, in some cases in excess of one billion reads per run. An excellent overview of the various technologies involved in current and near-term commercially available NGS instruments, as well as an excellent outline of the broad range of applications for these technologies is available (Metzker, 2010).

First-generation sequencing, also called Sanger sequencing, was developed by Sanger in 1977 and has seen many technological improvements along the way. This method now routinely results in a read length of on average 700 bases, but may be extended to 1,000 bases. Although very robust, first-generation sequencing is limited by the relatively small amounts of data that can be processed per unit of time, referred to as throughput, hence to costs. After several years of development in academia and industry, commercial second-generation DNA sequencing equipment became available in 2005. Technology platforms are now commercially available from Roche (454), Illumina/Solexa (GA and HiSEQ), Applied Biosystems (SOLiD), Helicos, Dover, Pacific Biosciences (SMRT) and Ion Torrent. These platforms differ in various aspects of their detailed technologies, but all manage to parallelize the sequencing reactions dramatically (Table 1). Currently, the Illumina/Solexa Genome Analyzer and its 2010 successor HiSEQ2000 seem to dominate the market. They achieve much higher throughput than Sanger sequencing by sequencing large numbers of DNA molecules in parallel. Tens of thousands of identical strands are anchored to a given location to be read in a process consisting of successive reactions (Metzker, 2010). The molecules to be sequenced are obtained by clonal amplification using PCR of individual DNA molecules. For this, different innovative high-throughput PCR approaches have been developed: emulsion PCR or bridge PCR. Emulsion PCR isolates individual DNA molecules along with primer-coated beads in aqueous droplets within an oil phase. It is used in 454 and SOLiD sequencing. Bridge PCR, in which fragments are amplified via primers attached to a solid surface, is used in the Illumina protocol. The clonal amplification results in a population of identical templates, each of which will undergo the same sequencing reaction. However, the clonal amplification is thought (or shown) to introduce (additional) errors and bias. The array of DNA anchor locations can have a high density of DNA fragments, leading to extremely high overall throughput and a resultant low cost per identified base. Due to the repetitive nature of the (PCR) procedures involved, the time to obtain results is usually several days and limits the average read length obtained. Another sequencing platform (Nano-Array) is based on existing, but fairly complex DNA manipulation and chemistry, followed by adsorption of self-assembling DNA molecules (so-called DNA nanoballs) onto photo-lithographically etched, surface-modified silicon nano-arrays with an ultra-dense grid-patterning. A mere US$ 4400 for sequencing a human genome was claimed (Drmanac et al., 2010), but it is currently unclear if and how this platform will be commercialized.

A somewhat different approach is found in the use of semiconductor chips with over 1.5 million (with numbers going up) ion sensors that establish electronic DNA sequencing. The Ion Torrent Personal Genomics Machine (PGM) sequencer reads DNA on a semiconductor chip by measuring the release of hydrogen ions as nucleotides get incorporated by DNA polymerase. The moment a nucleotide is incorporated into a strand of DNA, a hydrogen ion is

released. The charge from that ion changes the pH of the solution, which is detected by the ion sensor. The PGM could be considered the world's smallest solid-state pH meter. It calls each incorporated base and converts chemistry to digital information. Although sample preparation involves PCR and relies on polymerase-based sequencing-by-synthesis chemistry, the electronic detection obviates the need for lasers, cameras, or labeled nucleotides. This is reducing the costs of purchase (and running) considerably. The Ion Torrent PGM is being marketed from the beginning of 2011, so extensive user experience is not yet available.

Currently, single molecule sequencing, not requiring PCR for clonal propagation abd sometimes referred to as third generation sequencing, is thought to have a bright (both commercial and applied) future. Single molecule DNA sequencing extracts the maximum amount of information from a minimum of material. Technology is rapidly developing. The first single molecule sequencing platform from Helicos Biosciences (HeliScope) is on the market. It uses bright fluorophores and laser excitation to detect pyrosequencing events from individual DNA molecules fixed to a surface. The read length does not seem to offer significant improvements over Illumina and therefore it may have too little added value over Illumina. The Single Molecule Real Time (SMRT) sequencing technology (Eid et al., 2009) of Pacific Biosciences (PacBio RS) presents a single-molecule version of sequencing by synthesis (as in Sanger sequencing).

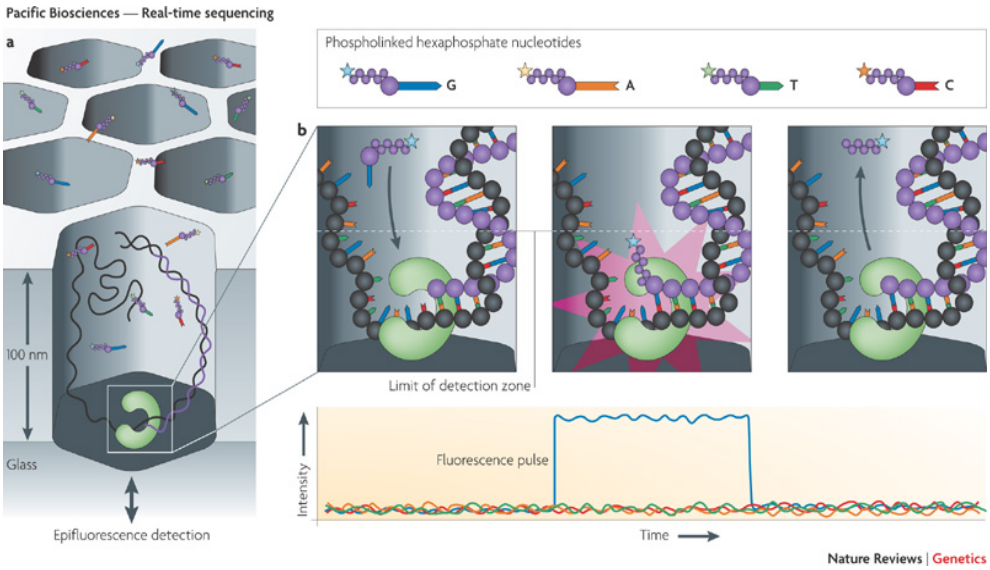



*Figure 3. Outline of single molecule real time DNA sequencing (Metzker, 2010).*

A single DNA polymerase molecule is present in a space that allows the detection of a single nucleotide of DNA being incorporated using fluorescent dyes (Figure 3). When a nucleotide is incorporated by the DNA polymerase, the fluorescent tag is cleaved off and the base call is made according to the kinetics and fluorescence of the dye. The PacBio RS is reported to be designed to produce read lengths greater than 1,000 bases on average with instances of over 10,000 bases. Interestingly, SMRT technology is able to detect DNA methylation without the need for bisulfite conversion as it can distinguish between C, 5mC and 5hmC (Flusberg et al., 2010). Further on the horizon of DNA sequencing developments seems nanopore sequencing that could yield even more data for yet less costs (Clarke et al., 2009). Sequencing a single molecule of DNA with a nanopore combines the potential for long read lengths with high speed, while obviating the need for PCR amplification. Detection relies on the electric signal that develops when DNA molecules are forced through a pore in a membrane by an electric field. With each base having a characteristic electrical signature, possibly employing electron tunneling (Huang et al., 2010a), movement though a pore can be used to analyze the sequence by reporting all of the signatures in a single read of a single molecule (Mirsaidov et al., 2010). Nanopore DNA sequencing is able to distinguish 5mC from cytosine (Clarke et al., 2009) and possibly also 5hmC from 5mC. If or when nanopore sequencing will result in new DNA sequencing platforms on the market is difficult to predict.

The production of large numbers of low-cost reads makes the NGS platforms described above useful for many applications. These include variant discovery by the re-sequencing of targeted regions of interest, or whole genomes, de novo assemblies of bacterial and eukaryotic genomes, cataloguing the transcriptomes of cells, tissues and organisms (RNA-seq) and species classification and/or gene discovery by metagenomics studies. For gene expression studies, microarrays are being replaced by sequencing-based methods, which identify and quantify transcripts without prior knowledge of a particular gene.

*Table 1. Overview of current DNA sequencing technologies.*

| | Method | Amplification | Throughput/ run | Read length (b) | Run time | Cost (€)/ Gb |
|---|---|---|---|---|---|---|
| Sanger (ABI3730) | Polymerase | PCR | 96 Kb | 1,000 | 2 hr | 2 M |
| Roche 454 | Polymerase | Emulsion PCR | 500 Mb | 400 | 10 hr | 20 K |
| Illumina HiSEQ | Polymerase | Bridge PCR | 400 Gb | 100 | 8 d | 200 |
| SOLiD v3 | Ligation | Emulsion PCR | 100 Gb | 50 | 7 d | 500 |
| Heliscope | Polymerase | None | 28 Gb | 30 | 8 d | ? |
| PacBio SMRT | Polymerase | None | 1 Gb | >1,000 | <2 hr | 5,000 |
| Ion Torrent | Chip | PCR | 20 Mb | 200 | 2 hr | 12 K |
| Oxford Nanopore | Electronic | None | ? | ? | ? | ? |

*Run is defined by the parallelization achieved by the individual apparatus. For ABI3730, it is 96 lanes; for HiSEQ it is 16 lanes. The run time should be taken into account. Some labor is included in the costs given, but no bio-informatics. All data are rough estimates for the situation in June 2011. Data given are collected from various sources (Gupta et al., 2010; Karow, 2010; Metzker, 2010), as well personal communication from Dr. E. Schijlen (Greenomics, Wageningen University DNA Sequencing Center).*

The sequence of whole genomes of related organisms is allowing large-scale comparative and evolutionary studies. Sequencing and re-sequencing of genomes-of-interest (human, crops) will contribute to the better understanding of the relationship between genotypes and phenotypes. In Arabidopsis, re-sequencing of the genomes of different accessions showed more structural variability than was expected (Santuari and Hardtke, 2010). The DNA data explosion generates obvious issues with respect to analysis, maintenance and storage. This will require concomitant developments in informatics and bioinformatics. In the future, it may be cheaper to re-sequence and analyze DNA than to store the information. For biology, this could imply an immense change (if not paradigm shift?) in the way of approaching science.

The new DNA sequencing technology is impacting biological research tremendously and the wide range of applications is growing rapidly. This is particularly the case for epigenetics research, in which the modern sequencing approaches are allowing analyses hitherto unthinkable, such as genome-wide profiling of epigenetic marks and chromatin structure using sequence-based methods.

## 2.2 Chromatin assays

Chromatin immunoprecipitation (ChIP) is a powerful technology to investigate protein-DNA interactions (Collas, 2010a). It is used to characterize modifications of chromatin-associated proteins or to identify in a genome-of-interest all the DNA-binding sites of a given DNA-binding protein-of-interest (Figure 4), for example a transcription factor. DNA and proteins are commonly reversibly cross-linked to attach proteins to their target DNA sequences. Usually formaldehyde is used, possibly in combination with a variety of long-range bi-functional cross-linkers, to freeze the in vivo situation. In native ChIP (NChIP), cross-linking is omitted. Subsequently, chromatin is isolated and fragmented, either by enzymatic digestion or by sonication of whole cells or nuclei. The protein–DNA complexes of

interest are immunoprecipitated from the supernatant (chromatin) using antibodies to the protein (or protein modification) of interest (Figure 4). The development of modification-specific antibodies, notably in combination with downstream high-throughput analysis, has paved the way for the assessment of the genome-wide distribution of histone modifications. Depending on the antibody, ChIP allows for studying chromatin-associated factors, histone modifications, histone variants and much more. Immunoprecipitated complexes are washed to remove non-specifically bound chromatin, the cross-link (if any) is reversed and the ChIP-enriched DNA is purified and analyzed by PCR, hybridization and/or cloning and sequencing (Collas, 2010a).

Chromatin immunoprecipitation in combination with microarrays (ChIP-chip, or ChIP-on-chip) is used to map DNA-binding proteins on genome-wide scale, but can also be used for nucleosome distribution and histone modification (Schones and Zhao, 2008). The analysis of ChIP samples is currently combined with second-generation sequencing in an approach known as ChIP-seq (Collas, 2010a). This technology is developing into the most powerful strategy currently available for ChIP analysis. It is used, for example, to generate extensive genome-wide chromatin state maps for specific cell types, such as embryonic stem cells or adipocytes (Mikkelsen et al., 2007; Mikkelsen et al., 2010), or to map transcription factor binding sites in a genome-wide approach (Kaufmann et al., 2010). In general, results of ChIP-seq analyses agree well with those obtained by other ChIP protocols (Collas, 2010a), but tend to present a higher resolution.

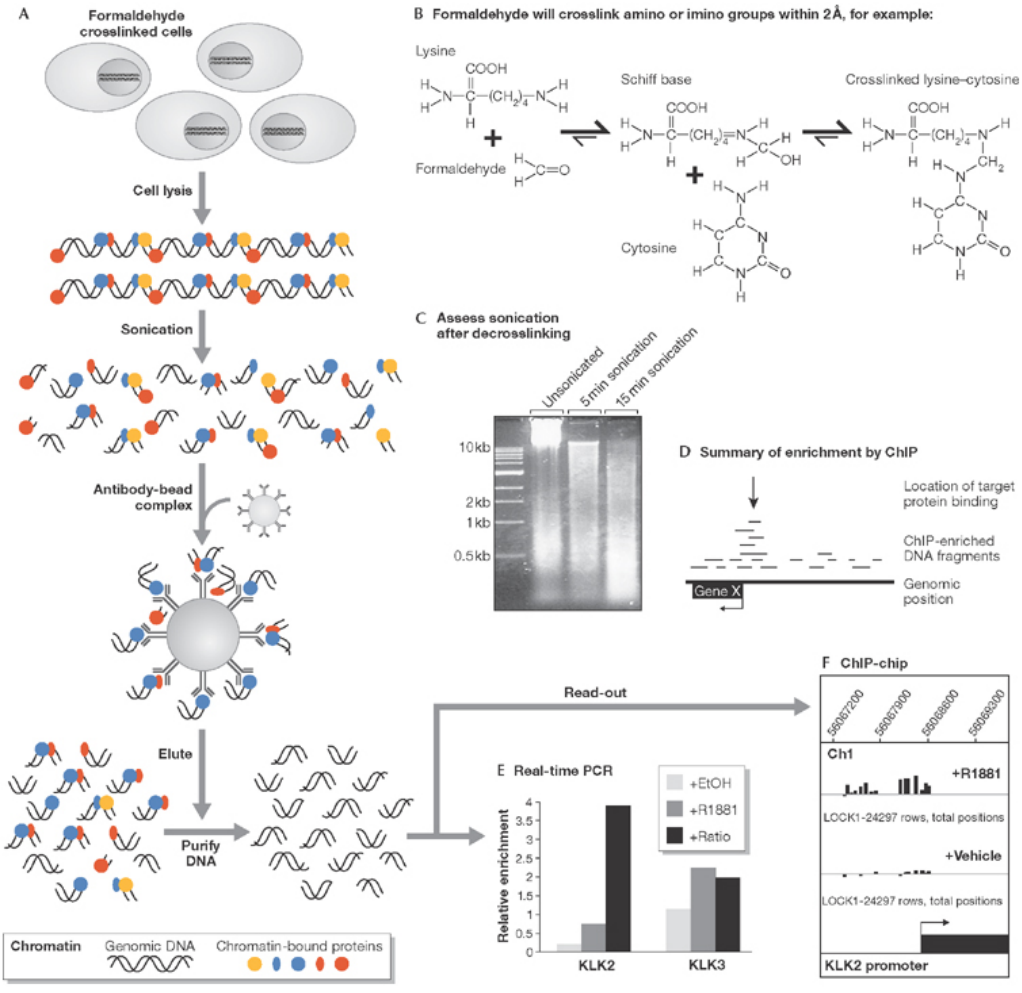



*Figure 4. Outline of chromatin immunoprecipitation strategies (Massie and Mills, 2008).*

Although appealing as concept and popular as technology, ChIP assays in practice generally involve cumbersome protocols, requiring long procedures, extensive sample handling and large amounts of biological material. The latter implies that results represent the situation in a population of many cells. As averaged snap-shot, the results thus obtained could include very different modifications in cells that are different (Schones and Zhao, 2008), but thought to be similar. It has also hampered the application of ChIP in case of rare cell samples. In addition, the current technology provides ample opportunities for technical errors and/or inconsistencies between replicates (Collas, 2010a). For example, the precise method of chromatin preparation is important and may introduce a lot of variability.

Another key issue in all assays based on immunological detection is the specificity of the antibodies used for protein precipitation. A recent study indicated that the quality of the currently available antibodies to detect post-translational modifications of histones may be suspect, supposedly due to structural changes brought about by neighboring modifications (Fuchs et al., 2011), confirming other reports that considerable numbers of commonly used antibodies raised against histone modifications fail specificity tests, such as in chromatin immunoprecipitation (Egelhofer et al., 2011). Of particular interest for the study of epigenetics is the co-occupancy of DNA binding proteins, either transcription factors or histone modifications, on the same stretch of DNA. For this issue, sequential ChIP analysis is used (Chaya and Zaret, 2004). Chromatin material resulting from the first ChIP is used as input for a second ChIP with another antibody. In combination with ChIP-seq, it has helped to pinpoint the existence of unstable nucleosomes in regions that were hitherto known as nucleosome-free regions (Jin et al., 2009).

Current technological improvements in ChIP assays predominantly aim to reduce the amount of material necessary for analysis and/or to speed up the protocol. This has resulted in a steady flow of smaller-scale alternatives, none of which seems to have become mainstream technology yet. Developments include the help of carrier chromatin (carrier ChIP), microChIP (μChIP), fast ChIP, matrix ChIP, the use of flow cytometry (ChIP-on-beads) and various other technologies (Collas, 2010a). The Fast ChIP assay shortens the procedure to essentially one day by incubation of antibodies with chromatin in an ultrasonic bath to increase the rate of antibody-protein binding and a resin-based DNA extraction procedure. Because of these technological developments, ChIP analyses have become feasible for small samples in case of for example suspected cancer or studies of early embryonic development. Of special interest are methods now available for combining ChIP-seq with small cell numbers (Goren et al., 2010). A proof-of-concept of a ChIP assay using lab-on-a-chip microfluidics may point the way to future automation and parallelization of this type of analysis (Wu et al., 2009).

## 2.3 DNA methylation assays

Covalent methylation of DNA at a cytosine (C) at position 5 (5-methylcytosine; 5mC) is an important (epi)genetic mark related to transcription and gene regulation. Such 5mC DNA methylation is likely the best and most extensively studied epigenetic modification. This is due in part to the technological developments available for such studies. Various excellent reviews on the various aspects of all techniques described in the literature are available (Tost, 2008; Gupta et al., 2010; Laird, 2010). Entire methylomes can now routinely be generated at single base-pair resolution (Laird, 2010). Here we will summarize the main technological developments and possibilities. No single method will be appropriate for every application. There are three currently main approaches to distinguish 5mC from unmethylated cytosine in DNA:
**(a)** restriction enzyme digestion
**(b)** affinity enrichment
**(c)** bisulfite conversion.

Nowadays, direct sequencing of meC could be an alternative. After such treatments, different analytical procedures can be followed, resulting in a large variation of techniques for determining DNA methylation patterns and profiles (Laird, 2010). Many of these methods are becoming obsolete, or can and are now combined with array- or second generation-sequencing-based approaches for genome-wide analyses. There exists also a wide diversity of DNA methylation analytical techniques that allow analyzing DNA methylation and its biological role on a genome-wide scale. A relatively newly identified covalent DNA modification is 5-hydroxymethylcytosine (5hmC), the epigenetic

relevance of which is not yet very clear. Possibly 5hmC is a step towards DNA demethylation. Methodologies that can distinguish 5hmC from 5mC at a genome-wide scale are in development. Moreover, PacBIO SMRT sequencing is able to detect and differentiate between 5mC and 5hmC (Flusberg et al., 2010).

