Haploinsufficient Transcription Factors in Myeloid Neoplasms

Tanner C. Martinez¹,² and Megan E. McNerney¹

¹Department of Pathology, Department of Pediatrics, Section of Hematology/Oncology, The University of Chicago Medicine Comprehensive Cancer Center, The University of Chicago, Chicago, Illinois, USA; email: megan.mcnerney@bsd.uchicago.edu
²Medical Scientist Training Program, The University of Chicago, Chicago, Illinois, USA

Keywords
hematopoietic stem cell, acute myeloid leukemia, myelodysplastic syndrome, transcription factor, haploinsufficiency, aneuploidy

Abstract
Many transcription factors (TFs) function as tumor suppressor genes with heterozygous phenotypes, yet haploinsufficiency generally has an underappreciated role in neoplasia. This is no less true in myeloid cells, which are normally regulated by a delicately balanced and interconnected transcriptional network. Detailed understanding of TF dose in this circuitry sheds light on the leukemic transcriptome. In this review, we discuss the emerging features of haploinsufficient transcription factors (HITFs). We posit that: (a) monoallelic and biallelic losses can have distinct cellular outcomes; (b) the activity of a TF exists in a greater range than the traditional Mendelian genetic doses; and (c) how a TF is deleted or mutated impacts the cellular phenotype. The net effect of a HITF is a myeloid differentiation block and increased intercellular heterogeneity in the course of myeloid neoplasia.
It takes two to make a thing go right.

—Rob Base and DJ E-Z Rock, “It Takes Two”

1. INTRODUCTION

By the classical genetic definition, malignancy was initially thought to be a dominant phenotype. Malignant cells fused with nonmalignant counterparts produced malignant hybrids (1). Subsequent work demonstrated that in some combinations the reverse is true; that is, fusions result in nonmalignant hybrids, creating the paradigm of a recessive mechanism of tumor suppression (2). This paradigm was formalized with the discovery of the first tumor suppressor gene, RB1. As described in Knudson’s two-hit hypothesis, both copies of a given tumor suppressor gene must be inactivated to cause cancer (3). Nearly 30 years later, this hypothesis was challenged by the finding that loss of only one copy of P27Kip (CDKN1B) is sufficient to promote the development of tumors (4). P27Kip thus founded the growing class of genes that act as haploinsufficient tumor suppressor genes and implicated the dose of tumor suppressors in oncogenesis. As discussed below, extensive tumor sequencing suggests that Knudson’s hypothesis may in fact apply only to a minority of tumor suppressor genes.

2. THE PARADIGM OF HAPLOINSUFFICIENCY

The classical definition of haploinsufficiency is the production of a dominant phenotype in a diploid organism after one allele of a gene undergoes a loss-of-function event (5). The context of the phenotype measured is critical and nuanced, as demonstrated in yeast. In genome-wide proliferation screens, the fraction of haploinsufficient yeast genes increases from 3% in nutrient-replete media to 20% in nutrient-deplete media. When morphological traits are also assessed, 33–59% of genes produce haploinsufficient phenotypes (6, 7). The extent of haploinsufficiency is harder to quantify in mammals; however, efforts to systematically knock out and phenotype all mouse genes reveal that a striking 42% of genes have significant phenotypes in heterozygous mice (8).

Surveys of healthy human genomes have revealed which human genes are among the most evolutionarily constrained, defined as genes that are the most sensitive to deleterious mutations. Almost half of coding genes have fewer loss-of-function variants than would be expected by chance, indicating a high degree of genetic constraint genome wide. Among constrained genes, haploinsufficient genes are the most conserved, as a deleterious variant in one allele is selected against (9). Transcription factors (TFs) appear more likely to produce haploinsufficient phenotypes when compared with other genes, and homeodomain-containing TFs are overrepresented among haploinsufficient transcription factors (HITFs) (10). Indicating the importance of finely tuned protein dose, genes predicted to produce haploinsufficient phenotypes are also likely to be triplosensitive, in that gain of an additional copy of the gene produces a dominant phenotype (11). However, measures characterizing the extent of haploinsufficiency genome wide in healthy individuals do not capture genes with deleterious alleles that do not come into play until after reproduction (e.g., BRCA1). Overall, these studies demonstrate the profound dose-sensitive nature of the human genome.

Just as unbiased sequencing studies have demonstrated greater haploinsufficiency across the genome than previously appreciated, similar sequencing studies of tumors highlight the inadequacy of the two-hit model of tumor suppressor genes. To measure haploinsufficient genes in cancer, Elledge and colleagues (12) performed a sophisticated pan-cancer analysis for patterns of loss-of-function mutations and gene deletions. They estimate that at least 30% of all genes exhibit haploinsufficiency. Notably, almost all sporadic tumor suppressor genes are predicted to produce haploinsufficient phenotypes. The correlation between haploinsufficiency and triplosensitivity is also observed in tumor suppressor genes. Overall, haploinsufficiency in cancer is much
more prevalent than previously appreciated. Furthermore, there is genetic pressure, even in cancer, to maintain dose-sensitive genes within a minimum and maximum range.

To highlight cardinal features of haploinsufficient genes in cancer, we turn to PTEN. While not a TF, PTEN ranks among the most well-studied proteins in this class. In prostate cancer, Pten heterozygous loss promotes proliferation; Pten homozygous loss leads to greater proliferation, but only after Trp53 inactivation to avoid cellular senescence (13, 14). Comparably, in hematopoietic malignancies, inactivation of both copies of PTEN results in hematopoietic stem cell depletion, and additional gene mutations are necessary to avoid this loss (14, 15). Even a small 20% reduction in PTEN protein levels can promote tumor development (16). Likewise, subtle, subdiploid reduction in the expression of the TF PU.1 (SPI1) is sufficient to promote a preleukemic state (17). Furthermore, in prostate cancer, mutation of PTEN is less common than copy number alteration of the locus containing PTEN, 10q23. This latter phenomenon may be related to the fact that in addition to PTEN, there are five other tumor suppressive genes at this locus that contribute to the phenotype in prostate cancer (18).

The above examples allow us to underscore some principles of HITFs, which we discuss further (Figure 1):

- **Principle 1:** One is enough, but sometimes two is better—HITFs may exhibit haploinsufficiency in that loss of one allele produces a phenotype, but in some instances loss of the second allele produces a greater or distinct effect.
- **Corollary to principle 1:** Myeloid neoplasms require a minimal myeloid cell identity—TFs involved in the terminal differentiation of myeloid cells (e.g., RUNX1, CEBPα, PU.1) are rarely mutated to zero remaining activity. This is not true of TFs that are not important in myeloid cell identity (e.g., TP53).
- **Principle 2:** There are more than three doses—phenotypes associated with the dose of a TF can occur on a continuum of TF activity and are not limited to the Mendelian diploid, haploid, and null levels.
- **Principle 3:** The monoallelic loss of multiple genes is cumulative—the concurrent losses of genes neighboring a HITF in the setting of aneuploidies and focal deletions can modulate the phenotype.

### 3. DOSE IN HEMATOPOIESIS

The classical, branching tree–based model of hematopoiesis presupposes discrete steps at which developing cells make unforeseen and binary fate choices. While a helpful tool in conceptualizing the hierarchy of development, this model fails to capture the heterogeneous and continuous nature of hematopoiesis (19). Prospective fractioning of hematopoietic stem cells (HSCs) on the basis of cell surface or cytoplasmic protein expression reveals the cell-intrinsic heterogeneity in key HSC functions, namely, cell cycle participation, lineage bias, and self-renewal capacity (20). HSC heterogeneity is further confirmed by retrospective lineage tracing models that demonstrate functional and transcriptional differences present in HSCs (21, 22). In other words, lineage biases are already present at the apex of hematopoiesis, are continuously accrued, and sometimes entail more than one route to arrive at a terminal differentiated cell (23, 24). TFs and coregulators are the likely origins of this cell-intrinsic HSC heterogeneity, as differentiation results from the proper expression and timing of TFs (25).

