

Transcription overhaul reannotation guide: Notes

Purpose: The goal in this guide is to provide an overview of the basic "functions" that researchers state for the various components of the basal transcription machinery, including the RNA polymerase, and the "regulatory" transcription factors. Where possible, we have listed typical statements given by researchers describing the "functions" of various transcription factors, and then given the corresponding GO term. Note that some researcher statements of "function" correspond to GO biological_process terms rather than to GO molecular_function terms.

Color coding: We have used color coding to distinguish the different types of recommendations for annotations that we have made. Molecular Function (MF) terms are highlighted with orange. Biological Process terms are highlighted with blue. Comments out of the scope of the transcription overhaul (txnOH) are highlighted in yellow. There are a couple places, highlighted in green, where we suspect a new term is warranted, but have not yet implemented it. Feel free to give us input on these.

Transcription factors - complexes vs single subunits: Many "transcription factors" are complexes, not single subunit gene products. Where the functional unit of action is the complex, the transcription factor activity terms listed here are appropriate for the function of the transcription factor complex. Since the "transcription factor activity" terms are in the MF ontology, it may often be appropriate to make annotations for all members of the complex using the "contributes to" qualifier. Some transcription factors contains subunits which, individually, possess catalytic activities. These activities should be annotated only to the subunits which possess them, not to all members of the complex.

On has_part relationships: Note that in this guide, we have indicated appropriate terms representing overall "complex" functions of a transcription factor, and some transcription factors have more than one molecular function. These functions generally involve one or more binding interactions. Following the decision at the 2010 Annotation Camp in Geneva to use the has_part relationship to represent when a "complex" function has binding as a component of how it functions, these terms have one or more has_part relationships to represent the appropriate binding interactions between the transcription factors and nucleic acids (DNA or RNA) or proteins (RNA polymerase, transcription factors). Note that due to the decision to use the has_part relationship, annotation to a transcription factor activity term does NOT automatically map up to binding of any kind. If you wish have the binding aspects of these functions appear in enrichment analyses, you may wish to also make direct annotations to the appropriate binding functions. When a single subunit is shown to possess binding activity individually, e.g. TBP binds to TATA containing DNA, you may want to annotate that subunit individually. For binding that occurs in the context of a complex, it might be appropriate to annotate all subunits of the complex using "contributes to". We have not indicated the individual binding terms. You can get indications of appropriate terms to consider by following the has_part relationships from the indicated terms and their parents.

RNA polymerase II basal transcription machinery

The basal transcription machinery is typically described as the minimum set of factors required for preinitiation complex (PIC) formation. For RNA polymerase II (RNAP II or Pol II), the basal machinery includes RNAP II itself as well as a number of "basal transcription factors" (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH), most of which are complexes (except for TFIIB). The basal transcription factors are often described as having multiple roles, many of which are related to binding and which occur in the context of _entire_ transcription factor, which again, is usually a complex. Some of the transcription factors also contain catalytic activities, however these activities are generally possessed by individual subunits. This section shows the summaries from two reviews and indicates the appropriate term(s) that correspond to each "function". Note that since researchers often use the word "function" in a looser way than the GO meaning of "molecular function" (indicated in orange), some of these "functions" correspond to terms in "biological process" (indicated in blue) in GO. Some of the indicated functions are catalytic activities contained by individual subunits within a given complex. These catalytic activities were outwith the scope of the transcription overhaul (txnOH). We have indicated that (in yellow), but also included a list of already existing catalytic terms to consider.

Factor	Protein composition	"Function" from Thomas & Chiang 2006 (Table 1)	from Sikorski & Buratowski (2009)	Suggested GO term(s)
TFIIA (2-3 subunits)	H. sapiens: p35 (alpha), p19 (beta) & p12 (gamma)	antirepressor	functions to counteract repressive effects of negative cofactors like NC2	may need a new MF term for this?
	S. cerevisiae: TOA1 & TOA2	stabilizes TBP-TATA complex		(contributes to) MF GO:0001129 - TBP-class protein binding RNA polymerase II transcription factor activity involved in preinitiation complex assembly
		coactivator	acts as a coactivator by interacting with activators and components of the basal initiation machinery	(contributes to) MF GO:0001128 - RNA polymerase II transcription coactivator activity involved in preinitiation complex assembly

