Transcriptional Addiction in Cancer

James E. Bradner, Denes Hnisz, and Richard A. Young

1Novartis Institutes for Biomedical Research, 181 Massachusetts Ave, Cambridge, MA 02139
2Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA
3Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Summary

Cancer arises from genetic alterations that invariably lead to dysregulated transcriptional programs. These dysregulated programs can cause cancer cells to become highly dependent on certain regulators of gene expression. Here we discuss how transcriptional control is disrupted by genetic alterations in cancer cells, why transcriptional dependencies can develop as a consequence of dysregulated programs, and how these dependencies provide opportunities for novel therapeutic interventions in cancer.

Introduction

Gene dysregulation is a hallmark of cancer. Recent progress in our understanding of transcription and its role in cancer pathogenesis suggest that many new insights will soon be leveraged for patient benefit. Thus, the bulk of the phenotypes, including those affecting their clinical progression and therapeutic responsiveness, are likely to be strongly regulated by the dysregulated versions of transcriptional programs operating within cancer cells. Increasingly, the molecular regulators of these programs, notably proteins involved in transcriptional control, are coming into view as attractive targets of a new generation of drugs that perturb their functions and thus the transcriptional programs that they govern.

By now, extensive cancer genome sequencing studies have revealed recurrent somatic mutations in tumor cells that affect nearly every DNA, RNA and protein component of normal transcriptional control. These findings provide insights into the genes whose alterations influence the cancer state and identify potential therapeutic targets. A number of excellent reviews describe these alterations that affect cell signaling, transcription factors, enhancer elements, chromatin regulators and chromosome structure (Garraway and Lander, 2013; Kandoth et al., 2013; Lawrence et al., 2014; Stratton et al., 2009; Sur and Taipale, 2016; Vogelstein et al., 2013; Watson et al., 2013).

An alternative approach to understanding cancer and identifying therapeutic targets is to discover the key components on which the dysregulated transcriptional programs depend in...
cancer cells (Figure 1). Such transcriptional dependencies are not typically identified by
cancer genome sequencing, but arise through focused mechanistic studies of gene control
programs operating in both normal and neoplastic cells. We describe our current views of the
transcriptional programs operating in normal cells, explain how these programs are altered
in tumor cells, and discuss recent insights into components of transcriptional control on
which certain cancer cells become dependent.

Transcriptional Programs in Normal Cells

Cell identity – more specifically the identity of one or another differentiated cell type – is
controlled in large part by the action of transcription factors (TFs) that recognize and bind
specific sequences in the genome and thereby regulate gene expression. While nearly half of
all the transcription factors encoded in the human genome are expressed in any one cell type
(Vaquerizas et al., 2009), a small number of master TFs, sometimes called lineage
regulators, are sufficient to establish control of the gene expression programs that define cell
identity (Buganim et al., 2013; Graf and Enver, 2009; Lee and Young, 2013; Morris and
Thus, the control of transcriptional programs that characterize normal differentiated cell
states is dominated by these master TFs, which are expressed at high levels in selected cell
types, tend to co-occupy most enhancers together with other master TFs, and typically
regulate their own genes through an autoregulatory loop that forms the core transcriptional
regulatory circuitry of a cell (Figure 2A) (Lee and Young, 2013). The master TFs of any one
cell type can be found at the enhancers of a majority of the active cell-type specific genes,
and may thus account for much of the organization of cell-type specific gene expression
programs.

The master TFs bind cooperatively to enhancer DNA elements and recruit coactivators and
the transcription apparatus (Bulger and Groudine, 2011; Levine et al., 2014; Long et al.,
2016; Malik and Roeder, 2010; Ong and Corces, 2011; Spitz and Furlong, 2012). These TFs
often activate transcription from the enhancer elements themselves (Figure 2B), producing
enhancer RNAs (eRNAs) that bind certain TFs and cofactors and contribute to enhancer
maintenance and dynamics (Lai et al., 2013; Li et al., 2016). Enhancers, which tend to be
cell-type specific because they are generally established by cell-type specific master TFs,
have been mapped in a broad spectrum of human tissue types by using epigenetic marks
associated with enhancer activity (ENCODE Project Consortium et al., 2012; Roadmap
Epigenomics et al., 2015).

Bound by master TFs, clusters of enhancers known as super-enhancers (SEs) regulate genes
that play prominent roles in cell identity or specialized cellular function (Chapuy et al.,
2013; Hnisz et al., 2013; Hnisz et al., 2015; Whyte et al., 2013). Enhancer-associated
proteins and RNAs, including TFs, co-factors such as Mediator, chromatin regulators,
signaling factors, RNA polymerase II (RNAPII), enhancer-associated chromatin marks
(H3K27Ac) and eRNAs, are all found at especially high density at the constituent enhancers
of SEs. The constituent enhancers of SEs physically associate with one another (Figure 2C)
(Dowen et al., 2014; Ji et al., 2016; Kieffer-Kwon et al., 2013) and can function as
independent or interdependent components of these large transcription-regulating complexes.
to drive high-level expression of their associated genes (Hah et al., 2015; Hay et al., 2016; Hnisz et al., 2015; Jiang et al., 2016; Shin et al., 2016).

Enhancers, and super-enhancers, become physically juxtaposed to target gene promoters by looping of the chromatin and, having become so, stimulate transcription from these promoters. Although enhancers can activate any gene, they are physically and functionally constrained to act within insulated neighborhoods (Figure 2D) (Hnisz et al., 2016a). Insulated neighborhoods are chromosomal loop structures formed by the interaction of two DNA sites bound by the CTCF protein and occupied by the cohesin complex. These chromosomal neighborhoods engender specific enhancer-gene interactions and are essential for normal gene activation and repression. The CTCF-CTCF loops that form insulated neighborhoods are the mechanistic basis of higher-order chromosome structures, such as topologically associating domains (TADs), and form a chromosome scaffold that is largely preserved throughout development (Gibcus and Dekker, 2013; Gorkin et al., 2014; Phillips-Cremins and Corces, 2013).