Imagining the three properties of HSC heterogeneity—self-renewal capacity, cell cycle participation, and lineage bias—as continuous spectra upon which HSCs can reside rather than as discrete phenotypes that cells either possess or do not possess is more consistent with the continuous model of hematopoiesis and the dose-dependent nature of TFs (Figure 2). Loss of one copy
HITF Principle 1: One is enough, but sometimes two is better

- Gene A\(^{+/+}\): Wild-type
- Gene A\(^{+/—}\): Haploinsufficient
- Gene A\(^{—/—}\): Null

Corollary to HITF Principle 1: Myeloid neoplasms require a minimal myeloid cell identity

- Wild-type
- Haploinsufficient
- Null

HITF Principle 2: There are more than three doses

- Wild-type
- Haploinsufficient
- Null

HITF Principle 3: The monoallelic loss of multiple genes is cumulative

- Focal deletion
- Chromosomal arm loss
- Chromosome loss

Figure 1

Principles of HITFs. (a) Some genes are haploinsufficient in that monoallelic loss in a diploid organism produces a phenotype. However, this does not preclude an increased effect or distinct phenotype upon a further reduction in activity or loss of the second allele. (b) Differentiation into the full panoply of myeloid cells is possible for progenitors with a fully intact myeloid transcriptional network (left). A block in myeloid differentiation is imposed with loss of a HITF involved in the differentiation of myeloid cells, and leukemogenesis can proceed with the acquisition of additional genetic aberrations (center). However, in the absence of myeloid cell identity, for example, through the total ablation of CEBP\(\alpha\) activity, myeloid leukemogenesis cannot proceed (right). (c) Dose has traditionally been conceptualized in the number of intact alleles of a gene. However, further evidence has given rise to a model of TF activity that extends beyond diploid, haploid, and null. Specifically, preleukemic states can arise in cells with slight, subdiploid alterations in dose of a HITF, and subhaploid but not null doses of a HITF efficiently drive tumorigenesis. (d) HITFs are only part of the story when it comes to the phenotype of deletion events that result in the loss of multiple genes. The ultimate cellular phenotypes of deletion events depend on the other codeleted genes.}

of a TF can reposition cells along these axes and potentially endow a cell with increased capacity for self-renewal and/or a block in lineage-specific differentiation. For example, monoallelic loss or mutation of the nonclustered, homeobox domain-containing TF CUX1 is found in approximately 10–15% of myeloid neoplasms (Table 1) (26, 27). Loss of CUX1 in mouse models promotes an expansion of hematopoietic stem and progenitor cells (HSPCs), progressive anemia, and an expansion of mature myeloid cells, leading to a lethal myelodysplastic syndrome (MDS) (28, 29).

In other words, the change in CUX1 dose influences HSPC decisions and moves the HSPCs along the axes. Similar changes can be conceptualized for the other TFs discussed in this review.
Myelodysplastic syndrome (MDS): hematologic neoplasms with clonal hematopoiesis, cytopenia(s), and abnormal cellular morphology that can be defined by morphology or recurrent genetic abnormalities

Enhancer: a cis-regulatory genomic element that influences the transcription of a target gene

Transcriptional noise: variations in transcription of a given gene between cells of the same type arising from the bursting and probabilistic nature of transcription initiation

Figure 2
Stem cell properties are continuous characteristics. Self-renewal, proliferation, and lineage bias are properties of hematopoietic stem cells that exist as continuous variables. These properties are dictated by the dose of HITFs. Therefore, alteration of the dose of a HITF can have profound effects on these properties. For example, GATA2 overexpression, which is found in primitive AMLs, results in excess self-renewal. In addition, loss of CUX1 results in increased proliferation and a myeloid bias at the expense of erythroid development. Abbreviations: AML, acute myeloid leukemia; HITF, haploinsufficient transcription factor.

4. MECHANISMS OF TRANSCRIPTION FACTOR HAPLOINSUFFICIENCY

4.1. Transcription Factor Dynamics

TFs actualize the final cell state by reciprocally reading and influencing the enhancer code to promote the expression of effector genes. Transcription is a dynamic and noisy process that depends on the genomic colocalization of many factors. Thus, fundamentally, transcription is a concentration-dependent process. The noise of transcription refers to the phenomenon where, rather than stable, consistent RNA polymerase II (Pol II) activity at a given gene, Pol II transcribes the gene in short, probabilistic bursts, which are influenced by enhancers and TFs (30). PU.1, GATA1, and GATA2 mRNA all exhibit noisy transcription with all stem cells exhibiting some expression of these genes, although not necessarily at the same time. Perhaps counterintuitively, these dynamic properties may increase the stability of stemness in the HSC pool as a whole by maintaining HSCs that are sampling every transcriptional state (31).

Transcriptional noise impacts HITFs on two ends: the transcription of the TF itself and transcription of the downstream target genes. TFs generally have low expression levels at baseline (32). Further loss of one TF allele portends particularly noisy expression, and the amount of the TF may even intermittently approach the null state as a result (33). While transcription is inherently noisy, monoallelic loss of a HITF amplifies this noise, resulting in an inconsistent transcriptome and increased intercellular heterogeneity (34). This variation is the precondition for selection, allowing a rare cell transcriptional state to be reinforced at the chromatin level, stabilizing an increase in fitness of a preleukemic cell (35, 36).

In summary, the loss of a HITF has two major ramifications for leukemogenesis (Figure 3): First, it provides the requisite myeloid differentiation block. Second, it increases the intercell heterogeneity by increasing the number of sampled transcriptomes, giving the cell an advantage in adapting to the environment.

In addition to increasing transcriptional noise, suboptimal TF concentrations impede the efficiency of TFs binding to cognate DNA sequences. TF dynamics are determined by search pattern,
Table 1  Frequency of HITF genetic aberrations and genetic aberrations affecting the dose of HITFs across the spectrum of myeloid diseases

<table>
<thead>
<tr>
<th>Gene</th>
<th>Genetic aberration</th>
<th>CH</th>
<th>CCUS</th>
<th>CMML</th>
<th>MPN</th>
<th>MDS</th>
<th>AML</th>
<th>t-MN</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEBPA</td>
<td>Mutation</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>2%</td>
<td>9%</td>
<td>4%</td>
</tr>
<tr>
<td>CUX1</td>
<td>Mutation</td>
<td>0.03–0.4%</td>
<td>3.4%</td>
<td>10%</td>
<td>0.2–3%</td>
<td>2–6%</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>−7/del(7q)</td>
<td>0.0002%</td>
<td>NR</td>
<td>1–10%</td>
<td>0%</td>
<td>2–16%</td>
<td>9–14%</td>
<td>33–50%</td>
</tr>
<tr>
<td>EGR1</td>
<td>Mutation</td>
<td>NR</td>
<td>NR</td>
<td>0%</td>
<td>NR</td>
<td>NR</td>
<td>0.5%</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>del(5q)</td>
<td>0.0003%</td>
<td>NR</td>
<td>1.5%</td>
<td>0.5%</td>
<td>4–15%</td>
<td>7–16%</td>
<td>49%</td>
</tr>
<tr>
<td>ETV6</td>
<td>Mutation</td>
<td>0.1%</td>
<td>0%</td>
<td>1.6%</td>
<td>3–6%</td>
<td>1%</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>−12/del(12p)</td>
<td>0.0001%</td>
<td>NR</td>
<td>1%</td>
<td>0.2%</td>
<td>3–5%</td>
<td>2–7%</td>
<td>12%</td>
</tr>
<tr>
<td>GATA1</td>
<td>Mutation</td>
<td>0–0.1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>&lt;1%</td>
<td>0.2%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>inv(3)</td>
<td>NR</td>
<td>NR</td>
<td>0.7%</td>
<td>NR</td>
<td>1–5%</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>−20/del(20q)</td>
<td>0.001%</td>
<td>NR</td>
<td>1%</td>
<td>0%</td>
<td>4–5%</td>
<td>2–6%</td>
<td>9%</td>
</tr>
<tr>
<td>RUNX1</td>
<td>Mutation</td>
<td>0–0.2%</td>
<td>3.4%</td>
<td>13.2%</td>
<td>0.5–2%</td>
<td>9–14%</td>
<td>6–12%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>t(8:21)</td>
<td>NR</td>
<td>NR</td>
<td>ED</td>
<td>ED</td>
<td>ED</td>
<td>1–4%</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>t(3;21)</td>
<td>NR</td>
<td>NR</td>
<td>ED</td>
<td>ED</td>
<td>ED</td>
<td>3–6%</td>
<td>1–2%</td>
</tr>
<tr>
<td></td>
<td>inv(16)</td>
<td>NR</td>
<td>NR</td>
<td>ED</td>
<td>ED</td>
<td>ED</td>
<td>2–4%</td>
<td>2–5%</td>
</tr>
<tr>
<td></td>
<td>+21</td>
<td>0.00025%</td>
<td>NR</td>
<td>1%</td>
<td>NR</td>
<td>2–4%</td>
<td>2–5%</td>
<td>11%</td>
</tr>
<tr>
<td>TP53</td>
<td>Mutation</td>
<td>2.6–4.4%</td>
<td>1.1%</td>
<td>1%</td>
<td>2–8%</td>
<td>6–13%</td>
<td>6–9%</td>
<td>25–37%</td>
</tr>
<tr>
<td></td>
<td>−17/del(17p)</td>
<td>0.0002%</td>
<td>NR</td>
<td>1%</td>
<td>2%</td>
<td>7%</td>
<td>7%</td>
<td>13%</td>
</tr>
<tr>
<td>PML-RARA</td>
<td>t(15;17)</td>
<td>NR</td>
<td>NR</td>
<td>ED</td>
<td>ED</td>
<td>ED</td>
<td>3–9%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Selected mutations that alter the dose of HITFs

