

Transcription Factors (from Wray et al Mol Biol Evol 20:1377)

Phenotype is affected by mutations in:

1. Structural region of a gene
 - ✓ Function of a protein is modified (structure/function relationship)
2. Regulatory region of a gene
 - ✓ When the protein is expressed (gene regulation)

Considerations of gene regulations

1. Change the regulation pattern of a gene can change phenotype
2. One transcription factor (TF) can affected multiple genes in a pathway
3. TF orthologs will regulate different organisms differently
4. Promoter contain module that can be changed to affect expression

Approaches to studying gene regulation

1. Mutants
 - Do induced mutants represent natural variation?
2. Expression patterns
 - ✓ Expression patterns of orthologs can differ among species
3. Expression levels
 - ✓ Phenotypic differences can result from changes in the amount of a protein

Effect of varying expression level

1. Spatial effects
 - ✓ Varying the amount of expression in a tissue can change phenotype
2. Cis-effects
 - ✓ Variation in expression level often related to changes in cis-elements
3. Inducibility
 - ✓ Alleles can be induced differentially

Levels of expression can vary at the:

1. mRNA level
2. Protein level

What amount of the variation is the result of “controlling region” variation???

1. Natural variation exists in promoters
 - ✓ Associated with phenotypic changes
2. Artificial selection of promoter sequences can change expression
 - ✓ Maize *tb* locus is an example
3. Promoter “elements” are conserved among species
 - ✓ Specific sequences important for gene expression
4. Variation in promoter sequence related to human disease susceptibility
 - ✓ Susceptibility to specific pathotypes related to promoter sequences

Transcription patterns are variable

1. Transcription initiation is the most important step in phenotypic expression
2. Regulation is at the gene not gene family level
 - ✓ Paralogs are independently regulated
3. Transcription is dynamic
 - ✓ Expression levels vary
 - ✓ Expression can fluctuate rapidly
 - ✓ Expression in neighboring cells can differ
4. Expression profiles vary among genes
 - ✓ Regulatory gene expression profile is inducible and highly variable
 - ✓ Housekeeping gene expression is generally constitutive but varies in response to stimuli and by cell type

Role of Controlling Regions (=Promoters) in Gene Expression

1. Promoters
 - ✓ Contain sequence motifs that bind factors that modulate gene expression
2. Constitutive (housekeeping) promoters
 - ✓ On by default,
 - ✓ Turned off in response to stimuli
3. Inducible promoters
 - ✓ Off by default
 - ✓ Turned on in response to stimuli
4. TF determine if genes are turned on or off

Promoters

1. Universal conserved features are not found
2. Common sequence motifs not found

Basal Gene Expression

1. Basal promoter
 - ✓ RNA polymerase complex binding site
 - ✓ Contains TATA box or initiator element
 - ✓ Null promoter exist
 - ✓ Lacks TATA box or initiator element
 - ✓ Multiple basal promoters exist for some genes
2. TATA-box binding protein (TBP)
 - ✓ First protein to bind the basal promoter
 - ✓ Other proteins guide TBP to the binding site
3. RNA polymerase holoenzyme complex
 - ✓ Complex interactions of proteins builds the transcription complex
4. Transcription start site
 - ✓ Begins about 30 bp downstream of site where the transcription complex
5. Translation start site
 - ✓ Begins about 10 – 10,000 bp from transcription start site
6. Basal promoters provides for minimal, low level of expression
 - ✓ Expression mediated by constitutively expressed general transcription factors

Modifying Basal Gene Expression Levels

1. TF binding to controlling regions required for full gene expression
 - ✓ TF are specific to cell types and stimuli conditions
 - ✓ Interaction of controlling regions and TF controls gene expression

Controlling Region TF Binding Sites

1. Binding sites are isolated in controlling region
 - ✓ Binding sites are embedded in regions to which no TF bind
2. Binding sites numbers
 - ✓ 10 – 50 binding sites for 5 –15 TF
3. Role of other sequences
 - ✓ Local, sequence-specific conformational changes can affect TF binding
 - ✓ AT-rich regions
 - ✓ Z-DNA
4. Spacing of binding sites
 - ✓ Partial overlap
 - ✓ 10s of kilobases