Normal cell states depend on signals received from the tissue microenvironments. Much of this contextual information is delivered by signaling pathways to SEs and, to a lesser extent, to typical enhancers (Figure 2E). SEs have been shown to integrate input from Wnt, TGFβ and LIF signaling pathways operating within embryonic stem cells (ESCs) (Hnisz et al., 2015) and BDNF and KCl signaling at c-Fos in neurons (Joo et al., 2016). This signal integration is thought to be a consequence of the ability of master TFs to recruit signal-activated TFs to enhancer sites previously established by the master TFs (Mullen et al., 2011; Trompouki et al., 2011). The extent to which extracellular and intracellular information is delivered to enhancers is underappreciated, due in part to the paucity of literature on nuclear signaling relative to cytoplasmic sequencing.

Enhancers that are responsive to various types of afferent signaling are thought to be highly dynamic because the activation activity of signaling TFs is linked to their destruction (Figure 2F). For example, activated TGFβ and BMP receptor kinases phosphorylate Smad transcription factors, which become fully functional transcriptional activators after being further phosphorylated by the transcriptional cyclin-dependent kinases CDK8 and CDK9. However, following their initial activation, these phosphorylated Smads are recognized by specific ubiquitin ligases operating in the nucleus, leading to their proteasome-mediated destruction (Alarcon et al., 2009). Remarkably, ubiquitylation and degradation of many transcriptional activators occurs at enhancer/promoter sites and can be required for efficient transcription (Geng et al., 2012; Thomas and Tyers, 2000).

Multiple cyclin-dependent kinases (CDKs), including CDK7, 8, 9, 12 and 13, are dynamic effectors of gene control and function, in part, by phosphorylating serine residues in the C-terminal domain of RNA polymerase II (Eick and Geyer, 2013). For example, CDK7 and CDK9 contribute to control of transcription initiation and elongation, respectively (Jonkers and Lis, 2015). Transcription factors may thus control initiation and/or elongation by their interactions with these CDKs.
DNA is packaged into nucleosomes, which consist of histones that are substrates for chromatin regulators that can modify various amino acid residues or bind in a modification-dependent manner to these histones (Figure 2G). The roles of these diverse chromatin regulators in gene control have been reviewed extensively elsewhere (Campos and Reinberg, 2009; Kouzarides, 2007; Piunti and Shilatifard, 2016; Soshnev et al., 2016; Tessarz and Kouzarides, 2014). The positioning of nucleosomes on DNA can also influence gene control, for example by limiting access of TFs to regulatory sequences, and ATP-dependent remodeling complexes influence transcriptional states by mobilizing nucleosomes (Kadox and Crabtree, 2015). The common functional theme of these regulators is that they facilitate maintenance of positive or negative gene expression states.

DNA methylation contributes to gene control at three levels. Methylation of enhancer and promoter sites can prevent TF binding and thus silences genes (Figure 2H) (Ziller et al., 2013). Methylation of CTCF loop anchor sites prevents CTCF binding and can thus alter insulated neighborhood structure (Ghirlando and Felsenfeld, 2016; Liu et al., 2016b). Cytosine methylation and hydroxymethylation present spatially positioned chemical motifs that can be recognized by chromatin-associated proteins (e.g., meCP2), thereby influencing transcriptional regulation (Baubec et al., 2013; Liu et al., 2016a; Mellen et al., 2012).

In summary, normal cells have transcriptional programs that are established and maintained by master transcription factors that regulate genes by binding specific enhancer elements, which in turn interact with genes within insulated neighborhoods. The maintenance of normal cell states depends on the tissue environment, and such information is delivered by signaling pathways ultimately to enhancers. Maintenance of cell identity and dynamics of cell states also depend on a large number of histone readers, writers and erasers, as well as regulators of DNA methylation, that together ensure chromatin states are appropriate for positive and negative gene regulation.

**Transcriptional dysregulation in cancer**

The complements of genetic alterations that collaborate to transform normal cells into neoplastic derivatives exhibit a high degree of tissue specificity. Tissuesspecific enhancers are structurally altered to drive expression of oncogenes (e.g. TMPRSS2-ERG in prostate cancer, IgH-locus alterations in B-cell malignancies, TCR-locus alterations in T-cell malignancies) (Sur and Taipale, 2016). In addition, many oncogenic signaling pathways require cell-specific chromatin contexts (e.g., NOTCH1 activation in T-cell but not B-cell leukemia, and EZH2 activation in B-cell but not T-cell lymphoma). Because many tissue-specific differentiation programs (specifically those that define cell identity) persist in cancer cells, it is clear that cancer arises from the collaborative interplay of oncogenic events acquired during multi-step tumor formation with the tissue-specifying gene expression programs that survive neoplastic progression and continue to influence cancer cell behavior.

This transcriptional dysregulation arises in cancer from disease-defining genetic alterations either indirectly, via mutation of signaling factors converging on transcriptional control, or directly, via genetic alterations in gene control factors themselves. Cancer-associated genetic alterations can affect proteins participating in nearly all levels of transcriptional control,
including trans-factors (transcription factors, signaling proteins, cofactors, chromatin regulators and chromosome structuring proteins) and cis-elements (enhancers, promoters and insulators) (Figure 3). Many excellent reviews have described an ever-expanding catalogue of these alterations (Bywater et al., 2013; Garraway and Lander, 2013; Kandoth et al., 2013; Lawrence et al., 2014; Stratton et al., 2009; Sur and Taipale, 2016; Vogelstein et al., 2013; Watson et al., 2013). Here we discuss a subset of these alterations, specifically those that lead to the most profound changes in the gene expression program, thereby driving the malignant cell state. We focus on these programs because they may direct the discovery and development of new classes of cancer therapeutics designed to target vulnerabilities of cancer cells, in particular their addiction to certain transcriptional programs.