To measure global DNA methylation, genomic DNA is hydrolyzed, followed by specific detection and quantification of the 5-methylcytosine content using analytical chemistry such as HPLC, MS, combinations of these (Rocha et al., 2010), or others (Tost, 2008). Obviously, such analyses do not provide any information about place or distribution of the 5mC residues. Although MS methods can be used for DNA methylation analysis of large sets of genes, it is currently not a method of choice for genome-wide, high-resolution surveys due to the limited throughput, high costs and demanding technical expertise. In the future, atomic force microscopy may generate suitable methods for DNA methylation analysis (Zhu et al., 2010).

### 2.3.1 Methylation-sensitive restriction enzymes

Some restriction enzymes allow discriminating methylated DNA from unmethylated DNA. Commonly used pairs in methylation analysis are *Hpa*II - *Msp*I, which both recognize CCGG, and *Sma*I - *Xma*I, which both recognize CCCGGG. Neither *Hpa*II nor *Sma*I can digest methylated DNA. Another enzyme is *Mcr*BC that recognizes two half sites of the form (G/A)mC at as far as 3 kb apart (optimal distance is 55–103 nucleotides) and only cleaves DNA containing 5mC. No enzymes are known that distinguish 5-hydroxymethylcytosine (5hmC) from 5mC or unmethylated residues.

Various enzyme-based genome-scale DNA methylation approaches exist (Laird, 2010), such as restriction landmark genome scanning (RLGS), or various PCR-based methods such as methylation-sensitive arbitrarily primed PCR (MS-AP-PCR) with several variants, methylation-sensitive representational difference analysis (MS-RDA), or amplification of inter-methylated sites (AIMS) and the like (Tost, 2008; Laird, 2010). These techniques generally involve a labor-intensive electrophoresis step, which explains why their use is decreasing. Newer technologies combine enzymatic methods with microarray-based analyses. Various commercial microarray platforms enable genome-scale DNA methylation analysis. Platforms comprise those of NimbleGen (Roche), Affymetrix, Agilent and Illumina. Each platform has its own characteristics and preferred use (Gupta et al., 2010). Examples of microarray-based assays include methylated CpG island amplification (MCA), combined with representational difference analysis (RDA) or array hybridization (MCAM). An alternative is differential methylation hybridization (DMH), modifications of which using the enzyme *Mcr*BC have developed into comprehensive high-throughput arrays for relative methylation (CHARM), or *Hpa*II fragment enrichment by ligation-mediated PCR (HELP).

Array-based methylation analyses are generally easy to perform, but often fail in case of repetitive sequences and depend on the representation of the regions-of-interest on the array. Yet, all restriction enzyme-based approaches obviously depend on the presence of suitable restriction sites in the regions-of-interest. To compensate for such issues, protocols combine methylation-specific restriction enzymes with NGS (Gupta et al., 2010). For example, methyl-MAPS (methylation mapping analysis by paired-end sequencing) is a method for genome-wide analysis of DNA methylation that has a single-CpG resolution and can address the methylation status of repeated sequences (Edwards et al., 2010). It combines digestion with five restriction endonucleases, isolation of the unmethylated fraction by limited digestion with *Mcr*BC and NGS DNA sequencing.

### 2.3.2 Affinity purification of methylated DNA

The specific chemical structure of 5mC and the presence of proteins that specifically recognize and bind to this covalently modified nucleotide (MeCP2 and MBD2) have yielded several rapid affinity-based methylated DNA enrichment methods. Enrichment of 5mC-containing regions for the genome-wide analysis of DNA methylation has developed into a powerful analytical technique. Various approaches are taken using ChIP-like technologies, either using antibodies that are specific for 5mC, or, in a more indirect approach, antibodies against high-affinity methyl-binding proteins. Such affinity purification includes meDNA binding domain proteins (MDB) or DNA immunoprecipitation (DIP) of meDNA with anti-meC antibodies (MeDIP, mDIP or mCIP) or variants thereof. These

enrichment approaches have also been adapted for high-throughput sequencing analysis. MeDIP is limited to regions with a CpG density of at least 2–3%. MeDIP is being increasingly used to map methylation profiles (the 'methylome') of promoters in a variety of organisms and cell types with moderate resolution. AutoMeDIP-seq presents a largely automated MeDIP protocol with library preparation steps for subsequent second-generation sequencing as accurate and high-throughput methylome mapping tool (Butcher and Beck, 2010). Such immunoprecipitation methods do not require methylation-dependent digestion of genomic DNA. However, due to the nature of the immunoprecipitation methods, it is currently not possible to provide methylation data at a single-nucleotide resolution.

Antibody-based enrichment is also an option for the analysis of 5hmC, because 5mC antibodies do not recognize 5hmC (Jin et al., 2010). An antibody against 5hmC, which can distinguish 5hmC from 5mC, can be used for a genome-wide 5hmC methylation analysis using a method similar to MeDIP in combination with an array or sequencing platform (Gupta et al., 2010). Alternatively, selective chemical labeling of 5hmC with modified glucose residues followed by immunoprecipitation allows determining the genome-wide distribution of 5hmC (Song et al., 2011).

### 2.3.3 Bisulfite conversion

Incubation of DNA with sodium bisulfite deaminates unmethylated cytosine to uracil but leaves methylated cytosine largely unaffected. This way, an epigenetic modification turns into a difference in DNA sequence. The change in the DNA sequence can be detected using a variety of methods, including PCR amplification followed by DNA sequencing. It has prompted the development of many DNA methylation and analysis techniques (Tost, 2008; Laird, 2010). Again, technologies have matured tremendously over the last years.

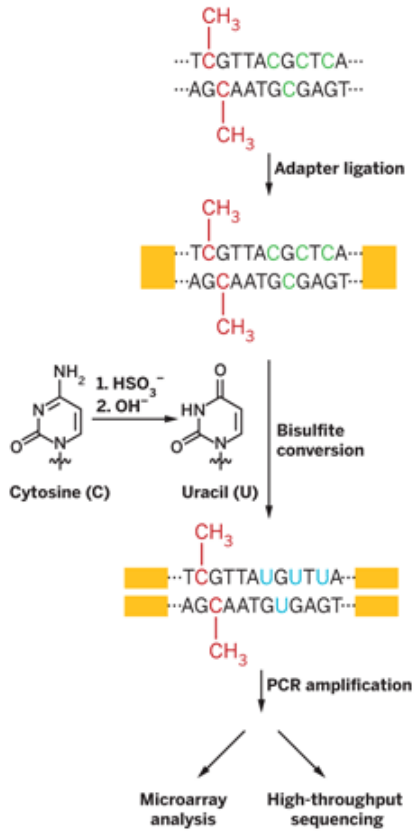



*Figure 5. Bisulfite conversion of DNA to detect methylated C residues (Cassiday, 2009).*

The use of bisulfite-converted DNA for methylation analysis (Figure 5) is surpassing every other methodology (Gupta et al., 2010). Various bisulfite-based methodologies are available and are capable of providing methylation data at a single-nucleotide resolution (Lister et al., 2009). However, bisulfite treatment cannot distinguish 5mC from 5hmC (Huang et al., 2010b; Nestor et al., 2010). After bisulfite treatment, the resulting genome is largely composed of only three nucleotides (A, G, and T). As a result, sequence complexity is reduced. This elicits various technical and analytical challenges. Hybridization assays, for example, tend to be less specific because of the reduced sequence complexity. Various platforms can be employed for analysis, such as in bisulfite methylation profiling (BiMP), as well as adjusted Illumina platforms (GoldenGate, Infinium). High-throughput sequencing is most likely the most appropriate technology for bisulfite assays, but also in this case the lower sequence complexity is an issue (Tost, 2008; Laird, 2010). Moreover, downstream analysis is cumbersome and requires specialized software. Various approaches for s-called reduced representation bisulfite sequencing (RRBS) have been proposed, but the expectation is that whole-genome shotgun bisulfite sequencing (WGSBS) is or will become the gold standard for DNA methylation analysis at the single base pair resolution, unless direct meC sequencing will develop to a straight-forward and reliable approach. Benchmark comparisons of different analytical techniques for DNA methylation in case-controlled studies showed all to produce accurate DNA methylation data (Beck, 2010), with a somewhat higher accuracy by bisulfite-based methods (Bock et al., 2010).

### 2.3.4 Alternative uses of DNA methylation

DNA methylation can be used as an alternative to ChIP in an approach that has become known as DamID (Van Steensel et al., 2001). DNA close to the target site of the protein-of-interest is methylated by a DNA adenine methyltransferase (Dam) fused to the DNA-binding protein-of-interest. The methylated sites are detected by digestion with a methyl-specific restriction enzyme (or otherwise, see above). The digestion products are analyzed for example by hybridization onto a microarray. DamID has been used to map binding sites for a variety of proteins. DamID and ChIP-on-chip approaches have been reported to yield very similar results (Van Steensel, 2005). DamID has the advantage that it does not require a specific antibody, but it has various other concerns (Orian et al., 2009).

## 2.4 Assays for histone modifications and histone variants

Genome-wide studies of histone modifications and nucleosome composition are performed with chromatin immunoprecipitation (ChIP), followed by microarray-based hybridization (ChIP-chip) or second-generation sequencing (ChIP-seq). It is to date the only technique allowing such studies (Schones and Zhao, 2008). ChIP-seq does not require any pre-selection of genomic regions. The resolution depends on the size of the chromatin fragments, the depth of sequencing as well as the quality of the antibody used. The technology allows assigning modifications to individual nucleosomes when using micrococcal nuclease. Histone variants are studied using the same technologies. Nucleosome positioning assays are generally based on the preferential cleavage of linker DNA over nucleosomal DNA by micrococcal nuclease, followed by tiling arrays or sequencing. Histones consist of a globular domain and a more flexible amino terminus, known as histone tail. The histone tail protrudes from the nucleosomal surface and can be modified by covalent acetylation, methylation, phosphorylation, ubiquitination, ribosylation, sumoylation, and biotinylation. Developments in mass spectrometry have contributed to the discovery of posttranslational modifications of histone proteins also in the globular domain.

A plethora of histone and histone related antibodies are commercially available. These include antibodies that specifically recognize the core histones and their variants modified by lysine acetylation, biotinylation, mono-, di- or tri-methylation; arginine mono- or di-methylation; serine or threonine phosphorylation and arginine substitution with citrulline. Antibodies also include histone variants and histone modifying enzymes as targets (Collas, 2010b).

## 2.5 Detection of small RNAs

Small RNAs are studied by regular RNA technology for isolation and enrichment, often adjusted for the small-sized nature of the molecules involved. Second-generation sequencing of RNA (RNA-seq) with focus on the small RNA complement of cells or tissues (sRNA-seq) for small RNA profiling is now more or less routine (Ozsolak and Milos, 2011). The depth of RNA-seq has contributed considerably to our knowledge on the small RNA world and appears to continue to identify new family members. A major limitation of RNA-seq is the apparent lack of correlation between the number of reads obtained for a given sRNA molecule, and their actual abundance (Linsen et al., 2009).

## 2.6 Assays for the assessment of nuclear architecture

The three-dimensional nuclear and chromatin architecture is studied with a variety of microscopic-molecular techniques. Main technologies include fluorescent in situ hybridization (FISH) and in vivo tagging of selected genomic loci, showing events within the nucleus of a single cell using microscopy (Van Steensel and Dekker, 2010). Over the past years, new techniques have started to offer views beyond the resolution limits of microscopy. Combined with genome-wide detection methods they allow the analysis of the nuclear architecture of entire genomes. The above approaches focus on physical interactions of genomic loci with relatively fixed nuclear structures or landmarks, such as the nuclear envelope or the nucleolus (Van Steensel and Dekker, 2010).

To investigate the spatial folding or bending of chromatin and chromosomes, the technique of chromosome (or chromatin) conformation capture (3C) is used (Louwers et al., 2009). This technology involves cross-linking chromatin, digestion of interacting regions with restriction enzymes, ligating the fragments, PCR amplification and hybridization or sequencing and counting the frequency by which two fragments are combined as parameter for nuclear distance (Figure 6). This way, one determines the relative frequency with which pairs of genomic loci are in direct physical contact. Ligation products present fragments that were sufficiently close to become cross-linked and give the population-averaged folding of the entire genome at a resolution of several kilobases.

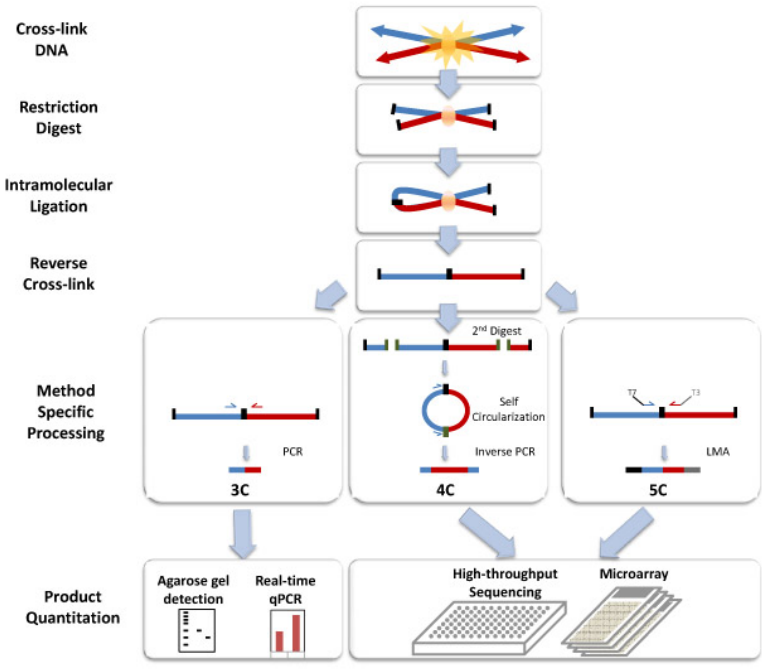



*Figure 6. Outline of the chromosome conformation capture (3C) technology to study the spatial organization of chromosomes (from http://en.wikipedia.org/wiki/Chromosome_conformation_capture).*

The technology known as 3C now consist of a rapidly developing family of related high(er) throughput technologies that use microarrays or high-throughput sequencing to analyze the ligation junctions. These variants are known as 4C (3C-on-chip or circular 3C with sequencing), 5C (3C-carbon copy), which combines 3C with multiplexed amplification to create libraries for sequencing, ChIP-loop (or 6C) as well as chromatin interaction analysis using paired-end tag sequencing (ChIA-PET; (Li et al., 2010). The latter two techniques include a ChIP step to selectively identify the ligation products that are bound to a protein-of-interest (Van Steensel and Dekker, 2010), either local or mote global. In the newest technologies, ligation junctions are marked with biotin, enabling purification with streptavidin prior to sequencing, thus allowing the analysis of chromatin interactions at more depth.

## 2.7 Epibioinformatics

The rise of the data volume due to the new high throughput technologies prompts for the concomitant development of appropriate data analysis and management methods. There is a growing need for software for data analysis and management. Notably the high-throughput technologies used in epigenetic research such as bisulfite sequencing, small RNA-seq and the various ChIP analyses outlined above require development and use of appropriate and user-friendly software and statistics, preferably using unified workflows and analytical pipelines.

The list of computational resources for the analysis of DNA methylation data, for example, is growing rapidly (Laird, 2010), as are resources (Huss, 2010) for other types of data (Brady and Provart, 2009). Generally, applications are diverse, quality is likely to be variable and use of the resources may require considerable, if not daunting, skills. Also adjusting genetic mapping technology to include epigenetic variation may be challenging for current computational approaches and infrastructures due to the amounts of data and computation required (Johannes et al., 2008). Although declaring the existence of (or need for) the field of 'epibioinformatics' may be somewhat overstated, future proper epigenetic analyses will benefit from accessible, valid and validated methods for analyses across laboratories and analytical platforms.

## 2.8 Concluding remarks

Epigenetic technology and research have progressed tremendously in recent years, notably because of the developments in DNA and RNA sequencing in terms of throughput and costs. Combined with chromatin-based analyses such as ChIP, these methods are generating impressive amounts of data helping to unlock the information present in the epigenetic complement of cells. The future will see further developments in sequencing and single-cell technology. This way, analyses of the epigenome of the single cell will come in reach (Glaser, 2010) and contribute to account for the anticipated large cell-to-cell differences in gene expression and possibly gene regulation. Major challenges for the future will focus on the interpretation and integration of the various technologies and data sets. Much is to be expected from the further integration of statistical genetics and epigenetics (Johannes et al., 2008), especially in combination with the expansion of the new and growing field of systems biology.

# 3. Epigenetic machinery and its molecular organization

As outlined above, in this report, four main categories or levels are distinguished in the molecular machinery of epigenetic modification:

- DNA (de)methylation
- protein (histone) modifications and histone variants
- RNA-mediated mechanisms
- higher-order chromatin organization.

Yet, all four epigenetic features interact with numerous positive and negative feedback mechanisms, including environmental signals. The challenge still is to understand in full what has been called 'the epigenetic triad' (Goodman et al., 2010). Including the phenomena of higher-order chromatin organization as part of epigenetics, as is done in this report, it would be more aptly called the 'epigenetic quadrangle'. The outcome of any molecular analysis will be the result of all such interactions. In the previous report (Nap and Geurts van Kessel, 2006), we have outlined the basics of these different categories of epigenetic modification. Here we will only recapitulate the most salient features of each level and concentrate on the new approaches and insights that have been presented in the literature since 2006.

## 3.1 DNA (de)methylation

DNA methylation involves the chemical modification of DNA by the addition of a methyl group to carbon-5 of the cytosine pyrimidine moiety (5mC). It is brought about by DNA methyltransferases (DNMTs) through the transfer of a methyl group from S-adenosyl methionine to DNA (Nap and Geurts van Kessel, 2006). DNA methylation is still by far the most widely studied epigenetic modification, also thanks to major technological improvements for large-scale and high-throughput analysis (see above). The new technology to determine DNA methylation status of individual nucleotides in DNA has resulted in several surprises that challenge the conventional views on the role of DNA methylation in the regulation of gene expression (Suzuki and Bird, 2008). The picture has -again- become much more complex. The overall occurrence of DNA methylation appears to be again more complex, and the functional links between DNA methylation and reduced (or silenced) transcription appears to be more subtle. Methylation is not just an effective mechanism to turn a gene off; rather it has become a way to intricately regulate the turning off and on of genes. Gene body methylation, for example, is positively associated with gene expression.

In flowering plants, methylated cytosines are detected in all sequence contexts, whereas CG methylation predominates in animals (Feng et al., 2010). In mammals, DNA methylation was thought to involve almost exclusively the cytosines in CpG dinucleotides, whereas in plants and (some) fungi, also methylation of cytosines in symmetrical CHG or asymmetrical CHH contexts (where H = A, C or T) occurs. Recently, however, more DNA methylation at CHG and CHH sites than was previously anticipated was reported in human stem cells (Lister et al., 2009). Such non-CpG methylation decreases during differentiation, but is restored in pluripotent stem cells (Laurent et al., 2010), suggesting an important role of this non-CG methylation in the origin and maintenance of the pluripotent state. The mechanism of CNN/CNG methylation in human is yet unclear (Laurent et al., 2010).

