

## MIT Open Access Articles

### *Transcriptional Regulation and Its Misregulation in Disease*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Lee, Tong Ihn, and Richard A. Young. "Transcriptional Regulation and Its Misregulation in Disease." *Cell* 152:6 (March 2013), pp. 1237-1251.

**As Published:** <http://dx.doi.org/10.1016/j.cell.2013.02.014>

**Publisher:** Elsevier

**Persistent URL:** <http://hdl.handle.net/1721.1/103908>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-NonCommercial-NoDerivs License





Published in final edited form as:

*Cell*. 2013 March 14; 152(6): 1237–1251. doi:10.1016/j.cell.2013.02.014.

## Transcriptional Regulation and its Misregulation in Disease

Tong Ihn Lee<sup>1</sup> and Richard A. Young<sup>1,2,\*</sup>

<sup>1</sup>Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA

<sup>2</sup>Department of Biology, Massachusetts

### Abstract

The gene expression programs that establish and maintain specific cell states in humans are controlled by thousands of transcription factors, cofactors and chromatin regulators. Misregulation of these gene expression programs can cause a broad range of diseases. Here we review recent advances in our understanding of transcriptional regulation and discuss how these have provided new insights into transcriptional misregulation in disease.

### Introduction

The key concepts of transcriptional control were established half a century ago in bacterial systems (Jacob and Monod, 1961). That pioneering work and many subsequent studies established that DNA binding transcription factors (also known as trans-factors) occupy specific DNA sequences at control elements (cis-elements) and recruit and regulate the transcription apparatus. In eukaryotic systems, there has been extensive study of specific transcription factors and their cofactors, the general transcription apparatus, and various chromatin regulators, leading to a present-day consensus model for selective gene control (Adelman and Lis, 2012; Bannister and Kouzarides, 2011; Bonasio et al., 2010; Conaway and Conaway, 2011; Fuda et al., 2009; Ho and Crabtree, 2010; Roeder, 2005; Spitz and Furlong, 2012; Taatjes, 2010; Zhou et al., 2012b).

Our knowledge of mammalian regulatory elements and the transcriptional and chromatin regulators that operate at these sites has increased considerably in the last decade. There have also been substantial advances in our understanding of the control of large portions of the gene expression program in embryonic stem cells (ESCs) and in a number of more differentiated cell types. In these relatively well-studied cells, for example, it is now understood that a small fraction of the hundreds of transcription factors that are present dominate the control of much of the active gene expression program (Graf, 2011; Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011).

The recent insights into control of cellular gene expression programs have had an important impact on our understanding of misregulation of gene expression in disease. Many different diseases and syndromes, including cancer, autoimmunity, neurological disorders, diabetes, cardiovascular disease and obesity, can be caused by mutations in regulatory sequences and in the transcription factors, cofactors, chromatin regulators and noncoding RNAs that

© 2013 Elsevier Inc. All rights reserved.

\*Correspondence: young@wi.mit.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

interact with these regions. New insights into the global effects of some of these mutations have recently emerged. These insights alter our view of the underlying cause of some diseases, and are the primary focus of this review.

We begin with a brief review of the basic features of human genes and the fundamentals of gene regulation. This leads to a discussion of cellular gene expression programs and the mechanisms involved in global regulation of transcription. We then describe how recent advances in our understanding of the control of gene expression have led to new insights into the mechanisms involved in misregulation of gene expression in various human diseases and disorders.

## Genes and Enhancer Elements

There are a remarkable variety and number of genes that are transcribed into protein-coding and non-coding RNA (ncRNA) species in mammalian cells (Table 1). The human genome is thought to contain approximately 20,000 protein-coding genes and at least as many ncRNA genes (Djebali et al., 2012). Functions have been determined or inferred for many of the protein-coding genes but less is understood about the functions of the ncRNA genes. Many of the ncRNAs contribute to control of gene expression through modulation of transcriptional or post-transcriptional processes (Bartel, 2009; Ebert and Sharp, 2012; Lee, 2012; Orom and Shiekhattar, 2011; Rinn and Chang, 2012; Wright and Ciosk, 2012). For example, the miRNAs, which are the best-studied of the various classes of ncRNAs, fine tune the levels of target mRNAs. Some of the long ncRNAs (lncRNAs) recruit chromatin regulators to specific regions of the genome and thereby modify gene expression and some apparently do not have a function but are simply a product of a transcriptional event that is itself regulatory (Latos et al., 2012).

Transcription factors typically regulate gene expression by binding enhancer elements and recruiting co-activators and RNA polymerase II to target genes (Lelli et al., 2012; Ong and Corces, 2011; Spitz and Furlong, 2012). Multiple transcription factors typically bind in a cooperative fashion to individual enhancers (Panne, 2008) and regulate transcription from the core promoters of nearby or distant genes through physical contacts that involve looping of the DNA between enhancers and the core promoters (Krivega and Dean, 2012). The core promoter elements, which include sites where transcription initiation occurs, can also be bound by certain transcription factors (Dikstein, 2011; Goodrich and Tjian, 2010).

Enhancers can be identified by profiling the locations of key transcriptional regulators genome-wide and testing whether these DNA elements are active in enhancer-reporter vectors, and a large population of embryonic stem cell (ESC) enhancers has been identified in this manner (Chen et al., 2008). Enhancers are occupied by nucleosomes with specific modifications and are sensitive to DNase treatment, and these features can be used to identify putative enhancers when the key transcriptional regulators are not known (Buecker and Wysocka, 2012; Thurman et al., 2012). Approximately 1 million putative enhancers have recently been identified in the human genome by using, in multiple cell types, a variety of high-throughput techniques that detect these features of enhancers (Dunham et al., 2012; Thurman et al., 2012). These putative enhancers provide a resource for identifying regions of the genome where sequence variation may impact factor binding and gene regulation, and thus contribute to disease. Recent studies suggest that a considerable portion of the genetic variation that is associated with disease occurs in these regulatory regions (Maurano et al., 2012).

## Transcriptional Control of Genes

Transcriptional regulation occurs at two interconnected levels: the first involves transcription factors and the transcription apparatus and the second chromatin and its regulators (Figure 1). We briefly discuss the fundamentals of transcriptional control in this order, noting recent advances and reviews where the reader can obtain more detailed information.

Transcription factors can be separated into two classes based on their regulatory responsibilities: control of initiation versus control of elongation (Adelman and Lis, 2012; Fuda et al., 2009; Rahl et al., 2010; Yankulov et al., 1994; Zhou et al., 2012b). This distinction is not absolute, as some transcription factors may contribute to control of both initiation and elongation. Transcription factors typically bind cofactors, which are protein complexes that contribute to activation (coactivators) and repression (corepressors) but do not have DNA-binding properties of their own. Most transcription factors are thought to contribute to transcription initiation and do so by recruiting coactivators. These coactivators include the Mediator complex, P300 and general transcription factors (Juven-Gershon and Kadonaga, 2010; Malik and Roeder, 2010; Sikorski and Buratowski, 2009; Taatjes, 2010). Recent studies have highlighted the importance of Mediator in integrating information from transcriptional activators, repressors, signaling pathways and other regulators during transcription initiation and during the switch to elongation (Berk, 2012; Borggreffe and Yue, 2011; Conaway and Conaway, 2011; Kagey et al., 2010; Kornberg, 2005; Lariviere et al., 2012; Malik and Roeder, 2010; Spaeth et al., 2011; Taatjes, 2010).

Once the recruited RNA polymerase II molecules initiate transcription, they generally transcribe a short distance, typically 20–50bp, and then pause (Figure 1) (Adelman and Lis, 2012). This process is controlled by the pause control factors DSIF and NELF, which are physically associated with the paused RNA polymerase II molecules. The paused polymerases may transition to active elongation through pause release or they may ultimately terminate transcription with release of the small RNA species. Pause release and subsequent elongation occurs through recruitment and activation of P-TEFb (Positive Transcription Elongation Factor b), which phosphorylates the paused polymerase and its associated pause control factors. P-TEFb can be brought to these sites in the form of a large complex called the super elongation complex (SEC) (Luo et al., 2012a; Smith et al., 2011b). Additional complexes, such as PAFc, also contribute to the regulation of elongation (Jaehning, 2010). Transcription factors such as c-Myc stimulate P-TEFb-mediated release of RNA polymerase II from these pause sites and thus contribute to the control of transcription elongation (Rahl et al., 2010).

Recent studies have provided new insights into cofactors that play important roles in DNA loop formation and maintenance, which are key to proper gene control. During transcription initiation, the DNA loop formed between enhancers and core promoter elements is stabilized by cohesin, which is recruited by the Nipbl cohesin-loading protein that is associated with Mediator (Kagey et al., 2010). The cohesin complex has circular dimensions capable of encircling two nucleosome-bound molecules of DNA. Reducing the levels of cohesin or Nipbl has the same adverse effect on transcription as reducing the levels of Mediator, so these cofactors apparently play a similarly important role in gene activity (Kagey et al., 2010). Although cohesin is recruited to active promoters, it also becomes associated with the DNA-binding factor CTCF, which has been implicated in formation of insulator elements. Thus, cohesin is thought to have roles in transcription activation at some genes and in silencing at others (Dorsett, 2011; Hadjur et al., 2009; Parelho et al. 2008; Phillips and Corces, 2009; Schmidt et al., 2010; Seitan and Merkenschlager, 2012; Wendt et al. 2008).

The fundamental unit of chromatin, the nucleosome, is regulated by protein complexes that can mobilize the nucleosome or modify its histone components (Figure 1). Gene activation is accompanied by recruitment of ATP-dependent chromatin remodeling complexes of the SWI/SNF family, which mobilize nucleosomes to facilitate access of the transcription apparatus and its regulators to DNA (Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011). In addition, there is recruitment, by transcription factors and the transcription apparatus, of an array of histone modifying enzymes that acetylate, methylate, ubiquitylate and otherwise chemically modify nucleosomes in a stereotypical fashion across the span of each active gene (Bannister and Kouzarides, 2011; Campos and Reinberg, 2009; Gardner et al., 2011; Rando, 2012; Zhu et al., 2013). These modifications provide interaction surfaces for protein complexes that contribute to transcriptional control. Enzymes that remove these modifications are also typically present at the active genes, producing a highly dynamic process of chromatin modification as RNA polymerase is recruited and goes through the various steps of initiation and elongation of the RNA species.

Repressed genes are embedded in chromatin with modifications that are characteristic of specific repression mechanisms (Beisel and Paro, 2011; Cedar and Bergman, 2012; Jones, 2012; Moazed, 2009; Reyes-Turcu and Grewal, 2012). One type of repressed chromatin, which contains nucleosome modifications generated by the Polycomb complex (e.g., histone H3K27me3), is found at genes that are silent but poised for activation at some later stage of development and differentiation (Orkin and Hochedlinger, 2011; Young, 2011). Another type of repressed chromatin is found in regions of the genome that are fully silenced, such as that containing retrotransposons and other repetitive elements (Feng et al., 2010; Lejeune and Allshire, 2011). The mechanisms that silence this latter set of genes can involve both nucleosome modification (e.g., histone H3K9me3) and DNA methylation.