**Trans-factors**

The TFs that are deregulated in cancer cells and have the potential to produce profound changes in gene expression programs can be considered to fall into three classes: master TFs involved in organizing cell identity, proliferation control TFs that amplify transcriptional output, and signaling TFs involved in dynamic changes in the control machinery occurring in response to extracellular signals. The activation of a master TF that is normally expressed in early development, such as the pluripotency TF OCT4, or a master TF that is normally expressed early in a specific lineage, such as TAL1 in T cells, can alter core regulatory circuitry and activate additional genes that are normally expressed in more embryonic states (Figure 4A) (Sanda et al., 2012). Genes encoding the MYC and TP53 proliferation control TFs, a classic oncogene and tumor suppressor gene, are among the most frequently mutated genes in cancer. MYC can have profound effects because it can function to amplify the entire gene expression program (Figure 4A) (Lin et al., 2012; Nie et al., 2012), and P53 is a powerful tumor suppressor because of its ability to arrest progress through the cell cycle or induce apoptosis (Lane and Levine, 2010). Dysregulation of signaling pathways is a common feature of cancer cells; a dysregulated signaling TF can profoundly change the gene expression program through its binding to enhancers occupied by master TFs (Figure 4A) (Mullen et al., 2011; Trompouki et al., 2011), and dysregulated signaling can even stimulate super-enhancer formation (Brown et al., 2014; Hnisz et al., 2015).

These TFs signal to RNAPII through transcriptional cofactors, defined here as regulatory components that do not bind directly to DNA in a sequence-specific manner. Exemplifying this class of transcriptional signaling proteins are the components of the Mediator complex, which is recruited to enhancer-promoter regions by TFs in the context of transcription activation (Allen and Taatjes, 2015; Kagey et al., 2010). Genetic alterations of Mediator complex-encoding genes are observed frequently in prostate cancer and in many uterine myomas (Allen and Taatjes, 2015; Barbieri et al., 2012; Makinen et al., 2011). Beyond these discrete diseases, the genome-wide activities of Mediator in gene control would be expected to function broadly in all cancer-associated transcription. Interestingly, few cancer-associated alterations are identified in the core RNAPII complex itself (Clark et al., 2016), suggesting that coordinated transcriptional signaling upstream from polymerase favors the neoplastic cell state more than alterations of this core complex.
Efficient transcriptional signaling from enhancers to promoters is often chromatin-dependent, being mediated by specialized transcriptional cofactors that physically associate with and/or biochemically modify the genome to reinforce gene activation or repression. Chromatin regulators function globally, so their dysregulation can also have profound effects on the gene expression program of cells (Jones et al., 2016). In some tumors, chromatin regulators have become fused to transcriptional cofactors, producing gene-specific effects, such as those observed in acute lymphoblastic leukemia cells with MLL-AF4 fusions (Figure 4B) (Guenther et al., 2008; Krivtsov et al., 2008).

Insulated neighborhoods contribute to proper positive and negative gene control, so alterations in chromosome-structuring proteins that establish and maintain insulated neighborhood boundaries would be expected to have profound effects on a cell’s overall gene expression program. Cancer genome sequencing has revealed that somatic mutations occur in CTCF and cohesin coding sequences in various solid tumors and leukemias (Lawrence et al., 2014), and it seems likely that these mutations contribute to oncogenesis by altering insulated neighborhoods throughout the genome, perhaps rendering chromatin generally more permissive to oncogenic transcriptional signaling (Figure 4C) (Viny et al., 2015).

Cis-elements

Cis-regulatory elements within the genome that contribute to cancer pathogenesis were first recognized in pioneering studies of cancer-associated chromosomal translocation [e.g. IgH-MYC in Burkitt’s lymphoma (Taub et al., 1982)]. Following early efforts in cancer genome sequencing, which concentrated on coding regions of the human genome, recent focused and genome-wide sequencing efforts have revealed frequent alteration of cis-elements in both solid and hematological tumors. Two types of cis-elements that play prominent roles in cancer biology - SEs and insulators – are discussed here.

In normal cells, SEs – which, as we noted above are large clusters of enhancers that bind high densities of transcriptional components - control genes that play prominent roles in specific cell identities. Tumor cells acquire SEs at oncogenic driver genes and they do so through many different mechanisms (Figure 5A) (Chapuy et al., 2013; Drier et al., 2016; Hnisz et al., 2013; Kennedy et al., 2015; Loven et al., 2013; Tomazou et al., 2015; Wang et al., 2015; Yang et al., 2015). The genetic mechanisms that lead to SE acquisition in cancer include DNA translocation (Affer et al., 2014; Drier et al., 2016; Groschel et al., 2014; Northcote et al., 2014; Walker et al., 2014), focal amplification (Hnisz et al., 2013; Shi et al., 2013; Zhang et al., 2016b), and nucleation by small INDELs that create master TF binding sites (Mansour et al., 2014). Additional, epigenomic mechanisms also lead to SE formation in cancer, such as those associated with oncogenic TF overexpression (Hnisz et al., 2013), the global function of oncogenic TF fusions (e.g. EWS-FLI) (Kennedy et al., 2015; Tomazou et al., 2015), and the consequences of upstream oncogenic signaling (e.g., RAS-dependent signaling to chromatin) (Nabet et al., 2015).

