Epigenetic regulation of gene expression

Beekman, R., Di Croce, L., Gopinathrao, G., Martin-Subero, JI., Matthews, L., May, B., Mukherji, M., Percipalle, P., Pfeifer, GP.

European Bioinformatics Institute, New York University Langone Medical Center, Ontario Institute for Cancer Research, Oregon Health and Science University.

The contents of this document may be freely copied and distributed in any media, provided the authors, plus the institutions, are credited, as stated under the terms of Creative Commons Attribution 4.0 International (CC BY 4.0) License. For more information see our license.

17/09/2020
Introduction

Reactome is a freely accessible, open-source, manually curated and peer-reviewed pathway database. Annotations of pathways are authored by expert biologists, in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. A system of evidence tracking ensures that all assertions are backed up by the primary literature. Reactome is used by clinicians, geneticists, genomics researchers, and molecular biologists to interpret the results of high-throughput experimental studies, by bioinformaticians seeking to develop novel algorithms for mining knowledge from genomic studies, and by systems biologists building predictive models of normal and disease variant pathways.

The development of Reactome is supported by grants from the US National Institutes of Health (P41 HG003751), University of Toronto (CFREF Medicine by Design), European Union (EU STRP, EMI-CD), and the European Molecular Biology Laboratory (EBI Industry program).

Literature references


Reactome database release: 73

This document contains 6 pathways (see Table of Contents)
Epigenetic regulation of gene expression

Stable identifier: R-HSA-212165

Compartments: nucleoplasm

Epigenetic processes regulate gene expression by modulating the frequency, rate, or extent of gene expression in a mitotically or meiotically heritable way that does not entail a change in the DNA sequence. Originally the definition applied only to heritability across generations but later also encompassed the heritable changes that occur during cellular differentiation within one organism.

Molecular analysis shows epigenetic changes comprise covalent modifications, such as methylation and acetylation, to DNA and histones. RNA interference has been implicated in the initiation of some epigenetic changes, for example transcriptional silencing of transposons. Proteins which bind to the modified DNA and histones are then responsible for repressing transcription and for maintaining the epigenetic modifications during cell division.

During differentiation, patterns of gene expression are established by polycomb complexes PRC1 and PRC2. PRC2 methylates histones and DNA to produce the initial marks of repression: trimethylated lysine-27 on histone H3 (H3K27me3) and 5-methylcytosine in DNA. PRC2, through its component EZH2 or, in some complexes, EZH1 trimethylates lysine-27 of histone H3. The H3K27me3 produced by PRC2 is bound by the Polycomb subunit of PRC1. PRC1 ubiquitinates histone H2A and maintains repression.

PRC2 and other epigenetic systems modulate gene expression through DNA methylation, the transfer of a methyl group from S-adenosylmethionine to the 5 position of cytosine in DNA by a family of DNA methyltransferases (DNMTs): DNMT1, DNMT3A, and DNMT3B.

In the reverse process TET1,2,3 and TDG demethylate DNA through the oxidation of the methyl group of 5-methylcytosine by TET enzymes and the excision of the oxidized product (5-formylcytosine or 5-carboxylcytosine) by TDG.

Ribosomal RNA (rRNA) genes are activated and deactivated according to the metabolic requirements of the cell. Positive epigenetic regulation of rRNA expression occurs through chromatin modifications produced by activators such as ERCC6 (CSB), the B-WICH complex, and histone acetylases such as KAT2B (PCAF). Negative epigenetic regulation of rRNA expression occurs through chromatin modifications pro-
duced by repressors such as the eNoSC complex, SIRT1, and the NoRC complex.

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Authorship</th>
<th>Editors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-02-09</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., May, B.</td>
</tr>
<tr>
<td>2009-07-06</td>
<td>Edited</td>
<td>May, B.</td>
</tr>
<tr>
<td>2014-02-26</td>
<td>Reviewed</td>
<td>Matthews, L.</td>
</tr>
</tbody>
</table>
Polycomb group proteins are responsible for the heritable repression of genes during development (Lee et al. 2006, Ku et al. 2008, reviewed in Simon and Kingston 2009, Margueron and Reinberg 2011, Di Croce and Helin 2013). Two major families of Polycomb complexes exist: Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2). PRC1 and PRC2 each appear to comprise sets of distinct complexes that contain common core subunits and distinct accessory subunits (reviewed in Nayak et al. 2011). PRC2, through its component EZH2 or, in some complexes, EZH1 produces the initial molecular mark of repression, the trimethylation of lysine-27 of histone H3 (H3K27me3). How PRC2 is initially recruited to a locus remains unknown, however cytosine-guanine (CpG) motifs and transcripts have been suggested. Different mechanisms may be used at different loci. The trimethylated H3K27 produced by PRC2 is bound by the Polycomb subunit of PRC1. PRC1 ubiquitinates histone H2A and maintains repression.

