Helix–loop–helix proteins and the advent of cellular diversity: 30 years of discovery

Cornelis Murre

Division of Biological Sciences, University of California at San Diego, La Jolla, California 92903, USA

Helix–loop–helix (HLH) proteins are dimeric transcription factors that control lineage- and developmental-specific gene programs. Genes encoding for HLH proteins arose in unicellular organisms >600 million years ago and then duplicated and diversified from ancestral genes across the metazoan and plant kingdoms to establish multicellularity. Hundreds of HLH proteins have been identified with diverse functions in a wide variety of cell types. HLH proteins orchestrate lineage specification, commitment, self-renewal, proliferation, differentiation, and homing. HLH proteins also regulate circadian clocks, protect against hypoxic stress, promote antigen receptor locus assembly, and program transdifferentiation. HLH proteins deposit or erase epigenetic marks, activate noncoding transcription, and sequester chromatin remodelers across the chromatin landscape to dictate enhancer–promoter communication and somatic recombination. Here the evolution of HLH genes, the structures of HLH domains, and the elaborate activities of HLH proteins in multicellular life are discussed.

More than 30 years ago, a group of DNA elements sharing conserved sequences was identified within the enhancers of the immunoglobulin heavy and light chain loci (Church et al. 1985; Ephrussi et al. 1985). In a series of elegant experiments, these distinct DNA segments, named E-box sites, were found to be protected from in vivo methylation in B-lineage cells. The E-box sites contained a signature core of six nucleotides: CANNTG.

Simultaneously, MyoD was identified. MyoD was an eye-opener with the remarkable ability to convert fibroblasts into myoblasts. MyoD was the first example of programmed transdifferentiation. MyoD induced the expression of hundreds of genes associated with skeletal muscle cell identity (Davis et al. 1987). Analysis of its primary amino acid sequence revealed that MyoD shared a region of similarity with the myc family of proteins. The shared feature was called the c-myc homology region (Tapscott et al. 1988).

In a parallel study, two proteins (E12 and E47) were identified. Both proteins bound to an E2-box site located in the intronic immunoglobulin light chain gene enhancer but differed from each other in the c-myc homology region (Murre et al. 1989a). Further investigation revealed that E12 and E47 were generated by differential splicing of exons encoding for the DNA-binding domain (Murre et al. 1989a; Yamazaki et al. 2018). This variation contributed to their unique affinities: E12 bound weakly to DNA, whereas E47 bound DNA with high affinity (Murre et al. 1989a; Sun and Baltimore 1991). Helical wheel analysis of both DNA-binding domains revealed two highly conserved amphipathic α helices separated by a loop domain. The conservation of the helices led to the hypothesis that the helix–loop–helix (HLH) motif functioned as a dimerization domain (Murre et al. 1989a). Indeed, simple mixing and binding experiments showed that the HLH region mediated homodimerization and heterodimerization (Murre et al. 1989a,b). Further analysis uncovered that, adjacent to the HLH region, a cluster of conserved basic amino acids mediated DNA binding (Davis et al. 1990). Hence, the name basic HLH (bHLH) proteins was established.

Characterization of HLH proteins

Thirty years have passed since those initial findings, and the field has witnessed the identification of ever-increasing additional HLH proteins. Based on expression patterns, dimerization selectivities, and DNA-binding specificities, the HLH proteins were segregated into distinct classes (Murre et al. 1989b). Class I HLH proteins, now categorized as E proteins, include E12, E47, E2-2, HEB, daughterless (*Drosophila melanogaster*), and HLH-2 (*Caenorhabditis elegans*). E proteins are abundantly expressed in many lineages and bind DNA as either homodimers or heterodimers. Class II HLH proteins include MyoD, myogenin, SCL, NeuroD1, NeuroD2, and members of the achaete–scute complex. This subset of HLH proteins binds DNA as either homodimers or heterodimers with E proteins