Mutations that alter insulator sequences of oncogene-containing insulated neighborhoods appear to make important contributions to the dysregulation of gene expression observed in some cancers (Figure 5B) (Flavahan et al., 2016; Hnisz et al., 2016b; Katainen et al., 2015).
Somatic mutations occur frequently and recurrently in loop anchors of oncogene-containing insulated neighborhoods in a broad spectrum of cancer cells. Thus, the CTCF DNA-binding motif in loop anchor regions is among the most altered human transcription factor-binding sequences in cancer cells (Hnisz et al., 2016b; Ji et al., 2016; Katainen et al., 2015). DNA hypermethylation occurs in some cancer cells, and tumor-specific DNA methylation has been implicated in the disruption of CTCF binding, alterations of chromosome structure and dysregulation of oncogene expression in a subset of gliomas (Flavahan et al., 2016). Moreover, chromosomal rearrangements that disrupt insulated neighborhoods can activate oncogenes without altering the sequences of the oncogenes themselves (Groschel et al., 2014; Hnisz et al., 2016b).

**Transcriptional addiction and cancer therapeutics**

The concept of oncogene addiction (Weinstein and Joe, 2006) refers to the behavior of cancer cells that exhibit an absolute dependence on oncogenes that were initially acquired during multi-step tumorigenesis and remain critical to the ongoing proliferation and viability of these cells long after they have progressed to a fully neoplastic state. As of late, this concept has been extended to include other changes acquired during tumor progression that fostered the early development of a tumor and continue to be absolutely essential to its continued growth. Included among these are the dysregulated transcriptional programs operating in certain tumor cells, yielding the concept of transcriptional addiction.

Various types of transcriptional addiction appear to operate in specific subsets of cancer. Thus, the majority of human cancers exhibit genetic amplification or transcriptional dysregulation of MYC, which is accompanied by an anabolic transcriptional response driving proliferation and metabolic adaptation (Beroukhim et al., 2010). Research in model systems of MYC addiction and withdrawal have validated addiction to this master regulatory transcription factor in solid and hematologic malignancies, both with and without structural changes to the MYC locus itself (Felsher and Bishop, 1999; Jain et al., 2002; Soucek et al., 2008; Soucek et al., 2013). As for many TFs, direct pharmacologic inhibition of MYC remains an elusive challenge in drug discovery. We discuss below the strategy of leveraging mechanistic insights into transcriptional dysregulation toward a more immediate therapeutic benefit in cancer.

**Direct inhibition of oncogenic TFs and cofactors**

Gene-targeted therapy has emerged as a paradigm of cancer medicine. Where available and where actionable, somatic alterations of driver oncogenes has provided strong guidance to the discovery and focused development of cancer therapeutics (Darnell, 2002; Pagliarini et al., 2015; Stuart and Sellers, 2009). Among the large number of somatically altered oncogenic transcription factors, only very few have been successfully approached by coordinated efforts in drug discovery. Most compounds of this class are, at best, experimental tools for the study of TFs in cell biology, including our own work to inhibit the NOTCH1 transactivation complex with constrained alpha helical peptides (Moellering et al., 2009). Nonetheless, in the longer term, small molecule agents directed at oncogenic TFs...
have the promise to confer significant clinical responses for patients bearing certain types of genetically defined cancers.

Among the most impactful examples of direct inhibition of an oncogenic TF is the development of all-\textit{trans} retinoic acid (ATRA) as therapy for acute promyelocytic leukemia (APML). ATRA was initially investigated owing to phenotypic evidence of differentiation in cultivated APML cells, an observation made without the guidance of cancer genetics. In 1985, a 5 year-old girl with anthracycline-refractory APML was administered ATRA at Shanghai Children’s Hospital, achieving a complete remission (CR) and ultimately long-term remission from ATRA with chemotherapy (Wang and Chen, 2008). A first case series reported by Zhen-Yi Wang three years later would establish ATRA as a highly effective therapy for APML, conferring CR in 23 of 24 patients as a single agent (Huang et al., 1988). Eleven years after the identification of the t(15;17) translocation by Rowley (Rowley et al., 1977), and seven years after the development of ATRA, Chen and colleagues reported that the reciprocal t(15;17) translocation encodes a novel oncogene comprising the gene encoding the retinoic acid receptor alpha (\textit{RARA}) fused to a second gene (\textit{PML}) specifying a novel Krüppel-like zinc finger protein (Chen et al., 1993). The ensuing decade of laboratory research would firmly establish ATRA as targeted therapy for APML via modulation and destabilization of the encoded chimeric PML-RARA oncogenic transcription factor (Wang and Chen, 2008). Notably, innovation in clinical investigation would ultimately pair ATRA with arsenic trioxide (also associated with PML-RARA degradation) in low- to intermediate-risk APML as effective first-line treatment, equivalent in efficacy to chemotherapy in this disease (Lo-Coco et al., 2013). APML exemplifies biology of transcriptional addiction, revealed through the effective development of transcriptional therapy with curative intent.

An emerging example of direct inhibition of an oncogenic transcriptional cofactor is bromodomain inhibition in carcinomas harboring fusions of BET bromodomain co-activators. Among the most aggressive subtypes of lung and head and neck cancer are tumors expressing the chimeric, oncogenic transcription factors BRD4-NUT or BRD3-NUT (so-called NUT midline carcinoma or NMC). NMC is a poorly differentiated, chemoresistant and aggressive malignancy that lacks effective, FDA-approved therapy. In 2010, we reported the first effective inhibitors of human bromodomains targeting the BET family (BRD2, BRD3 and BRD4), exemplified by the chemical probe JQ1 (Filippakopoulos et al., 2010). Using preclinical models of NMC, JQ1 was shown to displace BRD4 from chromatin, resulting in potent and irreversible squamous differentiation. In murine models harboring primary human NMC xenografts, JQ1 prompted a robust pro-differentiation and anti-proliferative response associated with durable responses by PET-CT imaging. Based on this rationale, drug-like derivatives of JQ1 have been transitioned to clinical investigation by our group and others. Early reports of index trials confirm unambiguous antitumor activity in advanced disease (Stathis et al., 2016).