<table>
<thead>
<tr>
<th>Gene</th>
<th>Genetic aberration</th>
<th>CH</th>
<th>CCUS</th>
<th>CMML</th>
<th>MPN</th>
<th>MDS</th>
<th>AML</th>
<th>t-MN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLL1</td>
<td>t(11q23;x)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>3%</td>
<td>3.3%</td>
</tr>
<tr>
<td>NPM1</td>
<td>Mutation</td>
<td>0–0.1%</td>
<td>0%</td>
<td>2.6%</td>
<td>0.7%</td>
<td>1.5–2%</td>
<td>22–27%</td>
<td>15%</td>
</tr>
<tr>
<td>FLT3</td>
<td>Mutation</td>
<td>0–0.1%</td>
<td>0%</td>
<td>2.6%</td>
<td>0.5%</td>
<td>1.5–5%</td>
<td>28–33%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Reference(s)


Abbreviations: AML, acute myeloid leukemia; CCUS, clonal cytopenia of undetermined significance; CH, clonal hematopoiesis; CMML, chronic myelomonocytic leukemia; ED, excluded on the basis of diagnostic criteria; HITF, haploinsufficient transcription factor; MDS, myelodysplastic syndrome; MPN, myeloproliferative neoplasm; NR, not reported; t-MN, therapy-related myeloid neoplasm.

search time, and on/off rates (25). The search pattern of a TF refers to how the protein moves about the nucleus to find target sequences. MYC, for instance, explores the entire nucleus, while other TFs, such as CTCF, linger in smaller areas containing a high concentration of binding sites (38, 39). TF search time refers to the interval between DNA binding events and can be influenced by mitigating factors such as the chromatin milieu (40). TFs spend the most time unbound and in search mode (25). Per the law of mass action, a reduction in TF molecules will lower the overall likelihood that the DNA targets are occupied by the HITF. While this conclusion is intuitive, the downstream consequences can be complex. For instance, TFs regularly interact with other TFs and/or cooperatively bind DNA. SOX2 is one such example and guides OCT4 to cobind DNA. Decreased SOX2 concentration not only curtails SOX2 DNA targeting but also increases the search time of OCT4 (41). Numerous TFs physically interact with one another and bind DNA in concert, suggesting that they could have linked search patterns as well. Prolonged searching for DNA binding sites could be further exacerbated in situations where the HITF is a pioneer...
Loss of a HITF results in a myeloid differentiation block and greater transcriptional plasticity. HSCs can develop normally by following one of the differentiation paths out of the bone marrow. Or, in a variation on Waddington’s landscape (37), HSCs can undergo monoallelic loss of a HITF, represented as falling through the trapdoor. The loss of a HITF has profound effects on the nuclear environment. Normally, TFs bridge DNA sequences at enhancers with the general transcription machinery to hold all of these molecules in close proximity to transcribe a gene. However, with decreased TF molecules present in the nucleus, clustering of all the requisite DNA sequences and transcriptional machinery happens less frequently. It is difficult to achieve a sufficiently high local concentration of a TF to move enhancers and genes into an environment permissive for transcription, which can result in genes failing to be relocated and thus transcribed. Consequently, increased transcriptional plasticity ensues as cells fail to maintain a consistent epigenome and transcriptome. Furthermore, such cells have an inability to properly differentiate to myeloid cells due to the dysregulated enhancer landscape. These two ramifications of HITF loss—increased transcriptional plasticity and a myeloid differentiation block—are depicted as the cell rolling around in the well, sampling many different transcriptomes, and being unable to escape down the path of normal myeloid differentiation. Abbreviations: HITF, haploinsufficient transcription factor; HSC, hematopoietic stem cell; TF, transcription factor.

Even in normal conditions, cells temporally manipulate the nuclear concentration of TFs to bring about distinct cellular fates. This phenomenon is exemplified by nuclear hormone receptors, which shuttle into the nucleus upon ligand binding, but also extends to other classes of TFs (42). For instance, the effects of the dosage-sensitive TF TP53 are conveyed by nuclear localization. Pulses of TP53 into the nucleus promotes the repair of DNA damage, whereas sustained TP53 nuclear localization causes senescence (43). In this example, the extent of DNA damage is communicated through the nuclear concentration of TP53, producing different outcomes. The developmental TF HES1 also exhibits distinct nuclear oscillation patterns and concentrations, producing different outcomes in neural progenitors (44). Nuclear factor kappa B (NFκB) signaling patterns in murine HSPCs similarly result in differential gene expression of developmental or apoptosis genes in response to inflammation (45). Given the finely tuned patterns necessary for these TFs to convey unique cell state information, it is unlikely that only one allele of a TF could achieve these
Transcription condensates: assemblies of high local concentrations of the machinery required to transcribe a gene (TFs, transcription coregulators, and RNA Pol II) and their subsequent dissolution.

4.2. Clustering the Characters

Context-dependent gene expression is largely mediated by enhancers (46). To influence transcription, an enhancer must have sustained interaction with a target promoter, and TFs bestow DNA sequence specificity on this interaction (41). One enhancer can regulate multiple genes, and multiple genes can be coordinately regulated in turn, a process bridged by TFs (41, 47, 48). These actors can be found in physical proximity to each other and have been described as transcriptional hubs or factories (48–51). More recently, liquid-liquid phase separation has been considered an etiology of nuclear subcompartmentalization (52). These biophysical structures may explain the observed clustering of the requisite transcriptional regulators and cis-regulatory elements that regulate Pol II activity. Much more work is warranted to understand the role of phase separation in transcriptional regulation. Nonetheless, in both the hub and condensate models, TFs migrate into subcompartments of the nucleus, increasing local TF concentration (49, 51, 53, 54).

KLF1 and EBF1 provide two examples of how TFs can be important for the relocation of target genes. In murine erythroid progenitors, KLF1 forms dynamic transcription factories. KLF1 shepherds together erythroid genes from across the genome, including hemoglobin genes, to coordinately drive transcription promoting erythropoiesis (49). Reiterating the importance of the proper nuclear localization of TFs and target genes, monoallelic loss of the B cell HITF EBF1 results in less efficient relocation of a target transgene out of heterochromatin, leading to decreased transgenic expression. In this scenario, there is insufficient EBF1 to pull the transgene into a nuclear environment more conducive for transcription (55). This finding supports a model of transcription where the TF must attain a certain local concentration at a given locus, which is associated with moving the gene into a permissive transcriptional environment. This model presupposes the ability to generate high local concentrations of a TF, a feat that is impaired in the setting of haploinsufficiency. The decreased frequency of activation of the transgene underscores the probabilistic nature of transcription and the buffering capacity of two alleles.

