toid cell lines also applies to erythroid cells, then the causative SNP either has a non–tissue-specific effect on transcription, or it acts at the posttranscriptional level, an issue that remains to be resolved.

Although the results are consistent with the long isoforms of BCL11A functioning to suppress HbF production in a dose-dependent manner, they do not distinguish between a direct effect on γ-globin gene expression and an indirect effect on the kinetics of erythropoiesis that might cause cells containing HbF to be overproduced (9). Additional findings of Sankaran et al. argue strongly in favor of the former model. For example, BCL11A was found to interact with multiple components of the nucleosome remodeling and histone deacetylase (NuRD) complex, which functions in transcriptional repression (10), as well as with GATA-1 and FOG-1, the major activating transcription factor and cofactor, respectively, of the erythroid lineage (11). Further, RNA interference specific for BCL11A in primary erythroid progenitors increased γ-globin mRNA expression and HbF production without inducing a generalized effect on cellular differentiation or altering expression of known globin gene transcription factors. Finally, chromatin immunoprecipitation revealed that BCL11A binds to multiple regions within the human β-globin locus that control silencing of the γ-globin gene (7).

Thus, BCL11A plays an important role in fetal-to-adult hemoglobin switching during normal erythropoiesis, and its expression in adult erythroid cells affects the amount of HbF produced. This raises questions about what controls the stage-specific shift between the long and short BCL11A isoforms, and the molecular mechanism by which BCL11A represses γ-globin gene expression. It is also not clear which genetic variant in BCL11A sets the expression level of this gene, nor how it functions in this capacity. Higher-resolution studies of BCL11A chromatin occupancy, functional characterization of the putative cis-regulatory elements containing these binding sites, and sequencing of the entire region of BCL11A that is in linkage disequilibrium with the SNP used to discover its relevance to HbF expression are needed to address these questions.

The findings of Sankaran et al. also have potential consequences for developing new therapies to treat sickle cell disease and other hemoglobinopathies. The inverse correlation between BCL11A and HbF expression, combined with the known ameliorative effect of HbF on the pathophysiology of sickle cell disease and β-thalassemia, suggests that inhibition of BCL11A expression or function could be an effective treatment for these disorders. The study also illustrates that, when experiments are appropriately designed, the initial findings of genome-wide association studies can be successfully followed up at a functional level.

References
3. V. G. Sankaran et al., Science 322, 1839 (2008); published online 4 December 2008 (10.1126/science.1165409).
that specify the physiological conditions or cell types where the gene will be expressed. The enhancer and basal promoter complexes functionally and physically interact to determine how often an RNA transcript is produced. The problem with this model is that enhancers can work over large distances of DNA in both directions, whereas basal elements are made up of short, low-complexity sequences (such as the TATA element) that appear frequently by chance. Transcription would likely initiate promiscuously throughout the genome without some way to suppress most transcription start sites.

The major mechanism for suppressing widespread transcription is to sequester potential transcription start sites by wrapping most of the genome in nucleosomes, subunits of chromatin in which DNA is coiled around histone proteins (5). Indeed, “real” transcription start sites are typically found in nucleosome-free regions generated by DNA sequences intrinsically resistant to nucleosome wrapping or by targeted modification and removal of nucleosomes to expose the underlying promoter sequences (6). Because transcription by RNA polymerase II through protein-coding sequences requires nucleosome displacement, eukaryotic cells even have mechanisms to quickly replace repressive nucleosomes in the wake of RNA polymerase II. Loss of this repression leads to cryptic RNAs initiating within transcribed regions (7). Therefore, functional eukaryotic promoters not only must attract RNA polymerase II, but also evade nucleosomal repression.

So what are we to make of these new, promoter-associated transcripts? The simplest view is that they arise from random, weak basal promoter elements that escape suppression. This idea is supported by the fact that the new RNAs are largely derived from DNA in nucleosome-free regions. Interestingly, Preker et al. (3) show that noneukaryotic DNA placed next to a strong mammalian promoter will also produce short divergent transcripts, indicating minimal sequence specificity. He et al. (2) also report antisense transcription—the synthesis of RNA complementary to the protein-coding RNA—initiating near 3′ ends of genes, another region that often has reduced nucleosome occupancy.

Although “backward” transcripts are less abundant than the “forward” (coding, full-length) transcripts produced by nearby transcription start sites, a clear correlation in their expression suggests that both types of RNAs respond to the same inducers of gene expression. Interestingly, the new papers find similar amounts of RNA polymerase II associated with the upstream and downstream transcription start sites. The mystery is why RNA polymerase II traveling in one direction can produce RNAs thousands of nucleotides long, whereas polymerases moving in the opposite direction don’t get very far. However, as noted by Core et al. (1), even at many of the transcription start sites that produce full-length RNAs, much of the RNA polymerase II that starts at the promoter does not effectively make it to the end of the gene. The RNA polymerase II molecules that accumulate near promoters are often referred to as “paused;” but it is unclear whether they await a positive signal to continue elongating (synthesizing RNA) or instead quickly terminate transcription soon after initiation. In either case, it appears that a rate-limiting step is escape of RNA polymerase II into processive, long-range elongation.

Even if the short promoter-associated RNAs simply result from incomplete suppression of cryptic initiation, it would be a mistake to assume that there is no associated function. The RNAs produced from these unanticipated transcription start sites may have some undiscovered role, but it is perhaps more likely that the act of transcription itself affects expression of the nearby gene. As suggested by several of the new papers, this could be mediated by transcription-coupled changes in the DNA topology or local chromatin structure. Recent studies in yeast have uncovered interesting regulatory relationships between closely spaced transcription start sites. In the case of the SER3 gene, RNA polymerase II initiating at an upstream transcription start site reads through the SER3 promoter to repress synthesis of the full-length messenger RNA (mRNA) (8). At several genes involved in nucleotide biosynthesis, the concentration of available nucleotides influences the choice between several possible transcription start sites. Although a subset of these is used when the cell needs full-length mRNA, the other transcription start sites produce short noncoding transcripts (9). In yeast, these and other cryptic unstable transcripts use an alternative transcription termination pathway that preferentially acts during early elongation (10, 11). This pathway targets cryptic unstable transcripts for rapid degradation by a complex of nucleases called the exosome. Interestingly, Preker et al. report that the mammalian promoter-associated RNAs are also exosome substrates, contributing to their lower levels.

As complete transcriptomes of cells are cataloged at increasingly finer levels of detail, the hope is that we will be able to discern the rules that determine where RNAs are made and how they are processed. However, we should remain open to the idea that expression of the genome may be rather sloppy, with many (perhaps even most) (12) initiation events generating nonproductive transcripts that are rapidly degraded. This “noise” could provide abundant raw material for evolution. A cryptic transcription start site upstream of the “correct” initiation site might produce an RNA with additional protein-coding sequence or altered translation efficiency. A minor transcription start site within a gene could produce a truncated protein variant that is targeted to a different subcellular location. If any of these events provide some selective advantage, over the course of time, the cryptic transcription start site could become an alternative one and eventually the real transcription start site. If a low amount of bidirectional transcription around transcription start sites were harmful, cells probably would have evolved additional mechanisms for further suppression. The prevalent nature of the short promoter-associated transcripts suggests that their synthesis may serve some functional role, but this remains to be proven.

References
1. L. J. Core, J. J. Waterfall, J. T. Lis, Science 322, 1845 (2008); published online 4 December 2008 (10.1126/science.1162228).
Gene Expression--Where to Start?
Stephen Buratowski

Science 322 (5909), 1804-1805.
DOI: 10.1126/science.1168805