

1.

a)

- b) TFIID is a complex of the TATA box binding protein (TBP) and a host of 13 other TAFs. TBP binds the genomic DNA at the TATA box along the minor groove and bends the DNA strand by about 80 degrees. The TAFs provide a significant amount of regulatory control, allowing the complex to respond to different activation signals. The TAFs have not been completely functionally annotated, but they have been shown to interact with activators, promoter elements, and to help TBP to bind to TATA-less promoters.

TFIIH has both kinase (phosphorylation of PolII CTD) and helicase (DNA melting) activity. The kinase activity (which is enhanced by TFIIIE *in vivo*) is essential for initiation of the transcription process and the transition to elongation. The helicase activity is essential for full DNA melting and promoter clearance once the transcription complex has established elongation.

TFIIF binds polymerase II and helps recruit it to the DABPolF complex, and also interacts with the non-template DNA. TFIIF function is related to the function carried out by the σ factor in prokaryotic transcription.

- c) Activators bind to particular regulatory sequences in the DNA. They also have one or more functional interfaces that allow them to interact with one or more TAFs and transcription factors. These interfaces can recruit general transcription factors or even the core polymerase itself. Often genes can respond to a variety of activators, allowing fine-tuned control of transcription activation in response to a wide variety of signals.

2.

The following lines of evidence suggest the importance of the B-linker in transcription, especially in promoter opening.

- Nuclear extracts were obtained from a yeast strain carrying mutations to the B gene. When these extracts were supplemented with wild-type TFIIB *in vitro*, promoter-dependent transcription occurred normally. When these extracts were instead supplemented with stable TFIIB variants with B-linker mutations, all promoter-dependent transcription was shut down. These B-linker mutations also caused lethality in yeast *in vivo*, confirming the importance of the B-linker in transcription.
- Using a transcription model based on the archaeal *Pfu* polymerase, it was determined that 4 different TFIIB variants with B-linker mutations were still able to form pre-initiation complexes (PICs), but were inactive in promoter-dependent transcription. This suggests that TFIIB has an important post-recruitment function, and that the B-linker is central to that function.
- Permanganate footprinting of ssDNA showed that B-linker mutants were unable to melt the promoter, confirming that TFIIB with B-linker mutations were still active except for DNA opening.
- It was also determined that mutations in the polymerase B-linker binding surface were also able to form stable PICs, but were inactive in promoter-dependent transcription. Control mutations to the polymerase outside the B-linker binding domain were as transcriptionally active as the wild-type complex.

The B-linker helix and the B-reader helix are connected by the B-reader loop and B-linker strand. Helix-loop-helix structures are common in DNA-binding domains, and perhaps this is the mechanism used by TFIIB. The B-linker would bind the DNA and melt the promoter, and the B-reader (connected by the flexible loop) would then be free to interact with the melted single-stranded DNA to locate the transcription start site.

3.

- a) Introduction of the 4 transcription factors *Oct3/4*, *Sox2*, *Klf3*, *c-Myc* into somatic cells elicits a 5-fold or more difference in the expression of 5,107 genes. Additionally, they noted that the histone methylation state of a variety of development-related genes (*Oct3/4*, *Sox2*, *Nanog*, *Gata6*, *Msx2*, *Pax6*, *Hand1*) in iPS cells was similar to that in hES cells. These suggest that the 4 TFs induce a large-scale cascading effect that drastically changes gene regulation and expression involving a substantial portion of the genome.
- b) My background is in genomics and bioinformatics, so my suggestions are definitely biased in that direction!
- i. The first approach I would use would be to do a Gene Ontology (GO) analysis of the 5,107 genes that are differentially expressed between somatic cells and pluripotent cells. GO analysis involves an enrichment analysis to identify functional annotations (molecular function annotations, biological process annotations, cellular component annotations) that are over-represented or under-represented in a subset (the 5,107 genes) in comparison to the whole (all 30k genes). The results from this analysis will not be fine-tuned, but trends in the data may provide some insights into the types of biological changes that are associated with the 4 transcription factors in question.
 - ii. The second approach I would use would be a ChIP-seq analysis. Using antibodies specific to one of the TFs, I would isolate the TFs from cell extract and sequence the DNA to which the TF was bound. This would enable identification of the sequence (or types of sequences) to which that TF binds, and subsequent bioinformatic analysis (such as motif finding). We could also perform directed mutation to identify which amino acids in each TF are required for specificity and/or affinity. This approach would help us characterize the TFs in terms of their interaction with the genomic DNA.
 - iii. Perhaps the most powerful (and most ambitious) approach I would use involves a systems biology and network analysis approach. I would introduce the 4 TFs into the somatic cell and then begin taking regular RNA extractions in order to derive time-series gene expression data. Given sufficient data, it is possible to infer a regulatory network from the expression data and determine which genes activate or suppress other genes. This would give a much broader picture of the function of the TFs and the genes with which they interact both directly and indirectly. Network analysis can also identify a variety of informative features of the regulatory network, such as important subnetworks and highly connected nodes (genes that interact with many other genes).