Features of TF Binding Sites

1. Size
 - ✓ Footprint (sequences covered by TF) is 10-20 bp
 - ✓ Direct binding site is 5-8 bp
 - ✓ Essential sequence is 4-6 bp
2. Site definition
 - ✓ Consensus sequence (although not all consensus sequences bind TF)
 - ✓ Biochemical activity (required to define a functional sequence)
3. Binding sites can overlap
 - ✓ TF pool determines which site is bound
 - ✓ Binding sites compete for a limited TF pool
4. Location
 - ✓ 100 basepairs to 100 kilobases from transcription start site
5. Functional location
 - ✓ >30 kb 5' of basal promoter
 - ✓ few kb of basal promoter
 - ✓ in 5' UTR
 - ✓ introns
 - ✓ >30 kb 3' of basal promoter
 - ✓ exon
 - ✓ other side of adjacent gene
6. Location constraints
 - ✓ Some sites are constrained to specific positions relative to transcription start site
7. Isolating binding sites effects
 - ✓ Insulator sequences limit TF interactions to specific basal promoters
 - ✓ TATA or TATA-less TF interaction specificity
 - ✓ Specific recruitment of TF at a specific sequence to interact with basal promoter
8. Multiple control

- ✓ On set of binding sets controls paralogs on opposite strands in opposite orientation
- ✓ Cross regulated sites share common binding sites

Abundance of Transcription Factors

1. TF are members of small to large multi-gene families
 - ✓ Arabidopsis
 - (CCAAT-DR1 Family) to 164 (C3H Family) paralog families
3. Result from gene duplication events
4. 12-15 unique DNA binding domains
 - ✓ Evolutionary conservation

Modular Nature of Transcription Factors

1. DNA binding domain
 - ✓ Localized
 - MADS-box or homeo domains
 - ✓ Dispersed
 - Zn-finger or leucine zipper domains
2. Protein-protein interaction domain
 - ✓ Binding to other proteins necessary for activation
3. Intracellular trafficking domains
 - ✓ Nuclear localization signal
4. Ligand binding domain
 - ✓ Steroid or hormone-binding domains
5. Evolutionary domain shuffling has occurred
 - ✓ Protein-protein interaction domain lost but DNA binding domain maintained

Transcription Factor DNA Binding Domain

1. Most bind the major groove of DNA
2. Domain sequence is highly conserved
 - ✓ Single amino acid mutations can alter significantly TF binding
3. TF binding specificity ranges from 3-5bp
4. Specificity may be increased by
 - ✓ Multiple binding domains
 - ✓ Domains that bind minor groove
 - ✓ Dimerization of two proteins, either homomeric or heteromeric
5. Binding is strong and highly specific
 - ✓ 5000 – 20,000 copies of TF needed for high binding specificity
6. Cofactor interactions increase specificity
 - ✓ Phosphorylation
7. Paralogs may have unique binding specificities

Transcription Factor Protein-Protein Interactions Modulate Gene Expression

1. Increase (or decrease) the frequency in which the transcription apparatus is built
 - ✓ Can recruit (or prevent recruitment) of apparatus components
2. Specific interactions necessary for effects to be realized
 - ✓ As homodimers
 - ✓ As heterodimers
 - ✓ As solo proteins
3. Neighboring effects
 - ✓ TF at one site can prevent cofactor from interacting with a neighboring site
4. Altering chromatin structure
 - ✓ Recruit other complexes that
 - ✓ Acetylate, deacetylate, methylate, or demethylate histones
 - ✓ Methylate or demethylate DNA
5. Create physical bends
 - ✓ Facilitates binding of other TF
6. Cofactors can bring TF and transcriptional apparatus together

Transcription Factor Activation or Repression of Transcription

1. Activation or repressor domains exists in TF
2. Action can be mediated through direct (or indirect via TAF) interaction with TBP

Transcription Factor Activation Depends Upon Specific Modifications and Interactions