## Control of Gene Expression Programs

The set of genes that are transcribed largely defines the cell. The gene expression program of a specific cell type includes RNA species from genes that are active in most cells (housekeeping genes) and genes that are active predominantly in one or a limited number of cell types (cell-type-specific genes). In embryonic stem cells (ESCs), for example, at least 60% of the protein-coding genes are transcribed into full length mRNA species, but only a minority are cell-type-specific and thus defining for ESCs (Assou et al., 2007). Mammals contain hundreds and possibly thousands of cell types, and most of these have yet to be studied with respect to the set of transcripts they contain. Thus, the terms “housekeeping” and “cell-type-specific” are relative rather than absolute and have yet to be precisely defined. Furthermore, the “transcriptome” of specific cells, derived from high-throughput sequencing, does not show a distinct boundary between “active” and “silent” genes, but rather a broad distribution of RNA levels that ranges from less than one RNA molecule/gene/cell to millions of RNA molecules/gene/cell, and it is not clear what level is functionally sufficient for each RNA species.

The particular set of transcription factors that are expressed in any one cell type controls the selective transcription of a subset of genes by RNA polymerase II, thereby producing the gene expression program of the cell. Studies of the transcription factors that are key to establishing and maintaining specific cell states suggest that only a small number of the transcription factors that are expressed in cells are necessary to establish cell-type-specific gene expression programs (Figure 2). For example, although more than half of the ~1200 genes encoding transcription factors show some evidence of transcription in ESCs, only a few of these transcription factors are needed to reprogram a broad range of cell types into induced pluripotent stem cells (iPSCs) with features essentially indistinguishable from ESCs (Graf, 2011; Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Yamanaka, 2012; Yeo

and Ng, 2013; Young, 2011). These ESC transcription factors, which include Oct4, Sox2 and Nanog, are expressed at high levels, bind regulatory elements associated with most active ESC genes, are involved in Polycomb-mediated repression of genes that specify other cell types, and positively regulate their own gene expression through interconnected autoregulatory loops (Figure 3)(Young, 2011). Activation of these endogenous interconnected autoregulatory loops may be key to cellular reprogramming by introduction of exogenous transcription factors. Other cell types express cell-type-specific, or lineage-specific, master transcription factors that are likely to share these key properties of the ESC master transcription factors.

Most of the transcription factors that are key to control of cell state and that can act as reprogramming factors are thought to control transcription initiation at the genes they regulate. For example, the ESC transcription factors Oct4 and Nanog bind to the P300 and Mediator coactivators (Chen et al., 2008; Kagey et al., 2010), which can then drive the formation of open chromatin and recruitment of the transcription apparatus. Similarly, many of the transcription factors that can reprogram or trans-differentiate cells, including MyoD, C/EBP $\beta$ , HNF1 $\alpha$ , HNF4  $\alpha$ , BRN2 and GATA4, bind to at least one of these coactivators (Borggreffe and Yue, 2011).

Recent studies have revealed that certain transcription factors can exert a broad effect on the gene expression programs of cells through elongation control (Figure 4). The c-Myc transcription factor can stimulate increased elongation from essentially the entire active gene expression program in diverse cell types (Lin et al., 2012; Nie et al., 2012; Rahl et al., 2010). The transcription factor AIRE functions to expand the set of genes that undergo RNA polymerase II pause release in specialized thymic stromal cells, allowing expression of the broad spectrum of self-antigens necessary to induce immune tolerance (Abramson et al., 2010; Giraud et al., 2012; Oven et al., 2007; Zumer et al., 2011). In hematopoiesis, the TIF1 $\gamma$  transcription factor controls erythroid cell fate by interacting with P-TEFb and regulating transcription elongation at a specific set of target genes (Bai et al., 2010). Development generally appears to be dependent on proper elongation control; the transcription elongation factor Tcea3 (TFIIS) contributes to the ability of ESCs to respond appropriately to differentiation cues (Park et al., 2012) and mutations in the P-TEFb repressor HEXIM cause gross developmental defects (Nguyen et al., 2012).

## Summary of Recent Advances in Gene Regulation

The key themes that have emerged from recent studies in transcriptional control and that are highlighted here are the following. Sequence variation in enhancers plays an important role in misregulation of gene expression and disease. A small number of key transcription factors dominate control of gene expression programs. Some transcription factors regulate transcription initiation while other factors control elongation, and factors that control this latter step can have profound effects on cell state. The Mediator coactivator complex integrates signals from diverse regulators and recruits cohesin complexes to active genes, which in turn contributes to both chromatin looping and gene activity. Diverse chromatin regulators mobilize nucleosomes and dynamically modify nucleosomes during active gene transcription and in gene silencing, and some chromatin regulators are regulated by lncRNAs. These advances in our understanding of sequences involved in gene control, transcriptional circuitry, the transcription apparatus and chromatin regulation have led to new insights into the mechanisms involved in misregulation of gene expression in various human diseases and disorders. We discuss some of these below.

## Misregulated Gene Expression in Disease

Many diseases and syndromes are associated with mutations in regulatory regions and in transcription factors, cofactors, chromatin regulators and noncoding RNAs (Table S1). These mutations can contribute to cancer, autoimmunity, neurological disorders, developmental syndromes, diabetes, cardiovascular disease and obesity, among others. We highlight here several insights into disease mechanisms that have emerged from advances in our understanding of gene regulation.

### Cancer

Recent studies have highlighted the link between disease-associated variants in regulatory DNA and breast cancer (Jiang et al., 2011), prostate cancer (Demichelis et al., 2012), colorectal cancer (Lubbe et al., 2012), renal cancer (Schodel et al., 2012), lung cancer (Liu et al., 2011), nasopharyngeal cancer (Yew et al., 2012) and melanoma (Huang et al., 2013; Horn et al., 2013). The genome instability that is a hallmark of cancer almost certainly contributes to further alter sequences in regulatory regions that can promote tumor progression.

Mutations in transcription factors have long been known to contribute to tumorigenesis, and recent studies indicate that overexpressed oncogenic transcription factors can alter the core autoregulatory circuitry of the cell. The oncogenic transcription factor TAL1, which is overexpressed in approximately half of the cases of T-cell acute lymphoblastic leukemia (T-ALL), forms an interconnected autoregulatory loop with several key transcription factor partners, and this circuitry contributes to the sustained activation of TAL1-regulated oncogenic program (Sanda et al., 2012). Thus, high levels of TAL1 produce a modified autoregulatory circuitry that drives the oncogenic program in T-ALL.

Most tumor cells depend on the transcription factor c-Myc, for their growth and proliferation (Littlewood et al., 2012). *MYC* is the most frequently amplified oncogene and the elevated expression of its gene product is associated with tumor aggression and poor clinical outcome. Elevated levels of c-Myc can promote tumorigenesis in a wide range of tissues. In tumor cells expressing high levels of c-Myc, the transcription factor accumulates in the promoter regions of most active genes, recruits the transcription elongation factor P-TEFb, and causes transcriptional amplification, producing increased levels of transcripts within the cell's gene expression program (Lin et al., 2012; Nie et al., 2012). Thus, rather than binding and regulating a new set of genes when overexpressed, c-Myc amplifies the output of the existing gene expression program (Figure 4). These results suggest that transcriptional amplification reduces rate-limiting constraints for tumor cell growth and proliferation.

Mutations in the Mediator coactivator complex have recently been implicated in the development of various tumors. Uterine leiomyomas, or fibroids, are benign tumors that affect millions of women. The *MED12* gene is altered in the majority of uterine leiomyomas and its expression is absent in many uterine leiomyosarcomas, the malignant counterparts of leiomyomas (Makinen et al., 2011a; Makinen et al., 2011b). *MED12* mutations also occur frequently in prostate cancer (Barbieri et al., 2012). *MED12* is part of the CDK module of the Mediator complex, and the CDK8 subunit of this module has been reported to act as an oncogene in both colon cancer and melanoma (Firestein et al., 2008; Kapoor et al., 2010; Morris et al., 2008). Mediator has roles in gene activation and repression, and can function both in transcription initiation and elongation, so further study is needed to establish how Mediator mutations contribute to these tumors. Alterations in cohesin expression and function have been noted in some cancer cells and there is speculation that cohesin misregulation may also contribute to development of various cancers, but direct evidence for

a role of cohesin in cancer remains to be established (Mannini and Musio, 2011; Xu et al., 2011).

Mutations in a variety of chromatin regulators have been implicated in development of cancer cells, and the normal functions of these regulators provide some clues to the mechanisms involved in altered gene expression. Loss of function mutations in several nucleosome remodeling proteins, including ARID1A, SMARCA4 (BRG1) and SMARCB1 (INI1) are associated with multiple types of cancer (Dawson and Kouzarides, 2012; Hargreaves and Crabtree, 2011; Tsai and Baylin, 2011; Wilson and Roberts, 2011), suggesting that defects in mobilizing nucleosomes near the promoters of active genes are involved. Similarly, various mutations in the Polycomb components EZH2 and SUZ12 and in the DNA methylation apparatus occur in multiple cancers, suggesting that in these instances it is the loss of proper gene silencing that contributes to tumorigenesis (Cedar and Bergman, 2012; Jones, 2012; Margueron and Reinberg, 2011; Mills, 2010). The majority of malignant melanomas overexpress SetDB1, a histone H3K9 methyltransferase that can contribute to gene activation or silencing, and this causes deregulation of HOX genes and accelerates melanoma (Ceol et al., 2011).

Gene fusions with the chromatin regulator MLL in leukemias are now known to alter transcription elongation (Luo et al., 2012b; Marschalek, 2010; Slany, 2009; Smith et al., 2011a). Several translocation partners of MLL are components of a super elongation complex (SEC) that includes P-TEFb and ELL proteins, which have also been shown to control transcription elongation (Lin et al., 2011; Lin et al., 2010; Luo et al., 2012a; Smith et al., 2011b). It is thought that translocation of any of the SEC subunits to the amino-terminal domain of MLL abnormally stabilizes the localization of the SEC at MLL target genes, including HOXA9 and HOXA10, which leads to excessive stimulation of RNA polymerase II into productive elongation at these genomic loci, which in turn contributes to aggressive acute leukaemia.