RNA polymerase II basal transcription machinery, continued

Factor	Protein composition	"Function" from Thomas & Chiang 2006 (Table 1)	from Sikorski & Buratowski (2009)	Suggested GO term(s)
TFIIB (1 subunit)	H. sapiens: p33	start site selection	directs accurate start site selection	BP GO:0001174 - transcriptional start site selection at RNA polymerase II promoter
	S. cerevisiae: SUA7	stabilizes TBP-TATA complex	stabilizes TFIID-promoter binding	MF GO:0001129 - TBP-class protein binding RNA polymerase II transcription factor activity involved in preinitiation complex assembly
		pol II/TFIIF recruitment	aids in recruitment of TFIIF/Pol II to the promoter	probably need a new term for recruitment of TFIIF/Pol II complex together
TFIID (14 subunits including TBP and TAFs)	H. sapiens: TBP + TAFs (TAF1-TAF14)	core promoter-binding factor	nucleates PIC assembly either through TBP binding to TATA sequences or TAF binding to other promoter sequences	(contributes to) MF GO:0001075 - sequence-specific core promoter binding RNA polymerase II transcription factor activity involved in preinitiation complex assembly
		coactivator	coactivator activity through direct interaction of TAFs and gene specific activators	(contributes to) MF GO:0001128 - RNA polymerase II transcription coactivator activity involved in preinitiation complex assembly
	protein kinase		annotate individual subunits as appropriate for catalytic activities (not addressed by txnOH)	
	ubiquitin-activating/conjugating activity			
	histone acetyltransferase			

RNA polymerase II basal transcription machinery, continued

Factor	Protein composition	"Function" from Thomas & Chiang 2006 (Table 1)	from Sikorski & Buratowski (2009)	Suggested GO term(s)
TFIIE (2 subunits)	H. sapiens: p56 (alpha) & p34 (beta)	recruits TFIIH	helps recruit TFIIH to promoters	(contributes to) MF GO:0001138 - TFIIH-class transcription factor recruiting transcription factor activity
		facilitates formation of an initiation-competent pol II		
	S. cerevisiae: TFA1 & TFA2		stimulates helicase and kinase activities of TFIIH	outwith scope of txnOH, consider BP annotations to regulate terms with TFIIH as a target in column 16
			binds ssDNA	MF GO:0003697 - single-stranded DNA binding (consider whether to annotate as complex or individual subunit)
			and is essential for promoter melting	BP GO:0001113 - transcriptional open complex formation at RNA polymerase II promoter
			involved in promoter clearance	BP GO:0001111 - promoter clearance from RNA polymerase II promoter

RNA polymerase II basal transcription machinery, continued

Factor	Protein composition	"Function" from Thomas & Chiang 2006 (Table 1)	from Sikorski & Buratowski (2009)	Suggested GO term(s)
TFIIF (2-3 subunits)	H. sapiens: RAP30 & RAP74	binds pol II and facilitates pol II recruitment to the promoter	tightly associates with RNAPII; enhances affinity of RNAPII for TBP-TFIIB-promoter complex	(contributes to) MF GO:0001139 - core RNA polymerase II recruiting transcription factor activity
		recruits TFIIE and TFIIH	necessary for recruitment of TFIIE/TFIIH to the PIC	(contributes to) MF GO:0001136 - TFIIE-class transcription factor recruiting transcription factor activity AND (contributes to) MF GO:0001138 - TFIIH-class transcription factor recruiting transcription factor activity
	S. cerevisiae: TFG1, TFG2, & TAF14	functions with TFIIB and pol II in start site selection	aids in start site selection	BP GO:0001111 - promoter clearance from RNA polymerase II promoter
		facilitates pol II promoter escape	and promoter escape	BP GO:0001111 - promoter clearance from RNA polymerase II promoter
		enhances the efficiency of pol II elongation	enhances elongation efficiency	GO:0032968 - positive regulation of transcription elongation from RNA polymerase II promoter

