Generic Transcription Pathway

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.

This is just an excerpt of a full-length report for this pathway. To access the complete report, please download it at the Reactome Textbook.

31/10/2022
Introduction

Reactome is open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations 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: 82

This document contains 15 pathways and 4 reactions (see Table of Contents)
OVERVIEW OF TRANSCRIPTION REGULATION:

Detailed studies of gene transcription regulation in a wide variety of eukaryotic systems has revealed the general principles and mechanisms by which cell- or tissue-specific regulation of differential gene transcription is mediated (reviewed in Naar, 2001. Kadonaga, 2004, Maston, 2006, Barolo, 2002; Roeder, 2005, Rosenfeld, 2006). Of the three major classes of DNA polymerase involved in eukaryotic gene transcription, Polymerase II generally regulates protein-encoding genes. Figure 1 shows a diagram of the various components involved in cell-specific regulation of Pol-II gene transcription.

Core Promoter: Pol II-regulated genes typically have a Core Promoter where Pol II and a variety of general factors bind to specific DNA motifs:

i: the TATA box (TATA DNA sequence), which is bound by the "TATA-binding protein" (TBP).

ii: the Initiator motif (INR), where Pol II and certain other core factors bind, is present in many Pol II-regulated genes.

iii: the Downstream Promoter Element (DPE), which is present in a subset of Pol II genes, and where additional core factors bind.

The core promoter binding factors are generally ubiquitously expressed, although there are exceptions to this.

Proximal Promoter: immediately upstream (5') of the core promoter, Pol II target genes often have a Proximal Promoter region that spans up to 500 base pairs (b.p.), or even to 1000 b.p.. This region contains a number of functional DNA binding sites for a specific set of transcription activator (TA) and transcription repressor (TR) proteins. These TA and TR factors are generally cell- or tissue-specific in expression, rather than ubiquitous, so that the presence of their cognate binding sites in the proximal promoter region programs cell- or tissue-specific expression of the target gene, perhaps in conjunction with TA and TR complexes bound in distal enhancer regions.
Distal Enhancer(s): many or most Pol II regulated genes in higher eukaryotes have one or more distal Enhancer regions which are essential for proper regulation of the gene, often in a cell or tissue-specific pattern. Like the proximal promoter region, each of the distal enhancer regions typically contain a cluster of binding sites for specific TA and/or TR DNA-binding factors, rather than just a single site.

Enhancers generally have three defining characteristics:

i: They can be located very long distances from the promoter of the target gene they regulate, sometimes as far as 100 Kb, or more.

ii: They can be either upstream (5') or downstream (3') of the target gene, including within introns of that gene.

iii: They can function in either orientation in the DNA.

Combinatorial mechanisms of transcription regulation: The specific combination of TA and TR binding sites within the proximal promoter and/or distal enhancer(s) provides a "combinatorial transcription code" that mediates cell- or tissue-specific expression of the associated target gene. Each promoter or enhancer region mediates expression in a specific subset of the overall expression pattern. In at least some cases, each enhancer region functions completely independently of the others, so that the overall expression pattern is a linear combination of the expression patterns of each of the enhancer modules.

Co-Activator and Co-Repressor Complexes: DNA-bound TA and TR proteins typically recruit the assembly of specific Co-Activator (Co-A) and Co-Repressor (Co-R) Complexes, respectively, which are essential for regulating target gene transcription. Both Co-A's and Co-R's are multi-protein complexes that contain several specific protein components.

Co-Activator complexes generally contain at least one component protein that has Histone Acetyl Transferase (HAT) enzymatic activity. This functions to acetylate Histones and/or other chromatin-associated factors, which typically increases that transcription activation of the target gene. By contrast, Co-Repressor complexes generally contain at least one component protein that has Histone De-Acetylase (HDAC) enzymatic activity. This functions to de-acetylate Histones and/or other chromatin-associated factors. This typically increases the transcription repression of the target gene.