**Targeting tissue identity and homeostasis**

Oncogenic events occur in the context of cell identity, which is established and maintained by TF-defined core regulatory circuits and signals from the tissue environment (Lee and
Targeting transcriptional identity has thus emerged as an important therapeutic strategy in cancer. Initial efforts have attempted to deplete the tissue type in which the tumor arose, or interfere with the signaling pathways that contribute to cell identity by connecting to cell-type-specific core regulatory circuits.

Tissue depletion may be accomplished surgically, as with excision of neoplastic and surrounding unaffected tissue in mastectomy or radical prostatectomy. Medical approaches also exist, as with anti-CD20 monoclonal antibody therapy or CD19-directed chimeric antigen receptor T-cell (CART) therapy to eradicate B-cell acute lymphoblastic leukemia (B-ALL). A consequence of B-cell depletion therapy is the temporary or permanent depletion of systemic, normal B-cells as a considerable side effect.

In contrast to tissue depletion, direct targeting of cell identity undermines the requisite interaction of oncogenic and cellular circuitry. The best-studied example of this strategy led to the development of nuclear hormone receptor antagonists. Nuclear hormone receptors facilitate differentiation and growth of discrete tissues during development. In post-pubertal adults, the estrogen receptor (ER) and androgen receptor (AR) TFs maintain tissue homeostasis and sex hormone responsive function of breast and prostate tissues, respectively, among other tissues. In low-grade breast and prostate cancer expressing ER and AR, respectively, these TFs are not oncogenes per se, and are genetically unaltered in the majority of patients. Indeed, these tumors are characterized each by diverse genetic alterations to oncogenes and tumor suppressors, the majority of which lack direct-acting agents (Kandoth et al., 2013; Lawrence et al., 2014; Vogelstein et al., 2013). The impactful and important contribution of anti-estrogen and anti-androgen therapy in these diseases, then, defines the opportunity of disrupting transcriptional pathways of cellular identity alone and in combination with oncogene-targeted therapy as available. Further, the broad use of glucocorticoids in lymphoid malignancies exemplifies the same conceptual strategy (Inaba and Pui, 2010), and establishes the feasibility of agonizing a core transcriptional pathway for therapeutic benefit.

**Context-specific transcriptional dependencies**

As discussed earlier, eukaryotic transcription is a dynamic network with multiprotein complexes collaborating as nodes of activating, repressing, remodeling and insulating function. More than a thousand human proteins contribute to gene control at the level of nuclear chromatin, spatially distributed to tens of thousands of sites in the human genome (ENCODE Project Consortium et al., 2012; Levine et al., 2014; Roadmap Epigenomics et al., 2015). Despite this complexity, specific oncogenic impulses can engender exceptional reliance on discrete protein complexes and even individual factors. The identification, validation and prosecution of these targets can yield important mechanistic insights and therapeutic opportunities (Table 1).

Indeed, our research on BET bromodomains began with a mechanistic hypothesis, specifically that BRD4 might mediate transcriptional addiction to MYC. Because MYC had been shown to play a key role in the transcriptional elongation operating at many genes, we hypothesized that BRD4 acted by mediating chromatin-dependent transcription elongation signaling to RNAP II. BRD4 possesses twin acetyl-lysine recognition domains.
(bromodomains), suggesting that BRD4 might localize to hyperacetylated regions of euchromatin occupied in cancer by MYC and associated lysine acetyltransferases. BRD4 binds the pTEFb elongation factor, establishing the possibility that BRD4 may function to facilitate elongation downstream of MYC function in cancer (Bisgrove et al., 2007; Rahl et al., 2010). Indeed, in models of MYC-addicted hematologic cancers, BRD4 localizes with MYC throughout the active genome, and contributes to the MYC-mediated proximal promoter pause release that enables elevated transcriptional elongation (Chapuy et al., 2013; Loven et al., 2013). These mechanistic studies and accompanying translational research establish BRD4 inhibition as a therapeutic strategy to inhibit MYC-dependent transcriptional signaling in multiple myeloma, diffuse large B-cell lymphoma and mixed lineage leukemia (MLL) (Chapuy et al., 2013; Dawson et al., 2011; Delmore et al., 2011; Zuber et al., 2011).

Of relevance to the present subject of targeting transcriptional addiction in cancer, these studies established that pharmacologic inhibition of a transcription factor or co-factor that is presumed to act widely on countless genes throughout the genome can nevertheless exert highly selective effects on gene control. Thus, the mechanistic investigation of this phenomenon identified regions of disproportionately high levels BRD4 occupancy in cis-regulatory regions of many genes that were associated with massive Mediator enrichment, i.e., super-enhancers (Loven et al., 2013). Moreover, BRD4 function has emerged as a transcriptional addiction in MYCN-amplified neuroblastomas (Puissant et al., 2013), a positive regulator of anti-apoptotic gene expression in AMLs (Dawson et al., 2011), and as a mediator of resistance to Notch pathway inhibition in T-ALLs (Yashiro-Ohtani et al., 2014).