The biochemical properties mediating the clustering of cis-regulatory elements and the general transcription machinery is an active area of investigation. Consequently, intrinsically disordered regions (IDRs) have received renewed attention for the ability of these protein domains to mediate clustering, condensation, and liquid-liquid phase separation. The overwhelming majority of TFs have IDRs, which lack fixed three-dimensional structure but can still convey interaction specificity (56, 57). Mutations in IDRs are common in cancer and enriched in known drivers of hematopoietic malignancy (58). The transactivation domains of TFs have repeatedly been shown to be IDRs, which form condensates with general transcription machinery, including Pol II, Mediator, and BRD4, leading to high local concentrations of the important players for transcription (51, 59–63). These findings lead to a model wherein the IDRs and DNA binding domains of TFs serve to couple the transcriptional machinery with DNA. As a result, the TFs are held in increased local concentration to facilitate repetitive DNA binding and transcriptional bursting (50, 52, 64, 65). The loss of one copy of a HITF decreases the probability of target gene compartmentalization and activation. Overall, sequestration of transcriptional players and processes within the diffusion space of transcription condensates reduces transcriptional noise, as important molecular players are more likely to bump into each other (66).

4.3. Models of Transcription Factor Action at Target Genes

Regardless of the physical relationship between TFs, enhancers, and transcription initiation, some genes are more sensitive than others to diminished TF levels, begging the question of how a TF
Acute myeloid leukemia (AML): the clonal expansion of malignant HSPCs defined by genetic aberration or extent of differentiation of blasts

Segregates in the nucleus under limiting conditions. For example, in murine HSPCs, heterozygosity of RUNX1 leads to a loss of binding to and expression of the Cebp locus but not the Cebpa locus, suggesting an underlying mechanism governing the actions of dose-limited RUNX1 (67). The differential sensitivity of some genes to the loss of a TF has led some investigators to conceptualize TFs as having either analog or digital outputs. RUNX1 in this case is functioning in a digital or binary manner; at a certain threshold of expression, RUNX1 either activates a gene target or does not. The TF ETV2 also exhibits digital effects. ETV2 can promote the development of both endothelial cells and erythroid cells, and this dichotomy is conveyed through different thresholds of binding and activation for genes of each lineage. Specifically, low levels of ETV2 are sufficient to promote expression of endothelial genes, but a much higher threshold must be met to promote erythroid gene expression (68). In contrast, NFκB TFs exhibit analog or gradient-like effects where the expression of target genes is directly proportional to the level of NFκB activation (69). At a limited number of loci tested, PU.1 also appears to act in an analog manner at loci related to proliferation (70). Interestingly, SOX9, which is mutated in devastating craniofacial developmental disorders, exhibits both analog and digital activity. Enhancers that exhibit analog activity are more likely to have a SOX9 palindrome motif and are associated with different pathways than digital enhancers, which are more likely to have a weak SOX9 motif and be bound by other TFs (71). This later work represents the most advanced analysis of a dose-dependent TF, accomplished through the use of a degron tag to generate a spectrum of TF doses. As such tools are further developed to manipulate TFs at a single-cell level in an inducible manner, we may find that more TFs resist simplified classification into analog and digital regulators.

4.4. Compensation

In addition to the biochemical sequestration that condensates provide, cells have other mechanisms to safeguard the robustness of the transcriptome. In some cases, however, these mechanisms may be insufficient to compensate for the monoallelic loss of a HITF. The long half-lives of mRNA molecules and posttranscriptional regulators are thought to buffer against the probabilistic nature of transcription. However, TFs are generally less stable on the mRNA and protein levels as compared with the rest of the genome (32). Hence, compensatory mechanisms are less efficacious for TFs.

Some proteins have closely related family members that can minimally compensate for the loss of one of the members. With respect to RUNX1, family members RUNX2 and RUNX3 can partially compensate for the loss of RUNX1 (72). In the setting of CEBPa loss, CEBPβ (CEBPB) can compensate in the setting of emergency granulopoiesis induced by inflammation (73). Yet, as a class, HITFs are less likely to have paralogs, contributing to the necessity of two alleles of these genes for proper development (74).

5. HOW TO LOSE A PARTNER IN SEVEN WAYS

Having considered the molecular mechanisms of TF action and the implications for haploinsufficiency and dose dependence, we now focus on the common means by which the abundance of a gene product can be altered and how the mechanism of gene inactivation can have a profound impact on the phenotype manifested (Figure 4). For example, the chromatin landscape of RUNX1-translocated acute myeloid leukemia (AML) is counterintuitively more akin to that of double-mutant CEBPa AML (CEBPαdm) than AMLs with RUNX1 point mutations (75). Below, we consider the different types of HITF loss and how the context impacts the phenotype.

5.1. Aneuploidy

Losing one copy of part or all of a chromosome is seemingly the most straightforward method to lose one allele of a TF; however, the implications of losing potentially hundreds of neighboring
a Aneuploidy

Examples: –7/del(7q), del(5q), del(20q), del(12p), –17/del(17p)

b Focal deletion

Examples: 7q22, 7q36

c Balanced translocations and inversions

Examples: t(15;17), inv(3), t(8;21), t(3;21), inv(16)

d Mutation

Examples: RUNX1, TP53, GATA1, GATA2

e Mutation of another gene

Examples: PU.1, RUNX1

f Change in isoform balance

Examples: CEBPA, RUNX1

g Epigenetic silencing

Example: GATA2

(Caption appears on following page)
genes in tandem as well have not been sufficiently appreciated (76). In most cases, aneuploidy is a detriment to the fitness of a cell—this is evinced in the finding that the more genes on a chromosome, the less likely cancer cells are to tolerate loss or gain of that chromosome (77). It is only under specific circumstances that aneuploidy confers an advantage. And while aneuploidy has been thought to be a late event of advanced-stage neoplasms, aneuploidy can be an early and founding event, as evidenced by aneuploidy in clonal hematopoiesis (78).

In general, any monosomy will have profound effects on the cell, including inducing ribosomal stress and gene expression changes that are not limited to the affected chromosome. The ensuing proteotoxic stress activates a TP53 response that can promote cell death (79). Consequently, aneuploidy is largely incongruous with intact TP53 signaling, likely due to proteotoxic stress, and this may be why AML and MDS with complex karyotypes often also present with TP53 mutations (80). That said, while there are common features of aneuploidies, there is variability to the cellular response to aneuploidy as well. In one report, cells harboring an identical aneuploid event can diverge with respect to cell cycle alterations, response to extracellular and intracellular signals, and gene expression (81). This aneuploidy-induced transcriptional plasticity likely also fosters increased drug resistance (82). Others have argued that there is extensive posttranscriptional compensation for aneuploidy in cancer. As a result, protein changes are not as dramatic as one might surmise, including for oncogenes and tumor suppressors (83). Nonetheless, an aneuploid event results in not only the loss of a HITF within a greater genetic deletion but also the concomitant cellular stress of aneuploidy.

A second distinction of aneuploidy-induced HITF loss is that several pathogenic genes may be simultaneously deleted en bloc. A contiguous gene syndrome (CGS) can arise in this event, which refers to the phenotypic consequences of the loss of multiple neighboring genes. It is technically challenging to experimentally model CGS; however, there is mounting evidence that multiple genes encoded on a chromosome arm can cooperatively drive cancer when lost (84). In some cases, CGS genes may converge on similar pathways; in others, genes may impact complementary pathways (85). It would be remiss to neglect codeleted genes when considering HITFs impacted by aneuploidy (HITF principle 3). In the remainder of this section, we highlight HITFs within recurrent aneuploid events in myeloid neoplasms. However, we acknowledge that aneuploidies and focal deletions have more complex phenotypes than solely the loss of the HITF alone.

**EGR1** encodes an ETS family TF and is frequently deleted in the context of del(5q). Consistent with del(5q) as a recurrent event in myeloid neoplasms arising in patients with a history...
of cytotoxic therapy [therapy-related myeloid neoplasms (t-MNs)], monoallelic loss of \textit{EGRI} is sufficient for myeloid neoplasia upon treatment with an alkylating agent (86). The loss of \textit{EGRI} conveys a fitness advantage by blunting the DNA damage and inflammation response in HSPCs (87). Highlighting principle 3, \textit{EGRI} cooperates with another 5q resident, \textit{APC}, encoding a Wnt signaling component, as well as \textit{Trp53} to promote de novo AML. In contrast, heterozygosity for only one or two of the three genes does not result in AML (88). This example highlights the aggregate phenotype that can arise due to cumulative haploinsufficiency and emphasizes the need to consider the impact of genes deleted concomitantly with the HITF.