Specific lncRNAs have recently been implicated in cancer progression. The ANRIL lncRNA mediates transcriptional repression of members of the INK4a/ARF/INK4b locus, which encode tumor suppressors whose repression is associated with various cancers (Aguilo et al., 2011; Popov and Gil, 2010). ANRIL functions by recruiting polycomb repressive complexes 1 and 2 (PRC1 and PRC2) and misregulation of ANRIL may lead to abnormal silencing of tumor suppressors and thus contribute to cancer progression (Kotake et al., 2011; Yap et al., 2010). Interestingly, genome-wide association studies have identified numerous polymorphisms that affect the expression and processing of ANRIL and are associated with increased susceptibility to an increasing variety of disease states, including multiple types of cancer, coronary artery disease and type 2 diabetes (Pasmant et al., 2011; Harismendy et al., 2011).

### Autoimmunity and Inflammation

Mutations in the autoimmune regulator (AIRE) protein cause type I autoimmune polyendocrinopathy syndrome. AIRE is a transcription factor whose role in promoting transcriptional elongation at genes with paused RNA polymerase II in the thymus explains why loss of AIRE function leads to autoimmune disease. Self-reactive T cells are normally eliminated during maturation in the thymus, due to the specialized ability of thymic stromal cells, and in particular medullary epithelial cells (MECs), to transcribe a large repertoire of genes encoding peripheral tissue antigens (Kyewski and Klein, 2006). This ectopic gene expression is controlled in a large part by AIRE, which is expressed almost exclusively in MECs. Mice and humans with an *AIRE* gene defect express only a fraction of the peripheral tissue antigens and develop immune infiltrates and autoantibodies directed at multiple peripheral tissues (Akirav et al., 2011; Gardner et al., 2009; Mathis and Benoist, 2009;



Metzger and Anderson, 2011). Recent studies have shown that AIRE interacts with P-TEFb and influences transcription elongation in primary MECs (Abramson et al., 2010; Giraud et al., 2012; Oven et al., 2007; Zumer et al., 2011). AIRE is physically associated with all the active genes in MECs, but has its greatest effect on genes that do not experience pause release in its absence (Figure 4). These results are consistent with the idea that AIRE causes the release of RNA polymerase II molecules that are nonproductively paused at the promoters of a broad spectrum of genes that are otherwise expressed only in peripheral tissues. Thus in MECs, AIRE's function is to expand the set of genes that undergo RNA polymerase II pause release.

Misregulation of the immune response transcriptional regulator NF- $\kappa$ B has been linked to inflammatory and autoimmune diseases, improper immune development and cancer. NF- $\kappa$ B is found in most cell types and is involved in cellular responses to stimuli such as infection and stress (Hayden and Ghosh, 2012). The transcription factor controls genes involved in inflammation, and is chronically active in inflammatory diseases such as inflammatory bowel disease, arthritis, sepsis, gastritis, asthma and atherosclerosis. Although most research into the mechanism of transcriptional activation by this and other regulators have focused on coactivator recruitment, evidence that NF- $\kappa$ B interacts with BRD4 and P-TEFb suggests that this ubiquitous regulator plays a role in elongation control at inflammatory genes during immune and stress responses (Barboric et al., 2001; Huang et al., 2009; Nowak et al., 2008). This view is supported by evidence that inhibitors of BRD4, which contributes to recruiting active P-TEFb, suppress expression of key inflammatory genes in activated macrophages and confer protection against lipopolysaccharide-induced endotoxic shock and bacteria-induced sepsis (Nicodeme et al., 2010).

### Developmental Disorders: Neurological

Mutations in various components of the Mediator coactivator have been linked to a variety of neurological disorders and other developmental deficiencies (Ding et al., 2008; Goh and Grants, 2012; Hashimoto et al., 2011; Kaufmann et al., 2010; Leal et al., 2009; Philibert et al., 2007; Risheg et al., 2007; Rump et al., 2011; Schwartz et al., 2007; Zhou et al., 2012a). Mutations in MED23 alter the interaction between enhancer-bound transcription factors and Mediator, leading to transcriptional dysregulation of mitogen-responsive immediate-early genes that affect brain development and plasticity. A similar defect in immediate-early gene expression is observed in cells from patients with another intellectual disability, Opitz-Kaveggia syndrome, which is caused by MED12 mutations. It would not be surprising to find that additional Mediator mutations contribute to neurological disorders, given the role of this coactivator in integrating information from transcriptional activators, repressors, and signaling pathways.

Heterozygous germline mutations in components of the SWI/SNF chromatin remodeling complex were recently identified in patients with various neurological syndromes whose common features are severe intellectual disability and speech delay (Hoyer et al., 2012; Santen et al., 2012a; Santen et al., 2012b; Tsurusaki et al., 2012; Van Houdt et al., 2012). These mutations were found in *SMARCB1*, *SMARCA4*, *SMARCA2*, *SMARCE1*, *ARID1A* and *ARID1B*. It has been suggested that that up to 3% of unexplained intellectual disability may be caused by mutations in genes encoding SWI/SNF components (Santen et al., 2012b). It is interesting to note that ARID1B component of human SWI/SNF interacts with elongin C (Li et al., 2010), a component of the SIII transcription elongation factor, which enhances transcription elongation by suppressing transient pausing of RNA polymerase II (Aso et al., 1995). Thus, alterations in SWI/SNF complexes have the potential to affect both chromatin remodeling and transcription elongation.

## Developmental Disorders: Cohesinopathies

Cohesinopathies are characterized by a wide variety of developmental defects, including growth and mental retardation, limb deformities, and craniofacial anomalies (Bose and Gerton, 2010; Liu and Krantz, 2008). This broad spectrum of phenotypes is now thought to be due to reduced cohesin loading and cohesin function in gene expression during development. A variety of cohesinopathies have been described, including Cornelia de Lange Syndrome and Roberts Syndrome, in which patients have mutations in the cohesin loading protein NIPBL or the proteins that constitute the cohesin complex. With recent evidence for roles of cohesin complexes in regulation of gene expression and DNA looping (Kagey et al., 2010; Kawauchi et al., 2009; Liu et al., 2009; Schaaf et al., 2009; Seitan et al., 2011), it has become apparent that these deficiencies lead to defects in transcriptional regulation and probably to the overall structure of chromatin in the nucleus of disease cells.

## Diabetes

Diabetes mellitus is a group of metabolic diseases in which a person has elevated blood sugar, either because the pancreas fails to produce adequate amounts of insulin, or because cells do not respond properly to the insulin that is produced. Mutations in pancreatic master transcription factors and the sequences they bind have been implicated in diabetes. The gene expression programs of pancreatic cells appear to be controlled by a small set of key transcription factors, including HNF1 $\alpha$ , HNF1 $\beta$ , HNF4 $\alpha$ , PDX1 and NeuroD1, some of which contribute to the interconnected autoregulatory circuitry of these cells (Odom et al., 2004). Mutations in any of these factors can result in various forms of maturity-onset diabetes of the young (MODY) (Maestro et al., 2007; Malecki, 2005). These mutations almost certainly have a deleterious effect on the interconnected autoregulatory circuitry formed by these factors and their target genes. The frequency of single nucleotide polymorphisms (SNPs) that are linked to defects in glucose homeostasis and diabetes is greatly enriched in the binding sites for these transcription factors (Maurano et al., 2012). This observation indicates that perturbations that affect the regulatory circuitry of pancreatic cells may contribute to diabetes. It also suggests that previously undiscovered regulatory networks and network architectures may be uncovered by incorporating information about disease-associated genetic variants and knowledge of the binding sites of diverse transcription factors.

## Cardiovascular Disease

Misregulated development of the cardiovascular system is among the most common class of congenital birth defects and diseases of the cardiovascular system are among the most prevalent clinical issues for adult populations (Bruneau, 2008; Kathiresan and Srivastava, 2012; Roger et al., 2012). It is well-established that loss of function mutations in certain transcription factors cause various cardiovascular deficiencies (Table S1), but new studies have highlighted the role that mutations in ncRNA species can play in cardiovascular diseases. Specific miRNAs have been implicated in both the promotion and inhibition of differentiation into cardiac lineages, cardiac hypertrophy, vascular differentiation and erythropoiesis (Han et al., 2011; Papageorgiou et al., 2012; Small and Olson, 2011). MicroRNAs have also been linked to causative and protective roles for multiple types of cardiovascular disease, including arrhythmia, fibrosis, hypertrophy due to high pressure and misregulation of cardiac energy metabolism (Callis et al., 2009; Care et al., 2007; Grueter et al., 2012; Luo et al., 2008; Thum et al., 2008; van Rooij et al., 2007; van Rooij et al., 2008; Yang et al., 2007). MicroRNAs are thought to fine-tune gene expression and thus the alterations in these cases are thought to lead to deficiencies in fine-tuning the cardiovascular gene expression program.

## Summary and Outlook

Several concepts have emerged from recent studies of gene expression programs in healthy and in disease cells. Genetic variation may contribute to disease largely through misregulation of gene expression. Mutations in the transcription factors that control cell state may impact the autoregulatory loops that are at the core of cellular regulatory circuitry, leading to the loss of a normal healthy cell state. Some transcription factors control RNA polymerase II pause release and elongation and when their expression or function is altered, can produce aggressive tumor cells (c-Myc) or some forms of autoimmunity (AIRE). Mutations in the coactivator complexes that integrate information from many transcription factors and contribute to DNA looping can cause a broad spectrum of developmental diseases. Alterations in specific chromatin regulators can contribute to development of cancer and many other diseases. Misregulation of noncoding RNAs can also contribute to disease. Additional insights into the role of transcriptional misregulation in human disease will require improved genome annotation, knowledge of the DNA sequences whose alterations contribute to disease, identification of the key transcriptional regulators of all cells of medical relevance, and further understanding of the roles of cofactors, chromatin regulators and ncRNAs.

It is essential to improve human genome annotation in order to more fully understand gene expression programs and their regulation, and thus gene misregulation in disease. As a first step, it would be ideal to identify the complete set of protein-coding and non-coding genes and to ascertain which of these are actively transcribed in specific cell types. There are considerable challenges associated with defining the complete set of expressed genes in any one type of mammalian cell. Such characterization has traditionally required large numbers of cells and for most primary cell types, it is challenging to obtain a homogeneous population of cells. While protein-coding genes can be recognized, at least in part, by the presence of a coding sequence, it is challenging to produce a complete and accurate annotation of ncRNA genes due to limitations in the read length of widely used sequence technologies and the short lifetime of many ncRNAs. Nonetheless, recent studies have identified a vast number and variety of ncRNAs in human cells, so there is promise that improved human genome annotation is at hand (Djebali et al., 2012). Furthermore, new technologies allow investigators to monitor RNA polymerase II molecules that are actively engaged in transcription (Core et al., 2008). Such approaches have recently provided evidence that most long ncRNA (lncRNA) species are the product of divergent transcription from the promoters of active protein coding genes (Sigova et al., 2012). This suggests that most active protein-coding genes in humans are actually divergently transcribed mRNA/lncRNA gene pairs.