Adaptor (Mediator) complexes: In addition to the co-activator complexes that assemble on particular cell-specific TA factors, there are at least two additional transcriptional co-activator complexes common to most cells. One of these is the Mediator complex, which functions as an "adaptor" complex that bridges between the tissue-specific co-activator complexes assembled in the proximal promoter (or distal enhancers). The human Mediator complex has been shown to contain at least 19 protein distinct components. Different combinations of these co-activator proteins are also found to be components of specific transcription Co-Activator complexes, such as the DRIP, TRAP and ARC complexes described below.

TBP/TAF complex: Another large Co-A complex is the "TBP-associated factors" (TAFs) that assemble on TBP (TATA-Binding Protein), which is bound to the TATA box present in many promoters. There are at least 23 human TAF proteins that have been identified. Many of these are ubiquitously expressed, but TAFs can also be expressed in a cell or tissue-specific pattern.

Specific Coactivator Complexes for DNA-binding Transcription Factors.

A number of specific co-activator complexes for DNA-binding transcription factors have been identified, including DRIP, TRAP, and ARC (reviewed in Bourbon, 2004, Blazek, 2005, Conaway, 2005, and Malik, 2005). The DRIP co-activator complex was originally identified and named as a specific complex associated with the Vitamin D Receptor member of the nuclear receptor family of transcription factors.
Similarly, the TRAP co-activator complex was originally identified as a complex that associates with the thyroid receptor (Yuan, 1998). It was later determined that all of the components of the DRIP complex are also present in the TRAP complex, and the ARC complex (discussed further below). For example, the DRIP205 and TRAP220 proteins were shown to be identical, as were specific pairs of the other components of these complexes (Rachez, 1999).

In addition, these various transcription co-activator proteins identified in mammalian cells were found to be the orthologues or homologues of the Mediator ("adaptor") complex proteins (reviewed in Bourbon, 2004). The Mediator proteins were originally identified in yeast by Kornberg and colleagues, as complexes associated with DNA polymerase (Kelleher, 1990). In higher organisms, Adapter complexes bridge between the basal transcription factors (including Pol II) and tissue-specific transcription factors (TFs) bound to sites within upstream Proximal Promoter regions or distal Enhancer regions (Figure 1). However, many of the Mediator homologues can also be found in complexes associated with specific transcription factors in higher organisms. A unified nomenclature system for these adapter / co-activator proteins now labels them Mediator 1 through Mediator 31 (Bourbon, 2004). For example, the DRIP205 / TRAP220 proteins are now identified as Mediator 1 (Rachez, 1999), based on homology with yeast Mediator 1.

Example Pathway: Specific Regulation of Target Genes During Notch Signaling:

One well-studied example of cell-specific regulation of gene transcription is selective regulation of target genes during Notch signaling. Notch signaling was first identified in Drosophila, where it has been studied in detail at the genetic, molecular, biochemical and cellular levels (reviewed in Justice, 2002; Bray, 2006; Schweisguth, 2004; Louvri, 2006). In Drosophila, Notch signaling to the nucleus is thought always to be mediated by one specific DNA binding transcription factor, Suppressor of Hairless. In mammals, the homologous genes are called CBF1 (or RBPJkappa), while in worms they are called Lag-1, so that the acronym "CSL" has been given to this conserved transcription factor family. There are at least two human CSL homologues, which are now named RBPJ and RBPJL.

In Drosophila, Su(H) is known to be bifunctional, in that it represses target gene transcription in the absence of Notch signaling, but activates target genes during Notch signaling. At least some of the mammalian CSL homologues are believed also to be bifunctional, and to mediate target gene repression in the absence of Notch signaling, and activation in the presence of Notch signaling.

Notch Co-Activator and Co-Repressor complexes: This repression is mediated by at least one specific co-repressor complexes (Co-R) bound to CSL in the absence of Notch signaling. In Drosophila, this co-repressor complex consists of at least three distinct co-repressor proteins: Hairless, Groucho, and dCtBP (Drosophila C-terminal Binding Protein). Hairless has been shown to bind directly to Su(H), and Groucho and dCtBP have been shown to bind directly to Hairless (Barolo, 2002). All three of the co-repressor proteins have been shown to be necessary for proper gene regulation during Notch signaling in vivo (Nagel, 2005).