In metazoans, transcriptional activity was associated with the absence of DNA methylation in so-called CpG islands. CpG islands are defined as regions of more than 200 bases with a GC content of at least 50% and a ratio of observed-to-expected CpG frequencies of at least 0.6 (Portela and Esteller, 2010). About 60% of human gene promoters are associated with such CpG islands and these promoters are usually unmethylated in normal cells. In general, CpG-island methylation is associated with gene silencing. However, a small proportion of CpG island-associated promoters (~6%) is methylated when active in a tissue-specific manner during early development or in differentiated tissues (Straussman et al., 2009). A next surprise relative to the understanding of DNA methylation in

2006 was the occurrence of notable DNA methylation outside CpG islands as means of transcriptional regulation. Regions of lower CpG density in close proximity (~2 kb) to CpG islands, now known as 'CpG island shores' (Irizarry et al., 2009) are involved in the regulation of transcriptional inactivation. Notably tissue-specific DNA methylation tends to occur not at CpG islands but at CpG island shores. Differentially methylated CpG island shores are sufficient to distinguish between different tissues.

Even more surprising was the finding, initially in the plant Arabidopsis, but later also in mammalian genomes (Suzuki and Miyazono, 2011), that about one third of all the genes in Arabidopsis carry methylated CpG residues in the transcribed region (Figure 7). Such DNA methylation tends to be present at the core of the transcription unit rather than the (5' or 3') end and is therefore called 'gene body methylation' (Zilberman et al., 2007). Whereas DNA methylation is generally considered as a mechanism to silence transcription, gene-body methylation correlates with higher expression levels than the expression levels of genes with either unmethylated promoter sequences or entirely unmethylated sequences. This way, gene body methylation is associated with transcriptional activation. The start of transcription is not affected and it may contribute to the prevention of spurious initiation of transcription (Zilberman et al., 2007). Gene body methylation is thought to be related to the efficiency of elongation. Gene body methylation is thought to be common in ubiquitously, but moderately expressed genes that classify as 'house-keeping' genes (Suzuki and Bird, 2008). As different ecotypes of Arabidopsis show differences in the status and distribution of gene body methylation (Vaughn et al., 2007), this epigenetic feature is apparently variable within the same species and may reflect some of the environmental history of the individual. Moreover, the pattern of gene body methylation in rice seems different from that in Arabidopsis. This indicates the existence of different functions of epigenetic pathways in different plant species, although within ecotypes of a species the methylation patterns are stably maintained for apparently very long periods (Zhang, 2008).

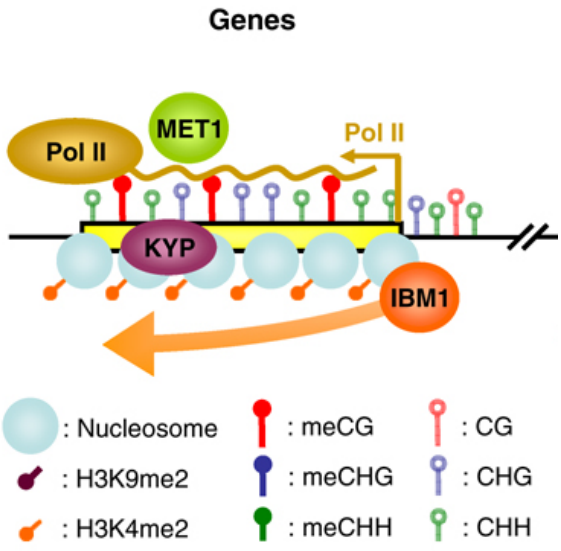



*Figure 7. Gene body methylation as part of the regulation of gene expression (Teixeira and Colot, 2009).*

Moreover, the relative stability of DNA methylation is being debated. It has become more likely that, at least for particular target genes, DNA methylation may be subject to a dynamic turnover (Hemberger and Pedersen, 2010), due to active DNA demethylation (Wu and Zhang, 2010). Possibly, the conversion of 5mC to 5hmC is a first step towards DNA demethylation (Hemberger and Pedersen, 2010). Active DNA demethylation may suggest that DNA methylation patterns have to be more actively maintained during cell division than previously assumed (Tahiliani et al., 2009). In Arabidopsis, demethylation primarily occurs at the 5' and 3' ends, a pattern opposite to the overall distribution of DNA methylation, editing the patterns of DNA methylation within the Arabidopsis genome (Penterman et al., 2007).

The overall scenario for the relationship between DNA (de)methylation and transcription can be summarized as follows (Portela and Esteller, 2010): promoters of genes are normally unmethylated at CpG islands, allowing transcription. The methylation patterns of CpG island shores add further flexibility to the regulation of transcription, notably with respect to tissue specificity. Methylation at a gene body appears to facilitate elongation and may help to prevent spurious transcription initiation. In line with the view that DNA methylation confers transcriptional silencing, also pseudogenes and repetitive sequences, notably transposons, tend to be hypermethylated, preventing chromosomal instability, translocations and gene disruption. In (human) disease, aberrant hypermethylation of CpG islands or CpG island shores leads to transcriptional inactivation, whereas the gene body tends to demethylate, allowing transcription to be initiated at several incorrect sites. In addition, the demethylation of repetitive sequences can occur, affecting chromosomal integrity.

The mechanisms by which DNA methylation can inhibit gene expression are manifold and extensive interplay exist between DNA methylation and other epigenetic features, such as histone modification (Portela and Esteller, 2010). DNA methylation can direct histone modifications and vice versa. Methylated DNA can promote the recruitment of methyl-CpG-binding domain (MBD) proteins that, in turn, recruit histone-modifying and chromatin-remodeling complexes. DNA methylation can also directly inhibit transcription by precluding the recruitment of DNA binding proteins to their target site. In contrast, unmethylated CpG islands generate a chromatin structure favorable for gene expression (Thomson et al., 2010). Moreover, as indicated before, DNA methylation and DNA methylation-associated proteins are involved in nuclear organization and in the establishment of specific chromosomal territories.

Yet another surprising finding with respect to DNA methylation was the apparently fairly widespread occurrence of a hydroxylated methyl group attached to cytidine, resulting in 5-hydroxymethyl-2-deoxycytidine (5hmC) (Figure 8). This modification constitutes 0.6% of all methyl modifications in Purkinje cells and 0.2% in brain cells, but it seems not to be present in cell lines (Kriaucionis and Heintz, 2009). The presence of 5hmC modifications in embryonic stem cells and brain neurons, which both have high plasticity, indicates that 5hmC is involved in the regulation of gene expression. Recently, however, 5hmC was demonstrated in more cells/cell types and its presence was suggested to have a relationship with ageing and neurodegenerative diseases (Song et al., 2011). As current bisulfite-based methylation assays do not allow the discrimination between 5mC and 5hmC, existing methylation maps may need to be revisited or detailed with new information (Beck, 2010). Only a subset of 5mC is converted to 5hmC, raising the possibility that cis-acting elements determine which 5mCs are targeted for conversion to 5hmC (Gupta et al., 2010). As DNMT1 and MeCP2 do not bind to 5hmC, the modification could be involved in fine-tuning the transcriptional repression machinery. The conversion of 5mC to 5hmC could be a first step towards DNA demethylation (Hemberger and Pedersen, 2010), or the creation of a change in DNA sequence (mutation) to create a single nucleotide polymorphism. This new DNA modification needs to be studied further to determine its role in epigenetic regulation and provide insight into the role of this modification in gene expression, cellular immortalization and transformation (Portela and Esteller, 2010).

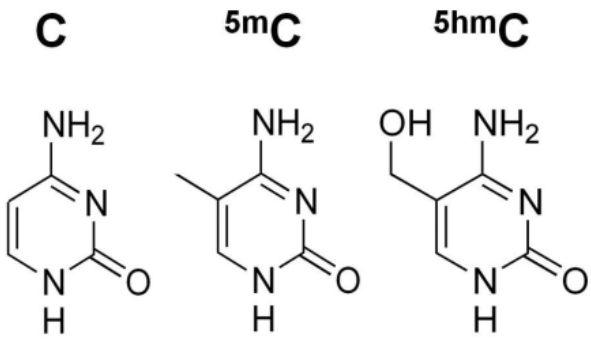



*Figure 8.* *A new component of DNA: 5-hydroxymethylcytosine (5hmC), compared to 5-methylcytosine (5mC) and cytosine (C) itself.*

Major players in the epigenetic transfer of DNA methylation are DNA methylases (DNMT). In mammals, three out of five members of the DNMT family have methyltransferase activity, two de novo and one for maintenance, although the latter has also de novo activity. The de novo DNMTs are highly expressed in embryonic stem (ES) cells and become down-regulated in differentiated cells (Portela and Esteller, 2010). They are seen as responsible for establishing the pattern of methylation during embryonic development and possibly during transgenerational epigenetic inheritance. The maintenance DNMT, DNMT1, has a 30- to 40-fold preference for hemi-methylated DNA and is thought to methylate hemi-methylated sites generated during replication. However, the division of tasks between de novo and maintenance methylation is not completely clear and may involve redundancy.

Moreover, it is not very clear how sequence specificity in DNA methylation is realized. It is generally assumed to involve other epigenetic features, such as small inhibitory (si)RNA-mediated, RNA-directed DNA methylation. There may be differences between symmetrical and asymmetrical methylation sites. The role and occurrence of RNA-directed DNA methylation (RdDM) to recruit DNMTs to catalyze de novo DNA methylation of genes and repeats is better documented in plants (Vrbsky et al., 2010) than in animals. In an Arabidopsis methyltransferase mutant, loss of CG methylation over generations triggers apparently stochastically a genome-wide activation of epigenetic mechanisms, such as RdDM, reduced expression of DNA demethylases and retargeting of a particular histone methylation mark. Results are interpreted to imply that DNA methylation is coordinating the epigenetic memory for stable transgenerational inheritance in plants (Mathieu et al., 2007).

## 3.2 Modifications involving histone proteins

Three different types of modifications involving (histone) proteins play a role in the molecular machinery of epigenetic modifications: histone modifications, nucleosome positioning and histone variant exchange. These will be summarized below. Relative to the understanding in 2006, only few surprises or new insights have been reported.

### 3.2.1 Histone modifications

In the nucleus, DNA is tightly folded around proteins to assemble the nucleosome, which comprises four different histone proteins, H2A, H2B, H3 and H4 (Nap and Geurts van Kessel, 2006). Posttranscriptional modification of these histone proteins plays a key role in epigenetic regulation. Various post-transcriptional modifications can occur at histone tails or at the folded core domains: acetylation, methylation, phosphorylation, ubiquitination, sumoylation and ADP-ribosylation, among others. All these modifications involve highly dynamic covalent binding of moieties to histone tails. Histone modification in the epigenome may act as a very dynamic process (Roudier et al., 2009).

Numerous enzymes that catalyze or remove these modifications have been analyzed in considerable detail and it is unlikely that the list is already complete (Portela and Esteller, 2010). The enzymes involved differ in substrate specificity. Histone methyltransferases, demethylases and kinases seem most specific to given histone subunits and tails (Chi et al., 2010), whereas histone acetyltransferases (HATs) and histone deacetylases (HDACs) generally modify various different residues. Histones can carry different modifications at different sites simultaneously. The particular combination of these tags may represent different types of chromatin. This has been proposed to constitute the so-called histone code (Nap and Geurts van Kessel, 2006). The combinatorial complexity of this code continues to be on the rise. Histone modifications can vary at the same site, the same histone tail and at different histone tails (Portela and Esteller, 2010). No less than 51 distinct 'chromatin states' based on the enrichment of specific combinations of histone modifications were distinguished to confer different biological activities (Ernst and Kellis, 2010). The particular histone modification distribution within such a chromatin type is predictive for gene expression. In Drosophila, five main chromatin types are defined (Filion et al., 2010). As is the case for DNA methylation, there are notable differences in histone modifications between promoter and gene body sequences (Karlic et al., 2010). Yet, the nature, dynamics and maintenance of the boundaries separating epigenetically distinct chromatin compartments are not yet clear.

Such histone modifications, in conjunction with the position of the nucleosomes along the DNA (see nucleosome positioning below), are involved in the regulation of many processes involving nuclear DNA: transcription, replication, repair and more (Nap and Geurts van Kessel, 2006) and may also have a role in (alternative) splicing (Ringrose, 2010). The relationship between transcription and the particular type(s) of modification is complex and seems to be context- as well as histone residue-dependent. In general, methylation at H3K4 and H3K36 is associated with transcriptional activation, whereas methylation at H3K9 and H3K27 is associated with transcriptional inactivation. Several transcriptional co-activators have HAT activity, whereas different transcriptional co-repressors exhibit HDAC activity. This indicates that HDACs function to reset chromatin by removing acetylation at active genes, whereas HATs are linked to transcriptional activation (Wang et al., 2008; Wang et al., 2009b).

Additional surprises with respect to the regulation of histone modifications have turned up in recent years. The presence of noncoding RNAs (ncRNAs) derived from heterochromatic regions has for example questioned the status of heterochromatin as transcriptionally inactive (Zaratiegui et al., 2007). In e.g. the yeast *Schizosaccharomyces pombe*, centromeric regions generate siRNAs that provide sequence specificity to the (RNA-induced) transcriptional silencing complex and recruit histone methylases involved in spread of the heterochromatin state. Similar observations exist in other organisms. In addition to covalent modifications, also clearing of histone marks was found to be involved in the regulation of gene expression (Santos-Rosa et al., 2009). In some instances, the tail of histone H3 is clipped after residue Ala21, removing the 21 N-terminal residues of the tail and all associated covalent post-transcriptional modifications, resulting in a massive clearing of histone marks. The extent and importance of histone tail clipping in the regulation gene expression is currently being investigated. Extensive interplay exists between histone modification and other epigenetic features, such as DNA methylation (Portela and Esteller, 2010). Histone methyltransferases, for example, recruit DNMTs for targeting DNA methylation to specific genomic regions (Zhao et al., 2009), thereby enforcing the silenced state established by the repressive histone marks and vice versa. Histone methyltransferases and de-methylases also regulate DNA methylation levels by modulating the stability of DNMT proteins (Wang et al., 2009a). Moreover, it has been reported that HDACs and HATs are both targeted to transcribed regions of active genes by phosphorylated RNA polymerase II (Wang et al., 2009b). The involvement of modifications of non-histone proteins in the regulation of gene expression adds yet another layer of complexity to the regulation of gene expression.

### 3.2.2 Nucleosome positioning

The positioning of nucleosomes to the DNA and concomitant packaging of DNA is thought to convey a next level in the regulation of transcription. It does not involve alterations in histone proteins, but plays a role in determining the accessibility of DNA for transcription (Nap and Geurts van Kessel, 2006). Nucleosomes including all modifications of the histones that constitute a particular nucleosome, can act as a barrier to transcription, blocking access of activators and transcription factors to the DNA, but can also interact with transcription factors to facilitate remodeling, or inhibit the elongation of transcripts by RNA polymerase. Large macromolecular complexes are known to move, destabilize, eject or restructure nucleosomes in an ATP hydrolysis–dependent manner. Such chromatin remodeling complexes share similar ATPase domains but differ in the composition of their unique subunits. The different complexes are seen as master regulators of gene expression, to promote chromatin assembly or disassembly, to repress or activate transcription and/or to participate in a variety of cellular processes including telomere regulation, chromosome segregation and DNA replication (Ho and Crabtree, 2010). Notably the position of nucleosomes around the transcription start site influences the initiation of transcription. Moreover, the 5' and 3' ends of genes possess nucleosome-free regions to provide space for the components of transcription (Cairns, 2009). It is tempting to speculate that nucleosome positioning may affect the gene body methylation discussed above (Chodavarapu et al., 2010). Genome-wide studies have, for example, revealed a correspondence between nucleosome positioning and intron-exon boundaries, suggesting a link between chromatin properties and splicing (Ringrose, 2010). Obviously, there exists an extensive interplay among the different members of the epigenetic machinery. Nucleosome remodeling is influenced by DNA methylation (and vice versa) and is also linked to specific histone modifications.

### 3.2.3 Histone variant exchange

Another level to modify the function of nucleosomes and chromatin entails the incorporation of different histone variants. These variant histone proteins differ from core histones in their tails, in their domain structure and in their key amino acids. Histone variants can affect nucleosome positioning and gene expression (Nap and Geurts van Kessel, 2006). For some variants, chromatin remodeling complexes are thought to be involved in the exchange of core histone variants. A particular chromatin remodeling complex is, for example, able to replace the H2A-H2B dimer of the nucleosome with the variant H2A.Z-H2B (Clapier and Cairns, 2009). Integration of epigenetic regulatory levels also occurs at the level of histone variant exchange. The incorporation of the histone variant H2A.Z, for example, protects genes against DNA methylation (Zilberman et al., 2008) and has emerged as key element in the regulation of genes, notably in relation to the environment (Deal and Henikoff, 2011). Exclusion of histone H2A.Z from methylated DNA is conserved between plants and animals (Zemach et al., 2010). Additionally, microRNAs (miRNAs) can affect histone variant replacement (Lal et al., 2009) or interact with chromatin remodeling complexes mediating the exchange of specific subunits (Yoo et al., 2009).

## 3.3 RNA-mediated mechanisms

Small RNAs (sRNAs) can direct and mediate epigenetic modifications of genomes (Nap and Geurts van Kessel, 2006). Recent years have seen an enormous expansion of the types and roles of members of the small RNA world in the regulation of gene expression. The numbers, types and roles of sRNAs have increased consistently and sRNAs are now known to be core components of signaling and interaction networks that mediate gene expression and epigenetic modifications in plants, animals and humans. Small RNAs probably affect nearly every biological process in every eukaryotic cell in a direct or indirect way (Czech and Hannon, 2011). A highly interactive network of small RNA-directed methylation of DNA and modification of histones and chromatin are involved in the control of transcription. Ample research has been devoted to the identification and functional characterization of components of the siRNA-directed DNA methylation pathway in plants (Simon and Meyers, 2010) to provide more insight into the underlying epigenetic phenomena. For example, the earlier discussed phenomenon of paramutation (Nap and Geurts van Kessel, 2006) is now also closely linked to a small RNA pathway (Erhard Jr et al., 2009; Arteaga-Vazquez and Chandler, 2010).

Regulatory sRNAs are short (approximately 20–24 nt in length), noncoding RNAs produced through the RNA interference (RNAi) pathway that involves multiple polymerases (Gao et al., 2010), among which some are apparently specific for plant systems. Transcriptome deep sequencing (RNA-seq) continues to lead to an expansion of the already extensive list of small RNAs, including snoRNA, miRNA, siRNA and, in plants, tasiRNA as well as natural *cis*-acting siRNAs. Small RNAs of generally less than 50 bases also include PIWI-interacting RNAs (piRNAs). The latter play a role in regulatory processes as diverse, yet apparently related, as transcriptional regulation, control of chromatin structure, heterochromatin formation and translational regulation (Caley et al., 2010).

MicroRNAs (miRNAs) and trans-acting small interfering (tasiRNAs) are typically ~21 nt in length and are involved in post-transcriptional silencing in complexes called RNA-induced silencing complexes (RISCs) consisting of proteins of the Argonaute family and other proteins (Czech and Hannon, 2011). The tasiRNAs typically induce cleavage of target mRNA with near-perfect complementarity. For targets with less-than-perfect complementarity, generally miRNAs are involved in either translational repression, mRNA destabilization and decay, or both (Djuranovic et al., 2011). In rice, a class of longer miRNAs (of 24 nt in length) was identified that directs DNA methylation to regulate gene regulation (Wu et al., 2010), indicating that the length of the small RNA may not be the best criterion for classification. Notably miRNAs have received a lot of attention and a wealth of (review) papers is available describing their characteristics, myriad roles and targets (Bartel, 2009; Fabian et al., 2010; Krol et al., 2010). The numbers of identified miRNAs has increased impressively over the years. The latest release of the miRNA repository miRBase (release 16, Sept 2010; www.mirbase.org) contains over 15.000 entries. Overall, it has become clear that miRNAs can use multiple, often seemingly contradictory, mechanisms to regulate their target mRNAs, and such mechanisms are also regulated in various ways. Numerous factors participate in miRNA biogenesis and also the regulators of miRNA biogenesis themselves are subject to control. These notions indicate that the roles of miRNAs are -or can be-

highly context-specific (Fabian et al., 2010; Krol et al., 2010) and may involve stochastic events. Notably, the search for miRNA targets is now in progress. Target gene identification is challenging and the rules of targeting are not completely understood yet (Thomas et al., 2010). In many diseases, notably cancer, the expression of miRNAs is deregulated and in many instances it was shown that miRNA profiling results in signatures that have -or can have- both diagnostic and prognostic value and offer various possibilities for therapeutics (Cho, 2010; Heneghan et al., 2010; Raber, 2010; Schaefer et al., 2010). The application of miRNA-based approaches in clinical practice is currently an area of very active research.