Knowledge of the sequence variation that contributes to disease is being gained at a rapid pace and this will improve our understanding of disease mechanisms and lead to new approaches to disease diagnosis and therapy. Several lines of evidence suggest that much of genetic variation contributes to disease through misregulation of gene expression. A substantial portion of the genomic sequences that are under positive selection are thought to be regulatory (Grossman et al., 2010). Disease-associated single nucleotide polymorphisms (SNPs) are enriched in regulatory regions (Ernst et al., 2011; Hindorff et al., 2009; Maurano et al., 2012). Many recent studies have identified links between disease-associated variants in regulatory DNA and a broad spectrum of human diseases, including cancer (Demichelis et al., 2012; Huang et al., 2013; Horn et al., 2013; Jiang et al., 2011; Liu et al., 2011; Lubbe et al., 2012; Schodel et al., 2012; Yew et al., 2012), congenital heart disease (Zhao et al., 2012), inflammatory lung disease (Han et al., 2012), multiple sclerosis (Alcina et al., 2012), Alzheimer's disease (Gaj et al., 2012), abdominal aortic aneurysm (Bown et al., 2011), amyotrophic lateral sclerosis (Iida et al., 2011) and coronary artery disease (Harismendy et

al., 2011). Disease-associated genetic variants in regulatory regions are most often found in regions that are utilized in a cell type-specific manner and are associated with diseases of the corresponding cell type (Ernst et al., 2011; Maurano et al., 2012). Thus, cell type-specific enhancer use can explain how genetic variants produce tissue-specific diseases. Disease-associated genetic variants can exist in regulatory regions that are very distant from the genes they control, but knowledge of the nature of loops between such distal enhancers and their target genes can explain how these distant variants affect a specific gene and its biological functions (Li et al., 2012; Maurano et al., 2012).

Genetic investigations and reprogramming studies suggest that only a small number of the hundreds of transcription factors that are expressed in cells are essential for establishing and maintaining the regulatory networks that produce specific cell states. If this holds true for most cell types, then it would be ideal to identify the key transcription factors for all cell types of medical relevance. It should be possible to identify these transcription factors if they have features identified for their counterparts in well-studied cells: relative high expression, occupancy of enhancers associated with a large fraction of active genes, and formation of interconnected autoregulatory loops. Discovering how gene expression programs are controlled in many different cell types should lead to further understanding of regulatory circuitry, facilitate cellular reprogramming, and accelerate the new field of regenerative medicine.

Cofactors and chromatin regulators are generally expressed in most cell types, but mutations in these genes often produce diseases or syndromes that exhibit tissue-specific disease phenotypes (Table S1). For example, defects in Mediator subunits contribute to nonsyndromic intellectual disability, Charcot-Marie-Tooth disease, Opitz-Kaveggia and Lujan syndromes, infantile cerebral and cerebellar atrophy and have been implicated in prostate cancer, in deficiencies in systemic energy homeostasis (Grueter et al., 2012) and in altered hair-cycling (Nakajima et al., 2012; Oda et al., 2012). Improved understanding of the interactions between cofactors and transcription factors and the mechanisms involved in information integration by these complex apparatuses will be valuable for understanding the mechanisms that produce tissue-specific phenotypes. Similarly, it will be important to further understanding the collaboration between the transcription apparatus and chromatin regulators in global control of gene expression programs. Recent efforts to target chromatin regulators for cancer therapy (Dawson et al., 2012) would benefit from a fuller understanding of the regulatory mechanisms and pathways that are impacted by these potential therapeutics.

Our future understanding of disease and the advance of personalized medicine will benefit from models of human transcriptional regulatory circuitry that integrate information about regulatory sequences and the key transcription factors, cofactors, chromatin regulators and ncRNAs that operate at regulatory sites. The development of these models should thus be among the priorities of biomedical research.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Our description of the themes highlighted in this review benefitted from discussions with Karen Adelman, Jay Bradner, Gerald Crabtree, Rudolf Jaenisch, Ian Krantz, Lee Lawton, David Levens, John Lis, Alex Marson, Matthias Merckenschlager, Alan Mullen, Duncan Odom, David Price, Peter Rahl, Robert Roeder, Ali Shilatifard, Phil Sharp, Alla Sigova, Alexander Stark, Dylan Taatjes and Leonard Zon. We thank David Orlando for help with data collation and analysis.

## References

- Abramson J, Giraud M, Benoist C, Mathis D. Aire's partners in the molecular control of immunological tolerance. *Cell*. 2010; 140:123–135. [PubMed: 20085707]
- Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet*. 2012; 13:720–731. [PubMed: 22986266]
- Aguilo F, Zhou MM, Walsh MJ. Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. *Cancer Res*. 2011; 71:5365–5369. [PubMed: 21828241]
- Akirav EM, Ruddle NH, Herold KC. The role of AIRE in human autoimmune disease. *Nat Rev Endocrinol*. 2011; 7:25–33. [PubMed: 21102544]
- Alcina A, Fedetz M, Fernandez O, Saiz A, Izquierdo G, Lucas M, Leyva L, Garcia-Leon JA, del Abad-Grau MM, Alloza I, et al. Identification of a functional variant in the KIF5A-CYP27B1-METTL1-FAM119B locus associated with multiple sclerosis. *J Med Genet*. 2013; 50:25–33. [PubMed: 23160276]
- Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*. 2006; 442:203–207. [PubMed: 16751777]
- Aso T, Lane WS, Conaway JW, Conaway RC. Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. *Science*. 1995; 269:1439–1443. [PubMed: 7660129]
- Assou S, Le Carrouer T, Tondeur S, Strom S, Gabelle A, Marty S, Nadal L, Pantesco V, Reme T, Hugnot JP, et al. A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells*. 2007; 25:961–973. [PubMed: 17204602]
- Bai X, Kim J, Yang Z, Juryec MJ, Akie TE, Lee J, LeBlanc J, Sessa A, Jiang H, DiBiase A, et al. TIF1gamma controls erythroid cell fate by regulating transcription elongation. *Cell*. 2010; 142:133–143. [PubMed: 20603019]
- Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011; 21:381–395. [PubMed: 21321607]
- Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stransky N, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*. 2012; 44:685–689. [PubMed: 22610119]
- Barboric M, Nissen RM, Kanazawa S, Jabrane-Ferrat N, Peterlin BM. NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol Cell*. 2001; 8:327–337. [PubMed: 11545735]
- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009; 136:215–233. [PubMed: 19167326]
- Beisel C, Paro R. Silencing chromatin: comparing modes and mechanisms. *Nat Rev Genet*. 2011; 12:123–135. [PubMed: 21221116]
- Berk AJ. Yin and yang of mediator function revealed by human mutants. *Proc Natl Acad Sci U S A*. 2012; 109:19519–19520. [PubMed: 23184968]
- Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. *Science*. 2010; 330:612–616. [PubMed: 21030644]
- Borggreffe T, Yue X. Interactions between subunits of the Mediator complex with gene-specific transcription factors. *Semin Cell Dev Biol*. 2011; 22:759–768. [PubMed: 21839847]
- Bose T, Gerton JL. Cohesinopathies, gene expression, and chromatin organization. *J Cell Biol*. 2010; 189:201–210. [PubMed: 20404106]
- Bown MJ, Jones GT, Harrison SC, Wright BJ, Bumpstead S, Baas AF, Gretarsdottir S, Badger SA, Bradley DT, Burnand K, et al. Abdominal aortic aneurysm is associated with a variant in low-density lipoprotein receptor-related protein 1. *Am J Hum Genet*. 2011; 89:619–627. [PubMed: 22055160]
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*. 2007; 128:1089–1103. [PubMed: 17346786]
- Bruneau BG. The developmental genetics of congenital heart disease. *Nature*. 2008; 451:943–948. [PubMed: 18288184]

- Buecker C, Wysocka J. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* 2012; 28:276–284. [PubMed: 22487374]
- Buganim Y, Itskovich E, Hu YC, Cheng AW, Ganz K, Sarkar S, Fu D, Welstead GG, Page DC, Jaenisch R. Direct reprogramming of fibroblasts into embryonic Sertoli-like cells by defined factors. *Cell Stem Cell.* 2012; 11:373–386. [PubMed: 22958931]
- Callis TE, Pandya K, Seok HY, Tang RH, Tatsuguchi M, Huang ZP, Chen JF, Deng Z, Gunn B, Shumate J, et al. MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J Clin Invest.* 2009; 119:2772–2786. [PubMed: 19726871]
- Campos EI, Reinberg D. Histones: annotating chromatin. *Annu Rev Genet.* 2009; 43:559–599. [PubMed: 19886812]
- Care A, Catalucci D, Felicetti F, Bonci D, Addario A, Gallo P, Bang ML, Segnalini P, Gu Y, Dalton ND, et al. MicroRNA-133 controls cardiac hypertrophy. *Nat Med.* 2007; 13:613–618. [PubMed: 17468766]
- Cedar H, Bergman Y. Programming of DNA methylation patterns. *Annu Rev Biochem.* 2012; 81:97–117. [PubMed: 22404632]
- Ceol CJ, Houvras Y, Jane-Valbuena J, Bilodeau S, Orlando DA, Battisti V, Fritsch L, Lin WM, Hollmann TJ, Ferre F, et al. The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature.* 2011; 471:513–517. [PubMed: 21430779]
- Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, Zhang Q, Yan G, Cui Q. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* 2013; 41:D983–986. [PubMed: 23175614]
- Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell.* 2008; 133:1106–1117. [PubMed: 18555785]
- Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. *Annu Rev Biochem.* 2009; 78:273–304. [PubMed: 19355820]
- Conaway RC, Conaway JW. Function and regulation of the Mediator complex. *Curr Opin Genet Dev.* 2011; 21:225–230. [PubMed: 21330129]
- Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science.* 2008; 322:1845–1848. [PubMed: 19056941]
- Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell.* 1987; 51:987–1000. [PubMed: 3690668]
- Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell.* 2012; 150:12–27. [PubMed: 22770212]
- Dawson MA, Kouzarides T, Huntly BJ. Targeting epigenetic readers in cancer. *N Engl J Med.* 2012; 367:647–657. [PubMed: 22894577]
- Demichelis F, Setlur SR, Banerjee S, Chakravarty D, Chen JY, Chen CX, Huang J, Beltran H, Oldridge DA, Kitabayashi N, et al. Identification of functionally active, low frequency copy number variants at 15q21.3 and 12q21.31 associated with prostate cancer risk. *Proc Natl Acad Sci U S A.* 2012; 109:6686–6691. [PubMed: 22496589]
- Dikstein R. The unexpected traits associated with core promoter elements. *Transcription.* 2011; 2:201–206. [PubMed: 22231114]
- Ding N, Zhou H, Esteve PO, Chin HG, Kim S, Xu X, Joseph SM, Friez MJ, Schwartz CE, Pradhan S, et al. Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol Cell.* 2008; 31:347–359. [PubMed: 18691967]
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. Landscape of transcription in human cells. *Nature.* 2012; 489:101–108. [PubMed: 22955620]
- Dorsett D. Cohesin: genomic insights into controlling gene transcription and development. *Curr Opin Genet Dev.* 2011; 21:199–206. [PubMed: 21324671]
- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489:57–74. [PubMed: 22955616]