In mammals, the same general pathway and mechanisms are observed, where CSL proteins are bifunctional DNA binding transcription factors (TFs), that bind to Co-Repressor complexes to mediate repression in the absence of Notch signaling, and bind to Co-Activator complexes to mediate activation in the presence of Notch signaling. However, in mammals, there may be multiple co-repressor complexes, rather than the single Hairless co-repressor complex that has been observed in Drosophila.

During Notch signaling in all systems, the Notch transmembrane receptor is cleaved and the Notch intracellular domain (NICD) translocates to the nucleus, where it there functions as a specific transcription co-activator for CSL proteins. In the nucleus, NICD replaces the Co-R complex bound to CSL, thus resulting in de-repression of Notch target genes in the nucleus (Figure 2). Once bound to CSL, NICD and CSL
proteins recruit an additional co-activator protein, Mastermind, to form a CSL-NICD-Mam ternary co-activator (Co-A) complex. This Co-R complex was initially thought to be sufficient to mediate activation of at least some Notch target genes. However, there now is evidence that still other co-activators and additional DNA-binding transcription factors are required in at least some contexts (reviewed in Barolo, 2002).

Thus, CSL is a good example of a bifunctional DNA-binding transcription factor that mediates repression of specific targets genes in one context, but activation of the same targets in another context. This bifunctionality is mediated by the association of specific Co-Repressor complexes vs. specific Co-Activator complexes in different contexts, namely in the absence or presence of Notch signaling.

**Literature references**


**Editions**

2008-02-26 Reviewed Freedman, LP.
Formation of ARC coactivator complex

Location: Generic Transcription Pathway

Stable identifier: R-HSA-212352

Type: binding

Compartments: nucleoplasm

ARC co-activator complex and assembly

The ARC co-activator complex is a subset of 18 proteins from the set of at least 31 Mediator proteins that, in different combinations, form "Adapter" complexes in human cells. Adapter complexes bridge between the basal transcription factors (including Pol II) and tissue-specific transcription factors (TFs) bound to sites within upstream Proximal Promoter regions or distal Enhancer regions (reviewed in Maston, 2006 and Naar, 2001).

The ARC complex was originally identified and named as a co-activator complex associated with transcription activator proteins (reviewed in Malik, 2005 and references therein). It was subsequently determined that many of the components of the ARC complex are also in the DRIP complex, and in the TRAP complex..

The ARC complex contains the following 14 proteins, which also are common to the DRIP and TRAP complexes: MED1, MED4, MED6, MED7, MED10, MED12, MED13, MED14, MED16, MED17, MED23, MED24, CDK8, CycC.

The ARC complex also contains 4 additional, ARC-specific components, which are now called: MED8, MED15, MED25, and MED 26 in the unified nomenclature scheme (Bourbon, 2004).

In addition, these various transcription co-activator proteins identified in mammalian cells were found to be the orthologues or homologues of the Mediator complex proteins in yeast, first identified by Kornberg and colleagues (Kelleher, 1990). The unified nomenclature system for these adapter / co-activator proteins now labels them Mediator 1 through Mediator 31 (Bourbon, 2004).

The order of addition of the ARC proteins during complex assembly is not fully determined, and may vary in different cell contexts. Therefore, ARC complex assembly is represented as a single reaction.

https://reactome.org
event, in which all 19 components assemble simultaneously into the ARC co-activator complex.

**Literature references**


**Editions**

2008-02-26 Reviewed Freedman, LP.
Formation of DRIP coactivator complex

**Location:** Generic Transcription Pathway

**Stable identifier:** R-HSA-212432

**Type:** binding

**Compartments:** nucleoplasm

---

DRIP co-activator complex and assembly

The DRIP co-activator complex is a subset of 14 proteins from the set of at least 31 Mediator proteins that, in different combinations, form "Adapter" complexes. Adapter complexes bridge between the basal transcription factors (including Pol II) and tissue-specific transcription factors (TFs) bound to sites within upstream Proximal Promoter regions or distal Enhancer regions (reviewed in Maston, 2006 and Naar, 2001).