\textit{CUX1}, a homeodomain-containing TF, has conserved dose sensitivity from \textit{Drosophila melanogaster} to humans. In flies, the homologue of \textit{CUX1}, Cut, has dose-dependent roles in the development of at least two organ systems (89, 90). Interestingly, up to four different doses of Cut are associated with different developmental fate decisions, suggesting that manipulation of the level of this TF is one way that developmental decisions are made. Different fates associated with different doses of \textit{CUX1} may be realized by the analog nature of \textit{CUX1} (91). The graded effects of \textit{CUX1} dose are true at the level of not only transcription but also the elicited phenotype. In mice, knockdown of \textit{CUX1} to an intermediate level results in a mild form of MDS with anemia, while further knockdown results in a more severe anemia, an expansion of mature myeloid cells, and a lethal form of MDS/myeloproliferative neoplasm (MPN) (29). Certain features unify \textit{CUX1} and its orthologues across model systems and human disease: Knockdown of \textit{CUX1} results in changes in proliferation and impaired differentiation (27, 29). In humans, \textit{CUX1} is deleted in the context of \(-7/del(7q)\) without silencing or mutation of the remaining allele (26, 27, 92). In de novo myeloid malignancies, \textit{CUX1} mutations or \(-7/del(7q)\) are more associated with MDS. In fact, in some cases of pediatric MDS, \(-7/del(7q)\) appears sufficient for the disease, as it is the sole somatic genetic change identified (93). Also, \(-7/del(7q)\) is strongly associated with t-MNs (94). One explanation for this association is that \textit{CUX1} has an early role in the DNA damage response; hence, \textit{CUX1} deficiency aberrantly allows continued HSPC proliferation in the setting of DNA damage stemming from cytotoxic therapy (95). The resultant clonal advantage of \textit{CUX1}-deficient cells ultimately leads to the development of an overt myeloid neoplasm. This mechanistic work highlights the role of the HITF levels in normal differentiation and how deviant levels promote transformation.

\textit{Del}(20q) results in the loss of \textit{MYBL2} (bMyb), a TF that regulates the differentiation and self-renewal of myeloid progenitors (96). Homozygous deletion of \textit{MYBL2} is embryonic lethal; however, haploinsufficiency drives a range of myeloid disorders, including MDS and MPN (97). Downregulation of \textit{MYBL2} is sufficient for murine HSPC expansion, which is consistent with \textit{MYBL2} downregulation as a common feature of MDS, regardless of 20q status (98). This finding indicates that \textit{del}(20q) is not the only route to downregulate \textit{MYBL2}.

\textit{Del}(12p) is a small interstitial deletion found in MDS with a complex karyotype. Due to its small size, \textit{del}(12p) is underreported in standard cytogenetic analyses (99). The ETS-family TF \textit{ETV6} is lost in \textit{del}(12p) events, and the remaining allele is generally not methylated or mutated, cementing this TF as a HITF (99, 100). \textit{ETV6} is a homodimeric transcriptional repressor required for the establishment of hematopoiesis and survival of adult HSCs (101). Like many of the other HITFs discussed here, including \textit{RUNX1}, \textit{CEBP}α, and \textit{GATA2}, \textit{ETV6} also harbors germline mutations (102). Similar to individuals carrying a germline \textit{RUNX1} mutation, patients with inherited \textit{ETV6} mutations have a high likelihood of developing thrombocytopenia due to defects in megakaryocyte and platelet maturation, underscoring the necessity of both alleles of these two genes for proper platelet generation (103).

Unlike the genes we have considered thus far, \textit{TP53} frequently harbors biallelic alterations. \textit{TP53} alterations include mutations, loss of heterozygosity, and deletions in the form of
-17/del(17p). As \( TP53 \) status is usually assessed amid full-blown malignancy, when biallelic alterations are common, \( TP53 \) can appear to abide by the two-hit hypothesis (104). Even so, there is ample evidence from mouse and human studies that monoallelic loss of \( TP53 \) is sufficient to promote leukemogenesis. \( Trp53 \) heterozygosity is conducive to murine HSPC clonal outgrowth after DNA damage (105). Likewise, heterozygosity for \( Trp53 \) is permissive for tumor formation in mice, and the remaining allele can remain intact and functional (106). In MDS, while many patients have monoallelic or biallelic \( TP53 \) mutations, a small subset of patients present with loss of 17p without mutation or loss of heterozygosity of the remaining allele (104). Notably, like other aneuploid events, del(17p) deletes additional tumor suppressor genes that independently drive tumorigenesis in mice (84). Hence, our understanding of del(17p) pathogenesis remains incomplete. Notwithstanding, these studies indicate that TP53 produces haploinsufficient phenotypes that may become more pronounced with biallelic loss (HITF principle 1).

### 5.2. Focal Deletions

In contrast to aneuploidy, focal deletions are smaller than a chromosome arm (77). Focal deletions can affect HITFs while potentially avoiding some of the fitness defects induced by aneuploidy. In fact, there is selection to delete tumor suppressor genes while preserving essential genes and oncogenes in focal deletions (107). As a result, focal copy number alterations are more potent in driving proliferation signatures as compared with chromosome arm level changes (108). That said, focal copy number alterations still correlate with mutations in DNA damage response genes. This latter finding implicates the DNA damage response pathway as a gatekeeper for genomic instability in general, as remarked upon above with \( TP53 \) mutations and aneuploidy (108).

Also analogous to aneuploidy, HITF principle 3 remains true for focal copy number changes. As an example, we return to del(7q). In addition to monosomy 7 and del(7q), smaller cryptic deletions can occur that are below the sensitivity of standard cytogenetics analysis. Higher-resolution approaches have revealed minimally deleted regions, including one spanning \( CUX1 \) on 7q22 (27, 109). While \( CUX1 \) loss is sufficient to cause myeloid disease, deletion of a 7q22 syntenic region in mice that does not include \( CUX1 \) also disrupted hematopoiesis (29, 110). These data implicate more than one haploinsufficient 7q22 gene. Furthermore, these findings of a CGS in a focal deletion extend to other copy number variations as well (111).

### 5.3. Balanced Translocations

Balanced translocations typically garner attention for the ensuing fusion product or overexpressed oncogene. For instance, \( t(15;17)(q22;q12) \) produces the PML-RARA fusion protein, pathognomonic for acute promyelocytic leukemia (APL). Less appreciated is the fact that such translocations inherently generate loss of function in the rearranged genes, in this case, PML and the TF RARA. Translocations can thereby result in three hits for the price of one. While the fusion product alone initiates leukemogenesis, haploinsufficiency of PML and RARA also further promotes disease (112, 113).

Similarly, inv(3)(q21q26.2) or \( t(3;3)(q21;q26.2) \) repositions the \( GATA2 \) distal enhancer away from \( GATA2 \) to instead drive oncogenic overexpression of the EVI1 TF (114). Inherently, this inversion or translocation also attenuates \( GATA2 \) expression. Indeed, \( GATA2 \) heterozygosity accelerates EVI1-mediated oncogenesis in mice (115). A minority (15%) of patients also have mutations of the nonrearranged \( GATA2 \) allele (HITF principle 1) (116). Thus, like \( t(15;17) \), inv(3) simultaneously produces both gain-of-function oncogenic events and loss-of-function events in HITFs.

\( RUNX1 \) is another HITF involved in balanced translocations. \( RUNX1 \) is a pleiotropic TF required for the initiation of definitive hematopoiesis and multilineage differentiation in adult hematopoiesis. However, \( RUNX1 \) does not work alone; it requires the cofactor CBF\( \beta \) (\( CBFB \)) to
achieve proper regulation of target genes (117). There are three ways that this critical interaction can be interrupted through translocation to promote the development of myeloid malignancy: Two are translocations involving RUNX1 and one is a translocation involving CBFB.