- Ebert MS, Sharp PA. Roles for microRNAs in conferring robustness to biological processes. *Cell*. 2012; 149:515–524. [PubMed: 22541426]
- Ernst J, Kheradpour P, Mikkelson TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011; 473:43–49. [PubMed: 21441907]
- Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011; 12:861–874. [PubMed: 22094949]
- Feng S, Jacobsen SE, Reik W. Epigenetic reprogramming in plant and animal development. *Science*. 2010; 330:622–627. [PubMed: 21030646]
- Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I, Freed E, Ligon AH, Vena N, Ogino S, et al. CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature*. 2008; 455:547–551. [PubMed: 18794900]
- Fuda NJ, Ardehali MB, Lis JT. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*. 2009; 461:186–192. [PubMed: 19741698]
- Gaj P, Paziewska A, Bik W, Dabrowska M, Baranowska-Bik A, Styczynska M, Chodakowska-Zebrowska M, Pfeffer-Baczk A, Barcikowska M, Baranowska B, et al. Identification of a late onset Alzheimer's disease candidate risk variant at 9q21.33 in Polish patients. *J Alzheimers Dis*. 2012; 32:157–168. [PubMed: 22785395]
- Gardner JM, Fletcher AL, Anderson MS, Turley SJ. AIRE in the thymus and beyond. *Curr Opin Immunol*. 2009; 21:582–589. [PubMed: 19833494]
- Gardner KE, Allis CD, Strahl BD. Operating on chromatin, a colorful language where context matters. *J Mol Biol*. 2011; 409:36–46. [PubMed: 21272588]
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*. 2006; 442:199–202. [PubMed: 16751776]
- Giraud M, Yoshida H, Abramson J, Rahl PB, Young RA, Mathis D, Benoist C. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc Natl Acad Sci U S A*. 2012; 109:535–540. [PubMed: 22203960]
- Goh YS, Grants JM. Mutations in the Mediator subunit MED23 link intellectual disability to immediate early gene regulation. *Clin Genet*. 2012; 81:430. [PubMed: 22129135]
- Goodrich JA, Tjian R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet*. 2010; 11:549–558. [PubMed: 20628347]
- Graf T. Historical origins of transdifferentiation and reprogramming. *Cell Stem Cell*. 2011; 9:504–516. [PubMed: 22136926]
- Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*. 2010; 327:883–886. [PubMed: 20056855]
- Grueter CE, van Rooij E, Johnson BA, DeLeon SM, Sutherland LB, Qi X, Gautron L, Elmquist JK, Bassel-Duby R, Olson EN. A cardiac microRNA governs systemic energy homeostasis by regulation of MED13. *Cell*. 2012; 149:671–683. [PubMed: 22541436]
- Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*. 2009; 460:410–413. [PubMed: 19458616]
- Han M, Toli J, Abdellatif M. MicroRNAs in the cardiovascular system. *Curr Opin Cardiol*. 2011; 26:181–189. [PubMed: 21464712]
- Han YJ, Ma SF, Wade MS, Flores C, Garcia JG. An intronic MYLK variant associated with inflammatory lung disease regulates promoter activity of the smooth muscle myosin light chain kinase isoform. *J Mol Med*. 2012; 90:299–308. [PubMed: 22015949]
- Hargreaves DC, Crabtree GR. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res*. 2011; 21:396–420. [PubMed: 21358755]
- Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature*. 2011; 470:264–268. [PubMed: 21307941]
- Harries LW. Long non-coding RNAs and human disease. *Biochem Soc Trans*. 2012; 40:902–906. [PubMed: 22817756]

- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 2012; 22:1760–1774. [PubMed: 22955987]
- Hashimoto S, Boissel S, Zarhrate M, Rio M, Munnich A, Egly JM, Colleaux L. MED23 mutation links intellectual disability to dysregulation of immediate early gene expression. *Science.* 2011; 333:1161–1163. [PubMed: 21868677]
- Hayden MS, Ghosh S. NF-kappaB, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev.* 2012; 26:203–234. [PubMed: 22302935]
- Hoyer J, Ekici AB, Ende S, Popp B, Zweier C, Wiesener A, Wohlleber E, Dufke A, Rossier E, Petsch C, et al. Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am J Hum Gen.* 2012; 90:565–572.
- HGNC Database, HUGO Gene Nomenclature Committee (HGNC). EMBL Outstation - Hinxton. European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton; Cambridgeshire, CB10 1SD, UK : Dec. 2012 [www.genenames.org](http://www.genenames.org)
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 2009; 106:9362–9367. [PubMed: 19474294]
- Ho L, Crabtree GR. Chromatin remodelling during development. *Nature.* 2010; 463:474–484. [PubMed: 20110991]
- Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, Nagore E, Hemminki K, Schadendorf D, Kumar R. TERT Promoter Mutations in Familial and Sporadic Melanoma. *Science.* 201310.1126/science.1230062
- Huang B, Yang XD, Zhou MM, Ozato K, Chen LF. Brd4 coactivates transcriptional activation of NF-kappaB via specific binding to acetylated RelA. *Mol Cell Biol.* 2009; 29:1375–1387. [PubMed: 19103749]
- Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly Recurrent TERT Promoter Mutations in Human Melanoma. *Science.* 2013 10/1126/science.1229259.
- Huang P, He Z, Ji S, Sun H, Xiang D, Liu C, Hu Y, Wang X, Hui L. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature.* 2011; 475:386–389. [PubMed: 21562492]
- Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, Srivastava D. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell.* 2010; 142:375–386. [PubMed: 20691899]
- Iida A, Takahashi A, Kubo M, Saito S, Hosono N, Ohnishi Y, Kiyotani K, Mushiroda T, Nakajima M, Ozaki K, et al. A functional variant in ZNF512B is associated with susceptibility to amyotrophic lateral sclerosis in Japanese. *Hum Mol Genet.* 2011; 20:3684–3692. [PubMed: 21665992]
- Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol.* 1961; 3:318–356. [PubMed: 13718526]
- Jaehning JA. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta.* 2010; 1799:379–388. [PubMed: 20060942]
- Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 2009; 37:D98–104. [PubMed: 18927107]
- Jiang Y, Shen H, Liu X, Dai J, Jin G, Qin Z, Chen J, Wang S, Wang X, Hu Z. Genetic variants at 1p11.2 and breast cancer risk: a two-stage study in Chinese women. *PLoS One.* 2011; 6:e21563. [PubMed: 21738711]
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012; 13:484–492. [PubMed: 22641018]
- Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol.* 2010; 339:225–229. [PubMed: 19682982]
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature.* 2010; 467:430–435. [PubMed: 20720539]



- Kajimura S, Seale P, Kubota K, Lunsford E, Frangioni JV, Gygi SP, Spiegelman BM. Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-beta transcriptional complex. *Nature*. 2009; 460:1154–1158. [PubMed: 19641492]
- Kapoor A, Goldberg MS, Cumberland LK, Ratnakumar K, Segura MF, Emanuel PO, Menendez S, Vardabasso C, Leroy G, Vidal CI, et al. The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. *Nature*. 2010; 468:1105–1109. [PubMed: 21179167]
- Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell*. 2012; 148:1242–1257. [PubMed: 22424232]
- Kaufmann R, Straussberg R, Mandel H, Fattal-Valevski A, Ben-Zeev B, Naamati A, Shaag A, Zenvirt S, Konen O, Mimouni-Bloch A, et al. Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. *Am J Hum Gen*. 2010; 87:667–670.
- Kawauchi S, Calof AL, Santos R, Lopez-Burks ME, Young CM, Hoang MP, Chua A, Lao T, Lechner MS, Daniel JA, et al. Multiple organ system defects and transcriptional dysregulation in the *Nipbl(+/-)* mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet*. 2009; 5:e1000650. [PubMed: 19763162]
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 2010; 465:182–187. [PubMed: 20393465]
- Kornberg RD. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci*. 2005; 30:235–239. [PubMed: 15896740]
- Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene*. 2011; 30:1956–1962. [PubMed: 21151178]
- Krivega I, Dean A. Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev*. 2012; 22:79–85. [PubMed: 22169023]
- Kywski B, Klein L. A central role for central tolerance. *Annu Rev Immunol*. 2006; 24:571–606. [PubMed: 16551260]
- Lariviere L, Seizl M, Cramer P. A structural perspective on Mediator function. *Curr Opin Cell Biol*. 2012; 24:305–313. [PubMed: 22341791]
- Latos PA, Pauler FM, Koerner MV, energin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, et al. Airn transcription overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science*. 2012; 338:1469–72. [PubMed: 23239737]
- Leal A, Huehne K, Bauer F, Sticht H, Berger P, Suter U, Morera B, Del Valle G, Lupski JR, Ekici A, et al. Identification of the variant Ala335Val of MED25 as responsible for CMT2B2: molecular data, functional studies of the SH3 recognition motif and correlation between wild-type MED25 and PMP22 RNA levels in CMT1A animal models. *Neurogenetics*. 2009; 10:275–287. [PubMed: 19290556]
- Lee JT. Epigenetic regulation by long noncoding RNAs. *Science*. 2012; 338:1435–1439. [PubMed: 23239728]
- Lejeune E, Allshire RC. Common ground: small RNA programming and chromatin modifications. *Curr Opin Cell Biol*. 2011; 23:258–265. [PubMed: 21478005]
- Lelli KM, Slattey M, Mann RS. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu Rev Genet*. 2012; 46:43–68. [PubMed: 22934649]
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*. 2012; 148:84–98. [PubMed: 22265404]
- Li XS, Trojer P, Matsumura T, Treisman JE, Tanese N. Mammalian SWI/SNF--a subunit BAF250/ARID1 is an E3 ubiquitin ligase that targets histone H2B. *Mol Cell Biol*. 2010; 30:1673–1688. [PubMed: 20086098]
- Lin C, Garrett AS, De Kumar B, Smith ER, Gogol M, Seidel C, Krumlauf R, Shilatifard A. Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes Dev*. 2011; 25:1486–1498. [PubMed: 21764852]