and is lineage-restricted. Class III proteins include c-myc, TFE3, sterol regulatory element-binding protein 1a (SREBP-1), and Mi. In addition to the HLH domain, they contain a leucine zipper dimerization domain and function as either transcriptional activators or repressors (Carroll et al. 2018). The majority of yeast and plant HLH proteins belong to this class of HLH proteins. Class IV HLH proteins include Mad, Max, and Mxi. They form heterodimers with c-myc and among each other to bind a distinct set of E-box sites (CACGTG or CATGTTG) (Carroll et al. 2018). Class V HLH proteins contain a HLH domain but lack a basic region. Upon interaction with bHLH proteins, they antagonize the DNA-binding activity of their bHLH partners (Benezra et al. 1990). Prominent members are the Id proteins Id1–4 and the *D. melanogaster* gene product Extramacrochaete (Emc). Class VI HLH proteins are characterized by the presence of a proline residue in their basic region. They include Hes and enhancer of split proteins. Class VI HLH proteins recognize a unique DNA sequence, named the N box (CACGCG or CACGAG), and function predominantly as transcriptional repressors by binding to Groucho (Paroush et al. 1994). Finally, class VII HLH proteins display as their defining feature multiple PAS domains that function as signaling sensors to light and oxygen. They include members of the HLH proteins that control the circadian clock (BMAL and CLOCK) and a response pathway to hypoxia (hypoxia-inducible factor α [HIF-α] and aryl hydrocarbon nuclear translocator [ARNT]) (Wang et al. 1995; Lowrey and Takahashi 2004; Huang et al. 2012). This class binds predominantly ACGTG or GCGTG core sequences and functions as transcriptional repressors.
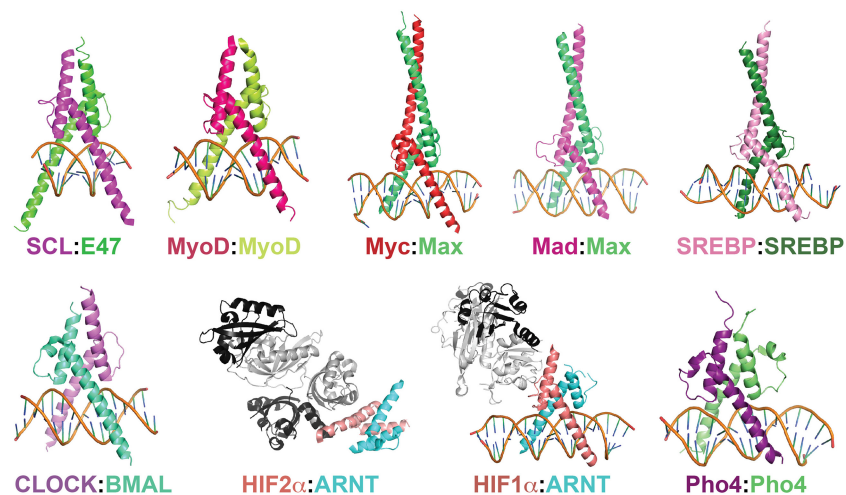
The structure of the HLH domain

The modeling suggested that the E-protein DNA-binding domain was folded into an HLH configuration. Detailed structural analysis soon followed. It validated the modeling. Crystal structures of E47 and MyoD HLH homodimers revealed two α helices connected by a loop domain as well as a basic region that contacted residues in the major groove (Fig. 1; Ellenberger et al. 1994; Ma et al. 1994). More specifically, E47 folded as a parallel four-helix bundle that intertwined, positioning the basic region in contact with each half of the E-box site (CAC or CAG). Two highly conserved amino acids—a glutamate and an arginine residue in the E47 DNA-binding region—mediated DNA-binding specificity. The glutamate residue directly contacted CA nucleotides, while arginine residues stabilized DNA binding. Dimerization united the two amphipathic helices common among all HLH proteins. The highly conserved hydrophobic interactions were stabilized through numerous van der Waals interactions, which facilitated dimerization.

The crystal structures of class III and class IV HLH domains and leucine zippers have also been elucidated. These classes bind DNA as quasisymmetric heterodimers, stabilized by hydrophobic and polar interactions that involve helix I and helix II as well as residues continuous with helix II located in the leucine zipper. The crystal structure of the Myc–Max heterodimer was particularly informative (Nair and Burley 2003). Myc and Max fold as two heterodimers that are oriented in a head-to-tail configuration, resulting in an anti-parallel four-helix bundle (Fig. 1; Nair and Burley 2003). The SREBP-1a, which modulates cholesterol metabolism, also contains an HLH domain that is continuous with a leucine zipper region (Fig. 1). The SREBP-1a domain is unusual, as it interacts with an asymmetric binding site. The asymmetric binding site recognition is instructed by a tyrosine rather than an arginine residue in the basic region, leading to a loss of polar interactions between the basic region and the major groove (Parraga et al. 1998).

Of particular interest are class VII HLH proteins that contain PAS domains. Prominent among them are CLOCK and BMAL1, which contain ACGTG or GCGTG core sequences in their binding sites. The crystal structure of the CLOCK/BMAL heterodimer revealed an unusual asymmetric heterodimer in which the HLH and PAS domains are intertwined to promote heterodimerization (Fig. 1; Huang et al. 2012). The HIFα–ARNT factors also

Figure 1. Representative structures for an ensemble of HLH domains. Note that for HIF:ARNT dimers, only the HLH domains are shown in color. Other domains are labeled in gray. The two chains are distinguished by dark versus light gray.

form HLH heterodimers and contain PAS domains. Recent crystal studies revealed that helix I and helix II of HIF-α and ARNT are extended and oriented such that the basic region interacts with residues in the major groove to permit interactions between the HIF-α PAS-A domain and the minor groove (Fig. 1; Wu et al. 2015). In the absence of DNA, contiguous interactions formed between the HIF-α HLH and PAS domains, whereas these domains were spatially segregated in ARNT (Fig. 1). The architecture of ARNT is distinct from CLOCK/BMAL heterodimers and may reflect a requirement to form stable HIF/ARNT heterodimers in both the absence and presence of DNA binding. The HIF/ARNT structure also uncovered five potential binding pockets for small ligands in the PAS domains. Notably, one such ligand, proflavine, bound to one of the pockets, consistent with its ability to interfere with HIF-α–ARNT heterodimerization. These findings revealed how HLH proteins can alter their structures in response to environmental signals.