Small interfering RNAs (siRNAs) are typically 24 nt in length and are involved in heterochromatin formation and transcriptional gene silencing. In plants, the siRNAs guide sequence-specific DNA and histone methylation through a pathway termed RNA-directed DNA methylation (RdDM), the details of which are becoming known in considerable detail now (Simon and Meyers, 2010). It involves converting single-stranded RNA (ssRNA) into double-stranded RNA (dsRNA) by RdRP and the involvement of various protein complexes (known as 'primary RdDM') to result in methylated DNA and histones. The primary siRNAs produced in this pathway can be amplified by a turnover mechanism in which a methylated DNA template is transcribed, producing aberrant or perhaps atypically processed RNA that results in additional siRNAs. The pathway allows for spreading into areas adjacent and beyond the primary target site via secondary siRNAs (known as 'secondary RdDM') and methylation. Establishment and maintenance of primary RdDM is independent of secondary RdDM (Daxinger et al., 2009). The occurrence and role of RdDM in metazoans is yet unclear. The siRNA-directed DNA methylation involves recruitment of (de novo) DNMTs, followed by siRNA-directed (de novo) DNA methylation at the DNA target region. Genome-wide views of cytosine methylation and snapshots of the state of chromatin in genomes show that epigenetic marks of silencing in a genome landscape are prominent in regions enriched in transposons, retro-elements, pericentromeric regions, and rRNA genes (Cokus et al., 2008; Lister et al., 2008; Lister et al., 2009).

There are strong correlations between small RNAs and DNA methylation. There is a 25-fold greater chance of identifying a methylcytosine at a siRNA-producing locus than finding a methylcytosine at a non-sRNA locus, indicating that siRNAs are driving DNA methylation and vice versa (Lister et al., 2009). Overall, siRNA-directed DNA methylation covers about 30% of the Arabidopsis genome. As two-third of the methylated loci are not associated with siRNAs, there should be substantial genomic cytosine methylation independent of small RNAs (Simon and Meyers, 2010). Deep sequencing of the small RNA complement in DNA methyltransferase and demethylation mutants showed altered siRNA populations, so both methylation and demethylation modulate sRNA levels (Lister et al., 2008), thereby further substantiating the dynamics of DNA methylation in gene regulation (Furner and Matzke, 2011). Comparative DNA methylation studies showed that gene body methylation is conserved between plants and animals, so (CpG) gene body methylation is an ancient property of eukaryotic genomes (Zemach et al., 2010). In contrast, selective methylation of transposons is not. A trend observed in several species is that the genes most likely to be methylated are modestly expressed, whereas the genes least likely to be methylated are at the extremes of transcriptional activity. The methylation patterns in rice resemble those in Arabidopsis, but differ from the early diverging land plants. Transposons and repeats are uniformly methylated in all of the plant species (Zemach et al., 2010). The green alga Chlamydomonas has a most unusual pattern of methylation, having non-CG methylation enriched in exons of genes rather than in repeats and transposons. DNA maintenance methylase has a conserved function in CG methylation in both transposons and gene bodies in the mouse, Arabidopsis, and zebrafish genomes (Feng et al., 2010).

Heterochromatin is typically composed of transposons, retrotransposons, and other repetitive elements that are maintained in a transcriptionally silent state usually attributed to methylation or post-translational histone modifications. In plants, a large portion of sRNAs originate from repeats and transposons, possibly reflecting the epigenetic 'architecture' of plant genomes and the RdDM pathway (Simon and Meyers, 2010).

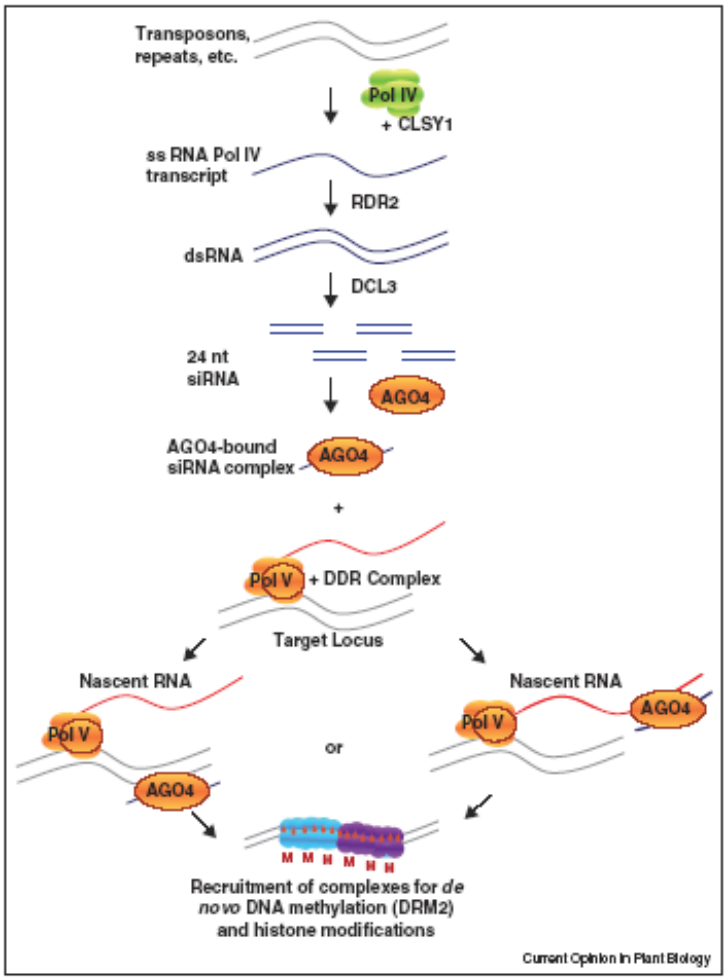



*Figure 9.* *Small RNAs have a direct role in heterochromatin formation (Simon and Meyers, 2010).*

Many studies showed that the RdDM pathway is directly involved in chromatin organization, by demonstrating a reduction in DNA methylation, a reduction or elimination of 5S-derived siRNAs, derepression of 5S rDNA genes, changes in chromatin compaction and differential silencing in mutant lines (Simon and Meyers, 2010). In wheat, maize and others, large numbers of sRNAs match transposable elements (TEs). As the wheat genome is composed for more than 80% of TEs, epigenetic silencing of TEs is thought to serve an important role in suppressing the mutagenic activity of TEs (Cantu et al., 2010). The growing number of whole-genome datasets now available for DNA methylation and histone modifications will allow the identification of inactive chromatin from its marks rather than from the presence of specific repeats or cytological observations. As such, it will allow a better definition of the characteristics of inactive or heterochromatic regions in genomes without repetitive characteristics, and will help to determine if and why some repeat elements lack the epigenetic marks of such regions.

The important role of sRNAs in epigenetic regulation was further emphasized by the unequivocal demonstration that both siRNA and miRNA are mobile signals that control gene expression in plant development (Martienssen, 2010). The sRNAs move from cell-to-cell to deliver signals specifying leaf and root developmental patterns, as well as epigenetic (re)programming and inheritance. Mobile 24 nt sRNAs direct DNA methylation in the genome of the recipient cell. Mobility depends on factors such as genomic locus, origin of the sRNA, and the cell type in which the sRNAs accumulate (Dunoyer et al., 2010; Molnar et al., 2010). In developing pollen, an interesting small RNA-mediated epigenetic restructuring was found to involve mobile sRNAs. In the vegetative nucleus (VN), TEs are transcribed specifically in the non-germ-line nucleus. The increase in TE transcripts does not result in inherited transposition effects since the VN does not contribute DNA to the embryo, but these TE transcripts stimulate the

production of sRNAs which appear to be mobilized to suppress transposons. In this way they are thought to protect the germ-line sperm cells (Slotkin et al., 2009). Similar epigenetic effects play a role in female gamete development (Olmedo-Monfil et al., 2010). Sexual reproduction in plants, therefore, depends on specific epigenetic reprogramming events that serve as a defense mechanism to prevent the incursion of transposons at a critical phase in the life cycle. Striking parallels are found in animal gametogenesis (Bourc'his and Voinnet, 2010). In view of all these recent developments in the small RNA world, we may still be in the beginning stages of comprehending the complexity of sRNA-mediated epigenetic phenomena (Simon and Meyers, 2010). The machinery involved in sRNA biosynthesis is not completely understood, nor is the regulation of sRNA-controlled methylation and formation of inactive chromatin, notably in relationship to environmental signals.

In addition to small non-coding RNAs, evidence is accumulating that also long noncoding RNAs (lncRNAs) are involved in the regulation of epigenetic marks across the epigenomic landscape (Caley et al., 2010). Generally, genomes undergo widespread transcription, most of which is noncoding. This transcription generates more than transcriptional noise and is supposed to have a functional role in guiding and assembling proteins responsible for epigenetic marks. In humans, the ENCODE project indicated that over 90% of the genome is transcribed, while only 1-2% is translated into protein. This observation is now confirmed in many species and its conservation suggests functional roles. The lncRNAs vary in length from 200 bases to an astonishing > 100 kb, are transcribed by RNA pol II, are often polyadenylated and undergo splicing (Caley et al., 2010). Because of their size, they may fold into complex structures that may combine RNA-RNA or RNA-DNA interactions with protein binding (Bonasio et al., 2010). They have roles in imprinting, enhancer function, chromatin structure, genome rearrangements and normal development. Also in plants data are accumulating that lncRNAs (>100 nt) are involved in the regulation of transcription, chromatin structure and nuclear organization. They may have a role in the phenotypic plasticity of plants. It is anticipated that lncRNA will be an important area of plant research soon (De Lucia and Dean, 2011). FISH experiments show that ncRNAs can coat whole chromosomes and that lncRNAs (or their formation) play a role in determining chromatin structure. This way, lncRNAs are emerging as yet another key regulator of the epigenome. Based on the chromatin signature for a transcription unit, over a thousand lncRNAs with regulatory potential were identified in the human genome (Caley et al., 2010). Complex networks of transcription factors, lncRNAs and chromatin modifying proteins interact in largely unknown ways to regulate transcriptional programs or convey transcriptional interference. Various mechanisms are proposed and may exist next to each other (Caley et al., 2010).

## 3.4 Higher-order chromatin organization

Notably this level of chromatin organization is not considered part of 'true' epigenetics by several authors. It is in general very difficult to establish with certainty that a given higher-order (or any order, for that matter) chromatin modification truly exerts an epigenetic function, or vice versa. It may, if any, exert a downstream effect (Ptashne, 2007), rather than act as an initiator. Moreover, the higher-order chromatin organization is the consequence of epigenetic regulation. As opinions in the scientific literature differ, we include in this report this type of chromatin organization in the realm of epigenetics (Feinberg, 2010).

Long-range chromatin interactions may have a memory function involved in the maintenance of active and silent states of gene expression in what is considered a part of the 'epigenetic memory' (Deng and Blobel, 2010). Changes in chromatin organization and sub-nuclear localization are thought to facilitate rapid re-induction of transcription as part of that memory. Data indicate that transcriptional re-activation of a gene following a period of transcriptional silence requires less stimulation (i.e., has a memory of the previous state) than first-time activation, although this could reflect an effect of epigenetic memory, rather than a driver of it (Deng and Blobel, 2010). In recent years, various relationships between genes and their surroundings have been confirmed or newly discovered by the study of genome-wide epigenetic marks, such as the existence of large multigene chromatin domains (Feinberg, 2010). In addition, large organized chromatin lysine modifications (known as LOCKs) organize a genome into very large domains and serve as a mechanism for functional organization of a genome. Molecular modeling has demonstrated that chromatin looping can contribute to the formation of chromosome territories (De Nooijer et al., 2009). There are frequent intra- and inter-chromosomal interactions, which were discovered through chromatin capture technologies combined with NGS sequencing (Lieberman-Aiden et al., 2009). Such interactions are

supposed to exhibit regulatory epigenetic functions. Chromosomes interact with each other, and gene-rich pairs of chromosomes show preferred associations with each other. Also, centromeres and telomeres tend to cluster. These are all examples of a bewildering variety and complexity of interactions among different genomic elements across chromosomes within a genome (van Steensel and Dekker, 2010), the regulatory role of which remains to be established (Deng and Blobel, 2010), especially since all such interactions are highly dynamic (Pawlowski, 2010). Differences in nuclear organization between small genome species and large genome species (Delgado et al., 2010) may suggest a role for physical constraints in maintaining proper nuclear function (Wako and Fukui, 2010).

Generally, a pronounced cell-to-cell variation of nuclear architecture within similar cell types is observed (Cremer and Cremer, 2010). This variation could reflect a major plasticity in large-scale chromatin organization (Sinclair et al., 2010), or indicate that different epigenomes and their proper functioning depend on differences in higher-order chromatin organization and nuclear architecture at large.

## 3.5 Concluding remarks

Since 2006, insights into the epigenetic machinery and its molecular organization have increased considerably. New players continue to be identified, such as those involved in RNA regulation and DNA modification. There is no reason to assume that all players and all roles have now been identified and, as a result, new players can be anticipated. Moreover, it has become clear that considerably more integration and cross-talk exists between different epigenetic phenomena than previously anticipated. As a result, it turns out that the regulation of gene expression in biological systems continues to increase in complexity. It is also clear that the epigenome does contain information that is crucial for proper gene expression and development. The precise function and relevance of the epigenetic machinery, however, still largely remains to be established in the context of individual epigenomes. In addition, there may be a hierarchy in levels of epigenetic regulation, but if and to what extent such a hierarchy truly exists is not yet clear. The integrated and complex nature of the various epigenetic regulatory networks indicate considerable redundancy in gene regulation in order to allow cells and organisms to develop properly and to respond appropriately to changes in internal or external signals or environments.

# 4. Applications and use of epigenetics

Epigenetic modifications of DNA and changes in chromatin structure play important roles in cellular differentiation and development. Considerable progress is being made in the unraveling of the epigenetic regulation, modifications in nuclear organization and chromosome packaging that affect normal development and natural variation. New technologies facilitate the analysis of genome-wide epigenetic changes and their role in cellular memory in somatic cells and gametes. It has also been found that epigenetic modifications play important roles in the development and progression of disease. They influence gene expression and can show even stronger correlations with phenotypes than genetic changes in the primary DNA sequences. In several ways, these advances in the understanding of epigenetic mechanisms are being employed in biotechnology and beyond.
A main area of application of epigenetics is affecting gene regulation within a given tissue or organism that is not aimed to reach the germ-line: applied mitotic epigenetic inheritance. This area of application is mainly found in mammalian (human) systems and focuses on health and disease, in the very broad sense. Applications involve both diagnostics, prognostics and advanced therapeutics, and several well-documented examples of success are available. In view of the widespread importance of epigenetic phenomena, various areas beyond disease are considering or developing epigenetic applications. Examples of the latter can be found in the fields of neurology and memory as well as food and health. A second area of applications aims at reaching the germ-line and affecting endogenous gene expression stably over multiple generations: applied meiotic or transgenerational epigenetic inheritance. These applications are currently only found in the realm of plant systems. Although the occurrence of transgenerational epigenetic inheritance is well documented in plants, successes of its application are few, if any. In this chapter, we will give examples of applications of epigenetic inheritance, without the claim to be complete or exhaustive.

## 4.1 Applications of epigenetic inheritance in mammalian (human) systems

The major areas of application of mitotic epigenetic inheritance as introduced above are based on small RNA, notably miRNA, and on chromatin-based modifications.

### 4.1.1 miRNA-based approaches

The growing role and importance of miRNAs in cancer adds to their attraction as agents in clinical practice. As biomarkers, they are helpful for diagnostic, prognostic and monitoring purposes, such as for drug responses. Furthermore, they are obvious targets for novel therapeutic strategies. To aid molecular diagnostics, miRNAs can be detected and quantified in frozen tissues, but also in formalin-fixed paraffin-embedded tissues, as well as in serum or plasma samples. As further advantage, miRNA isolations generally require only small biopsy specimens. This way, miRNAs represent a major improvement over other approaches (Ferracin et al., 2010; White et al., 2011).
Various high-throughput analyses indicate that miRNA expression patterns are different in almost all human tumor types compared to the normal tissue counterpart (Fabbri, 2010). These differences translate to tumor-specific miRNA signatures and allow diagnosing the tissue of origin of the cancer (cancer classification). In some cases, also specific tumor subtypes can be identified and their expression profile can be effective in the identification of the tissue of origin of a metastasis to help decide the most appropriate treatment. In addition, miRNAs can serve as prognostic biomarkers of human cancers. Because some miRNAs are present in the blood of cancer patients, miRNAs can be used for non-invasive diagnostics and prognostics (Fabbri, 2010). In prostate cancer as well as breast cancer, the presence of miRNAs in the blood appeared to correlate with the occurrence of metastatic tumors. When the diagnostic relevance of circulating miRNAs is validated in large patient cohorts, the respective cancers and/or metastasis can be detected very early in routine blood examinations (Asslaber et al., 2010; Ferracin et al., 2010), although, as of yet, the clinic tends to be somewhat reluctant.

Insight into the role of miRNAs in other disorders, such as in neurodegenerative disorders and Alzheimer's disease (Mouton-Liger et al., 2010), vascular disease (Jamaluddin et al., 2011), blood poisoning (sepsis) (Wang et al., 2010), or ageing (Chen et al., 2010; Grillari and Grillari-Voglauer, 2010) is very likely to yield future diagnostic and/or prognostic biomarkers. The first ever miRNA-based diagnostic test was designed to differentiate between pancreatic cancer and pancreatitis, which often have similar symptoms, reporting accuracies of 95% (Wahid et al., 2010). Another miRNA-based test proposed for regulatory approval, is able to distinguish between two types of lung cancer. Studies in *C. elegans* have shown, however, how complicated the role of miRNAs can be: miRNAs both promote and antagonize the life span of this organism (De Lencastre et al., 2010). If suitable targets can be identified, miRNA-based therapy is expected to modulate/improve the response to chemotherapy (Rossbach, 2011).

Currently, miRNAs and/or miRNA targeting are in the process of being developed into therapy. Strategies considered are (Rossbach, 2011):

(**a**) **over-expression**: increased concentration of one or more targeted miRNAs by introduction of native miRNAs or miRNA-mimics that are improved in terms of target specificity, (bio)chemical longevity or otherwise. In such a strategy, reduced growth of cancer cells was obtained by the introduction of miRNA mimics (Bader et al., 2010).

(**b**) **knockdown:** reduced concentration of one or more miRNAs by introduction of antisense molecules or so-called antagomirs. Several types of synthetic analogs or chemical modifications for anti-miRNA antisense molecules are evaluated, again aiming at a high specificity and/or stability (Rossbach, 2011). So-called locked-nucleic-acid (LNA) miRNAs were among the first to be tested in clinical trials.

(**c**) **indirect modulation:** use of therapeutic drugs that modulate miRNA expression indirectly. Currently this option should be considered as a theoretical possibility.

Several pre-clinical and clinical trials based on miRNAs are currently in progress. The clinical trial website (www.clinicaltrials.gov) currently (Mar 2011) documents 61 of such studies related to miRNAs. The first clinical trials used a LNA-based antisense molecule (against miR-122) for the treatment of hepatitis C (Wahid et al., 2010). Despite such progress, there is a general lack of appropriate delivery systems with sufficient specificity and efficacy, be it viral, using liposomes or otherwise (Rossbach, 2011).