- Lin C, Smith ER, Takahashi H, Lai KC, Martin-Brown S, Florens L, Washburn MP, Conaway JW, Conaway RC, Shilatifard A. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol Cell*. 2010; 37:429–437. [PubMed: 20159561]
- Lin CY, Loven J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA. Transcriptional amplification in tumor cells with elevated c-Myc. *Cell*. 2012; 151:56–67. [PubMed: 23021215]
- Littlewood TD, Kreuzaler P, Evan GI. All things to all people. *Cell*. 2012; 151:11–13. [PubMed: 23021211]
- Liu G, Gramling S, Munoz D, Cheng D, Azad AK, Mirshams M, Chen Z, Xu W, Roberts H, Shepherd FA, et al. Two novel BRM insertion promoter sequence variants are associated with loss of BRM expression and lung cancer risk. *Oncogene*. 2011; 30:3295–3304. [PubMed: 21478907]
- Liu J, Krantz ID. Cohesin and human disease. *Annu Rev Genomics Hum Genet*. 2008; 9:303–320. [PubMed: 18767966]
- Liu J, Zhang Z, Bando M, Itoh T, Deardorff MA, Clark D, Kaur M, Tandy S, Kondoh T, Rappaport E, et al. Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol*. 2009; 7:e1000119. [PubMed: 19468298]
- Lubbe SJ, Pittman AM, Olver B, Lloyd A, Vijaykrishnan J, Naranjo S, Dobbins S, Broderick P, Gomez-Skarmeta JL, Houlston RS. The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. *Oncogene*. 2012; 31:3777–3784. [PubMed: 22158048]
- Luo X, Lin H, Pan Z, Xiao J, Zhang Y, Lu Y, Yang B, Wang Z. Down-regulation of miR-1/miR-133 contributes to re-expression of pacemaker channel genes HCN2 and HCN4 in hypertrophic heart. *J Biol Chem*. 2008; 283:20045–20052. [PubMed: 18458081]
- Luo Z, Lin C, Guest E, Garrett AS, Mohaghegh N, Swanson S, Marshall S, Florens L, Washburn MP, Shilatifard A. The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. *Mol Cell Biol*. 2012a; 32:2608–2617. [PubMed: 22547686]
- Luo Z, Lin C, Shilatifard A. The super elongation complex (SEC) family in transcriptional control. *Nat Rev Mol Cell Biol*. 2012b; 13:543–547. [PubMed: 22895430]
- Maestro MA, Cardalda C, Boj SF, Luco RF, Servitja JM, Ferrer J. Distinct roles of HNF1beta, HNF1alpha, and HNF4alpha in regulating pancreas development, beta-cell function and growth. *Endocr Dev*. 2007; 12:33–45. [PubMed: 17923767]
- Makinen N, Heinonen HR, Moore S, Tomlinson IP, van der Spuy ZM, Aaltonen LA. MED12 exon 2 mutations are common in uterine leiomyomas from South African patients. *Oncotarget*. 2011a; 2:966–969. [PubMed: 22182697]
- Makinen N, Mehine M, Tolvanen J, Kaasinen E, Li Y, Lehtonen HJ, Gentile M, Yan J, Enge M, Taipale M, et al. MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science*. 2011b; 334:252–255. [PubMed: 21868628]
- Malecki MT. Genetics of type 2 diabetes mellitus. *Diabetes Res Clin Pract*. 2005; 68(Suppl1):S10–21. [PubMed: 15955369]
- Malik S, Roeder RG. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet*. 2010; 11:761–772. [PubMed: 20940737]
- Mannini L, Musio A. The dark side of cohesin: the carcinogenic point of view. *Mutat Res*. 2011; 728:81–87. [PubMed: 22106471]
- Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature*. 2011; 469:343–349. [PubMed: 21248841]
- Marro S, Pang ZP, Yang N, Tsai MC, Qu K, Chang HY, Sudhof TC, Wernig M. Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell*. 2011; 9:374–382. [PubMed: 21962918]
- Marschalek R. Mixed lineage leukemia: roles in human malignancies and potential therapy. *FEBS J*. 2010; 277:1822–1831. [PubMed: 20236311]
- Mathis D, Benoist C. Aire. *Annu Rev Immunol*. 2009; 27:287–312. [PubMed: 19302042]
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012; 337:1190–1195. [PubMed: 22955828]

- Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. *Cell*. 2012; 148:1172–1187. [PubMed: 22424228]
- Metzger TC, Anderson MS. Control of central and peripheral tolerance by Aire. *Immunol Rev*. 2011; 241:89–103. [PubMed: 21488892]
- Mills AA. Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins. *Nat Rev Cancer*. 2010; 10:669–682. [PubMed: 20865010]
- Mo YY. MicroRNA regulatory networks and human disease. *Cell Mol Life Sci*. 2012; 69:3529–3531. [PubMed: 22926413]
- Moazed D. Small RNAs in transcriptional gene silencing and genome defence. *Nature*. 2009; 457:413–420. [PubMed: 19158787]
- Morris EJ, Ji JY, Yang F, Di Stefano L, Herr A, Moon NS, Kwon EJ, Haigis KM, Naar AM, Dyson NJ. E2F1 represses beta-catenin transcription and is antagonized by both pRB and CDK8. *Nature*. 2008; 455:552–556. [PubMed: 18794899]
- Nakajima T, Inui S, Fushimi T, Noguchi F, Kitagawa Y, Reddy JK, Itami S. Roles of MED1 in Quiescence of Hair Follicle Stem Cells and Maintenance of Normal Hair Cycling. *J Invest Dermatol*. 2012; 133:354–360. [PubMed: 22931914]
- Ng HH, Surani MA. The transcriptional and signalling networks of pluripotency. *Nat Cell Biol*. 2011; 13:490–496. [PubMed: 21540844]
- Nguyen D, Krueger BJ, Sedore SC, Brogie JE, Rogers JT, Rajendra TK, Saunders A, Matera AG, Lis JT, Uguen P, et al. The Drosophila 7SK snRNP and the essential role of dHEXIM in development. *Nucleic Acids Res*. 2012; 40:5283–5297. [PubMed: 22379134]
- Nicodeme E, Jeffrey KL, Schaefer U, Beinke S, Dewell S, Chung CW, Chandwani R, Marazzi I, Wilson P, Coste H, et al. Suppression of inflammation by a synthetic histone mimic. *Nature*. 2010; 468:1119–1123. [PubMed: 21068722]
- Nie Z, Hu G, Wei G, Cui K, Yamane A, Resch W, Wang R, Green DR, Tessarollo L, Casellas R, et al. c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell*. 2012; 151:68–79. [PubMed: 23021216]
- Nowak DE, Tian B, Jamaluddin M, Boldogh I, Vergara LA, Choudhary S, Brasier AR. RelA Ser276 phosphorylation is required for activation of a subset of NF-kappaB-dependent genes by recruiting cyclin-dependent kinase 9/cyclin T1 complexes. *Mol Cell Biol*. 2008; 28:3623–3638. [PubMed: 18362169]
- Oda Y, Hu L, Bul V, Elalieh H, Reddy JK, Bikle DD. Coactivator MED1 ablation in keratinocytes results in hair-cycling defects and epidermal alterations. *J Invest Dermatol*. 2012; 132:1075–1083. [PubMed: 22189783]
- Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, et al. Control of pancreas and liver gene expression by HNF transcription factors. *Science*. 2004; 303:1378–1381. [PubMed: 14988562]
- Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine. Johns Hopkins University; Baltimore, MD: 2012. World Wide Web URL: <http://omim.org/>
- Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet*. 2011; 12:283–293. [PubMed: 21358745]
- Orkin SH, Hochedlinger K. Chromatin connections to pluripotency and cellular reprogramming. *Cell*. 2011; 145:835–850. [PubMed: 21663790]
- Orom UA, Shiekhattar R. Noncoding RNAs and enhancers: complications of a long-distance relationship. *Trends Genet*. 2011; 27:433–439. [PubMed: 21831473]
- Oven I, Brdickova N, Kohoutek J, Vaupotic T, Narat M, Peterlin BM. AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. *Mol Cell Biol*. 2007; 27:8815–8823. [PubMed: 17938200]
- Pang ZP, Yang N, Vierbuchen T, Ostermeier A, Fuentes DR, Yang TQ, Citri A, Sebastiano V, Marro S, Sudhof TC, et al. Induction of human neuronal cells by defined transcription factors. *Nature*. 2011; 476:220–223. [PubMed: 21617644]
- Panne D. The enhanceosome. *Curr Opin Struct Biol*. 2008; 18:236–242. [PubMed: 18206362]

- Parelho V, Hadjir S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*. 2008; 132:422–433. [PubMed: 18237772]
- Park KS, Cha Y, Kim CH, Ahn HJ, Kim D, Ko S, Kim KH, Chang MY, Ko JH, Noh YS, et al. Transcription elongation factor Tcea3 regulates the pluripotent differentiation potential of mouse embryonic stem cells, via the Lefty1- Nodal-Smad2 pathway. *Stem Cells*. 2012;10.1002/stem.1284
- Papageorgiou N, Tousoulis D, Androulakis E, Siasos G, Briasoulis A, Vogiatzi G, Kampoli AM, Tsiamis E, Tentolouris C, Stefanadis C. The role of microRNAs in cardiovascular disease. *Curr Med Chem*. 2012; 19:2605–2610. [PubMed: 22489721]
- Pasmant E, Sabbagh A, Vidaud M, Bieche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J*. 2011; 25:444–448. [PubMed: 20956613]
- Philibert RA, Bohle P, Secrest D, Deaderick J, Sandhu H, Crowe R, Black DW. The association of the HOPA(12bp) polymorphism with schizophrenia in the NIMH Genetics Initiative for Schizophrenia sample. *Am J Med Genet B Neuropsychiatr Genet*. 2007; 144B:743–747. [PubMed: 17299734]
- Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell*. 2009; 137:1194–1211. [PubMed: 19563753]
- Popov N, Gil J. Epigenetic regulation of the INK4b-ARF-INK4a locus: in sickness and in health. *Epigenetics*. 2010; 5:685–690. [PubMed: 20716961]
- Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA. c-Myc regulates transcriptional pause release. *Cell*. 2010; 141:432–445. [PubMed: 20434984]
- Rando OJ. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr Opin Genet Dev*. 2012; 22:148–155. [PubMed: 22440480]
- Reyes-Turcu FE, Grewal SI. Different means, same end-heterochromatin formation by RNAi and RNAi-independent RNA processing factors in fission yeast. *Curr Opin Genet Dev*. 2012; 22:156–163. [PubMed: 22243696]
- Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012; 81:145–166. [PubMed: 22663078]
- Risheg H, Graham JM Jr, Clark RD, Rogers RC, Opitz JM, Moeschler JB, Peiffer AP, May M, Joseph SM, Jones JR, et al. A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. *Nat Genet*. 2007; 39:451–453. [PubMed: 17334363]
- Roeder RG. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett*. 2005; 579:909–915. [PubMed: 15680973]
- Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, et al. Executive summary: heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation*. 2012; 125:188–197. [PubMed: 22215894]
- Ruepp A, Kowarsch A, Theis F. PhenomiR: microRNAs in human diseases and biological processes. *Methods Mol Biol*. 2012; 822:249–260. [PubMed: 22144204]
- Rump P, Niessen RC, Verbruggen KT, Brouwer OF, de Raad M, Hordijk R. A novel mutation in MED12 causes FG syndrome (Opitz-Kaveggia syndrome). *Clin Genet*. 2011; 79:183–188. [PubMed: 20507344]
- Sanda T, Lawton LN, Barrasa MI, Fan ZP, Kohlhammer H, Gutierrez A, Ma W, Tatarek J, Ahn Y, Kelliher MA, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell*. 2012; 22:209–221. [PubMed: 22897851]
- Santen GW, Aten E, Sun Y, Almomani R, Gilissen C, Nielsen M, Kant SG, Snoeck IN, Peeters EA, Hilhorst-Hofstee Y, et al. Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat Genet*. 2012a; 44:379–380. [PubMed: 22426309]
- Santen GW, Kriek M, van Attikum H. SWI/SNF complex in disorder: SWItching from malignancies to intellectual disability. *Epigenetics*. 2012b; 7:1219–1224. [PubMed: 23010866]
- Sayed D, Abdellatif M. MicroRNAs in development and disease. *Physiol Rev*. 2011; 91:827–887. [PubMed: 21742789]