RNA polymerase II basal transcription machinery, continued

Factor	Protein composition	"Function" from Thomas & Chiang 2006 (Table 1)	from Sikorski & Buratowski (2009)	Suggested GO term(s)
TFIIH (10 subunits)	H. sapiens: p89/XPB, p80/xPD, p62, p52, p44, p40/CDK7, p38/Cyclin H, p34, p32/MAT1, & p8/TFB5	ATPase activity for transcription initiation and promoter clearance; helicase activity for promoter opening	ATPase/helicase activity for promoter opening and promoter clearance	annotate individual subunits as appropriate for catalytic activities (not addressed by txnOH)
				BP GO:0001113 - transcriptional open complex formation at RNA polymerase II promoter
		facilitates transition from initiation to elongation		BP GO:0001111 - promoter clearance from RNA polymerase II promoter
	S. cerevisiae: CCL1, KIN28, RAD3, SSL1, SSL2, TFB1, TFB2, TFB3, TFB4, & TFB5	transcription-coupled nucleotide excision repair	helicase activity for transcription coupled DNA repair	outwith scope of txnOH, but consider BP GO:0006283 - transcription-coupled nucleotide-excision repair, or child terms
		kinase activity for phosphorylating pol II CTD	kinase activity required for phosphorylation of RNAP II CTD	annotate individual subunits as appropriate for catalytic activities (not addressed by txnOH)
	E3 ubiquitin ligase activity			
RNAP II or pol II (12 subunits)	H. sapiens: RPB1-RPB12	transcription initiation, elongation, termination	catalyzes transcription of all mRNAs and a subset of noncoding RNAs including [most] snoRNAs and miRNAs	(contributes to) MF GO:0001055 - RNA polymerase II activity
				BP GO:0006366 - transcription from RNA polymerase II promoter
	S. cerevisiae: RPO21/RPB1, RPB2, RPB3, RPB4, RPB5, RPO26/RBP6, RPB7, RPB8, RPB,9 RPB10, RPB11, & RPC10/RPB12	recruitment of mRNA capping enzymes		outwith scope of txnOH, not currently represented in GO
		transcription-coupled recruitment of splicing and 3' end processing factors		BP GO:0034402 - recruitment of 3' end processing factors to RNA polymerase II holoenzyme complex
		CTD phosphorylation, glycosylation, and ubitination		[occurs on RNAP II, not something RNAP II does]

RNA polymerase II cofactors

A number of transcription cofactors are characterized as being involved in or required for transcription from RNAP II promoters. A cofactor is defined as "interacting selectively and non-covalently with a regulatory transcription factor and also with the basal transcription machinery in order to modulate transcription." The phrase "cofactor activity" is invoked a lot in the literature, and it was difficult to pin down exactly how they function beyond this general definition. I think it may turn out that there are multiple different types of cofactors that act in slightly different ways. Some of the basal transcription factors, e.g. TFIIA and TFIID, are described as having "cofactor activity". These two large complexes have not been classified as part of the basal transcription machinery, and have been purified numerous different ways with slightly differing compositions. In addition, there is some evidence that these complexes may be involved in "basal" transcription as well as in "activator-dependent" transcription, adding to the confusion. Nevertheless, these complexes are described as having "cofactor activity", as defined above. Thus we have represented that level of detail, including the positive (coactivator) and negative (corepressor) versions of cofactors. If our understanding advances, it may be appropriate to make more specific terms representing more precise mechanisms of action. The functions given for the cofactor complexes in this section of the guide are derived only from the Sikorski & Buratowski review; again we have listed appropriate MF (orange) or BP terms (blue) that correspond with the described functions.

Factor	Protein composition	"Function" from Thomas & Chiang 2006 (table 1)	from Sikorski & Buratowski (2009)	Suggested GO term(s)
Mediator (at least 24 subunits)	multiple complexes have been identified; functional differences between them are not always clear		bridges interaction between activators and basal factors	(contributes to) MF GO:0001128 - RNA polymerase II transcription coactivator activity involved in preinitiation complex assembly
			stimulates both activator dependent and basal transcription	BP GO:0006357 - regulation of transcription from RNA polymerase II promoter
			required for transcription from most RNAPII dependent promoters	BP GO:0006366 - transcription from RNA polymerase II promoter
SAGA (20 subunits)	multiple complexes have been identified; functional differences between them are not always clear		interacts with activators, histone H3, and TBP	(contributes to) MF GO:0001128 - RNA polymerase II transcription coactivator activity involved in preinitiation complex assembly
			histone acetyltransferase activity	annotate individual subunits as appropriate for catalytic activities (not addressed by txnOH)
			deubiquitinating activity	