### 4.1.2 Chromatin-based modifications

Human disease, notably cancer, is characterized by genome-wide spread changes in their genomic DNA methylation and histone modification patterns. Epigenetic therapies aim to restore the normal epigenetic modification patterns largely through the inhibition of known epigenetic modifiers, such as DNA methyltransferases (DNMT), histone deacetylases (HDAC), histone methyltransferases, and de-methylases (Mund and Lyko, 2010). For each class of inhibitors, progress is reported and each may develop into actual drugs. The clinical trial website (see above) currently documents 161 of such trials, 136 of which deal with inhibition of epigenetic modifiers and nine with chromatin-based modifications. Cytosine analogs are strong inhibitors of DNMTs and are approved for the treatment of a pre-leukemic bone marrow disorder (Mund and Lyko, 2010). Several clinical trials are in progress (Yang et al., 2010), notably in combination therapy with HDAC inhibitors. Although the ability of such DNMT inhibitors to reduce global DNA methylation is clear, the association between the clinical response and the extent of demethylation is less obvious (Yang et al., 2010). Moreover, as the effect concerns global demethylation, targeted delivery and prevention of side-effects are important issues.

The histone deacetylase (HDAC) inhibitors suberoylanilide hydroxamic acid (SAHA), and a structurally related other inhibitor, are both approved for the treatment of T cell lymphoma. SAHA exerts pharmacological activity against different tumor cell lines in vitro and induces cell cycle arrest in breast cancer cell lines. Withdrawal causes the cells to start dividing again. HDAC inhibitors are frequently associated with transcriptional reactivation of tumor suppressor genes, such as p21. There are currently at least 20 structurally different HDAC inhibitors under investigation in clinical trials, either in monotherapy or in combination therapy with anticancer drugs (Mund and Lyko, 2010). In such combinations, synergistic effects are not uncommon and such effects may increase the clinical use of HDAC inhibitors.

HDAC inhibitors, like SAHA, are often highly specific, but its target enzymes may not be specific for histones and may, therefore, affect many proteins not involved in epigenetic regulation. Analysis of lysine acetylation sites identified no less than 3,600 acetylation sites on 1,750 proteins that in theory could be inhibited by HDAC (Choudhary et al., 2009). As HDAC inhibitors induce hyperacetylation of many non-histone protein substrates, the effects of HDAC inhibitors in cancer therapies may in fact not be mediated through epigenetic changes in histone acetylation at all (Mund and Lyko, 2010). In addition to histone acetylation, histone methylation notably of lysine residues is associated with epigenetic regulation. Chaetocin inhibits specific lysine histone methyltransferases, but its effects on tumor growth are unclear. Over-expression of a catalytically inactive version of a H3K9 methyltransferase showed growth inhibition in cancer cells. This may represent a future epigenetic drug-based strategy. Currently also the compound 3-deazaneoplanocin A (DZNep) is tested. This drug promotes the degradation of yet another histone methyltransferase. Histone methylation is dynamic and can be enzymatically reversed by histone demethylases, such as the lysine-specific demethylase 1 (LSD1). As such, also LSD1 may be a candidate target for therapeutic intervention for various types of cancer (Mund and Lyko, 2010).

### 4.1.3 Genome-wide or chromosome-based approaches

The rapid growth of DNA sequencing and associated epigenome-wide profiling technologies is currently generating new diagnostic and prognostic applications, but it is difficult to predict how soon these technologies will develop into therapeutic applications.

### 4.1.4 Other approaches

The variation in (trans)gene expression between individual transformants known as 'position effect' is supposed to be due to the influence of neighboring chromatin. The use of so-called chromatin boundaries, such as insulators, matrix-associated regions (known as MAR elements) and more, was shown to reduce the variation between individual transformants upon genetic transformation considerably (Bushey et al., 2008). The use of such elements may become attractive in mammalian systems (Girod et al., 2007; Harraghy et al., 2008). In mammalian systems, the use of DNA insulators that are thought to block chromatin-associated repressors is used to obtain notably higher production of therapeutic proteins in cell lines, notably in combination with MAR elements (Kwaks and Otte, 2006; Otte et al., 2007). Generally, however, large differences exist between the chromatin elements used and the results obtained. In most cases, the precise relationship with epigenetic phenomena has not been studied in any detail.

Another strategy of targeted epigenetic modification involves the use of artificial transcription factors to modify chromatin (Blancafort and Beltran, 2008). Artificial transcription factors made of zinc finger (ZF) domains have been developed to regulate specific genes and phenotypes both in cells and whole organisms. They may be based on epigenetic modifications that influence transcription factor-DNA interactions and can for example be used for targeted gene silencing by chimeric DNA methyltransferases (Blancafort and Beltran, 2008). Chimeric transcription factors fused with DNA methyltransferase sequences were shown to target DNA methylation to specific DNA sequences as method to repress the expression of target genes (Li et al., 2007). Possibly other enzymes of the epigenetic machinery can be targeted as well. Such epigenetic modifications with artificial transcription factors made of ZF domains seem feasible and may see more applications in the future. It could be advantageous to combine such applications with zinc finger nucleases that allow targeted gene editing (Doyon et al., 2011). The latter are already entering the clinic (Ledford, 2011).

### 4.1.5 Future targets for mitotic epigenetic application

In addition to the examples above, several other epigenetic phenomena are considered for application, but overall successes are sparse. In human medicine, there is great interest in stem cells for therapeutics, cloning and other applications. Mammalian stem cells have a large cellular plasticity that is progressively lost upon development. Stem cells are, therefore, proposed as reporter systems for undesired or unfavorable epigenetic modifications or

epimutations by 'epigenome disruptors' that may affect gene expression, alter cell fate and/or poise the cell for disease (Hemberger and Pedersen, 2010). This could be achieved by studying loci (such as imprinted genes) that undergo epigenetic alterations in normal development, and monitoring their response to potentially disruptive agents (Hemberger and Pedersen, 2010). In addition, it has been found that fully differentiated cell types can be reprogrammed into induced pluripotent stemm cells (iPSC) by temporary over-expression of only four transcription factors, thereby questioning the concept of epigenetic stability of the differentiated cell (Hanna et al., 2010).

Epigenetic profiling has shown that pluripotency is associated with hypomethylation of many gene promoters and with particular histone modifications (Hemberger and Pedersen, 2010). Epigenome profiling could therefore be used to screen and possibly select for such characteristics. Yet, stem cells, either natural or induced, are epigenetically heterogeneous and dynamic. Notably cell culture (as 'the environment') induces extensive alterations, related in part to the dynamic nature of DNA methylation and demethylation (Hemberger and Pedersen, 2010). Moreover, differentiation of cultured iPSCs is somewhat biased towards the cell type from which they were derived, suggesting the presence of some epigenetic memory (Kim et al., 2010). The potential future use of stem cells or iPSCs as reporters for epigenetic phenomena will therefore be challenging (Hemberger and Pedersen, 2010).

### 4.1.6 Future targets for meiotic epigenetic applications

In mammalian systems, there has been and there is quite some debate about the occurrence, hence use, of meiotic or transgenerational epigenetic inheritance. Although evidence is accumulating that it is likely to exist, actual applications seem far away. In the future, in the area of stem cell research for cloning and or in the area of food and diet, the first applications will be defined and investigated.

## 4.2 Applications of epigenetic inheritance in plant systems

In plant systems, different applications of epigenetic inheritance, of phenomena that are supposed to be related to epigenetic inheritance, are proposed in the literature.

### 4.2.1 Somaclonal variation

In plants, spontaneous modifications of visible plant features are well known in cultivation. For example in crops that are vegetatively propagated on an industrial scale, the appearance of aberrant phenotypes is relatively common (King et al., 2010). Such a phenomenon may have an epigenetic background that perhaps could be handled better with the help of epigenetic profiling and screening. Upon regeneration from a somatic cell, plants show variable phenotypes, a phenomenon generally known as somaclonal variation. In view of the role of epigenetics, however, the term 'somaclonal variation' is now used to indicate genetic changes (mutations) in the DNA sequence. The combination of genetic and/or epigenetic changes or phenomena may contribute to, or be at the heart of, the variability seen in tissue culture, such as rejuvenation, habituation and a variety of morphological deviations in different plant species (Smulders and de Klerk, 2011). In general, epigenetic profiling of off-types is only just beginning. In micro-propagated trees, variation in level of DNA methylation was associated with morphological variation (Valledor et al., 2007). Treatment of Brassica varieties with DNA methylation inhibitors generated a phenotypic variability similar to the variability observed as somaclonal variation in tissue culture (King, 1995). Explant type and regeneration pathway influenced the DNA methylation status in Arachis (Pacheco et al., 2008).

If such observations and correlations turn out to be wide-spread, epigenetic profiling may help to identify and/or select cells which suffer least from epigenetic variation, when uniformity and stability are desired, or show most variation, when the variation is thought advantageous for future breeding (King et al., 2010). In many species (or cultivars within a given species), reproducible regeneration is cumbersome (or impossible), suggesting that the epigenotype of the cells challenged to regenerate resist such regeneration. Epigenetic profiling and selection,

and/or deliberate changes of such an epigenotype, could help enlarge the pool of species that can be routinely regenerated from a given cell. As regeneration is a condition for genetic engineering, epigenetic screening and selection could help in developing genetic transformation protocols for recalcitrant plant species. Along the same line of reasoning, it can be hypothesized that reproductive barriers for fertilization are under epigenetic control and that epigenetics plays a role in the prevention of gene flow between species. If so, strategies targeting the epigenome may help to allow cross-fertilization of species that normally would not cross-hybridize (Comai, 2005; Ishikawa and Kinoshita, 2009; Kimatu and Bao, 2010).

### 4.2.2 Environmental tolerance

Any phenotype results from both its genotype and the environment. The latter can, or is thought to, affect the epigenome and the various epigenetic marks outlined above. In maize, for example, the planting distance was shown to influence DNA methylation and the effects were different between inbred strains and hybrids (King et al., 2010). In rice, spaceflight was found to induce heritable alterations in DNA methylation and gene expression (Ou et al., 2009; Ou et al., 2010). Various papers relate DNA methylation to various environmental stresses, such as water deficit, salt, metals or heat (Smulders and de Klerk, 2011). In Arabidopsis, stress-mediated release of gene silencing correlated with alterations in histone occupancy and histone acetylation (Lang-Mladek et al., 2010). Also tissue culture and callus formation may present typical cases of environmental stress (Desjardins, 2007) that could be alleviated by proper screening and selection after epigenetic profiling. Factors involved in chromatin remodeling were shown to be associated with environmental stress in Arabidopsis (Mlynárová et al., 2007). Breeding focusing on epigenetic phenomena could contribute to a larger environmental tolerance of crops in such a way that crops may adjust to adverse environmental conditions more easily. The markers and strategies for such breeding, either with or without genetic engineering will, however, need more research and understanding to allow its application.

### 4.2.3 Other approaches

In transgenic plants, there is considerable variation in (trans)gene expression between individual transformants. This so-called 'position effect' is supposed to be due to the influence of neighboring chromatin. The use of chromatin boundaries, such as insulators, matrix-associated regions (MAR elements) and more, was shown to reduce the variation between individual transformants upon genetic transformation considerably (Mlynárová et al., 1996; Bushey et al., 2008). However, the approach has not seen much application. Generally, large differences exist between the chromatin elements used and the results obtained. In most cases, the precise relationship with epigenetic phenomena has not been studied in any detail. The transgenic use of chromatin modifying elements implies obvious changes in the DNA sequence, violating the main characteristics of true epigenetic phenomena.

### 4.2.4 Transgenerational inheritance

In plants, the occurrence of transgenerational epigenetic inheritance is now well established (Johannes et al., 2009; Lang-Mladek et al., 2010), although mainly in relation to DNA methylation and methylation mutants. Other types of epigenetic changes may be more difficult to analyze in the context of transgenerational inheritance. The application of transgenerational epigenetic inheritance aims at introducing changes in the expression of endogenous genes that give a desired and stable phenotype in the organism of interest and its offspring, without changes in the primary DNA sequence. The challenge is to establish the phenotype in the offspring: for applications this should be sufficiently stable and reliable. Any application of mitotic epigenetic inheritance that happens to survive meiosis either intentional or stochastic becomes transgenerational. It could be argued that the applications towards cloning, regeneration or reduced position effects into whole plants should not only generate proper organisms (mitosis), but also generate sexually propagated offspring (meiosis). We will here give the notable applications in plants.

*- Removal of the epigenetic inducer*

Most applications published require or involve deliberate changes in the DNA sequence (hence, not involving epigenetics) at some point in time. However, when crossing out the inducing locus, or by using recombination systems such as Cre-*lox* for removal, it is possible to generate plant material with changes in the phenotype without changes in the primary DNA sequence (hence, epigenetics). Key issue in such applications of transgenerational epigenetic inheritance is the stability: is the epigenetic change and accompanying phenotype maintained over generations after the trigger is removed.

A major application of epigenetics entails gene silencing, either unintentionally (with complex integrations) or deliberately using inverted repeats (IRs), based on dsRNA and RNA-directed DNA methylation (RdDM) of similar sequences. The latter is related to DNA as well as to histone methylation. A wealth of publications is dealing with different approaches, genes, host plants and construct types, often yielding different results. IRs of target promoters directly introduced into an expression cassette generally mediate de novo DNA methylation by simultaneous activation of post-transcriptional gene silencing (PTGS) and RdDM. In tobacco, a transgene with an IR of a target promoter in an intron resulted in methylation, but did not induce PTGS. Fully homogeneous methylation patterns were established successfully in primary transformants (*cis*-RdDM), and in the first generations after crossing the trigger with a sensor construct (*trans*-RdDM). This result is in contrast to previous findings, in which more than three successive generations were required and methylation was not homogeneous (Dalakouras et al., 2009). Inverted repeat-induced promoter silencing is considered a technology for achieving potent and heritable gene silencing in plants. The DNA methylation triggered could be maintained by DNA methyltransferases to allow inheritance of gene silencing in subsequent generations in the absence of the inducing transgene. As discussed before (Nap and Geurts van Kessel, 2006), it could be feasible to induce such a change with an incoming virus (e.g. followed by meristem culture to get rid of the virus) or any other transient inducer. Although conceptually appealing, there are currently only a few examples of such approaches in the scientific literature, including a very recent one (Kanazawa et al., 2011). This may indicate that the transient inducer approach does not work very well, and/or is not stable enough. The recent success (Kanazawa et al., 2011) may, however, be the trigger for such technology to take off.

Although in this and several other cases silencing has been achieved (Eamens et al., 2008), only few cases have been reported with heritable silencing in the absence of the trigger. The latter is required for the occurrence of transgenerational epigenetic inheritance. Moreover, the heritable silencing concerns transgenes (Jones et al., 2001) and tends to fail in case of endogenous genes (Kanazawa et al., 2011). It may not be coincidental that the genes associated with such maintenance of DNA methylation in the absence of the trigger (Jones et al., 2001) are also the genes used in the epiRILs studies outlined below. In general however, upon removal of the trigger, the silenced phenotype gets lost, either immediately, or gradually.

The first clear example of the use of transgenerational epigenetic silencing in plants with respect to endogenous promoters is provided in a very recent report targeting promoters in petunia and tomato using a virus (CMV)-based vector (Kanazawa et al., 2011). Silencing was found to depend on the function of the virus-encoded 2b protein which is involved in the transport of siRNA to the nucleus. Plant-specific asymmetric DNA methylation was meiotically maintained in the absence of the RNA trigger for RdDM. Accumulation of siRNA in shoot meristem promoted by the viral system used may account for the heritable epigenetic changes. Different promoters in different plants gave similar, but not identical, results and it is not clear yet how histone modification and/or DNA methylation interconnect in these systems. These latter results indicate that acquired epigenetic changes upon targeting endogenous promoters can be heritable in plants without the presence of the inducer. This approach might result in modified plants that do not carry a transgene while still having the resultant desired phenotype (Kanazawa et al., 2011). Such strategies could offer approaches to avoid regulations for genetic modification (GM).

*- EpiRILS: the combination of genetics and epigenetics*

Much information about the potential use and stability of transgenerational epigenetic inheritance comes from the use of so-called epigenetic recombinant inbred lines, or epiRILS, in Arabidopsis. Two different research groups used different populations to generate mosaic epigenomes after inbreeding for multiple generations. To generate epiRILS, a plant line carrying a mutation in epigenetic regulation is crossed to its corresponding isogenic wild-type

line (Figure 10). Use was made of mutants in a DNA methyltransferase (*met*1), and a nucleosome-remodeling ATPase, required for both DNA and histone H3K9 methylation, which is characteristic of heterochromatin (*ddm*1). Upon self-fertilization, the heterozygote progeny carrying only the wild-type alleles is used as progenitors of multiple, independent recombinant inbred lines through repeated inbreeding and single-seed descent. As a result, the independent recombinant inbred lines carry different patchworks of epi-modifications inherited from the mutant parent. As the mutant parent is isogenic with the wild-type inbred parent, only epigenetic variation in methylation status is present, but no variation in the primary DNA sequence. In such populations, all sequence-based markers (SNP, AFLP etc.) will turn out to be obsolete. Mapping the epigenetic variation will require the novel methods of mapping epialleles (differential DNA methylation) as described above.

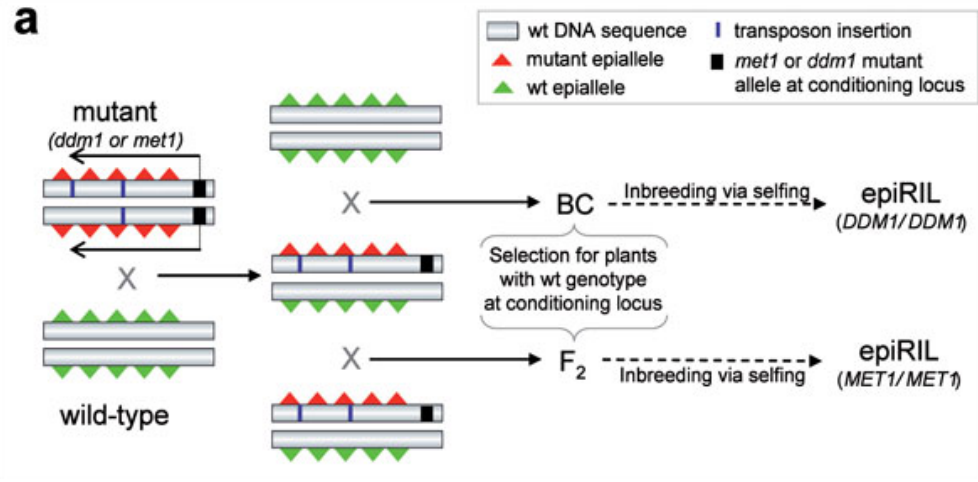



*Figure 10. Strategy to generate epiRILs in Arabidopsis (Johannes and Colomé-Tatché, 2011).*

This epiRIL material is currently being analyzed in considerable detail and the analyses are not finalized yet. For application purposes, the epiRILs should show stable inheritance of multiple parental DNA methylation variants (epi-alleles) over at least eight generations, as well as considerable variation and high heritability for flowering time and plant height (Johannes et al., 2009). It appears that such playing with the epigenome unlocks novel variation that is of interest for, and could be used in, plant breeding. However, the reduced methylation also incurred activation of transposon activity and resulted in additional genome instability. It remains to be seen how much of the phenotypic variability is due to transposon activity. Moreover, the existence of stable epi-alleles in populations complicates traditional association and linkage studies considerably. Theoretical suggestions on how to deal with such complications have been put forward (Johannes et al., 2008). Although this is an exciting area of research that may result in future applications, it is currently too early to say whether (commercial) applications will be feasible and, if so, on what time scale.