- Schaaf CA, Misulovin Z, Sahota G, Siddiqui AM, Schwartz YB, Kahn TG, Pirrotta V, Gause M, Dorsett D. Regulation of the Drosophila Enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. *PLoS One*. 2009; 4:e6202. [PubMed: 19587787]
- Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res*. 2010; 20:578–588. [PubMed: 20219941]
- Schodel J, Bardella C, Sciesielski LK, Brown JM, Pugh CW, Buckle V, Tomlinson IP, Ratcliffe PJ, Mole DR. Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat Genet*. 2012; 44:420–425. S421–422. [PubMed: 22406644]
- Schwartz CE, Tarpey PS, Lubs HA, Verloes A, May MM, Risheg H, Friez MJ, Futreal PA, Edkins S, Teague J, et al. The original Lujan syndrome family has a novel missense mutation (p. N1007S) in the MED12 gene. *J Med Genet*. 2007; 44:472–477. [PubMed: 17369503]
- Seal RL, Gordon SM, Lush MJ, Wright MW, Bruford EA. genenames.org: the HGNC resources in 2011. *Nucleic Acids Res*. 2011; 39:D514–519. [PubMed: 20929869]
- Seitan VC, Hao B, Tachibana-Konwalski K, Lavagnoli T, Mira-Bontenbal H, Brown KE, Teng G, Carroll T, Terry A, Horan K, et al. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature*. 2011; 476:467–471. [PubMed: 21832993]
- Seitan VC, Merckenschlager M. Cohesin and chromatin organisation. *Curr Opin Genet Dev*. 2012; 22:93–100. [PubMed: 22155130]
- Sekiya S, Suzuki A. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature*. 2011; 475:390–393. [PubMed: 21716291]
- Sigova AA, Mullen AC, Moline B, Gupta S, Orlando DA, Guenther MG, Almada AA, Lin C, Burge CB, Sharp PA, Giallourakis CC, Young RA. Divergent transcription of lncRNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A*. 2012; 109:1221904110
- Sikorski TW, Buratowski S. The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol*. 2009; 21:344–351. [PubMed: 19411170]
- Slany RK. The molecular biology of mixed lineage leukemia. *Haematologica*. 2009; 94:984–993. [PubMed: 19535349]
- Small EM, Olson EN. Pervasive roles of microRNAs in cardiovascular biology. *Nature*. 2011; 469:336–342. [PubMed: 21248840]
- Smith E, Lin C, Shilatifard A. The super elongation complex (SEC) and MLL in development and disease. *Genes Dev*. 2011a; 25:661–672. [PubMed: 21460034]
- Smith ER, Lin C, Garrett AS, Thornton J, Mohaghegh N, Hu D, Jackson J, Saraf A, Swanson SK, Seidel C, et al. The little elongation complex regulates small nuclear RNA transcription. *Mol Cell*. 2011b; 44:954–965. [PubMed: 22195968]
- Spaeth JM, Kim NH, Boyer TG. Mediator and human disease. *Semin Cell Dev Biol*. 2011; 22:776–787. [PubMed: 21840410]
- Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012; 13:613–626. [PubMed: 22868264]
- Taatjes DJ. The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends Biochem Sci*. 2010; 35:315–322. [PubMed: 20299225]
- Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Non-coding RNAs: regulators of disease. *J Pathol*. 2010; 220:126–139. [PubMed: 19882673]
- Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006; 126:663–676. [PubMed: 16904174]
- Thum T, Gross C, Fiedler J, Fischer T, Kissler S, Bussen M, Galuppo P, Just S, Rottbauer W, Frantz S, et al. MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature*. 2008; 456:980–984. [PubMed: 19043405]
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012; 489:75–82. [PubMed: 22955617]

- Tsai HC, Baylin SB. Cancer epigenetics: linking basic biology to clinical medicine. *Cell Res.* 2011; 21:502–517. [PubMed: 21321605]
- Tsurusaki Y, Okamoto N, Ohashi H, Kosho T, Imai Y, Hibi-Ko Y, Kaname T, Naritomi K, Kawame H, Wakui K, et al. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet.* 2012; 44:376–378. [PubMed: 22426308]
- Van Houdt JK, Nowakowska BA, Sousa SB, van Schaik BD, Seuntjens E, Avonce N, Sifrim A, Abdul-Rahman OA, van den Boogaard MJ, Bottani A, et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat Genet.* 2012; 44:445–449. S441. [PubMed: 22366787]
- van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science.* 2007; 316:575–579. [PubMed: 17379774]
- van Rooij E, Sutherland LB, Thatcher JE, DiMaio JM, Naseem RH, Marshall WS, Hill JA, Olson EN. Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc Natl Acad Sci U S A.* 2008; 105:13027–13032. [PubMed: 18723672]
- Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature.* 2010; 463:1035–1041. [PubMed: 20107439]
- Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature.* 2011; 474:390–394. [PubMed: 21572438]
- Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol.* 2011; 21:354–361. [PubMed: 21550244]
- Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature.* 2008; 451:796–801. [PubMed: 18235444]
- Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer.* 2011; 11:481–492. [PubMed: 21654818]
- Wright JE, Ciosk R. RNA-based regulation of pluripotency. *Trends Genet.* 2012; 28:1016–1027. [PubMed: 22101016]
- Xie H, Ye M, Feng R, Graf T. Stepwise reprogramming of B cells into macrophages. *Cell.* 2004; 117:663–676. [PubMed: 15163413]
- Xu H, Tomaszewski M, McKay MJ. Can corruption of chromosome cohesion create a conduit to cancer? *Nat Rev Cancer.* 2011; 11:199–210. [PubMed: 21326324]
- Yamanaka S. Induced pluripotent stem cells: past, present, and future. *Cell Stem Cell.* 2012; 10:678–684. [PubMed: 22704507]
- Yang B, Lin H, Xiao J, Lu Y, Luo X, Li B, Zhang Y, Xu C, Bai Y, Wang H, et al. The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. *Nat Med.* 2007; 13:486–491. [PubMed: 17401374]
- Yankulov K, Blau J, Purton T, Roberts S, Bentley DL. Transcriptional elongation by RNA polymerase II is stimulated by transactivators. *Cell.* 1994; 77:749–759. [PubMed: 8205623]
- Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell.* 2010; 38:662–674. [PubMed: 20541999]
- Yeo JC, Ng HH. The transcriptional regulation of pluripotency. *Cell Res.* 2013; 23:20–32. [PubMed: 23229513]
- Yew PY, Mushiroda T, Kiyotani K, Govindasamy GK, Yap LF, Teo SH, Lim PV, Govindaraju S, Ratnavelu K, Sam CK, et al. Identification of a functional variant in SPLUNC1 associated with nasopharyngeal carcinoma susceptibility among Malaysian Chinese. *Mol Carcinog.* 2012; 51(Suppl 1):E74–82. [PubMed: 22213098]
- Young RA. Control of the embryonic stem cell state. *Cell.* 2011; 144:940–954. [PubMed: 21414485]
- Zhao JY, Yang XY, Gong XH, Gu ZY, Duan WY, Wang J, Ye ZZ, Shen HB, Shi KH, Hou J, et al. Functional variant in methionine synthase reductase intron-1 significantly increases the risk of

congenital heart disease in the Han Chinese population. *Circulation*. 2012; 125:482–490. [PubMed: 22179537]

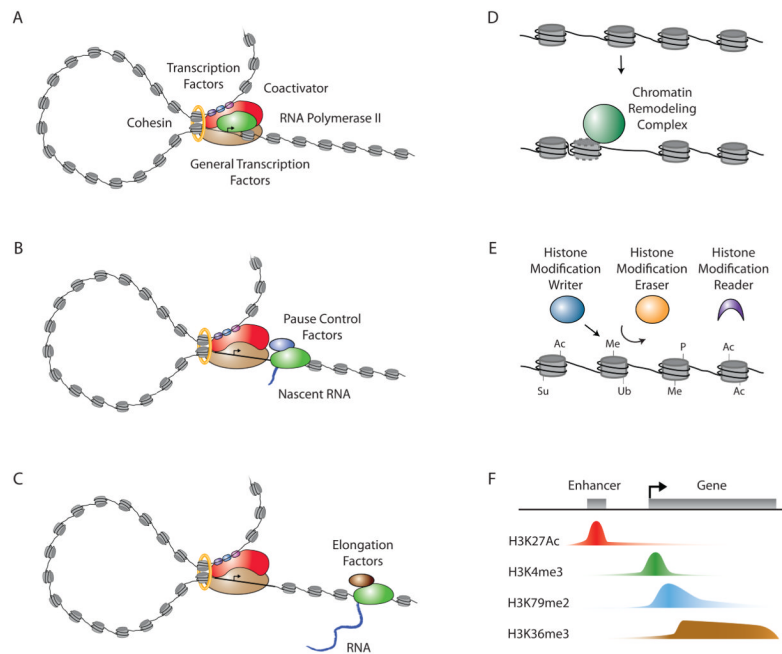
Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature*. 2008; 455:627–632. [PubMed: 18754011]

Zhou H, Spaeth JM, Kim NH, Xu X, Friez MJ, Schwartz CE, Boyer TG. MED12 mutations link intellectual disability syndromes with dysregulated GLI3-dependent Sonic Hedgehog signaling. *Proc Natl Acad Sci U S A*. 2012a; 109:19763–19768. [PubMed: 23091001]

Zhou Q, Li T, Price DH. RNA polymerase II elongation control. *Annu Rev Biochem*. 2012b; 81:119–143. [PubMed: 22404626]

Zhu J, Adli M, Zou JY, Verstappen G, Coyne M, Zhang X, Durham T, Miri M, Deshpande V, De Jager PL, et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*. 2013; 152:1–13.