The first translocation is the t(8;21)(q22;q22.1) translocation, which fuses the RUNX1 DNA-binding Runt homology domain (RHD) to the Nervy-homology region (NHR) domains of ETO (RUNX1T1). The NHR recruits the repressive nuclear receptor corepressor (NCoR) complex (118). Consequently, this translocation results in a neomorphic RUNX1 fusion protein that represses myeloid differentiation genes it normally activates, such as PU.1, GATA1, and CEBPα (119–121). The second translocation is t(3;21), which is most commonly found in t-MNs; t(3;21) fuses the RHD of RUNX1 to the TF EVI1. RUNX1-EVI1 also results in the loss of CEBPA expression, and thus the promyeloid differentiation program, but uniquely upregulates GATA2, producing a more primitive AML (122, 123). Both translocations create a potent oncoprotein and simultaneously disrupt the involved RUNX1 allele. The derivative RUNX1 cannot interact with CBFb to drive proper gene expression and is effectively dysfunctional. The remaining wild-type RUNX1 is in competition with the oncogenic RUNX1 fusion for DNA binding sites, resulting in a finely tuned transcriptional network optimized for cellular fitness (124). This is made apparent upon loss of the wild-type RUNX1 allele, as knockdown of RUNX1 in cells expressing RUNX1-ETO causes apoptotic cell death due to the overactivity of RUNX1-ETO (125). The remaining wild-type RUNX1 activity is seemingly subhaploid but absolutely critical, highlighting the dose-dependent actions of RUNX1 (HITF principle 2).

The final translocation disrupting RUNX1 dose is inv(16)(p13q22), which rearranges CBFB, the gene encoding the obligate RUNX family TF cofactor, CBFb. Inv(16) fuses CBFB to MYH11 (SMMHC); the resultant CBFb-SMMHC oncoprotein retains the ability to bind RUNX1. CBFb-SMMHC actually has greater affinity for RUNX1 than CBFb alone and co-opts RUNX1 to drive the leukemogenic program (126). Wild-type RUNX1 activity is diminished but not abolished as a result. In fact, some residual RUNX1 activity is essential for the survival of this AML (125, 127, 128). Further suggesting reduced RUNX1 dose as important in this disease, monoallelic loss of Runx1 in mice phenocopies inv(16) AML with respect to granulocyte maturation failure (127). In addition to disrupting RUNX1 function, CBFb-SMMHC reduces TP53 dose by indirectly mediating the inappropriate deacetylation and inactivation of TP53 (129). Thus, the mechanism of pathogenesis of inv(16) AML is the result of the oncogenic fusion protein compounded by the interference with two dose-dependent TFs, TP53, and RUNX1.

### 5.4. Mutations

Mutations are not created equal and can have widely varying effects on protein function, so we must consider the details of individual mutations to understand how they affect the dose of a given TF.

RUNX1 is among the most frequently mutated TFs in AML and MDS (130). Patients with familial platelet disorder have inherited loss-of-function mutations in RUNX1, which results in cell-intrinsic impairment of megakaryocyte differentiation and an increased risk of myeloid malignancy (131). Expectedly, the loss of one copy of RUNX1 in induced pluripotent stem cells (iPSCs) results in compromised megakaryocyte differentiation (132). RUNX1 mutations exist on a functional spectrum. N-terminal mutations result in pure loss of function, as they result in truncations or mutations that disrupt the ability to bind DNA or CBFb. Mutations in the DNA binding domain (RHD) are weakly dominant negative. Finally, mutations in the C-terminal transactivation domain are strongly dominant negative (80% reduction in RUNX1 activity) (133). The precise mechanism of strong dominant negative mutations is unclear and could be multifactorial.
Underscoring the exquisite dose sensitivity of RUNX1, loss of one copy of RUNX1 or a weak dominant negative mutation (R174Q) are both associated with thrombocytopenia and decreased megakaryocyte differentiation in iPSC culture. However, only the strong dominant negative mutation results in an expansion of myeloid progenitors and increased genomic instability (134). Further emphasizing the dose-dependent nature of RUNX1, myeloid malignancies with lower RUNX1 activity have a higher likelihood of secondary leukemic progression (135). The spectrum of RUNX1 activity and associated phenotypes capture the essence of HITF principle 2.

GATA1 is an X-linked gene, so while it cannot be considered classically haploinsufficient, it still exhibits dose-dependent effects. This characteristic is featured in the GATA1 knockdown mouse, which displays impaired erythropoiesis and megakaryopoiesis (136). Total GATA1 ablation causes erythroid progenitor apoptosis, while a partial decrease promotes proliferation and a leukemia-like disease, underscoring the dose-dependent nature of GATA1 (137). A second example of this is apparent in Down syndrome. Of infants with Down syndrome, characterized by trisomy 21, 10% develop a self-limiting leukemia-like disease termed transient abnormal myelopoiesis (TAM). Twenty percent of TAM cases will evolve to acute megakaryoblastic leukemia (AMKL). Virtually every case of TAM and AMKL has a mutation in GATA1 (138). GATA1 mutations in the context of trisomy 21 are sufficient for TAM but not for AMKL, in which other mutations are acquired. Generally, the mutations in GATA1 are 5′ nonsense mutations, allowing protein synthesis to restart at a 3′ start codon to create a shorter isoform, GATA1s (139). The GATA1s isoform lacks part or all of the N-terminal transactivation domain, reducing the binding of an important cofactor, FOG1, ultimately leading to a TF with less transactivation capability. In this way, incomplete loss GATA1 activity is pathogenic in TAM and AMKL.

A second GATA family TF member, GATA2, is also a HITF. Loss of one copy of murine Gata2 results in reduced HSC numbers, increased HSC proliferation, and impaired self-renewal upon transplant (140). Conversely, overexpression of GATA2 results in a differentiation block, suggesting that GATA2 is also triplosensitive (141). The vast majority of mutations in GATA2 are predicted to generate a loss-of-function allele, via either nonsense or missense mutations in one of the two zinc finger domains (142). Mutations in the first GATA2 zinc finger domain create a loss-of-function allele due to the inability of the mutant protein to bind DNA. These mutations are associated with CEBPaDM in myeloid neoplasms (143). In a mouse model of CEBPaDM and a first zinc finger domain mutation of Gata2, the Gata2 mutation causes decreased accessibility of myeloid TF binding sites, while both TFs increase erythroid TF expression (144). This mechanistic study provides an explanation for the bilineage leukemia that arises from synergistic and concurrent mutations in CEBPA and GATA2.

We return to consider how TP53 is different than the other TFs discussed in this review. TP53 mutations are enriched in older patients, t-MNs, and complex karyotype malignancies (145). In myeloid neoplasms, loss-of-function, dominant negative, and gain-of-function TP53 mutations have been described (105, 146, 147). All of these mutations result in differing degrees of TP53 impairment. There is an inverse relationship between TP53 activity and the ensuing fitness advantage in malignancies. Cells with the least amount of TP53 activity have the greatest fitness advantage, and cells can accordingly gain a further advantage by removing residual TP53 activity. For instance, of MDS cases with TP53 aberrations, one-third are monoallelic and two-thirds are biallelic (104). Monoallelic TP53 mutations do not connote the poor prognosis that biallelic mutations carry. In some patients with monoallelic TP53 mutations, MDS transforms to secondary AML with concomitant additional TP53 loss. With regards to the clonal structure, patients with monoallelic TP53 mutations have greater clonal diversity than patients with biallelic TP53 mutations, and biallelic TP53 mutations are always in the dominant clone, further indicating the relative advantage of leukemic cells with biallelic TP53 mutations (HITF principle 1). These results...
buttress the notion that monoallelic TP53 loss is sufficient to drive MDS but is not synonymous with biallelic TP53 disease.

5.5. Mutation of Another Gene

While rarely mutated itself, the ETS family TF PU.1 is often a concomitant casualty of other genetic aberrations. The centrality of PU.1 in the myeloid transcriptional network means that many common genetic aberrations in AML affect PU.1 dose. Functionally, PU.1 is required for HSCs to generate myeloid progenitors and for myeloid differentiation beyond the granulocyte monocyte progenitor stage (148). In addition to orchestrating differentiation, PU.1 is a regulator of HSC proliferation and protein synthesis. Consequently, loss of PU.1 increases HSC cycling and decreases self-renewal (70). The mouse models of PU.1 illustrate the dose-dependent aspects of this TF. Heterozygous PU.1-deficient mice have various hematopoietic defects but do not spontaneously develop AML (149). However, an 80% reduction in PU.1 by homozygous deletion of an enhancer upstream of PU.1 results in a transplantable AML (150). Interestingly, heterozygous deletion of this enhancer produces a 35% reduction in PU.1 and a preleukemic state, which progresses to an overt AML on a DNA-repair-deficient background (17). These mouse models demonstrate that varying PU.1 doses from slight to more severe reductions are associated with unique leukemic and preleukemic states (HITF principle 2). In fact, low levels of PU.1 activity are required to form competent leukemic stem cells, as mutation of the RUNX1 binding sites in a Pu.1 enhancer results in a severe decrease in leukemic stem cell activity and delayed onset of RUNX1-ETO-driven leukemia (72).