The DRIP complex was originally identified and named as a co-activator complex associated with the Vitamin D Receptor member of the nuclear receptor family of transcription factors (Rachez, 1998). It was later determined that all of the components of the DRIP complex were also in the TRAP complex, and the ARC complex.

The DRIP complex contains the following 14 proteins, which also are common to the ARC and TRAP complexes: MED1, MED4, MED6, MED7, MED10, MED12, MED13, MED14, MED16, MED17, MED23, MED24, CDK8, CycC.

All of the DRIP adapter complex components are present in the ARC adapter complex, but the ARC complex also has 4 additional components (Rachez, 1999). These ARC-specific components are now called: MED8, MED15, MED25, and MED 26 in the unified nomenclature scheme (Bourbon, 2004).

Similarly, all 14 of the DRIP adapter complex components are present in the TRAP adapter complex, but the TRAP complex also has 4 additional components (Bourbon, 2004). These TRAP-specific components are now called: MED20, MED27, MED30, and MED 31 in the unified nomenclature scheme.

In addition, these various transcription co-activator proteins identified in mammalian cells were found to be the orthologues or homologues of the Mediator complex identified in yeast, first identified by Kornberg and colleagues (Kelleher, 1990).

https://reactome.org
Literature references


Editions

2008-02-26 Reviewed Freedman, LP.
TRAP co-activator complex and assembly

The TRAP co-activator complex is a subset of 18 proteins from the set of at least 31 Mediator proteins that, in different combinations and in different contexts, form specific co-activator or "Adapter" complexes in human cells. These complexes bridge between the basal transcription factors (including Pol II) and tissue-specific transcription factors (TFs) bound to sites within upstream Proximal Promoter regions or distal Enhancer regions (reviewed in Maston, 2006 and Naar, 2001).

The TRAP complex was originally identified and named as a co-activator complex associated with the Thyroid Hormone Receptor member of the nuclear receptor family of transcription factors (Yuan, 1998). It was later determined that many of the components of the TRAP complex are also in the DRIP complex, and in the ARC complex.

The TRAP complex contains the following 14 proteins, which also are common to the DRIP and ARC complexes: MED1, MED4, MED6, MED7, MED10, MED12, MED13, MED14, MED16, MED17, MED23, MED24, CDK8, CycC.

The TRAP complex also contains 4 additional components, which are now called: MED20, MED27, MED30, and MED 31 in the unified nomenclature scheme (Bourbon, 2004).

In addition, these various transcription co-activator proteins identified in mammalian cells were found to be the orthologues or homologues of the Mediator complex proteins in yeast, first identified by Kornberg and colleagues (Kelleher, 1990). The unified nomenclature system for these adapter / co-activator proteins now labels them Mediator 1 through Mediator 31 (Bourbon, 2004).

The order of addition of the TRAP proteins during complex assembly is not fully determined, and may vary in different cell contexts. Therefore, TRAP co-activator complex assembly is represented as a single
reaction event, in which all 18 components assemble simultaneously into the TRAP co-activator complex.

**Literature references**


THE NOTCH-HLH TRANSCRIPTION PATHWAY:

Notch signaling was first identified in Drosophila, where it has been studied in detail at the genetic, molecular, biochemical and cellular levels (reviewed in Justice, 2002; Bray, 2006; Schweisguth, 2004; Louvri, 2006). In Drosophila, Notch signaling to the nucleus is thought always to be mediated by one specific DNA binding transcription factor, Suppressor of Hairless. In mammals, the homologous genes are called CBF1 (or RBPJkappa), while in worms they are called Lag-1, so that the acronym "CSL" has been given to this conserved transcription factor family. There are at least two human CSL homologues, which are now named RBPJ and RBPJL.