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-02-09</td>
<td>Authored</td>
<td>Gopinathrao, G., May, B.</td>
</tr>
<tr>
<td>2010-04-06</td>
<td>Edited</td>
<td>May, B.</td>
</tr>
<tr>
<td>2014-02-12</td>
<td>Reviewed</td>
<td>Di Croce, L.</td>
</tr>
</tbody>
</table>
TET1,2,3 and TDG demethylate DNA

**Location:** Epigenetic regulation of gene expression

**Stable identifier:** R-HSA-5221030

**Compartments:** nucleoplasm

About 2-6% of all cytosine residues and 70-80% of cytosine residues in CG dinucleotides in mammalian cells are methylated at the 5 position of the pyrimidine ring. The cytosine residues are methylated by DNA methyltransferases after DNA replication and can be demethylated by passive dilution during subsequent replication or by active modification of the 5-methylcytosine base. Cytosine demethylation is developmentally regulated: one wave of demethylation occurs in primordial germ cells and one wave occurs by active demethylation in the male pronucleus after fertilization.

Some mechanisms of active demethylation remain controversial, however progressive oxidation of the methyl group of 5-methylcytosine followed by base excision by thymine DNA glycosylase (TDG) has been reproducibly demonstrated in vivo (reviewed in Wu and Zhang 2011, Franchini et al 2012, Cadet and Wagner 2013, Kohli and Zhang 2013, Ponnaluri et al. 2013, Rasmussen and Helin 2016). Ten-eleven translocation proteins TET1, TET2, and TET3 are dioxygenases that first oxidize 5-methylcytosine to 5-hydroxymethylcytosine (5-hmC) (Tahiliani et al. 2009, Ito et al. 2010), which is found in significant quantities and specific genomic locations in stem cells and neurons (Kinney and Pradhan 2013). TET proteins can further oxidize 5-hmC to 5-formylcytosine (5-fC) and then 5-carboxylcytosine (5-caC) (He et al. 2011, Ito et al. 2011). G:5-fC and G:5-caC base pairs are recognized by TDG, which excises the 5-fC or 5-caC and leaves an abasic site.

TET1 in mouse is expressed in neurons and its expression depends on neuronal activity (Guo et al. 2011, Kaas et al. 2013, Zhang et al. 2013). TET1 is also found in embryonic stem cells (Ficz et al. 2011, Koh et al. 2011, Wu et al. 2011) and in primordial germ cells of mice, where it plays a role in erasure of imprinting (Yamaguchi et al. 2013). TET3 is expressed in oocytes and zygotes of mice and is required for demethyla-
tion in the male pronucleus (Gu et al. 2011, Iqbal et al. 2011). TET2 is the most highly expressed TET family protein in hemopoietic stem cells and appears to act as a tumor suppressor. TET2 is also expressed in embryonic stem cells (Koh et al. 2011).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-12-29</td>
<td>Authored, Edited</td>
<td>May, B.</td>
</tr>
<tr>
<td>2014-01-29</td>
<td>Reviewed</td>
<td>Pfeifer, GP.</td>
</tr>
<tr>
<td>2014-02-21</td>
<td>Reviewed</td>
<td>Mukherji, M.</td>
</tr>
</tbody>
</table>
Positive epigenetic regulation of rRNA expression

Location: Epigenetic regulation of gene expression

Stable identifier: R-HSA-5250913

Compartments: nucleoplasm

Transcription of rRNA genes is controlled by epigenetic activation and repression according to the metabolic requirements of the cell (reviewed in Percipalle and Farrants 2006, McStay and Grummt 2008, Goodfellow and Zomerdijk 2012, Grummt and Langst 2013). Depending on the growth state of the cell, about half of the approximately 400 rRNA genes are expressed and these have the modifications characteristic of active chromatin: unmethylated DNA and acetylated histones. Repressed genes generally have methylated DNA and histone H3 methylated at lysine-9. Regulators of activation include ERCC6 (CSB), histone acetylases such as KAT2B (PCAF), and the B-WICH complex. Dysregulation of RNA polymerase I transcription plays a role in disease (reviewed in Hannan et al. 2013).

The B-WICH complex positively regulates rRNA expression by remodeling chromatin and recruiting histone acetyltransferases that modify histones to transcriptionally active states

ERCC6 (CSB) and EHMT2(G9a) positively regulate rRNA expression by ERCC6 recruiting the histone methyltransferase EHMT2 (also known as G9a) which dimethylates histone H3 at lysine-9 within the transcribed regions of rRNA genes.

ERCC6 (CSB) and KAT2B (PCAF) positively regulate rRNA expression by ERCC6 recruiting the histone acetyltransferase KAT2B to the promoter where KAT2B acetylates histone H4 at several lysine residues and histone H3 at lysine-9. The acetylated chromatin facilitates the assembly of RNA polymerase I initiation complex.