Many paths to myeloid neoplasms pass through PU.1, given its requirement in myeloid differentiation and cell cycle control. RUNX1 normally regulates PU.1 expression in a dynamic and lineage-specific manner; thus, alteration of RUNX1 dose has repercussions for PU.1 levels (151). RUNX1 fusion proteins, for instance, not only disrupt the normal function of RUNX1, as discussed above, but also directly reduce PU.1 expression. There are two active ways that RUNX1 fusion proteins reduce PU.1 dose: exploitation of an antisense promoter normally used to downregulate PU.1 during T cell development, and inactivation of a long noncoding RNA that normally facilitates PU.1 expression (152, 153). In addition, some N-terminal RUNX1 mutations abrogate RUNX1 interaction with and recruitment of the activating H3K4 methyltransferase MLL (KMT2A) to the PU.1 promoter (154). Moving beyond RUNX1, PU.1 levels are low in APL and increase after ATRA-induced differentiation; in fact, overexpression of PU.1 is sufficient to promote differentiation in this model (155). Mutations in CEBPA result in the downregulation of PU.1, and the FLT3 internal tandem duplication mutation results in the downregulation of PU.1 as well as the downregulation of CEBPα (72, 156–158). Although different groups have found conflicting results, mutation of the nucleolar protein NPM1 results in a strong nuclear export signal and can drag PU.1 along with it into the cytoplasm (159, 160). Finally, a reduction in PU.1 expression accelerates the development of AML when combined with loss of the DNA demethylase TET2. The residual PU.1 is not sufficient to overcome the aberrant methylation at PU.1 binding sites at critical myeloid enhancers (161). The acquisition of a mutation in a TET2-deficient cell that results in the loss of PU.1 expression could rapidly shift the transcriptional balance toward myeloid differentiation and cell cycle dysregulation (162, 163). Overall, reduction in PU.1 dose is a common consequence of mutations in myeloid genes, underscoring the fundamental role of a stable myeloid transcriptional network in preventing leukemogenesis.

5.6. Change in Isoform Balance

TFs often have multiple isoforms—products with different functions that arise from alternative splicing or posttranslational regulation. CEBPα is one such TF with multiple isoforms containing
distinct activities. CEBPα is required for granulocyte differentiation, and, apropos of the corollary to HITF principle 1, homozygous Cebpα knockout mice do not develop AML, even in the presence of a strong oncogene (73). The full-length CEBPα p42 protein contains a C-terminal basic leucine zipper DNA binding domain and an N-terminal transactivation domain, the latter of which is omitted in the shorter p30 isoform (164). Both isoforms are normally present, and their abundance is regulated translationally. As a result of this unique arrangement, different CEBPα mutations have distinct results. N-terminal mutations result in the production of p30 only, and C-terminal mutations disrupt the ability of CEBPα to bind DNA or homodimerize. Most tumors with CEBPA mutations have biallelic mutations of both types (165). The combination of these two mutations shifts the balance of isoforms toward p30 homodimers and away from normally predominant p42 homodimers (164). The shift to entirely p30 homodimers resulting from two N-terminal mutations is sufficient for leukemogenesis (166). However, the combination of an N-terminal mutation and a C-terminal mutation results in leukemia with faster onset, indicating C-terminal mutations might be more potent in suppressing the myeloid transcriptional network (167). These results have several implications. Importantly, biallelic mutations do not recapitulate the CEBPα null state, suggesting that there is a minimal dose of CEBPα that is required to form leukemic cells. Secondly, the mechanism by which leukemic cells alter the dose of CEBPα is partly due to an N-terminal mutation shifting the isoform balance toward the production of p30 homodimers. This later mechanism is exacerbated by further impairing p30 CEBPα activity with a C-terminal mutation. Of note, CEBPα isoforms can be deregulated through less direct means as well, as occurs with del(5q) (168).

While RUNX1 can be inactivated, it can also be amplified by somatic or germline gains of chromosome 21, on which RUNX1 is encoded. In TAM, a neoplasm dependent on RUNX1, the apparent gain in RUNX1 expression through trisomy 21 actually translates into a loss of RUNX1 activity through dysregulation of the three RUNX1 isoforms. RUNX1b and RUNX1c are similar, full-length isoforms; however, RUNX1a is a primate-specific, shorter isoform that lacks the C-terminal transactivation domain (117). Compared with the other isoforms, RUNX1a has a higher affinity for DNA and is thought to act in a dominant negative manner, similar to the impact of mutations in the RUNX1 transactivation domain (169). RUNX1a expression is elevated in TAM, likely to due to the inability of the aberrant GATA1s isoform to repress RUNX1a, ultimately resulting in a decrease in wild-type RUNX1 activity (170). In addition, RUNX1a recruits other transcriptional regulators to inhibit megakaryocyte terminal differentiation and promote the proliferation of megakaryocytic blasts (170). Gain of chromosome 21 is not the only way that RUNX1 isoforms can become dysregulated. A specific mutation in a splicing factor, SRSF2^{P95H}, also promotes the expression of RUNX1a (171). Hence, like CEBPα, RUNX1 dose can also be altered by a change in the balance of isoforms to promote abnormal proliferation and differentiation.

### 5.7. Epigenetic Silencing

DNA methylation is one of many ways to inhibit the expression of genes, including HITFs. AML co-opts the DNA methylation machinery to augment tumorigenesis in concert with driver mutations, and, therefore, AML can be subdivided on the basis of DNA methylation signatures (172). In an unbiased look at diverse AMLs, GATA2 showed allele-specific expression in 60% of patients, indicating that one allele had been epigenetically silenced (173). This is particularly pronounced in the case of CEBPα^{DM} AML, where 95% of cases show allele-specific expression of GATA2. In leukemias with a concomitant GATA2 mutation, the wild-type allele is more likely to be silenced (HITF principle 1). Given the synergistic effects of CEBPα^{DM} and GATA2 mutation, epigenetic silencing is another route to decrease wild-type GATA2 and amplify the potency of a GATA2 mutation. Epigenetic silencing as a mechanism of decreasing HITF dose is not limited to GATA2 and applies to other HITFs such as CEBPA (174).
6. THERAPEUTIC OPTIONS

Haploinsufficiency opens at least two therapeutic windows: decreasing the activity of the residual protein to kill leukemic cells reliant on that protein or boosting the activity of the remaining wild-type protein to restore differentiation. These options can potentially avoid the toxic effects of nonspecific chemotherapy by specifically targeting cells containing the mutant allele. In an example of the first approach, knocking down the wild-type RUNX1 allele or inhibiting CBFβ binding impairs AML growth in RUNX1 mutant xenografted mice (175). This example ought to inspire further investigation into molecules that can be given therapeutically to accomplish this goal of removing residual HITF activity in mutant cells.

Using a tetracycline-inducible (and reversible) small hairpin RNA transgene, knockdown of CUX1 expression results in de novo MDS/MPN or an alkylating agent–induced t-MN in mice (29, 95). In diseased mice, tetracycline withdrawal restores CUX1 levels and normalizes blood parameters. This example provides a genetic proof of principle for augmenting the wild-type allele as a method to treat disease driven by a HITF. Further research is warranted to ascertain methods that will allow the restoration of normal HITF protein expression.

Concordant with HITF principle 3, focal deletions and aneuploidies provide an additional opportunity to exploit the essentiality of concurrently deleted genes. Also described as collateral lethality, this concept assumes a therapeutic window for targeting cells that are haploinsufficient for an essential gene. Lenalidomide is a successful example of this. Lenalidomide and its derivatives are highly efficacious in the treatment of MDS with isolated del(5q) because lenalidomide mediates the degradation of CK1α, an essential protein encoded on 5q. Reducing CK1α levels to null results in TP53-mediated apoptosis (176). This same strategy could be applied to other recurrent chromosomal deletions in myeloid malignancies.