Several features of TF gene regulation are thought to contribute to the exceptional sensitivity of large SEs to BRD4 inhibition in tumor cells. TFs and their mRNAs generally have short half-lives, so the genes that encode key TFs may evolve SEs in order to maintain a high transcriptional output of these short-lived proteins (Figure 6A). Many TFs bind and auto-regulate the genes that encode them, so that disruption of TF levels may have an especially pronounced effect on auto-regulated SE control of these genes (Figure 6A). The cooperative features of enhancer components may make SEs especially vulnerable to inhibition of enhancer factors (Figure 6B). Genes encoding tumor cell master TFs acquire especially large SEs, with exceptionally high densities of enhancer factors, and this may further contribute to the exceptional vulnerability of SE-driven oncogenic TFs seen with inhibition of BRD4 (Bhagwat et al., 2016; Chapuy et al., 2013; Loven et al., 2013; Shi and Vakoc, 2014). SE-driven transcription may also be especially sensitive to disruptions in cooperative interactions that contribute to the 3D chromatin architecture at SEs (Dowen et al., 2014; Ji et al., 2016; Kieffer-Kwon et al., 2013).

The transcriptional CDKs have likewise emerged as compelling targets for transcriptional therapy in cancer. CDK7 is a key component of the general transcription initiation apparatus, which governs RNAPII activity by phosphorylating residues on its C-terminal domain. ChIP-seq data indicates that CDK7 densely occupies the SEs that drive expression of oncogenes in a wide variety of cancers, including T-ALL (Kwiatkowski et al., 2014), NSCLC (Christensen et al., 2014), neuroblastoma (Chipumuro et al., 2014), and triple-negative breast cancer (Wang et al., 2015). Small molecule inhibition of CDK7 by THZ1
leads to rapid loss of transcripts for TFs that contribute to oncogenesis and whose expression is driven by SEs in these cells, consistent with the idea that these SE-driven TF genes are especially vulnerable to inhibition of CDK7. The preferential loss of SE-driven TF gene expression, along with expression of genes involved in the DNA damage response, also occurs when tumor cells are treated with inhibitors of CDK12 and CDK13 (Zhang et al., 2016a). Interestingly, inhibition of the CDK8 and CDK19, components of Mediator that contribute to transcriptional repression, leads to hyperactivation of SE-driven genes in AML, and this is as deleterious to the leukemia cells as BRD4 inhibition, which has the opposite effect on these SE-driven genes (Pelish et al., 2015). These and other results show that the leukemia cells are addicted to a specific level of SE-associated gene transcription (Pelish et al., 2015).

The tumor-promoting effect of fusion oncogenes can produce actionable transcriptional dependencies. For example, many AML and MLL leukemias contain oncogenic fusions between MLL and AF genes; the AF part of the fusion product binds the DOT1L histone methyltransferase, which plays important roles in the control of transcriptional elongation by dimethylating histone H3 on lysine 79 (H3K79) (Cai et al., 2015; Deshpande et al., 2013). The MLL-AF fusions thus recruit DOT1L to MLL-target genes, including the TF genes HOXA9 and MEIS1, and the hypermethylation of histones leads to aberrant expression of those genes, which in turn drive leukemogenesis (Figure 6C) (Chen and Armstrong, 2015). These leukemias are especially vulnerable to genetic and chemical perturbation of DOT1L, and DOT1L inhibitors are now in clinical trials for these malignancies (Daigle et al., 2013).

Evidence that tumor cells can develop additional dependencies on specific transcriptional and chromatin regulators, rendering them susceptible to transcriptional drugs, has emerged from studies of drug resistance. Treatment of tumor cells with various anti-cancer agents can select for populations of drug-tolerant cells that have become dependent on specific chromatin regulators such as histone demethylases (KDM5A) and deacetylases (HDACs) (Figure 6D) (Sharma et al., 2010). The drug-tolerant tumor cells can in turn be ablated with histone deacetylase inhibitors, establishing a paradigm of combination therapy using inhibitors of chromatin regulators against drug resistance (Sharma et al., 2010).

**Targeting chromosomal neighborhoods and associated addictions**

As described above, insulated neighborhoods, whose ends are established by CTCF anchor sites, are frequently perturbed in cancer genomes by somatic mutations or aberrant DNA methylation; these perturbations have been implicated in the activation of cellular proto-oncogenes (Flavahan et al., 2016; Hnisz et al., 2016b; Katainen et al., 2015). Cancer cells containing aberrantly methylated neighborhood anchors can develop transcriptional dependencies that are amenable to small molecule therapeutics (Flavahan et al., 2016). The recent development of genetic and epigenetic editing technologies that can manipulate or repair chromosomal neighborhoods suggests novel therapeutic strategies for such modified cancer genomes (Amabile et al., 2016; Doudna and Charpentier, 2014; Liu et al., 2016b).

As an example, the expression of active genes can be reduced by disrupting the boundaries of its insulated neighborhood, which can, in turn, cause enhancers within a neighborhood to loop to other gene targets outside the neighborhood (Dowen et al., 2014). Targeted
methylation of a neighborhood anchor site with a dCas9-DNA-methyltransferase-3 fusion protein has been shown to disrupt the neighborhood through the eviction of CTCF (Liu et al., 2016b). Thus, targeted methylation of an oncogene enhancer and/or the anchor sites of oncogene-containing neighborhoods might lead to selective downregulation of oncogene expression (Figure 6E).

In glioma cells that harbor mutations in the IDH1 gene, hypermethylation of insulated neighborhood anchor sites occurs, which leads to loss of CTCF binding at those anchors and activation of oncogenes that were previously silent in the intact neighborhoods (Flavahan et al., 2016). This suggests that restoration of the perturbed neighborhood boundaries might lead to silencing of the oncogene. Targeted DNA 5-hydroxy-methylation with a dCas9-TET fusion protein has recently been demonstrated (Amabile et al., 2016; Liu et al., 2016b), and this strategy could be used to restore an insulated neighborhood whose anchor site has been disrupted by aberrant DNA methylation (Figure 6E).