CSL is an example of a bifunctional DNA-binding transcription factor that mediates repression of specific target genes in one context, but activation of the same targets in another context. This bifunctionality is mediated by the association of specific Co-Repressor complexes vs. specific Co-Activator complexes in different contexts, namely in the absence or presence of Notch signaling.

In Drosophila, Su(H) represses target gene transcription in the absence of Notch signaling, but activates target genes during Notch signaling. At least some of the mammalian CSL homologues are believed also to be bifunctional, and to mediate target gene repression in the absence of Notch signaling, and activation in the presence of Notch signaling.

Notch Co-Activator and Co-Repressor complexes: This repression is mediated by at least one specific corepressor complexes (Co-R) bound to CSL in the absence of Notch signaling. In Drosophila, this corepressor complex consists of at least three distinct corepressor proteins: Hairless, Groucho, and dCtBP (Drosophila C-terminal Binding Protein). Hairless has been show to bind directly to Su(H), and Groucho and dCtBP have been shown to bind directly to Hairless (Barolo, 2002). All three of the co-repressor proteins have been shown to be necessary for proper gene regulation during Notch signaling in vivo (Nagel, 2005).
In mammals, the same general pathway and mechanisms are observed, where CSL proteins are bifunctional DNA binding transcription factors (TFs), that bind to Co-Repressor complexes to mediate repression in the absence of Notch signaling, and bind to Co-Activator complexes to mediate activation in the presence of Notch signaling. However, in mammals, there may be multiple co-repressor complexes, rather than the single Hairless co-repressor complex that has been observed in Drosophila.

During Notch signaling in all systems, the Notch transmembrane receptor is cleaved and the Notch intracellular domain (NICD) translocates to the nucleus, where it there functions as a specific transcription co-activator for CSL proteins. In the nucleus, NICD replaces the Co-R complex bound to CSL, thus resulting in de-repression of Notch target genes in the nucleus. Once bound to CSL, NICD and CSL proteins recruit an additional co-activator protein, Mastermind, to form a CSL-NICD-Mam ternary co-activator (Co-A) complex. This Co-A complex was initially thought to be sufficient to mediate activation of at least some Notch target genes. However, there now is evidence that still other co-activators and additional DNA-binding transcription factors are required in at least some contexts (reviewed in Barolo, 2002).

Mammalian CSL Corepressor Complexes: In the absence of activated Notch signaling, DNA-bound CSL proteins recruit a corepressor complex to maintain target genes in the repressed state until Notch is specifically activated. The mammalian corepressor complexes include NCOR complexes, but may also include additional corepressor proteins, such as SHARP (reviewed in Mumm, 2000 and Kovall, 2007). The exact composition of the CSL NCOR complex is not known, but in other pathways the "core" NCOR corepressor complex includes at least one NCOR protein (NCOR1, NCOR2, CIR), one Histone Deacetylase protein (HDAC1, HDAC2, HDAC3, etc), and one TBL1 protein (TBL1X, TBL1XR1) (reviewed in Rosenfeld, 2006). In some contexts, the core NCOR corepressor complex may also recruit additional corepressor proteins or complexes, such as the SIN3 complex, which consists of SIN3 (SIN3A, SIN3B), and SAP30, or other SIN3-associated proteins. An additional CSL - NCOR binding corepressor, SHARP, may also contribute to the CSL corepressor complex in some contexts (Oswald, 2002). The CSL corepressor complex also includes a bifunctional cofactor, SKIP, that is present in both CSL corepressor complexes and CSL co-activator complexes, and may function in the binding of NICD and displacement of the corepressor complex during activated Notch signaling (Zhou, 2000).