Zumer K, Plemenitas A, Saksela K, Peterlin BM. Patient mutation in AIRE disrupts P-TEFb binding and target gene transcription. *Nucleic Acids Res*. 2011; 39:7908–7919. [PubMed: 21724609]

**Figure 1.****Transcriptional regulation**

A. Formation of a pre-initiation complex. Transcription factors bind to specific DNA elements (enhancers) and to coactivators, which bind to RNA polymerase II, which in turn binds to general transcription factors at the transcription start site (arrow). The DNA loop formed between the enhancer and the start site is stabilized by cofactors such as the Mediator complex and cohesin.

B. Initiation and pausing by RNA polymerase II. RNA polymerase II begins transcription from the initiation site, but pause control factors cause it to stall some tens of base pairs downstream.

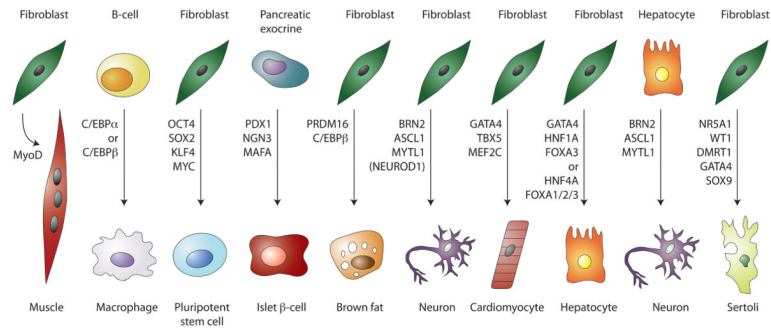
C. Pause release and elongation. Various transcription factors and cofactors recruit elongation factors, including P-TEFb, which phosphorylates the pause release factors and polymerase, allowing elongation to proceed.

D. Chromatin structure is regulated by ATP-dependent remodeling complexes that can mobilize the nucleosome, allowing regulators and the transcription apparatus increased access to DNA sequences.

E. Transcriptional activity is influenced by proteins that modify and bind the histone components of nucleosomes. Some proteins add modifications (writers), some remove modifications (erasers) and others bind via these modifications (readers). The modifications include acetylation (Ac), methylation (Me), phosphorylation (P), sumoylation (Su) and ubiquitination (Ub).

F. Histone modifications occur in characteristic patterns associated with different transcriptional activities. As an example, the characteristic patterns observed at actively transcribed genes are shown for histone H3 lysine 27 acetylation (H3K27Ac), histone H3 lysine 4 trimethylation (H3K4me3), histone H3 lysine 79 dimethylation (H3K79me2) and histone H3 lysine 36 trimethylation (H3K36me3).

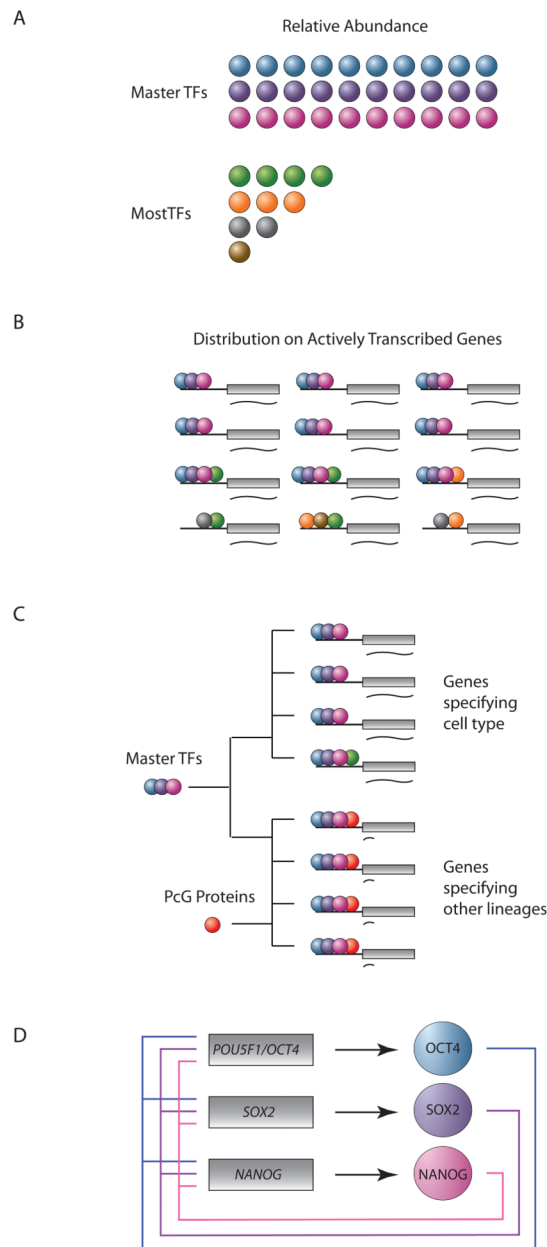




**Figure 2.**

Master transcriptional regulators and reprogramming factors

Transcription factors that have dominant roles in the control of specific cell states and that are capable of reprogramming cell states when ectopically expressed in various cell types (Buganim et al., 2012; Davis et al., 1987; Huang et al., 2011; Ieda et al., 2010; Kajimura et al., 2009; Marro et al., 2011; Pang et al., 2011; Sekiya and Suzuki, 2011; Takahashi and Yamanaka, 2006; Vierbuchen et al., 2010; Xie et al., 2004; Zhou et al., 2008).

**Figure 3.**

Features of master transcription factors of ES cells that likely extend to other cell types

A. Master transcription factors are expressed at high levels relative to other transcription factors (30,000 – 300,000 molecules/cell).

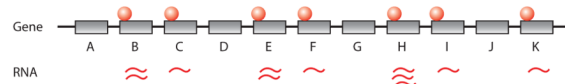
B. Master transcription factors dominate control of the gene expression program by forming enhancers that are associated with most active ESC genes.

C. Master transcription factors positively regulate transcription of cell type-specifying genes and, together with Polycomb group proteins, negatively regulate the expression of genes that specify other cell types.

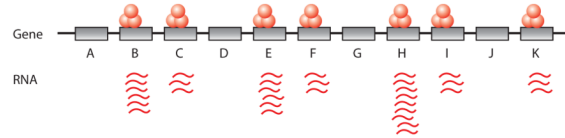
D. Master transcription factors (circles) positively regulate their own genes (boxes), forming interconnected autoregulatory loops.

### Transcription Amplification

Cell Type X (low c-Myc levels)

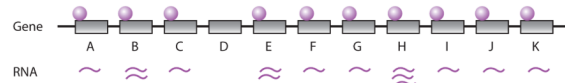


Cell Type X (high c-Myc levels)

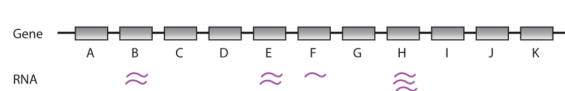


### Expanded Pause Release

Thymic cells with wildtype AIRE

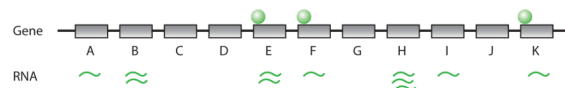


Thymic cells with mutant AIRE

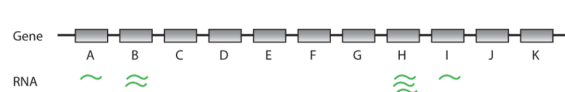


### Specific Pause Release

Erythroid cells with wildtype TIF1gamma



Erythroid cells with mutant TIF1gamma



**Figure 4.**

Global alterations in gene expression programs through transcription elongation.

Top panel: Transcriptional amplification of the gene expression program. The transcription factor c-Myc stimulates increased elongation of most actively transcribed genes, producing increased levels of transcripts for most genes in the gene expression program of the cell.

Middle panel: Expanded pause release extends the gene expression program. In some cells, RNA polymerase will initiate transcription at some genes but fails to transition to elongation. AIRE stimulates pause release for many of these initiated genes, thus producing transcripts for many genes that are normally expressed only in peripheral tissues.

Bottom panel: Specific pause release. Some elongation factors stimulate pause release at specific sets of genes that are important for a particular cell's function.

**Table I****Human Genes**

<b>Class</b>	<b>Abbreviation</b>	<b>Function</b>	<b>Estimated #</b>
Messenger RNA	mRNA	Protein-coding	20,078 <sup>1</sup>
Ribosomal RNA	rRNA	Structural and functional component of ribosome	531 <sup>1</sup>
Transfer RNA	tRNA	Translational adapter molecule	512 <sup>2</sup>
Small nuclear RNA	snRNA	Processing of pre-mRNA	1,923 <sup>1</sup>
Small nucleolar RNA	snoRNA	Processing of rRNA, tRNA and snRNA	1,529 <sup>1</sup>
Antisense RNA	aRNA	Gene regulation	4,424 <sup>1</sup>
Long noncoding RNA	lncRNA	Gene regulation	12,933 <sup>1</sup>
MicroRNA	miRNA	Translational inhibition and mRNA degradation	>500 <sup>3</sup>
Small interfering RNA	siRNA	Post-transcriptional gene silencing	N/A <sup>4</sup>
Piwi-interacting RNA	piRNA	Protect genome integrity	N/A <sup>5</sup>
Enhancer RNA	eRNA	Unknown	Variable <sup>6</sup>

<sup>1</sup>GENCODE (Harrow et al., 2012)

<sup>2</sup>HGNC Database (Seal et al., 2011; HGNC Database)

<sup>3</sup>2000–3000 putative miRNA have been annotated (Harrow et al., 2012), but the majority are not validated.

<sup>4</sup>Human endogenous siRNA are rare and have not been systematically identified.

<sup>5</sup>piRNA have not been systematically identified, although estimates indicate that hundreds of piRNAs are derived from each of over 100 loci (Aravin et al., 2006; Brennecke et al., 2007; Girard et al., 2006).

<sup>6</sup>eRNAs are generated from active enhancers, thus the number of eRNAs depends on the set of active enhancers in a cell (Kim et al., 2010; Wang et al., 2011). Current estimates indicate that 25–80% of active enhancers generate eRNAs.