1. Post-translational modifications such as phosphorylation necessary
2. Activation and repression domains may reside in same protein
 - ✓ Specific functional activity depends upon cofactor involvement
3. A TF can act as a repressor if it blocks the binding site of a TF activator
 - ✓ This interaction can have a downstream effect on other expression steps

Cooperative-binding and Interaction of Transcription Factors

1. Precise spacing required for some interactions involving TF
 - ✓ Nucleosome (40 bp multiples) or decondensed DNA (10 bp multiples) distances
 - ✓ Interactions with chromatin remodeling complexes may have a moderate distance requirement
2. Bending and looping supports interactions
 - ✓ Removes distance specificity requirement

Role of Functional Modules

1. Functional modules can have several functions
 - ✓ Initiate transcription
 - ✓ Boost transcription rate
 - ✓ Mediate extracellular signals
 - ✓ Repress transcription
 - ✓ Insulate on module from another (insulator function)
 - ✓ Bring other modules into contact with basal promoter
 - ✓ Integrate other module status into a global expression pattern

Additive and Epistatic Interactions of Transcription Factors

1. Modifying one TF and its module interaction can additively reduce the phenotype
2. Modifying insulator, tethering, or integrator TF functions is epistatic
3. Proper expression, recruitment, and modular association of TF is necessary for full phenotypic expression

A Transcription Family Has Multiple Target Genes

1. Because of the limited number of TF, a single TF may interact with 10s to 100s of genes
2. *Drosophila eve* and *ftz* regulate the majority of genes in the genome
3. The function of TF networks may genes
4. Mutations can be modulated by the effects of other downstream genes

The Genome Is Significantly Involved in Gene Regulation

1. The number of promoter sequences is equal to the number of protein coding sequences
2. Transcription regulation a major function of the genome

Arabidopsis Transcription Factors (information from: <http://arabidopsis.med.ohio-state.edu/AtTFDB/>) Total = 1430 transcription factors.

Family	Description	Number
ABI3VP1 Family		18
Alfin-like Family		7
AP2-EREBP Family	AP2(Apetala) /EREBP (ethylene-responsive element binding protein) transcription factors	120
ARF Family	Auxin response transcription factors	22
ARR-B Family	Type B regulators of phosphorealy signaling	15
bHLH Family	Basic helix-loop-helix transcription factors	146
bZIP Family	Basic leucine zipper transcription factors	70
C2C2-CO-like Family	CONSTANS-like transcription factors	30
C2C2-Dof Family	Dof transcription factors	36
C2C2-Gata Family	GATA motif binding transcription facotrs	28
C2C2-YABBY Family	YABBY transcription factors	5
C2H2 Family	Zinc finger protein 1 transcription factors	98
C3H Family		164
CCAAT-DR1 Family		2
CCAAT-HAP2 Family	CCAAT-binding transcription factors related to yeast HAP2 family	10
CCAAT-HAP3 Family	CCAAT-binding transcription factors related to yeast HAP3 family	10
CCAAT-HAP5 Family	CCAAT-binding transcription factors related to yeast HAP5 family	13
CPP Family		8
E2F-DP Family	G1/S phase regulatory transcription factors	8
EIL Family	Ethylene insensitive 3 transcription factors	6
G2-like Family	Maize golden 2 transcription factors	4
GRAS Family	SCARECROW-like transcription factors	25
GRF Family	Rice GROWTH-REGULATING FACTOR transcription factors	9
HB Family	Homeobox transcription factors	66
HSF Family	Heat shock transcription factors	21
MADS Family	MADS box transcription factors	100
MYB Family	MYB-binding domain transcription factors	137
MYB-related Family	MYB-related-binding domain transcription factors	9
NAC Family	NAC auxin signal transducing transcription factors	90
Orphan Family		3
SBP Family	SPL1 transcription factors	16
TCP Family	TCP transcription factors	26
Trihelix Family	GT-1 and GT-2 trihelix binding domain transcription factors	29
TUB Family	Tubulin transcription factors	10
WRKY Family	WRKY transcription factors	69