Mammalian CSL Coactivator Complexes: Upon activation of Notch signaling, cleavage of the transmembrane Notch receptor releases the Notch Intracellular Domain (NICD), which translocates to the nucleus, where it binds to CSL and displaces the corepressor complex from CSL (reviewed in Mumm, 2000 and Kovall, 2007). The resulting CSL-NICD "binary complex" then recruits an additional coactivator, Mastermind (Mam), to form a ternary complex. The ternary complex then recruits additional, more general co-activators, such as CREB Binding Protein (CBP), or the related p300 coactivator, and a number of Histone Acetyltransferase (HAT) proteins, including GCN5 and PCAF (Fryer, 2002). There is evidence that Mam also can subsequently recruit specific kinases that phosphorylate NICD, to downregulate its function and turn off Notch signaling (Fryer, 2004).

Combinatorial Complexity in Transcription Cofactor Complexes: HDAC9 has at least 7 splice isoforms, with some having distinct interaction and functional properties. Isoforms 6 and 7 interact with NCOR1. Isoforms 1 and 4 interact with MEF2 (Sparrow, 1999), which is a specific DNA-binding cofactor for a subset of HLH proteins. Isoform 3 interacts with both NCOR1 and MEF2. Although many HDACs only have one or two isoforms, this complexity for HDAC9 illustrates the level of transcript complexity and functional specificity that such "general" transcriptional cofactors can have.

**Literature references**


**Editions**

2008-02-09  Authored, Edited  Caudy, M.
A classic example of bifunctional transcription factors is the family of Nuclear Receptor (NR) proteins. These are DNA-binding transcription factors that bind certain hormones, vitamins, and other small, diffusible signaling molecules. The non-liganded NRs recruit specific corepressor complexes of the NCOR/SMRT type, to mediate transcriptional repression of the target genes to which they are bound. During signaling, ligand binding to a specific domain the NR proteins induces a conformational change that results in the exchange of the associated CoR complex, and its replacement by a specific coactivator complex of the TRAP / DRIP / Mediator type. These coactivator complexes typically nucleate around a MED1 coactivator protein that is directly bound to the NR transcription factor.

A general feature of the 49 human NR proteins is that in the unliganded state, they each bind directly to an NCOR corepressor protein, either NCOR1 or NCOR2 (NCOR2 was previously named "SMRT"). This NCOR protein nucleates the assembly of additional, specific corepressor proteins, depending on the cell and DNA context. The NR-NCOR interaction is mediated by a specific protein interaction domain (PID) present in the NRs that binds to specific cognate PID(s) present in the NCOR proteins. Thus, the human NRs each take part in an NR-NCOR binding reaction in the absence of binding by their ligand.

A second general feature of the NR proteins is that they each contain an additional, but different PID that mediates specific binding interactions with MED1 proteins. In the ligand-bound state, NRs each take part in an NR-MED1 binding reaction to form an NR-MED1 complex. The bound MED1 then functions to nucleate the assembly of additional specific coactivator proteins, depending on the cell and DNA context, such as what specific target gene promoter they are bound to, and in what cell type.

The formation of specific MED1-containing coactivator complexes on specific NR proteins has been well-characterized for a number of the human NR proteins (see Table 1 in (Bourbon, 2004)). For example, binding of thyroid hormone (TH) to the human TH Receptor (THRA or THRb) was found to result in the recruitment of a specific complex of Thyroid Receptor Associated Proteins - the TRAP coactivator complex - of which the TRAP220 subunit was later identified to be the Mediator 1 (MED1) homologue.
Similarly, binding of Vitamin D to the human Vitamin D3 Receptor was found to result in the recruitment of a specific complex of D Receptor Interacting Proteins - the DRIP coactivator complex, of which the DRIP205 subunit was later identified to be human MED1.

**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-11-20</td>
<td>Authored</td>
<td>Caudy, M.</td>
</tr>
<tr>
<td>2009-05-27</td>
<td>Edited</td>
<td>Caudy, M.</td>
</tr>
<tr>
<td>2009-08-29</td>
<td>Reviewed</td>
<td>Freedman, LP.</td>
</tr>
</tbody>
</table>
KRAB-ZNF / KAP Interaction

Location: Generic Transcription Pathway

Stable identifier: R-HSA-975040

Type: binding

Compartments: nucleoplasm