The identification of oncogenes activated by perturbations of insulated neighborhood CTCF anchor sites by aberrant methylation has revealed transcriptional dependencies of these tumor cells. For example, the IDH1-mutant glioma cells described above overexpress the PDGFRA gene, due to hypermethylation of the CTCF anchor site of the insulated neighborhood containing PDGFRA, which disrupts a boundary of the neighborhood and allows enhancers located elsewhere to activate PDGFRA. The growth of these cells is inhibited by small molecules that block PDGFRA, while gliomas containing a wild type IDH1 gene and an intact PDGFR neighborhood are unaffected by such inhibition (Flavahan et al., 2016). In this case, the dependence of the cancer cells in these tumors on PDGFRA was not identified by cancer genome sequencing (which revealed the IDH1 mutation), but by hypothesis-driven study of the effect of IDH1-associated hypermethylation on transcriptional control in these tumor cells.

Inhibition of DNA methyltransferases (DNMTs) by 5-azanucleoside drugs such as 5-azacitidine have broad impacts on gene expression. Though toxic in high doses, these compounds have shown therapeutic benefit at optimized low doses, and are approved for the treatment of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) (Jones et al., 2016). These drugs were notable for their ability to alter the identities of cultured mammalian cells (Taylor and Jones, 1979) and are thought to activate tumor suppressor genes that have been silenced by DNA methylation and to reverse the tumor promoting effects of cancer-specific genetic alterations (Jones et al., 2016). With our new understanding of the roles of insulated neighborhoods and DNA methylation in gene control, it is possible that a major effect of these drugs is to alter neighborhood structures and thus the gene expression programs to which the tumor cells are addicted.

Future challenges

The fields of cancer biology and transcriptional biology are rapidly maturing and converging, establishing a compelling role for transcriptional dysregulation in cancer. We think it is imperative to commit to more aggressive study of the mechanisms that produce transcriptional addiction and learn how to exploit these for new therapies. Despite
tremendous progress, there exist significant challenges ahead before these nascent insights can be broadly leveraged for patient benefit.

Where feasible, targeting oncogenic TFs in cancer (e.g., PML-RARA) has demonstrated profound clinical benefit. However, therapeutics capable of disrupting oncogenic TFs are lacking. We need a new science of transcriptional therapeutics and a generational commitment to the pursuit of the high-hanging fruit. Oncogenic MYC remains a holy grail of cancer therapy, yet there are as yet no compounds that directly target this TF. One solution may be new platforms of discovery chemistry capable of identifying and optimizing TF-directed compounds, delivery solutions to bring biomolecules nimbly across cell membranes and, quite plausibly, entirely new classes of agents. As an example, the general chemical strategy for drug-induced degradation of targeted proteins provides one path to this goal (Figure 6F) (Winter et al., 2015).

Where known, targeting tissue-specifying master TFs in cancer (ER/PR, AR) has profound clinical benefit. However, our knowledge of tissue-specifying TFs and dependencies remains limited. Here there is a need for more epigenome-based insights into cancer core regulatory circuitry, elaborated from primary human tumors. Some important insights into specific TF dependencies have recently emerged using such approaches [e.g., OCA-B in DLBCL (Chapuy et al., 2013); LMX1A in a medulloblastoma subtype (Lin et al., 2016)].

The development of clinically useful inhibitors of various components of the transcriptional apparatus will require further mechanistic understanding of the role of transcriptional addiction in cancer pathogenesis. As an example, CDK9 inhibition is highly active in CLL; at present, however, the drugs that have been developed exhibit only a narrow therapeutic index and no biomarkers of pharmacologic efficacy exist. Clarity on the mechanisms that create such dependencies may provide solutions such as therapeutic synergies.

A ubiquitous challenge of targeted cancer therapy is the emergence of tumor cells resistant to the therapeutic agent. While emerging evidence suggests that transcriptional inhibitors can suppress the emergence of drug resister cells when combined with other therapeutic agents (Sharma et al., 2010; Yashiro-Ohtani et al., 2014), recent studies indicate that tumor cells can develop resistance against the transcriptional inhibitors (Fong et al., 2015; Rathert et al., 2015; Shu et al., 2016). Future investigations will need to evaluate the effects of combination therapies that include transcriptional inhibitors.

While genome structural alterations have long been understood to have roles in cancer, the science of genome structure and function is still in its infancy. More complete mechanistic understanding of insulators, CTCF, and regional roles of chromatin-associated complexes is needed in cancer. Careful dissection of genome structure and its regulators in experimental and translational model systems will surely reveal new targets.

Improved understanding of dysregulated transcription in the context of the cancer cell heterogeneity present in individual tumors will be important for the development of effective therapies. However, precise measures of this heterogeneity at the level of the epigenome is currently limited by the fact that the generation of genome-wide maps of DNA-associated proteins often requires samples of >10^6 cells. As a consequence, current analyses obscure
the heterogeneity created by intermingled clonal subpopulations within individual tumors that may differ from one another in important ways. The need to understand transcriptional control in, for example, tumor-initiating cells, AKT-low dormant cells, and metastatic carcinoma cells generated by epithelial-mesenchymal transitions, will require single-cell integrated epigenomic analyses. Recent insights into intratumoral heterogeneity in primary glioblastoma, obtained using single-cell RNA-seq, highlight this need (Patel et al., 2014).

Finally, we need to study the dynamics of transcriptional control in the most relevant model system of cancer – patients. We do not presently have the capability to study the dynamic effects of treatment in order to elucidate mechanisms of response and resistance. Indeed, future technologies that reveal epigenetic responses to targeted therapy are likely to provide novel, critically important mechanistic insights that will greatly benefit patients.