*- Induction of epigenetic variation*

Treatment of Brassica varieties with DNA methylation inhibitors have generated a phenotypic variability that proved heritable over various sexual generations (King, 1995), but such examples are rare.

*- Selection for epigenetic variation*

A curious application of epigenetic inheritance in plants has been the selection for energy use efficiency that determines seed yield in canola. Individual plants and their self-fertilized progenies from an isogenic canola population were recursively selected for respiration intensity and used to generate populations with distinct physiological and agronomical characteristics (Hauben et al., 2009). Resulting populations were found to be genetically identical, but epigenetically different. The global DNA methylation and histone modification patterns were altered and heritable, as were the physiological and agronomical characteristics (notably crop yield) of the selected lines (Hauben et al., 2009). The yield increases observed could be attributed to (mitochondrial) heterosis effects. These results suggest that selection allows fixation of desired phenotypes in populations with particular epigenomic states. This would

allow shaping the epigenome for a trait of interest. Such selection for epigenetic variation would demonstrate that epigenetic changes occur in populations without genetic modification. Yet, in view of the results of previous classical breeding efforts, the results presented are quite surprising. Therefore, it will be necessary to study and understand the underlying molecular mechanism(s) in greater detail. Notably, cause and effect relationships need to be unraveled. Also, it should be investigated how wide-spread these phenomena are. Possibly the results are specific for Brassica, which generally demonstrates exceptional morphological and genomic plasticity in terms of segmental chromosomal duplications and polyploidy (King et al., 2010).

## 4.3 Concluding remarks

Epigenetic drugs for human application are on the market and in development, but their specificity is still a major challenge. Epigenetic profiling for the diagnosis and monitoring of cancer and other disorders is likely to become standard practice soon, and this profiling may be of help to select for additional applications in the future as well. The application of induced pluripotent stem cells (iPSCs) for diagnostic and (somatic) therapeutic purposes clearly requires further investigation. Applications of epigenetic modification in plant systems develop slowly but steadily. Active use of epigenetic inheritance has, however, been rare, possibly because of the complex, redundant, dynamic and largely reversible nature of epigenetic gene regulation. Epigenetic information in plant breeding is likely to be worthwhile, if not essential, but will require significant improvements in analytical procedures and biological understanding. To account for cell-to-cell variation, development of single-cell analyses and understanding may be required for applications becoming sufficiently specific, targeted and stable.

# 5. Epigenetics in the biosafety assessment of genetically modified organisms

The previous chapters have shown that epigenetic mechanisms are fully intertwined with the regulation of gene expression. These mechanisms add layers of flexibility and control to gene expression. As a result of the increased understanding of that control of gene regulation, the issue now arises whether epigenetic regulation should be included as separate issue in the safety assessment of genetically modified organisms, and if so, how that should be organized. In the context of the epigenetic machinery outlined in the previous chapters, an issue to be addressed is the relationship between genetic modification and epigenetic effects. Epigenetic changes could be the result of the incoming transgene affecting the surrounding chromatin (*cis* epi-effects), or the expression of endogenous genes (*trans* epi-effects), whereas the effects should persist in the absence of the transgene. Also, endogenous epigenetic mechanisms could affect the expression of the transgene itself to result in undesired outcomes. In addition, applications should be considered in which epigenetic changes are the deliberate target of the genetic modification (epigenetic engineering). These categories will be outlined in much more detail below. Issues to be considered are, for example, can an incoming transgene be an epigenetic disruptor, either *cis* or *trans*, with unintended effects that are undesirable from a safety perspective? Can epigenetic mechanisms result in unpredicted or unintended effects with respect to the expression of an introduced transgene? If so, what are the consequences for biosafety assessments? Moreover, can the epigenome be target for deliberate chance that is not reflected in the DNA of the organism? Are current procedures sufficient or are new or adjusted procedures necessary? Can any of the new technologies of epigenomic screening help to alleviate or simplify existing procedures? How early and reliable can putative epigenetic effects, if any, be spotted?

First of all, in any discussion or safety assessment, it should be clear from the very beginning what exactly is meant with the term or concept 'epigenetics'. It may generally not be worth arguing over definitions, but scientists, regulators, managers, politicians, layman and other people involved in the discussions should at least know what each participant in discussions or evaluations means with this particular term. In preparing this report, through studying the literature and through personal communications, it has become clear that different people have different opinions about the scope and contents of epigenetics. In this report, we decided to adhere to the definition that was used in our previous report (Nap and Geurts van Kessel, 2006). Epigenetics is here defined as 'changes in gene function that are mitotically and/or meiotically heritable and do not entail a change in DNA sequence' (Section 1.1). It is in any case recommended that any future documents on epigenetics and biosafety assessments are very clear and explicit in their definitions. In addition, care must be taken not to generalize too easily: as in the biosafety assessment of genetic engineering, also epigenetic modification should be considered on an individual, case-by-case basis.

## 5.1 Biosafety assessment in mammalian (human) systems

The main area of application of epigenetics in mammalian systems targets gene regulation within a given tissue or organism that is not aimed to reach the germ-line. Such applied mitotic epigenetic inheritance does not involve genetic modification. It focuses on health and disease involving diagnostics, prognostics and advanced therapeutics. The specificity and side-effects of epigenetic drugs in human medicine are obviously topics of concern in pharmacotoxicology (Goodman et al., 2010), but as these drugs will not (or are not supposed to) enter the germ-line (no genetic modification intended or allowed), and may be the last resort for terminally ill patients, the type of considerations to be taken into account are incomparable to the issues with respect to GM plants and the environment. Absence of effects in the germ-line is, however, an issue that would seem to deserve more future consideration. In view of the widespread importance of epigenetic gene regulation, various areas beyond disease are considering or developing epigenetic applications, such as the fields of neurology and memory, as well as food and health. These applications will also not involve genetic modification.

With the advent of genetic engineering in notably stem cell applications, attention for biosafety in that area is on the rise. Especially the presence of transgene is considered to affect the epigenome (*cis* epi-effects, see above). In current mammalian stem cell modification, it has therefore been proposed (Papapetrou et al., 2011) to define regions in the genome that can serve as 'safe harbors' for integration of a single vector copy. In such regions, the integrated material is (expected to be) adequately expressed without perturbing endogenous gene structure or function. The five criteria that were proposed are: distance of at least 50 kb from the 5' end of any gene; distance of at least 300 kb from any cancer related gene; distance of at least 300 kb from any miRNA gene; location outside a transcription unit; location outside ultra-conserved regions. Yet, the experimental evidence for applying such criteria seems sparse and alternative criteria do exist (DeKelver et al., 2010).

In the remainder of the discussion, we will focus on the considerations for the safety assessment of epigenetic effects in plant systems.

## 5.2 Epigenetics and biosafety assessment of GM plants

The relationships between biosafety assessment, epigenetics and genetic modification can be considered on at least two different levels:
(**a**) epigenetic changes as result of the incoming transgene (Section 5.2.1)
(**b**) epigenetic changes as deliberate target of the genetic modification (Section 5.2.2).

These two distinct levels of consideration will be discussed below. It should be noted that in such discussions, similar to gene-based biosafety assessments, a case-by-case approach would be the preferred strategy.

In view of the current system of genetically modified plant biosafety analysis and of the body of knowledge reflected in current biosafety assessment, it would seem reasonable to classify potential epigenetic effects as result of the incoming transgene in the category 'unintended effects' of genetic modification, unless the epigenome is the explicit target of the genetic modification (see below). Therefore, in the context of this report, we feel that it is necessary to define more clearly the concept of 'unintended effect'. A thoughtful overview of the concept of unintended effects in genetically modified (GM) crops and their products was presented in the literature as early as 2004 (Cellini et al., 2004), with an updated and somewhat different perspective published in 2008 (Deng et al., 2008). Unintended effects comprise any biologically relevant difference relative to what was intended with, or expected from, the genetic modification, including effects that are not intended but predictable on theoretical grounds. It is, therefore, the opposite of the expected or intended effect and the latter may also be used to define unintended effects. The expected or intended effect can operationally be defined as: the novel gene is present in a single copy (or a known number of copies), with known genomic location, is only expressed at the right time and place, generates only the protein of interest, yields only the phenotype of interest and that phenotype has no known adverse effects on the environment or on other characteristics of the organism, notably when used in food or feed preparation (Cellini et al., 2004). Unintended effects may generate knowledge of the biological system and in hindsight they could be explained and/or considered predictable. Unintended effects do not need to be undesired, depending on the specifics of the outcome of the effect. It is feasible, although rare, that the effect, although unintended and unexpected, should be considered an added bonus.

In the context of biosafety, unintended effects that are undesired are the issue for assessment. The attention for -and concerns about- unintended effects has resulted in an avalanche of publications developing and discussing complicated concepts as 'substantial equivalence', the 'precautionary principle', as well as the potential value of profiling with 'omics' technologies to detect unintended effects either *a priori* or through surveillance. Technologies that have been considered (and/or advocated) include metabolomics, proteomics and transcriptomics (Kok et al., 2008). For example, one could consider to measure expression and/or activity of all genes/enzymes responsible for known epigenetic mechanisms (DNMTs. HDACs, HATs, HMTs etc.). Transgene-associated differences in such epigenetic factors may be detected, but for proper biosafety assessments these require the demonstration of a relationship (either direct causal or even indirect) with the transgene (Batista and Oliveira, 2010). Defining levels for concern in relationship to adverse outcomes will require a definition of the normal epigenome under a variety of

conditions and a definition of what changes from normal are considered adverse in which context. Any difference that could imply an unintended effect should therefore be considered for its statistical significance. To this end, it will be necessary to understand what is normal and what constitutes an epigenetic effect associated with an adverse outcome. At the moment, it is not clear what the normal spectrum is of epigenetic variability in cells/tissues/ organisms, nor how much of the epigenetic variability is environmentally induced, or how much of it is transferred to offspring or, in ecological considerations, to wild relatives. As a result, a baseline for normality cannot yet be defined. The challenges in such technologies and proper interpretation of the results obtained (Batista et al., 2008) are therefore considerable (Davies, 2010). In view of current knowledge and all the methodological and statistical issues that are considered in the framework of the application of 'omics' technologies in biosafety assessment (Davies, 2010), it is currently unlikely that such profiling will help to predict or explain putative epigenetic effects of an incoming transgene, although current epigenetic profiling technologies can serve as excellent tools for scientific investigations and will be of help to explore epigenetics for future application purposes (see Chapter 4).

### 5.2.1 Epigenetic changes as result of an incoming transgene.

The incoming transgene could give rise to undesired or unintended effects on gene expression. In more detail, three possibilities can be distinguished with respect to potential epigenetic changes as result of the incoming transgene:
**(a)** effect of the transgene on surrounding chromatin (*cis* epi-effects)
**(b)** effect of the transgene on the expression of endogenous genes (*trans* epi-effects)
**(c)** effect of epigenetic mechanisms on the expression of the transgene

These separate possibilities will be discussed below.

*- Effect of the transgene on surrounding chromatin (*cis *epi-effects)*
The presence of the incoming transgene could change the local chromatin characteristics and that change could have a notable influence on the expression of neighboring genes and/or the phenotype of the resulting organism. The position effect applications as described above (Mlynárová et al., 1996) show the influence of the DNA neighborhood, such as chromatin boundaries, on the expression of a transgene; likewise the presence of transgene could affect its surroundings and influence the expression of nearby genes. Such unintended effects are part of the current assessment protocols. In order for such effects to be considered epigenetic, they should persist after the transgene has left. There are no examples in the literature where it is shown that such effects, if any, occur and persist upon removal of the transgene.

*- Effect of the expression of the transgene on the expression of endogenous genes (*trans *epi-effects)*
The incoming transgene or its expression product could act as an 'epigenetic disruptor' and affect the expression of any endogenous gene (irrespective of any sequence similarity to the transgene) by any of the epigenetic mechanisms outlined above. The issue for biosafety in connection with epigenetics is whether such effects persist in the absence of the inducing transgene (transient GM). As long as the observed changes depend on the presence (or expression) of the transgene as inducer and revert to the original situation in the absence of the transgene or its expression product, they can and should not be considered epigenetic in terms of biosafety assessments. It is expected, that the larger majority of such unintended changes, if any, will depend on the physical presence or expression of the transgene. In such cases, he unintended changes, if any, are covered by the current biosafety assessment and monitoring protocols. Only in case the unintended changes in endogenous gene expression persist (over cell divisions and/or sexual reproduction) after the transgene has left the genome (for example, by outcrossing), the changes should be considered epigenetic. In Section 5.2.2., the situation is considered that such an epigenetic change is deliberate, here the situation is considered that the change is unintended. This situation is obviously the most challenging issue for regulation, although it seems likely that if any undesirable phenotype would occur, it would be detected sufficiently early in the research phase of a GM project or be detected in the general surveillance protocols.

The expression of the incoming transgene could for example interfere with the (small or large) RNA complement in the cell. For example, the transgene-encoded RNA may exhibit similarity to an existing and co-occurring miRNA and by RNA/miRNA interaction titrate away the level of this unintended target to such an extent that it affects the normal function of the target mRNA. Thus it may, indirectly, affect endogenous gene expression and regulation. Such an unintended change in miRNA concentration could have effects that go beyond the effects of the transgene. When the affected situation and/or de result of that situation persists in tissues or cells when the transgene is no longer expressed (but still present), or no longer present, it may reflect a true case for unintended epigenetic effects. Such situations could also relate to moving miRNAs, for example in crops that are grafted on rootstocks (such as could be the case in wine or some ornamentals) on genetically modified root stocks. Such effects have not been described in the literature. Effects, when occurring and adverse, are likely to be detected sufficiently early in the research and development phase, including field trials, of any anticipated commercial release.

Convincing data have yet to be put forward in the scientific literature showing the prolonged influence of a transgene (or its expression) as an unintended epigenetic disruptor on one or more endogenous genes. Targeted analyses of defined genetic modification events employing epigenome-wide expression profiling in combination with whole genome tiling arrays (and/or analysis in combination with the genome sequence) to analyze the occurrence (and consequences) of such effects have so far not been presented in the literature and are therefore beyond the scope of the present report. As outlined above, such epigenome profiling is difficult due to the highly redundant and reversible nature of the epigenetic mechanisms involved, as well as all methodological and statistical issues that need to be considered (Davies, 2010). We are not aware of examples in the literature that allow the conclusion of the actual occurrence of such effects, apart from the example of the targeted application of RdDM discussed above (Kanazawa et al., 2011).

*- Effect of epigenetic mechanisms on the expression of the transgene*

Endogenous epigenetic mechanisms could affect the intended expression of the transgene. Examples of this possibility include inadvertent gene silencing of the transgene introduced, for example due to complex integrations (RNA-based silencing), low expression due to unfavorable integration sites (chromatin-based silencing) or other effects on the expression of the transgene (Mlynárová et al., 1996; Eamens et al., 2008; Matzke et al., 2009). These effects are well known in the literature and are reviewed extensively. In fact, the unexpected silencing of a gene involved in the color of a Petunia flower in a field trial (Meyer et al., 1992) has initiated in part the framework of biosafety regulation we now have. There are anecdotal reports of gradual loss of gene activity of transgenes over time, over generations or in different environments. The deliberate use of RNAi as gene silencing method includes examples of inadvertent loss of gene silencing (hence reactivation of the silenced gene) over time, over generations or in different environments. In such cases, the unintended effect on (trans)gene expression is generally interpreted as the result from epigenetic mechanisms. The inadvertent loss of the intended phenotype associated with the incoming transgene is a typical example of the 'producer's risk' and in order to prevent loss of markets, it is in the direct interest of producers that plants have the intended phenotype. Generally, such material will not be developed into a commercial product for release. In case the plant material is already on the market, the mandatory post-market monitoring and general stewardship of the product by the producer will be sufficient to detect (and withdraw) such events.

Overall, it can be concluded that from the point of view of biosafety analysis, epigenetic mechanisms and effects are already taken into account in the current biosafety assessment and monitoring procedures of GM plants. Post-market monitoring of any commercially released GM plants (general surveillance) aiming at identifying unanticipated adverse effects on human health or the environment which could arise directly or indirectly from the GM plants is mandatory in the EU (Bartsch et al., 2007). In the context of current biosafety assessment of GM plants, notably unintended effects (also known as pleiotropic effects), are subject of analysis and monitoring in the research and development phase before commercial release, including field trials, as well as after release. This would cover any epigenetic change translating into an unintended and undesired phenotype.

### 5.2.2 Epigenetic changes as deliberate target of genetic modification

Earlier, this option was proposed as 'epigenetic engineering' and possible applications were given (Nap and Geurts van Kessel, 2006). Despite its theoretical possibilities and obvious attractiveness, however, as yet only one clear example using RNA-directed DNA methylation or RdDM, discussed above, has now been put forward in the literature (Kanazawa et al., 2011). Such strategies are presented as means to avoid GM regulations. The fact that so little examples are put forward in the literature may indicate that the approach does not work as good or stable as is necessary for (commercial) application. It may also indicate that there is currently no confidence in industry that the approach will indeed help to evade GM regulations. Whether the recent report (Kanazawa et al., 2011) will spark new applications remains to be seen. It should be noted that the RdDM technology and its ramifications is currently subject of assessment in an EU Working Group 'New Breeding Technologies' for future consideration in EU procedures. Therefore, its putative consequences for regulation are considered beyond the scope of the present report. Although the use of chromatin boundaries, targeted RNAi and other examples mentioned as applications in Chapter 4, do target directly or indirectly the epigenetic machinery, their approach is essentially genetic in nature and will require current regulation. The generation of additional variation in plants via epiRILs is particularly interesting, although the possible genomic instability as result of induced transposon activation may preclude (commercial) applications. It will be interesting to see whether similar variation can be introduced with components of the epigenetic machinery other than DNA methylation. In the future, epiRILS may be accomplished in crops of interest without genetic modification, although it will require mutagenesis to generate -or screening to identify- the desired DNA methylation mutant necessary for the approach. The regulatory status of such mutation strategies is also considered in the EU Working Group evaluating new breeding technologies.

## 5.3 Concluding remarks

Epigenetics is an important aspect of gene regulation. Yet, on the basis of current knowledge and literature, the likelihood that an incoming transgene generates an epigenetic modification that is maintained in a cell/organism without that transgene and gives rise to unwanted characteristics would seem sufficiently small against the background of epigenetic change and variability generated by development, environmental stress and/or other sources of epigenetic variation. An important issue for the regulation and biosafety assessment of putative epigenetic modifications is whether such modifications could have putatively adverse consequences that require additional tests or monitoring. As of yet, putative consequences as outlined in this report will be noted through the current biosafety assessment and monitoring procedures. If the incoming transgene will result in an undesired phenotype due to epigenetic mechanisms, the existing frameworks are in place to detect such an adverse phenotype. It is considered sufficiently unlikely that such putative epigenetic changes will not be noticed in the procedures now in place.

Epigenetic effects, if directly targeted, should translate in notable effects on gene expression that have undesired phenotypic consequences and the same would seem to be true for unintended epigenetic effects. Current biosafety assessments are appropriate and sufficient to detect such effects. The continued development of epigenetic technology, notably at the whole genome level, will help to understand and in the future possibly modify the epigenomic status of genomes to obtain phenotypes of interest. The field is evolving at a very rapid pace and a great deal still needs to be learned prior to be able to consider any added value of comprehensive epigenetic evaluations into future biosafety assessments.