7. CONCLUSION

We hope to have conveyed the importance of dose in a broad sense. Traditionally, dose has been thought of diploid, haploid, and null, but many examples exist showing the pathogenic importance of non-Mendelian gene levels. In addition, there are ways to affect the dose of a HITF beyond its deletion or mutation. The myeloid cell identity is dependent on the complex transcriptional network brought about by the TFs discussed in this review, so it follows that changing the dose of one TF has secondary effects on the dose of other TFs. This balance brings added complexity to considering the genetic aberrations of a particular tumor. For example, if a patient presenting with a TET2 mutation in the form of clonal hematopoiesis acquires another mutation that changes the dose of PU.1, there is a risk of progression to a myeloid neoplasm (161). A tumor with a FLT3ITD mutation also has an altered dose of CEBPα, and this is likely a source of the myeloid differentiation block associated with the disease. While myeloid cancers typically have fewer mutations than solid tumors, the recurrent mutations have multiple effects on the transcriptional network. A better understanding of this complexity will clarify why some patients respond to a given therapy and will open up treatment strategies for those who do not.

SUMMARY POINTS

1. A large portion of the coding genome is under genetic constraint due to haploinsufficiency, which includes almost all transcription factors (TFs) and tumor suppressor genes.
2. TFs function to influence the contact time between enhancers and target promoters to regulate the expression of target genes.
3. Loss of one allele of a gene can produce distinct effects that can be augmented by the loss of the second allele.

4. Given that myeloid neoplasms arise in cells with myeloid identity, TFs involved in myeloid cell identity are rarely mutated to zero residual activity. This is in contrast to TP53, the function of which is often completely lost in malignant cells.

5. There are TF doses not defined by strict Mendelian genetics that are consequential for the development of the preleukemic and leukemic states.

6. When one haploinsufficient transcription factor is altered, additional TFs are impacted due to the interwoven nature of the myeloid transcriptional network.

FUTURE ISSUES

1. What is the role of transcription factor (TF) dose in normal hematopoiesis?
2. How do TF dynamics change for hematopoietic TFs in the haploinsufficient state?
3. How do mutations in or loss of the transactivation domain and/or intrinsically disordered region (IDR) change TF dynamics and impact haploinsufficient phenotypes?
4. How do mutations in IDRs of TFs change their interaction partners and ability to interact with promoters and enhancers?
5. How do cancer cells optimize the levels of a given TF to best drive the oncogenic state, and how is this altered by environmental stresses such as inflammation and chemotherapeutic pressure?
6. How do genes concomitantly lost with haploinsufficient transcription factors (HITFs) modulate the disease phenotype in the context of a contiguous gene syndrome?
7. What mechanisms of normal TF regulation can be exploited to therapeutically manipulate the dose of a HITF to treat disease?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We would like to thank Matthew R. M. Jotte and Madhavi Senagolage for their feedback on this manuscript. We would like to thank Alivia Gao for her help in illustrating the figures. T.C.M. is supported by the National Heart, Lung, and Blood Institute (F30HL163992). M.E.M. is supported by the National Institutes of Health (R01 HL166184, R01 HL142782, R01 CA231880, and P30 CA014599), the Leukemia and Lymphoma Society Scholar Award, the V Foundation for Cancer Research Pediatric Cancer Research V Scholar All-Star Award, and the Cancer Research Foundation Fletcher Scholar Award.

LITERATURE CITED

9. This paper highlights the extent of genetic constraint in the human genome due to haploinsufficiency.


Access provided by 73.9.100.221 on 01/30/24. See copyright for approved use.

Biochemistry

175(7):1842–55.e16

Cell Rep


Mol. Cell


PNAS


Mol. Cell


Blood


Mol. Cell

55. This paper demonstrates how monoallelic loss of a HITF can influence the nuclear position of target genes.

Cell Stem Cell


Mol. Syst. Biol.


Mol. Cell


Oncogenesis


Science


592 Martinez • McNerney


---

161. This paper demonstrates the importance of minor changes in TF dose for the progression of leukemia.


170. This paper demonstrates how gain of a chromosome can negatively affect the dose of a TF.


Contents

Chance and Opportunity: A Personal Story
Abul K. Abbas .......................................................... 1

Diffuse Pleural Mesothelioma: Advances in Molecular Pathogenesis, Diagnosis, and Treatment
Christopher A. Febres-Aldana, Rachel Fanaroff, Michael Offin,
Marjorie G. Zauderer, Jennifer L. Sauter, Soo-Ryum Yang, and Marc Ladanyi ........ 11

Dynamic Multiplex Tissue Imaging in Inflammation Research
Stefan Uderhardt, Georgiana Neag, and Ronald N. Germain 43

Antibody and B Cell Responses to SARS-CoV-2 Infection
and Vaccination: The End of the Beginning
Katbarina Röltgen and Scott D. Boyd 69

Genetics and Pathogenesis of Dystonia
Mirja Thomsen, Lara M. Lange, Michael Zech, and Katja Lohmann 99

Update on Epithelial-Mesenchymal Plasticity in Cancer Progression
Rosa Fontana, Aida Mestre-Farrera, and Jing Yang 133

Control of Cell Death in Health and Disease
Nobuhiko Kayagaki, Joshua D. Webster, and Kim Newton 157

Role of the Microenvironment in Glioma Pathogenesis
Maya Anjali Jayaram and Joanna J. Phillips 181

Within-Host Evolution of Bacterial Pathogens in Acute
and Chronic Infection
John P. Dekker .......................................................... 203

Neutrophils in Physiology and Pathology
Alejandra Aroca-Crevillén, Tommaso Vicanolo, Samuel Ocadia,
and Andrés Hidalgo .................................................. 227

Genome Instability and DNA Repair in Somatic
and Reproductive Aging
Stephanie Panier, Siyao Wang, and Björn Schumacher 261
Hypoxia-Induced Signaling in Gut and Liver Pathobiology
*Sumeet Solanki and Yatrik M. Shab* .......................................................... 291

Pediatric Cholestatic Diseases: Common and Unique Pathogenetic Mechanisms
*Harry Sutton, Saul J. Karpen, and Binita M. Kamath* ............................................... 319

Neurodegenerative Disease Tauopathies
*Benjamin C. Creekmore, Ryobei Watanabe, and Edward B. Lee* .................. 345

Epigenomic Characterization of Lymphoid Neoplasms
*Marti Duran-Ferrer and José Ignacio Martín-Subero* .................................. 371

Cancer as a Disease of Development Gone Awry
*Ben Z. Stanger and Geoffrey M. Wahl* .................................................. 397

Comparative Pathogenesis of Severe Acute Respiratory Syndrome Coronaviruses
*Jingshu Zhang, Melanie Rissmann, Thijs Kuiken, and Bart L. Haagmans* ............. 423

Acetaminophen Hepatotoxicity: Paradigm for Understanding Mechanisms of Drug-Induced Liver Injury
*Hartmut Jaeschke and Anup Ramachandran* ........................................ 453

Clonal Hematopoiesis, Inflammation, and Hematologic Malignancy
*Rashmi Kanagal-Shamanna, David B. Beck, and Katherine R. Calvo* .................. 479

ENPP1 in Blood and Bone: Skeletal and Soft Tissue Diseases Induced by ENPP1 Deficiency
*Carlos R. Ferreira, Thomas O. Carpenter, and Demetrios T. Braddock* ............ 507

Toward Explainable Artificial Intelligence for Precision Pathology
*Frederick Klauschen, Jonas Dippel, Philipp Keyl, Philipp Jurmeister, Michael Bockmayr, Andreas Mock, Oliver Buchstab, Maximilian Alber, Lukas Ruff, Grégoire Montavon, and Klaus-Robert Müller* .................. 541

Haploinsufficient Transcription Factors in Myeloid Neoplasms
*Tanner C. Martinez and Megan E. McNerney* ........................................ 571

Errata

An online log of corrections to *Annual Review of Pathology: Mechanisms of Disease* articles may be found at http://www.annualreviews.org/errata/pathmechdis