Acknowledgments

We are grateful to Brian Abraham, Brad Bernstein, Daniel Day, Matthew Fearer, Isaac Klein, Nick Kwiatkowski, Charles Lin, Marc Mansour, Peter Rahl, Jussi Taipale, William Tansey, Marc Timmers, Chris Vakoc and Robert Weinberg for critical comments. Supported by an NIH Grant HG002668 (R.A.Y.), an Erwin Schrödinger Fellowship (J3490) from the Austrian Science Fund (D.H.) and a Margaret and Herman Sokol Postdoctoral Award (D.H.). R.A.Y. and J.E.B are founders and shareholders of Syros Pharmaceuticals. R.A.Y. is a founder and shareholder of Marauder Therapeutics. J.E.B. is a shareholder and executive of Novartis Pharmaceuticals, and a shareholder of Acetylon Pharmaceuticals. Syros, Novartis and Acetylon are discovering and developing therapeutics directed at transcriptional pathways in cancer.

References


---

Cell. Author manuscript; available in PMC 2018 February 09.


Cell. Author manuscript; available in PMC 2018 February 09.
Figure 1. Genetic changes and dysregulated gene expression programs lead to cancer cell state
The path to a cancer cell involves genetic alterations that lead to changes in the gene expression program. The dysregulated program can create dependencies on transcriptional regulators that make the tumor cells more sensitive to inhibition of these regulators than normal cells.
Figure 2. Key features of transcriptional regulation of gene expression programs
A) In this model of cell-type specific core regulatory circuitry, master TFs co-regulate their own genes, forming an interconnected autoregulatory loop (left), as well as those of those of many other cell-type specific genes (right).
B) Enhancers are DNA elements bound by multiple TFs that recruit coactivators (such as Mediator) and RNA polymerase II, which can initiate transcription within enhancer sequences to produce enhancer RNAs (eRNAs).
C) SE constituents are physically connected. Interactions among SE constituent enhancers, and between SEs and a target gene are indicated by red arcs.

D) Human genome contains over 10,000 insulated neighborhoods, which are produced by multimerization of CTCF bound to two sites and reinforced with cohesin; enhancer-gene interactions occur predominantly within these neighborhoods.

E) SEs are enriched for signaling transcription factors, and their associated genes, which tend to play prominent roles in cell identity, are thus especially responsive to signaling.

F) Signaling TFs and other TFs that play important regulatory roles are themselves regulated by ubiquitin and proteasome-mediated destruction.

G) Chromatin regulators that act through “writing”, “erasing” or “reading” histone modifications.

H) DNA methylation can contribute to gene control by causing loss of TF binding at enhancers and loss of CTCF at insulated neighborhood loop anchors.
Figure 3. Components of gene control altered in cancer

Examples of three types of cis-factors (enhancers, promoters and insulators) and various trans-factors (transcription factors, cofactors, chromatin regulators, RNA polymerase II and histones) that acquire recurrent somatic mutations in cancer cells.
Figure 4. Common mechanisms of dysregulation of gene expression programs in cancer involving transcription factors

A) Model of a transcriptional regulatory circuit controlling the gene expression program in normal cells is shown top left. Dysregulation of gene expression programs in cancer cells can occur through dysregulation an oncogenic master TFs (bottom left), dysregulation of a transcriptional amplifier (e.g. MYC) (bottom right), and dysregulated signaling (top right).

B) Model of aberrant chromatin modification affecting gene expression programs in cancer cells.
C) Model of effect of cohesin mutations and CTCF mutations on enhancer-promoter interactions and insulated neighborhoods in cancer cells.
Figure 5. Common mechanisms of dysregulation of gene expression programs in cancer involving cis-factors

A) Mechanisms leading to the acquisition of super-enhancers to drive oncogenes in cancer cells: translocation of an existing super-enhancer, focal amplification of an enhancer element, and nucleation of super-enhancer through somatic insertion of transcription factor binding sites.

B) Activation of silent proto-oncogenes by somatic mutations that disrupt insulated neighborhood anchor sites.

**Figure 6. Drugging transcriptional dependencies**

A) Model of the features of super-enhancer–associated oncogene control that contribute to transcriptional dependencies.

B) Model of the mechanistic basis of higher transcriptional activity and vulnerability of super-enhancers: co-operative interactions between the co-activators recruited to these sites.

C) Therapeutic strategy to inhibit transcriptional dependencies due to aberrant recruitment of DOT1L to driver genes in MLL-AF fusion leukemias.
D) Therapeutic strategy to attack small populations of drug-tolerant tumor cells with histone deacetylase inhibitors.
E) Therapeutic strategy to downregulate oncogenes through manipulation and repair of insulated neighborhoods.
F) Targeted degradation of transcriptional regulators using degronimids that recruit E3-ubiquitin ligase for proteasome-mediated degradation.
Table 1

Examples of transcriptional dependencies that occur as a consequence of transcriptional dysregulation

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Component of transcriptional control perturbed by genetic change</th>
<th>Genetic change</th>
<th>Transcriptional dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>T cell leukemia</td>
<td>Master transcription factor</td>
<td>TAL1 overexpression</td>
<td>CDK7</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>Enhancer/Super-enhancer</td>
<td>IgH-MYC translocation</td>
<td>BRD4</td>
</tr>
<tr>
<td>Glioma</td>
<td>Insulated neighborhood anchor</td>
<td>IDH1 mutation</td>
<td>PDGFRA</td>
</tr>
<tr>
<td>Mixed lineage leukemia</td>
<td>Chromatin regulator</td>
<td>MLL-AF9 fusion</td>
<td>DOT1L BRD4</td>
</tr>
<tr>
<td>T cell leukemia</td>
<td>Signaling factor</td>
<td>NOTCH1 mutation</td>
<td>BRD4</td>
</tr>
</tbody>
</table>