# Acknowledgements

The authors of this report wish to thank the COGEM supervisory committee, consisting of Prof.dr. J.J.M. Dons (chairman, Wageningen; COGEM), Dr. F. van Leeuwen (NKI, Amsterdam), Dr. M. Stam (SILS, Amsterdam), Dr. D.C.M. Glandorf (RIVM/Bureau GGO) and B. Erkamp, MSc. (COGEM) for guidance, constructive criticism on earlier versions of this report and support, Prof. dr. H. de Jong, Dr. C. van de Wiel, Dr. E. Schijlen (Wageningen) as well as Dr. D. de Bruijn (Nijmegen) for information, help, as well as constructive comments on earlier versions of this report.

# Literature

This report is for the larger part based on (numerous) reviews from the scientific literature. These reviews allow delving deeper into the subject and retrieve the primary research literature.


Allis, C.D. (2007).
Epigenetics. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Anon. (2010).
Making a mark (editorial). Nature Biotechnology 28, 1031.

Arteaga-Vazquez, M.A. and Chandler, V.L. (2010).
Paramutation in maize: RNA-mediated trans-generational gene silencing.
Current Opinion in Genetics and Development 20, 156-163.

Asslaber, D., Piñón Hofbauer, J., Greil, R. and Egle, A. (2010).
MicroRNAs as biomarkers for the diagnosis and prognosis of human cancer.
Journal of Nucleic Acids Investigation 1, e14.

Bader, A.G., Brown, D. and Winkler, M. (2010).
The promise of microRNA replacement therapy. Cancer Research 70, 7027-7030.

Bartel, D.P. (2009).
MicroRNAs: target recognition and regulatory functions. Cell 136, 215-233.

Bartsch, D., Gathmann, A., Hartley, S., Hendriksen, N.B., Hails, R., Lheureux, K., Kiss, J., Mesdagh, S., Neemann, G., Perry, J., Renckens, S., Schiemann, J. and Sweet, J. (2007).
First EFSA experiences with monitoring plans.
Journal fur Verbraucherschutz und Lebensmittelsicherheit 2, 33-36.

Batista, R. and Oliveira, M. (2010).
Plant natural variability may affect safety assessment data.
Regulatory Toxicology and Pharmacology 58, S8-S12.

Batista, R., Saibo, N., Lourenço, T. and Oliveira, M.M. (2008).
Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. Proceedings of the National Academy of Sciences of the United States of America 105, 3640-3645.

Beck, S. (2010).
Taking the measure of the methylome. Nature Biotechnology 28, 1026-1028.

Berger, S.L., Kouzarides, T., Shiekhattar, R. and Shilatifard, A. (2009).
An operational definition of epigenetics. Genes and Development 23, 781-783.

Bird, A. (2007).
Perceptions of epigenetics. Nature 447, 396-398.

Blancafort, P. and Beltran, A.S. (2008).
Rational design, selection and specificity of Artificial Transcription Factors (ATFs): the influence of chromatin in target gene regulation. Combinatorial Chemistry and High Throughput Screening 11, 146-158.

Bock, C., Tomazou, E.M., Brinkman, A.B., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H.G. and Meissner, A. (2010).
Quantitative comparison of genome-wide DNA methylation mapping technologies.
Nature Biotechnology 28, 1106-1114.

Bonasio, R., Tu, S. and Reinberg, D. (2010).
Molecular signals of epigenetic states. Science 330, 612-616.

Bourc'his, D. and Voinnet, O. (2010).
A small-RNA perspective on gametogenesis, fertilization, and early zygotic development.
Science 330, 617-622.

Brady, S.M. and Provart, N.J. (2009).
Web-queryable large-scale data sets for hypothesis generation in plant biology. Plant Cell 21, 1034-1051.

Bushey, A.M., Dorman, E.R. and Corces, V.G. (2008).
Chromatin Insulators: regulatory mechanisms and epigenetic inheritance. Molecular Cell 32, 1-9.
Butcher, L.M. and Beck, S. (2010).
AutoMeDIP-seq: a high-throughput, whole genome, DNA methylation assay. Methods 52, 223-231.
Cairns, B.R. (2009).
The logic of chromatin architecture and remodelling at promoters. Nature 461, 193-198.
Caley, D.P., Pink, R.C., Trujillano, D. and Carter, D.R.F. (2010).
Long noncoding RNAs, chromatin, and development. The Scientific World Journal 10, 90-102.
Cantu, D., Vanzetti, L.S., Sumner, A., Dubcovsky, M., Matvienko, M., Distelfeld, A., Michelmore, R.W. and Dubcovsky, J. (2010).
Small RNAs, DNA methylation and transposable elements in wheat. BMC Genomics 11, e408.
Cassiday, L. (2009).
Mapping the epigenome. Chemical and Engineering News 87, 11-16.
Cellini, F., Chesson, A., Colquhoun, I., Constable, A., Davies, H.V., Engel, K.H., Gatehouse, A.M.R., Karenlampi, S., Kok, E.J., Leguay, J.J., Lehesranta, S., Noteborn, H., Pedersen, J. and Smith, M. (2004).
Unintended effects and their detection in genetically modified crops.
Food and Chemical Toxicology 42, 1089-1125.
Chaya, D. and Zaret, K.S. (2004).
Sequential chromatin Immunoprecipitation from animal tissues. Methods in Enzymology 376, 361-372.
Chen, L.H., Chiou, G.Y., Chen, Y.W., Li, H.Y. and Chiou, S.H. (2010).
MicroRNA and aging: a novel modulator in regulating the aging network.
Ageing Research Reviews 9, SUPPL. S59-S66.
Chi, P., Allis, C.D. and Wang, G.G. (2010).
Covalent histone modifications-miswritten, misinterpreted and mis-erased in human cancers.
Nature Reviews Cancer 10, 457-469.
Cho, W.C. (2010).
MicroRNAs as therapeutic targets for lung cancer. Expert Opinion on Therapeutic Targets 14, 1005-1008.
Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., Casero, D., Bernal, M., Huijser, P., Clark, A.T., Krämer, U., Merchant, S.S., Zhang, X., Jacobsen, S.E. and Pellegrini, M. (2010).
Relationship between nucleosome positioning and DNA methylation. Nature 466, 388-392.
Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V. and Mann, M. (2009).
Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science 325, 834-840.
Clapier, C.R. and Cairns, B.R. (2009).
The biology of chromatin remodeling complexes. Annual Review of Biochemistry 78, 273-304.
Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S. and Bayley, H. (2009).
Continuous base identification for single-molecule nanopore DNA sequencing.
Nature Nanotechnology 4, 265-270.
Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M. and Jacobsen, S.E. (2008).
Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning.
Nature 452, 215-219.
Collas, P. (2010a).
The current state of chromatin immunoprecipitation. Molecular Biotechnology 45, 87-100.
Collas, P. (2010b).
Programming differentiation potential in mesenchymal stem cells. Epigenetics 5, 476-482.
Comai, L. (2005).
The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6, 836-846.

Cremer, T. and Cremer, M. (2010).
Chromosome territories. Cold Spring Harbor Perspectives in Biology 2, a003889.
Czech, B. and Hannon, G.J. (2011).
Small RNA sorting: matchmaking for argonautes. Nature Reviews Genetics 12, 19-31.
Dalakouras, A., Moser, M., Zwiebel, M., Krczal, G., Hell, R. and Wassenegger, M. (2009).
A hairpin RNA construct residing in an intron efficiently triggered RNA-directed DNA methylation in tobacco. Plant Journal 60, 840-851.
Davies, H. (2010).
A role for 'omics' technologies in food safety assessment. Food Control 21, 1601-1610.
Daxinger, L. and Whitelaw, E. (2010).
Transgenerational epigenetic inheritance: more questions than answers. Genome Research 20, 1623-1628.
Daxinger, L., Kanno, T., Bucher, E., Van Der Winden, J., Naumann, U., Matzke, A.J.M. and Matzke, M. (2009).
A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. EMBO Journal 28, 48-57.
De Lencastre, A., Pincus, Z., Zhou, K., Kato, M., Lee, S.S. and Slack, F.J. (2010).
MicroRNAs both promote and antagonize longevity in *C. elegans*. Current Biology 20, 2159-2168.
De Lucia, F. and Dean, C. (2011).
Long non-coding RNAs and chromatin regulation. Current Opinion in Plant Biology 14, 168-173.
De Nooijer, S., Wellink, J., Mulder, B. and Bisseling, T. (2009).
Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. Nucleic Acids Research 37, 3558-3568.
Deal, R.B. and Henikoff, S. (2011).
Histone variants and modifications in plant gene regulation. Current Opinion in Plant Biology 14, 116-122.
DeKelver, R.C., Choi, V.M., Moehle, E.A., Paschon, D.E., Hockemeyer, D., Meijsing, S.H., Sancak, Y., Cui, X., Steine, E.J., Miller, J.C., Tam, P., Bartsevich, V.V., Meng, X., Rupniewski, I., Gopalan, S.M., Sun, H.C., Pitz, K.J., Rock, J.M., Zhang, L., Davis, G.D., Rebar, E.J., Cheeseman, I.M., Yamamoto, K.R., Sabatini, D.M., Jaenisch, R., Gregory, P.D. and Urnov, F.D. (2010).
Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Research 20, 1133-1142.
Delgado, M., Frescatada, M., Ribeiro-Varandas, E., Viegas, W. and Jones, R.N. (2010).
Dynamics of functional heterochromatic domains in the plant interphase nucleus. Cytogenetic and Genome Research 129, 64-71.
Deng, P.J., Zhou, X.Y., Yang, D.Y., Hou, H.L., Yang, X.K., Li, Y.H., Yang, Y.C., Wang, X.L., Fang, S.S., Wu, S.Q. and Liu, J. (2008).
The definition source. manifestation of and assessment of unintended effects in genetically modified plants. Journal of the Science of Food and Agriculture 88, 2401-2413.
Deng, W. and Blobel, G.A. (2010).
Do chromatin loops provide epigenetic gene expression states?
Current Opinion in Genetics and Development 20, 548-554.
Desjardins, Y. (2007).
How 'Micropropagation-Omics' can contribute to a better understanding of phenomena taking place in plant tissue culture. Acta Horticulturae 748, 39-53.
Djuranovic, S., Nahvi, A. and Green, R. (2011).
A parsimonious model for gene regulation by miRNAs. Science 331, 550-553.
Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia, D.F., Miller, J.C., Urnov, F.D., Gregory, P.D. and Holmes, M.C. (2011).
Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures.
Nature Methods 8, 74-79.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K.P., Baccash, J., Borcherding, A.P., Brownley, A., Cedeno, R., Chen, L., Chernikoff, D., Cheung, A., Chirita, R., Curson, B., Ebert, J.C., Hacker, C.R., Hartlage, R., Huser, B., Huang, S., Jiang, Y., Karpinchyk, V., Koenig, M., Kong, C., Landers, T., Le, C., Liu, J., McBride, C.E., Morenzoni, M., Morey, R.E., Mutch, K., Perazich, H., Perry, K., Peters, B.A., Peterson, J., Pethiyagoda, C.L., Pothuraju, K., Richter, C., Rosenbaum, A.M., Roy, S., Shafto, J., Sharanhovich, U., Shannon, K.W., Sheppy, C.G., Sun, M., Thakuria, J.V., Tran, A., Vu, D., Zaranek, A.W., Wu, X., Drmanac, S., Oliphant, A.R., Banyai, W.C., Martin, B., Ballinger, D.G., Church, G.M. and Reid, C.A. (2010).
Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327, 78-81.

Dunoyer, P., Schott, G., Himber, C., Meyer, D., Takeda, A., Carrington, J.C. and Voinnet, O. (2010).
Small RNA duplexes function as mobile silencing signals between plant cells. Science 328, 912-916.

Eamens, A., Wang, M.B., Smith, N.A. and Waterhouse, P.M. (2008).
RNA silencing in plants: yesterday, today, and tomorrow. Plant Physiology 147, 456-468.

Edwards, J.R., O'Donnell, A.H., Rollins, R.A., Peckham, H.E., Lee, C., Milekic, M.H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., Gingrich, J.A., Haghighi, F., Nutter, R. and Bestor, T.H. (2010).
Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Research 20, 972-980.

Egelhofer, T.A., Minoda, A., Klugman, S., Lee, K., Kolasinska-Zwierz, P., Alekseyenko, A.A., Cheung, M.S., Day, D.S., Gadel, S., Gorchakov, A.A., Gu, T., Kharchenko, P.V., Kuan, S., Latorre, I., Linder-Basso, D., Luu, Y., Ngo, Q., Perry, M., Rechtsteiner, A., Riddle, N.C., Schwartz, Y.B., Shanower, G.A., Vielle, A., Ahringer, J., Elgin, S.C.R., Kuroda, M.I., Pirrotta, V., Ren, B., Strome, S., Park, P.J., Karpen, G.H., Hawkins, R.D. and Lieb, J.D. (2011).
An assessment of histone-modification antibody quality. Nature Structural and Molecular Biology 18, 91-94.

Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H. and Nadeau, J.H. (2010).
Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics 11, 446-450.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. and Turner, S. (2009).
Real-time DNA sequencing from single polymerase molecules. Science 323, 133-138.

Erhard Jr, K.F., Stonaker, J.L., Parkinson, S.E., Lim, J.P., Hale, C.J. and Hollick, J.B. (2009).
RNA polymerase IV functions in paramutation in *Zea mays*. Science 323, 1201-1205.

Ernst, J. and Kellis, M. (2010).
Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature Biotechnology 28, 817-825.

Fabbri, M. (2010).
MiRNAs as molecular biomarkers of cancer. Expert Review of Molecular Diagnostics 10, 435-444.

Fabian, M.R., Sonenberg, N. and Filipowicz, W. (2010).
Regulation of mRNA translation and stability by microRNAs. Annual Review of Biochemistry 79, 351-379.

Falk, R. (2010).
What is a gene?-Revisited. Studies in History and Philosophy of Science Part C 41, 396-406.

Feinberg, A.P. (2010).
Epigenomics reveals a functional genome anatomy and a new approach to common disease. Nature Biotechnology 28, 1049-1052.

Felsenfeld, G. and Groudine, M. (2003).
Controlling the double helix. Nature 421, 448-453.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., Ukomadu, C., Sadler, K.C., Pradhan, S., Pellegrini, M. and Jacobsen, S.E. (2010).
Conservation and divergence of methylation patterning in plants and animals.
Proceedings of the National Academy of Sciences of the United States of America 107, 8689-8694.

Ferracin, M., Veronese, A. and Negrini, M. (2010).
Micromarkers: MiRNAs in cancer diagnosis and prognosis.
Expert Review of Molecular Diagnostics 10, 297-308.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J. and van Steensel, B. (2010).
Systematic protein location mapping reveals five principal chromatin types in Drosophila cells.
Cell 143, 212-224.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J. and Turner, S.W. (2010).
Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature Methods 7, 461-465.

Fuchs, S.M., Krajewski, K., Baker, R.W., Miller, V.L. and Strahl, B.D. (2011).
Influence of combinatorial histone modifications on antibody and effector protein recognition.
Current Biology 21, 53-58.

Furner, I.J. and Matzke, M. (2011).
Methylation and demethylation of the Arabidopsis genome. Current Opinion in Plant Biology 14, 137-141.

Gao, Z., Liu, H.L., Daxinger, L., Pontes, O., He, X., Qian, W., Lin, H., Xie, M., Lorkovic, Z.J., Zhang, S., Miki, D., Zhan, X., Pontier, D., Lagrange, T., Jin, H., Matzke, A.J.M., Matzke, M., Pikaard, C.S. and Zhu, J.K. (2010).
An RNA polymerase II-and AGO4-associated protein acts in RNA-directed DNA methylation.
Nature 465, 106-109.

Girod, P.A., Nguyen, D.Q., Calabrese, D., Puttini, S., Grandjean, M., Martinet, D., Regamey, A., Saugy, D., Beckmann, J.S., Bucher, P. and Mermod, N. (2007).
Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells.
Nature Methods 4, 747-753.

Glaser, V. (2010).
Analyzing the genome one cell at a time. Genetic Engineering and Biotechnology News 30, i21.

Goodman, J.I., Augustine, K.A., Cunnningham, M.L., Dixon, D., Dragan, Y.P., Falls, J.G., Rasoulpour, R.J., Sills, R.C., Storer, R.D., Wolf, D.C. and Pettit, S.D. (2010).
What do we need to know prior to thinking about incorporating an epigenetic evaluation into safety assessments? Toxicological Sciences 116, 375-381.

Goren, A., Ozsolak, F., Shoresh, N., Ku, M., Adii, M., Hart, C., Gymrek, M., Zuk, O., Regev, A., Milos, P.M. and Bernstein, B.E. (2010).
Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA.
Nature Methods 7, 47-49.

Grillari, J. and Grillari-Voglauer, R. (2010).
Novel modulators of senescence, aging, and longevity: small non-coding RNAs enter the stage.
Experimental Gerontology 45, 302-311.

Gupta, R., Nagarajan, A. and Wajapeyee, N. (2010).
Advances in genome-wide DNA methylation analysis. BioTechniques 49, iii-xi.

Halfmann, R. and Lindquist, S. (2010).
Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits.
Science 330, 629-632.

Hanna, J.H., Saha, K. and Jaenisch, R. (2010).
Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell 143, 508-525.

Harraghy, N., Gaussin, A. and Mermod, N. (2008).
Sustained transgene expression using MAR elements. Current Gene Therapy 8, 353-366.

Hauben, M., Haesendonckx, B., Standaert, E., Van Der Kelen, K., Azmi, A., Akpo, H., Van Breusegem, F., Guisez, Y., Bots, M., Lambert, B., Laga, B. and De Block, M. (2009).
Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield.
Proceedings of the National Academy of Sciences of the United States of America 106, 20109-20114.

Hemberger, M. and Pedersen, R. (2010).
Epigenome disruptors. Science 330, 598-599.

Heneghan, H.M., Miller, N. and Kerin, M.J. (2010).
MiRNAs as biomarkers and therapeutic targets in cancer. Current Opinion in Pharmacology 10, 543-550.

Ho, L. and Crabtree, G.R. (2010).
Chromatin remodelling during development. Nature 463, 474-484.

Huang, S., He, J., Chang, S.A., Zhang, P.M., Liang, F., Li, S.Q., Tuchband, M., Fuhrmann, A., Ros, R. and Lindsay, S. (2010a).
Identifying single bases in a DNA oligomer with electron tunnelling. Nature Nanotechnology 5, 868-873.

Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R. and Rao, A. (2010b).
The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS ONE 5, e8888.

Huss, M. (2010).
Introduction into the analysis of high-throughput-sequencing based epigenome data.
Briefings in Bioinformatics 11, 512-523.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J.B., Sabunciyan, S. and Feinberg, A.P. (2009).
The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics 41, 178-186.

Ishikawa, R. and Kinoshita, T. (2009).
Epigenetic programming: the challenge to species hybridization. Molecular Plant 2, 589-599.

Jamaluddin, M.S., Weakley, S.M., Zhang, L., Kougias, P., Lin, P.H., Yao, Q. and Chen, C. (2011).
MiRNAs: roles and clinical applications in vascular disease. Expert Review of Molecular Diagnostics 11, 79-89.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Felsenfeld, G. (2009).
H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nature Genetics 41, 941-945.

Jin, S.G., Kadam, S. and Pfeifer, G.P. (2010).
Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Research 38, e125 (gkq223).

Johannes, F. and Colomé-Tatché, M. (2011).
Concerning epigenetics and inbreeding. Nature Reviews Genetics 12, 376-377.

Johannes, F., Colot, V. and Jansen, R.C. (2008). Epigenome dynamics: a quantitative genetics perspective.
Nature Reviews Genetics 9, 883-890.

Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuisson, J., Heredia, F., Audigier, P., Bouchez, D., Dillmann, C., Guerche, P., Hospital, F. and Colot, V. (2009).
Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genetics 5, e1000530.

Jones, L., Ratcliff, F. and Baulcombe, D.C. (2001).
RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. Current Biology 11, 747-757.