Catalytic activities found in subunits of basal txn factors

The transcription overhaul dealt with the functions of transcription factors as a whole. Thus, for the basal transcription factors, most of which are complexes, we dealt only with the functions of the complexes. Some of the basal transcription factors and cofactors contain subunits which individually have catalytic activities, e.g. kinase, helicase, histone acetyltransferase, etc. This section indicates some of the terms which exist for these catalytic activities. These terms, or their child terms, may be appropriate for annotation of the individual subunits which possess them. Note that the transcription overhaul did not address any of these terms, that most of these are not specific to transcription. As this was not the focus of the transcription overhaul, we have selected some high level terms as a guide; you should consider if there are more specific terms appropriate for a given gene.

Factor	Protein composition	"Function" from Thomas & Chiang 2006 (table 1)	from Sikorski & Buratowski (2009)	Suggested GO term(s)
		protein kinase		GO:0004672 - protein kinase activity (or child terms)
		kinase activity for phosphorylating pol II CTD	kinase activity required for phosphorylation of RNApII CTD	MF GO:0008353 - RNA polymerase II carboxy-terminal domain kinase activity
		ATPase activity [for transcription initiation and promoter clearance]	ATPase/helicase activity [for promoter opening and promoter clearance]	GO:0016887 - ATPase activity
		helicase activity [for promoter opening]		GO:0004386 - helicase activity
		histone acetyltransferase		GO:0004402 - histone acetyltransferase activity
		ubiquitin-activating/ conjugating activity		GO:0004842 - ubiquitin-protein ligase activity
		E3 ubiquitin ligase activity		
		deubiquitinating activity		GO:0004843 - ubiquitin-specific protease activity

RNA polymerase II "regulatory" transcription factors

Type of factor, by DNA binding	Examples	Note	Suggested GO term(s)
<p>Factors which bind DNA sequences in a sequence specific manner (generally upstream of and) proximal to the core promoter. Note that proximal is used loosely, as opposed to "distal enhancers". These are the classic "regulatory transcription factors" which bind to a specific DNA sequence.</p>	<p>Mammalian: feel free to suggest some</p>	<p>This term is the most general, requiring only that you know that it binds to a specific sequence and in doing so, modulates transcription</p>	<p>MF GO:0000982 - RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity</p>
		<p>A great many of the "regulatory" transcription factors act by improving the rate of PIC formation. In practice though, David has usually not seen evidence for this.</p>	<p>MF GO:0001074 - RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in preinitiation complex assembly</p>
	<p>S. cerevisiae: GAL4, GCN4, etc.</p>	<p>In practice, David has been finding it easy to find evidence for a positive effect, or a negative effect, on transcription</p>	<p>MF GO:0001077 - RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription</p>
			<p>MF GO:0001078 - RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription</p>

RNA polymerase II "regulatory" transcription factors, cont.

Type of factor, by DNA binding	Examples	Note	Suggested GO term(s)
Factors which bind enhancer DNA sequences distal to the core promoter		David and I kept these enhancer terms because, particularly for mammalian, people talk about "distal enhancers" that function independently of distance or orientation, which is what the term appears to have been originally intended for. However, David has come across cases where the term enhancer is used very loosely to mean things other than these distal regulatory sites. Sometimes for the same site, some groups indicate that it is proximal to the core promoter, while others refer to it as an enhancer. Karen has come across use of the word "enhancer" to refer to the UAS region in <i>S. cerevisiae</i> while other researchers refer to it in ways that would suggest that it is part of the core promoter proximal region. Thus, we think we may need to refine the "enhancer binding" terms in GO. Be cautious of using these terms unless you are really sure that it is an enhancer; this may take reading more than a single paper. Please feel free to provide input on improving the usage of "enhancer", sample papers, etc.	MF GO:0003705 - sequence-specific enhancer binding RNA polymerase II transcription factor activity