Literature references


### Editions

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-01-29</td>
<td>Authored, Edited</td>
<td>May, B.</td>
</tr>
<tr>
<td>2015-11-07</td>
<td>Reviewed</td>
<td>Percipalle, P.</td>
</tr>
</tbody>
</table>
Transcription of rRNA genes is controlled by epigenetic activation and repression (reviewed in McStay and Grummt 2008, Goodfellow and Zomerdijk 2012, Grummt and Langst 2013). About half of the roughly 400 rRNA genes are expressed and these have the modifications of active chromatin: unmethylated DNA and acetylated histones. Repressed genes generally have methylated DNA and histone H3 methylated at lysine-9. Regulators of repression include the eNoSC complex, SIRT1, and the NoRC complex.

SIRT1 negatively regulates rRNA expression as a subunit of the eNoSC complex, which deacetylates histone H3 and dimethylates lysine-9 of histone H3 (H3K9me2).

NoRC negatively regulates rRNA expression by shifting a nucleosome near the start of rRNA transcription into a more repressive location and recruiting Histone Deacetylase 1 and 2 (HDAC1, HDAC2) and DNA Methyltransferase 1 and 3b (DNMT1, DNMT3b).

**Literature references**


<table>
<thead>
<tr>
<th>Editions</th>
<th>Date</th>
<th>Role</th>
<th>Author/Reviewer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2014-01-29</td>
<td>Authored, Edited</td>
<td>May, B.</td>
</tr>
<tr>
<td></td>
<td>2014-01-31</td>
<td>Reviewed</td>
<td>May, B.</td>
</tr>
</tbody>
</table>
DNA methylation

**Location:** Epigenetic regulation of gene expression

**Stable identifier:** R-HSA-5334118

**Compartments:** nucleoplasm

Methylation of cytosine is catalyzed by a family of DNA methyltransferases (DNMTs): DNMT1, DNMT3A, and DNMT3B transfer methyl groups from S-adenosylmethionine to cytosine, producing 5-methylcytosine and homocysteine (reviewed in Klose and Bird 2006, Ooi et al. 2009, Jurkowska et al. 2011, Moore et al. 2013). (DNMT2 appears to methylate RNA rather than DNA.) DNMT1, the first enzyme discovered, preferentially methylates hemimethylated CG motifs that are produced by replication (template strand methylated, synthesized strand unmethylated). Thus it maintains existing methylation through cell division. DNMT3A and DNMT3B catalyze de novo methylation at unmethylated sites that include both CG dinucleotides and non-CG motifs.

DNA from adult humans contains about 0.76 to 1.00 mole percent 5-methylcytosine (Ehrlich et al. 1982, reviewed in Klose and Bird 2006, Ooi et al. 2009, Moore et al. 2013). Methylation of DNA occurs at cytosines that are mainly located in CG dinucleotides. CG dinucleotides are unevenly distributed in the genome. Promoter regions tend to have a high CG-content, forming so-called CG-islands (CGIs), while the CG-content in the remaining part of the genome is much lower. CGs tend to be unmethylated, while the majority of CGs outside CGIs are methylated. Methylation in promoters and first exons tends to repress transcription while methylation in gene bodies (regions of genes downstream of the promoter and first exon) correlates with transcription (reviewed in Ehrlich and Lacey 2013, Kulis et al. 2013). Proteins such as MeCP2 and MBDs specifically bind 5-methylcytosine and may recruit other factors.

Mammalian development has two major episodes of genome-wide demethylation and remethylation (reviewed in Zhou 2012, Guibert and Weber 2013, Hackett and Surani 2013, Dean 2014). In mice about 1 day after fertilization the paternal genome is actively demethylated by TET proteins together with thymine DNA glycosylase and the maternal genome is demethylated by passive dilution during replication, however methylation at imprinted sites is maintained. The genome has its lowest methylation level about 3.5 days post-fertilization. Remethylation occurs by 6.5 days post-fertilization. The second demethyla-
tion-remethylation event occurs in primordial germ cells of the developing embryo about 12.5 days post-fertilization. DNMT3A and DNMT3B, together with the non-catalytic DNMT3L, play major roles in the re-methylation events (reviewed in Chen and Chan 2014). How the methyltransferases are directed to particular regions of the genome remains an area of active research. The mechanisms at each locus may differ in detail but a connection between histone modifications and DNA methylation has been observed (reviewed in Rose and Klose 2014).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Author/Editorship</th>
<th>Verifier/Editor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-02-21</td>
<td>Authored, Edited</td>
<td>May, B.</td>
</tr>
<tr>
<td>2014-07-24</td>
<td>Reviewed</td>
<td>Beekman, R., Martín-Subero, JL.</td>
</tr>
</tbody>
</table>
**Table of Contents**

- Introduction
- Epigenetic regulation of gene expression
  - PRC2 methylates histones and DNA
  - TET1,2,3 and TDG demethylate DNA
- Positive epigenetic regulation of rRNA expression
- Negative epigenetic regulation of rRNA expression
- DNA methylation

Table of Contents