Kanazawa, A., Inaba, J.I., Shimura, H., Otagaki, S., Tsukahara, S., Matsuzawa, A., Kim, B.M., Goto, K. and Masuta, C. (2011).
Virus-mediated efficient induction of epigenetic modifications of endogenous genes with phenotypic changes in plants. Plant Journal 65, 156-168.

Karlic, R., Chung, H.R., Lasserre, J., Vlahoricek, K. and Vingron, M. (2010).
Histone modification levels are predictive for gene expression.
Proceedings of the National Academy of Sciences of the United States of America 107, 2926-2931.

Karow, J. (2010).
Ion Torrent reveals initial specs for PGM; early-access user provides first feedback. www.genomeweb.com/sequencing - In sequence.

Kaufmann, K., Muiño, J.M., Østerås, M., Farinelli, L., Krajewski, P. and Angenent, G.C. (2010).
Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nature Protocols 5, 457-472.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., Yabuuchi, A., Takeuchi, A., Cunniff, K.C., Hongguang, H., McKinney-Freeman, S., Naveiras, O., Yoon, T.J., Irizarry, R.A., Jung, N., Seita, J., Hanna, J., Murakami, P., Jaenisch, R., Weissleder, R., Orkin, S.H., Weissman, I.L., Feinberg, A.P. and Daley, G.Q. (2010).
Epigenetic memory in induced pluripotent stem cells. Nature 467, 285-290.

Kimatu, J.N. and Bao, L. (2010).
Epigenetic polymorphisms could contribute to the genomic conflicts and gene flow barriers resulting to plant hybrid necrosis. African Journal of Biotechnology 9, 8125-8133.

King, G.J. (1995).
Morphological development in *Brassica oleracea* is modulated by *in vivo* treatment with 5-azacytidine. Journal of Horticultural Science 70, 333-342.

King, G.J., Amoah, S. and Kurup, S. (2010).
Exploring and exploiting epigenetic variation in crops. Genome 53, 856-868.

Kok, E.J., Lehesranta, S.J., Van Dijk, J.P., Helsdingen, J.R., Dijksma, W.T.P., Van Hoef, A.M.A., Koistinen, K.M., Kärenlampi, S.O., Kuiper, H.A. and Keijer, J. (2008).
Changes in gene and protein expression during tomato ripening - consequences for the safety assessment of new crop plant varieties. Food Science and Technology International 14, 503-518.

Kriaucionis, S. and Heintz, N. (2009).
The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. Science 324, 929-930.

Krol, J., Loedige, I. and Filipowicz, W. (2010).
The widespread regulation of microRNA biogenesis, function and decay. Nature Reviews Genetics 11, 597-610.

Kwaks, T.H.J. and Otte, A.P. (2006).
Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. Trends in Biotechnology 24, 137-142.

Laird, P.W. (2010).
Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics 11, 191-203.

Lal, A., Pan, Y., Navarro, F., Dykxhoorn, D.M., Moreau, L., Meire, E., Bentwich, Z., Lieberman, J. and Chowdhury, D. (2009).
MiR-24-mediated downregulation of H2AX suppresses DNA repair in terminally differentiated blood cells. Nature Structural and Molecular Biology 16, 492-498.

Lang-Mladek, C., Popova, O., Kiok, K., Berlinger, M., Rakic, B., Aufsatz, W., Jonak, C., Hauser, M.T. and Luschnig, C. (2010).
Transgenerational Inheritance and resetting of stress-induced loss of epigenetic gene silencing in Arabidopsis. Molecular Plant 3, 594-602.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Sung, K.W.K., Rigoutsos, I., Loring J. and Wei, C.L. (2010).
Dynamic changes in the human methylome during differentiation. Genome Research 20, 320-331.

Ledford, H. (2011).
Targeted gene editing enters clinic. Nature 471, 16.

Li, F., Papworth, M., Minczuk, M., Rohde, C., Zhang, Y., Ragozin, S. and Jeltsch, A. (2007).
Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. Nucleic Acids Research 35, 100-112.

Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Mohamed, Y.B., Ooi, H.S., Tennakoon, C., Wei, C.L., Ruan, Y. and Sung, W.K. (2010).
ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing.
Genome Biology 11, R22.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S. and Dekker, J. (2009).
Comprehensive mapping of long-range interactions reveals folding principles of the human genome.
Science 326, 289-293.

Linsen, S.E.V., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R.K., Fritz, B., Wyman, S.K., de Bruijn, E., Voest, E.E., Kuersten, S., Tewari, M. and Cuppen, E. (2009).
Limitations and possibilities of small RNA digital gene expression profiling. Nature Methods 6, 474-476.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008).
Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133, 523-536.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B. and Ecker, J.R. (2009).
Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315-322.

Louwers, M., Splinter, E., van Driel, R., de Laat, W. and Stam, M. (2009).
Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C).
Nature Protocols 4, 1216-1229.

Martienssen, R. (2010).
Small RNA makes its move. Science 328, 834-835.

Massie, C.E. and Mills, I.G. (2008).
ChIPping away at gene regulation. EMBO Reports 9, 337-343.

Mathieu, O., Reinders, J., Čaikovski, M., Smathajitt, C. and Paszkowski, J. (2007).
Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell 130, 851-862.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B. and Matzke, A.J. (2009).
RNA-mediated chromatin-based silencing in plants. Current Opinion in Cell Biology 21, 367-376.

Metzker, M.L. (2010).
Sequencing technologies - the next generation. Nature Reviews Genetics 11, 31-46.

Meyer, P., Linn, F., Heidmann, I., Meyer z, A.H., Niedenhof, I. and Saedler, H. (1992).
Endogenous and environmental factors influence 35S promoter methylation of a maize A1 gene construct in transgenic petunia and its colour phenotype. Molecular and General Genetics 231, 345-352.

Mikkelsen, T.S., Xu, Z., Zhang, X., Wang, L., Gimble, J.M., Lander, E.S. and Rosen, E.D. (2010).
Comparative epigenomic analysis of murine and human adipogenesis. Cell 143, 156-169.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S. and Bernstein, B.E. (2007).
Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553-560.

Mirsaidov, U.M., Wang, D.Q., Timp, W. and Timp, G. (2010).
Molecular diagnostics for personal medicine using a nanopore.
Wiley Interdisciplinary Reviews-Nanomedicine and Nanobiotechnology 2, 367-381.

Mlynárová, L., Nap, J.P. and Bisseling, T. (2007).
The SWI/SNF chromatin-remodeling gene AtCHR12 mediates temporary growth arrest in *Arabidopsis thaliana* upon perceiving environmental stress. Plant Journal 51, 874-885.

Mlynárová, Ľ., Keizer, L.C.P., Stiekema, W.J. and Nap, J.P. (1996).
Approaching the lower limits of transgene variability. Plant Cell 8, 1589-1599.

Molnar, A., Melnyk, C.W., Bassett, A., Hardcastle, T.J., Dunn, R. and Baulcombe, D.C. (2010).
Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells.
Science 328, 872-875.

Mouton-Liger, F., Paquet, C. and Hugon, J. (2010).
Biogenesis and regulation of microRNA: implication in Alzheimers disease. Future Neurology 5, 839-850.
Mund, C. and Lyko, F. (2010).
Epigenetic cancer therapy: proof of concept and remaining challenges. BioEssays 32, 949-957.
Nap, J.P. and Geurts van Kessel, A. (2006).
Epigenetics in context: literatuurrapport in opdracht van de COGEM. (Wageningen: Plant Research International).
Nestor, C., Ruzov, A., Meehan, R.R. and Dunican, D.S. (2010).
Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxy-methylcytosine in DNA. BioTechniques 48, 317-319.
Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vazquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., Slotkin, R.K., Martienssen, R.A. and Vielle-Calzada, J.P. (2010).
Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature 464, 628-632.
Orian, A., Abed, M., Kenyagin-Karsenti, D. and Boico, O. (2009).
DamID: a methylation-based chromatin profiling approach. Methods in Molecular Biology 567, 155-169.
Otte, A.P., Kwaks, T.H.J., Van Blokland, R.J.M., Sewalt, R.G.A.B., Verhees, J., Klaren, V.N.A., Siersma, T.K., Korse, H.W.M., Teunissen, N.C., Botschuijver, S., Van Mer, C. and Man, S.Y. (2007).
Various expression-augmenting DNA elements benefit from STAR-select, a novel high stringency selection system for protein expression. Biotechnology Progress 23, 801-807.
Ou, X., Long, L., Zhang, Y., Xue, Y., Liu, J., Lin, X. and Liu, B. (2009).
Spaceflight induces both transient and heritable alterations in DNA methylation and gene expression in rice (*Oryza sativa* L.). Mutation Research 662, 44-53.
Ou, X., Long, L., Wu, Y., Yu, Y., Lin, X., Qi, X. and Liu, B. (2010).
Spaceflight-induced genetic and epigenetic changes in the rice (*Oryza sativa* L.) genome are independent of each other. Genome Canada 53, 524-532.
Ozsolak, F. and Milos, P.M. (2011).
RNA sequencing: advances, challenges and opportunities. Nature Reviews Genetics 12, 87-98.
Pacheco, G., Cardoso, S.R.S., Gagliardi, R.F., Valls, J.F.M., Ferreira, P.C.G., Cardoso, M.A. and Mansur, E. (2008).
Genetic and epigenetic analyses of in vitro-grown plants of *Arachis villosulicarpa* Hoehne (*Leguminosae*) obtained from seed explants through different regeneration pathways.
Journal of Horticultural Science and Biotechnology 83, 737-742.
Papapetrou, E.P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L.M.S., Kadota, K., Roth, S.L., Giardina, P., Viale, A., Leslie, C., Bushman, F.D., Studer, L. and Sadelain, M. (2011).
Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. Nature Biotechnology 29, 73-81.
Pawlowski, W.P. (2010).
Chromosome organization and dynamics in plants. Current Opinion in Plant Biology 13, 640-645.
Pearson, H. (2006).
What is a gene? Nature 441, 398-401.
Penterman, J., Zilberman, D., Jin, H.H., Ballinger, T., Henikoff, S. and Fischer, R.L. (2007).
DNA demethylation in the Arabidopsis genome.
Proceedings of the National Academy of Sciences of the United States of America 104, 6752-6757.
Portela, A. and Esteller, M. (2010).
Epigenetic modifications and human disease. Nature Biotechnology 28, 1057-1068.
Ptashne, M. (2007).
On the use of the word 'epigenetic'. Current Biology 17, R233-R236.
Ptashne, M., Hobert, O. and Davidson, E. (2010).
Questions over the scientific basis of epigenome project. Nature 464, 487.
Raber, M.N. (2010).
Emerging techniques for determining the tissue of origin in patients with unknown primary carcinoma.
Expert Opinion on Medical Diagnostics 4, 455-457.
Riddihough, G. and Zahn, L.M. (2010).
What is epigenetics? Science 330, 611.

Ringrose, L. (2010).
How do RNA sequence, DNA sequence, and chromatin properties regulate splicing?
F1000 Biology Reports 2, no 74.
Rocha, M.S., Castro, R., Rivera, I., Kok, R.M., Smulders, Y.M., Jakobs, C., De Almeida, I.T. and Blom, H.J. (2010).
Global DNA methylation: comparison of enzymatic- and non-enzymatic-based methods.
Clinical Chemistry and Laboratory Medicine 48, 1793-1798.
Rossbach, M. (2011).
Therapeutic implications of microRNAs in human cancer. Journal of Nucleic Acids Investigation 1, e3.
Roudier, F., Teixeira, F.K. and Colot, V. (2009).
Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. Trends in Genetics 25, 511-517.
Russo, V.E.A., Martienssen, R.A. and Riggs, A.D. (1996).
Epigenetic Mechanisms of Gene Regulation - Cold Spring Harbor Monograph Series 32. (New York: Cold Spring Harbor Laboratory Press).
Santos-Rosa, H., Kirmizis, A., Nelson, C., Bartke, T., Saksouk, N., Cote, J. and Kouzarides, T. (2009).
Histone H3 tail clipping regulates gene expression. Nature Structural and Molecular Biology 16, 17-22.
Santuari, L. and Hardtke, C.S. (2010).
The case for resequencing studies of *Arabidopsis thaliana* accessions: mining the dark matter of natural genetic variation. F1000 Biology Reports 2, no 85.
Schaefer, A., Stephan, C., Busch, J., Yousef, G.M. and Jung, K. (2010).
Diagnostic, prognostic and therapeutic implications of microRNAs in urologic tumors.
Nature Reviews Urology 7, 286-297.
Schones, D.E. and Zhao, K. (2008).
Genome-wide approaches to studying chromatin modifications. Nature Reviews Genetics 9, 179-191.
Simon, S.A. and Meyers, B.C. (2010).
Small RNA-mediated epigenetic modifications in plants. Current Opinion in Plant Biology 14, 1-8.
Sinclair, P., Bian, Q., Plutz, M., Heard, E. and Belmont, A.S. (2010).
Dynamic plasticity of large-scale chromatin structure revealed by self-assembly of engineered chromosome regions. Journal of Cell Biology 190, 761-776.
Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A. and Martienssen, R.A. (2009).
Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136, 461-472.
Smulders, M.J.M. and de Klerk, G.J. (2011).
Epigenetics in plant tissue culture. Plant Growth Regulation 63, 137-146.
Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T.J., Zhang, B., Godley, L.A., Hicks, L.M., Lahn, B.T., Jin, P. and He, C. (2011).
Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.
Nature Biotechnology 29, 68-75.
Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z. and Cedar, H. (2009).
Developmental programming of CpG island methylation profiles in the human genome.
Nature Structural and Molecular Biology 16, 564-571.
Suzuki, H.I. and Miyazono, K. (2011).
Emerging complexity of microRNA generation cascades. Journal of Biochemistry 149, 15-25.
Suzuki, M.M. and Bird, A. (2008).
DNA methylation landscapes: provocative insights from epigenomics. Nature Reviews Genetics 9, 465-476.
Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. and Rao, A. (2009).
Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930-935.
Teixeira, F.K. and Colot, V. (2009).
Gene body DNA methylation in plants: A means to an end or an end to a means. EMBO Journal 28, 997-998.
Thomas, M., Lieberman, J. and Lal, A. (2010).
Desperately seeking microRNA targets. Nature Structural and Molecular Biology 17, 1169-1174.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., Turner, D.J., Illingworth, R. and Bird, A. (2010).
CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464, 1082-1086.

Tost, J. (2008).
Methods for the genome-wide and gene-specific analysis of DNA methylation levels and patterns.
In: Epigenetics, J. Tost, ed. (Academic Press), pp. 63-103.

Valledor, L., Hasbún, R., Meijón, M., Rodríguez, J.L., Santamaría, E., Viejo, M., Berdasco, M., Feito, I., Fraga, M.F., Cañal, M.J. and Rodríguez, R. (2007).
Involvement of DNA methylation in tree development and micropropagation.
Plant Cell, Tissue and Organ Culture 91, 75-86.

Van Steensel, B. (2005).
Mapping of genetic and epigenetic regulatory networks using microarrays.
Nature Genetics 37, SUPPL S18-S24.

Van Steensel, B. and Dekker, J. (2010).
Genomics tools for unraveling chromosome architecture. Nature Biotechnology 28, 1089-1095.

Van Steensel, B., Delrow, J. and Henikoff, S. (2001).
Chromatin profiling using targeted DNA adenine methyltransferase. Nature Genetics 27, 304-308.

Vaughn, M.W., Tanurdzic, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P.D., Dedhia, N., McCombie, W.R., Agier, N., Bulski, A., Colot, V., Doerge, R.W. and Martienssen, R.A. (2007).
Epigenetic natural variation in *Arabidopsis thaliana*. PLoS Biology 5, 1617-1629.

Venkatachalam, R., Ligtenberg, M.J., Hoogerbrugge, N., de Bruijn, D.R., Kuiper, R.P. and Geurts van Kessel, A. (2010).
The epigenetics of (hereditary) colorectal cancer. Cancer Genetics and Cytogenetics 203, 1-6.

Vrbsky, J., Akimcheva, S., Watson, J.M., Turner, T.L., Daxinger, L., Vyskot, B., Aufsatz, W. and Riha, K. (2010).
siRNA-mediated methylation of Arabidopsis telomeres. PLoS Genetics 6, 1-12.

Wahid, F., Shehzad, A., Khan, T. and Kim, Y.Y. (2010).
MicroRNAs: synthesis, mechanism, function, and recent clinical trials.
Biochimica et Biophysica Acta - Molecular Cell Research 1803, 1231-1243.

Wako, T. and Fukui, K. (2010).
Higher organization and histone modification of the plant nucleus and chromosome.
Cytogenetic and Genome Research 129, 55-63.

Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., Gaudet, F., Li, E. and Chen, T. (2009a).
The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation.
Nature Genetics 41, 125-129.

Wang, J.F., Yu, M.L., Yu, G., Bian, J.J., Deng, X.M., Wan, X.J. and Zhu, K.M. (2010).
Serum miR-146a and miR-223 as potential new biomarkers for sepsis.
Biochemical and Biophysical Research Communications 394, 184-188.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W. and Zhao, K. (2009b).
Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes.
Cell 138, 1019-1031.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q. and Zhao, K. (2008).
Combinatorial patterns of histone acetylations and methylations in the human genome.
Nature Genetics 40, 897-903.

White, N.M.A., Fatoohi, E., Metias, M., Jung, K., Stephan, C. and Yousef, G.M. (2011).
Metastamirs: a stepping stone towards improved cancer management.
Nature Reviews Clinical Oncology 8, 75-84.

Wu, A.R., Hiatt, J.B., Lu, R., Attema, J.L., Lobo, N.A., Weissman, I.L., Clarke, M.F. and Quake, S.R. (2009).
Automated microfluidic chromatin immunoprecipitation from 2,000 cells.
Lab on a Chip - Miniaturisation for Chemistry and Biology 9, 1365-1370.

Wu, S.C. and Zhang, Y. (2010).
Active DNA demethylation: many roads lead to Rome. Nature Reviews Molecular Cell Biology 11, 607-620.

Yang, X., Lay, F., Han, H. and Jones, P.A. (2010).
Targeting DNA methylation for epigenetic therapy. Trends in Pharmacological Sciences 31, 536-546.
Yoo, A.S., Staahl, B.T., Chen, L. and Crabtree, G.R. (2009).
MicroRNA-mediated switching of chromatin-remodelling complexes in neural development.
Nature 460, 642-646.
Youngson, N.A. and Whitelaw, E. (2008).
Transgenerational epigenetic effects. Annual Review of Genomics and Human Genetics 9, 233-257.
Zaratiegui, M., Irvine, D.V. and Martienssen, R.A. (2007).
Noncoding RNAs and gene silencing. Cell 128, 763-776.
Zemach, A., McDaniel, I.E., Silva, P. and Zilberman, D. (2010).
Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328, 916-919.
Zhang, X. (2008).
The epigenetic landscape of plants. Science 320, 489-492.
Zhao, Q., Rank, G., Tan, Y.T., Li, H., Moritz, R.L., Simpson, R.J., Cerruti, L., Curtis, D.J., Patel, D.J., Allis, C.D., Cunningham, J.M. and Jane, S.M. (2009).
PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. Nature Structural and Molecular Biology 16, 304-311.
Zhu, R., Howorka, S., Pröll, J., Kienberger, F., Preiner, J., Hesse, J., Ebner, A., Pastushenko, V.P., Gruber, H.J. and Hinterdorfer, P. (2010).
Nanomechanical recognition measurements of individual DNA molecules reveal epigenetic methylation patterns. Nature Nanotechnology 5, 788-791.
Zilberman, D., Coleman-Derr, D., Ballinger, T. and Henikoff, S. (2008).
Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456, 125-129.
Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. and Henikoff, S. (2007).
Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics 39, 61-69.