Despite their intricate differences in structure, the overall architecture of HLH proteins displays deep conservation. Their shared features raise questions about how HLH domains associated with multicellular life relate to unicellular HLH proteins. Pho4 is an HLH protein in yeast that regulates the expression of genes encoding for enzymes involved in the biosynthesis of phospholipids and amino acids. The crystal structure of the Pho4 HLH domain showed a four-helix bundle that binds the E-box site as a homodimer (Fig. 1; Shimizu et al. 1997). Except for a short helical region in the loop domain, the general structure of Pho4 appeared remarkably similar to mammalian HLH domains (Fig. 1). Thus, the HLH domain is an ancient DNA-binding domain that arose in unicellular organisms, and, although it expanded and diversified across the animal and plant kingdoms, its structural features essentially remained the same.

HLH genes in the light of evolution

In 1964, Theodosius Dobzhansky noted "Nothing in biology makes sense except in the light of evolution" (Dobzhansky 1964). HLH proteins are no exception. Whereas prokaryotes lack HLH genes, HLH genes have been identified in all eukaryotes, including yeast, fungi, metazoans, and plants. In *Sacchromyces cerevisiae*, eight HLH gene products regulate the expression of genes involved in glycolysis and in phospholipid, phosphate, and amino acid biosynthesis (Robinson and Lopes 2000). In multicellular organisms, the number of HLH genes greatly expanded and diversified (Fig. 2). Detailed sequencing analysis of genomes representing the main metazoan and plant lineages has helped to define when and how HLH families originated and to what extent they are related. These studies showed that an initial diversification of the HLH genes dates to pre-Cambrian times (600 million years ago), prior to emergence of the metazoans (Degnan et al. 2009). This initial diversification was followed by a second massive expansion of HLH genes early in metazoan evolution, culminating in the diversification of bilatarians (animals

with symmetric body plans) and cnidarians (animals associated with radial symmetric body plans).

Deep phylogenetic analysis of metazoan HLH genes revealed a large ensemble of genes that could be segregated into 44 different families (Fig. 2; Simionato et al. 2007). The vast majority of these genes originated in protostomes (animals with bilateral symmetry and three germ layers that undergo spiral cleavage during cell division) and deuterostomes (animals with bilateral symmetry and three germ layers that undergo radial cleavage during cell division). These genes may also have been present in urbitlateria (the hypothetical last common ancestor of the bilaterians living in pre-Cambrian times). Several of these genes are members of the E2A, MyoD, Twist, and achaete–scute families, indicating that these HLH proteins all radiated from ancestors that emerged prior to the diversification of bilaterians and nonbilaterians. During the evolution of bilaterians, the HLH gene pool essentially remained the same.

To obtain mechanistic insight into how HLH genes were retained, gained, or lost during early animal evolution, HLH genes in the genomes of Lophotrochozoans (a highly diverse group of organisms that belong to the bilaterians) were analyzed (Bao et al. 2017). The Lophotrochozoans include the annelids, brachiopods, and the molluscs, a heterogeneous phylum of at least 100,000 species. Analysis revealed that gene duplication played a key role in generating HLH gene diversity. HLH genes that acquired new functions (paralogs) were frequently linked as clusters, some of which represented the "remains" of HLH genes dating back >540 million years ago, prior to the separation of the annelid and mollusc lineages. Others more recently duplicated from ancestral HLH genes and acquired related yet new activities. Notably, HLH genes in the molluscs generally appeared to undergo very few changes at the family level but displayed substantial diversification within families as a result of gene duplication (Bao et al. 2017).

In plants, HLH diversification dates back to 400 million years ago (Fig. 2). A mere four HLH proteins were identified in the genomes of chlorophytes and red algae, whereas a staggering 100–170 HLH genes composing 26 single branches (clades) are present in land plants (Pires and Dolan 2010a). Twenty clades were found in ancestors that preceded extant mosses and vascular plants (Fig. 2; Pires and Dolan 2010b). The establishment of HLH clades early in plant evolution suggests rapid diversification of ancestral species, possibly associated with the movement into new habitats. Although HLH gene gains and losses occurred throughout subsequent plant evolution, the majority of these genes remained conserved, giving rise to HLH genes associated with the genomes of modern plants.

At first glance, the dramatic expansion and diversification of HLH proteins appear to correlate with the generation of cellular diversity. The initial increase in the number of metazoan HLH genes occurred in parallel with the evolution of multicellularity (Fig. 2; Simionato et al. 2007). Analysis of HLH genes in model organisms such as humans, mice, and rats have revealed massive expansion in the number of HLH genes that occurred during
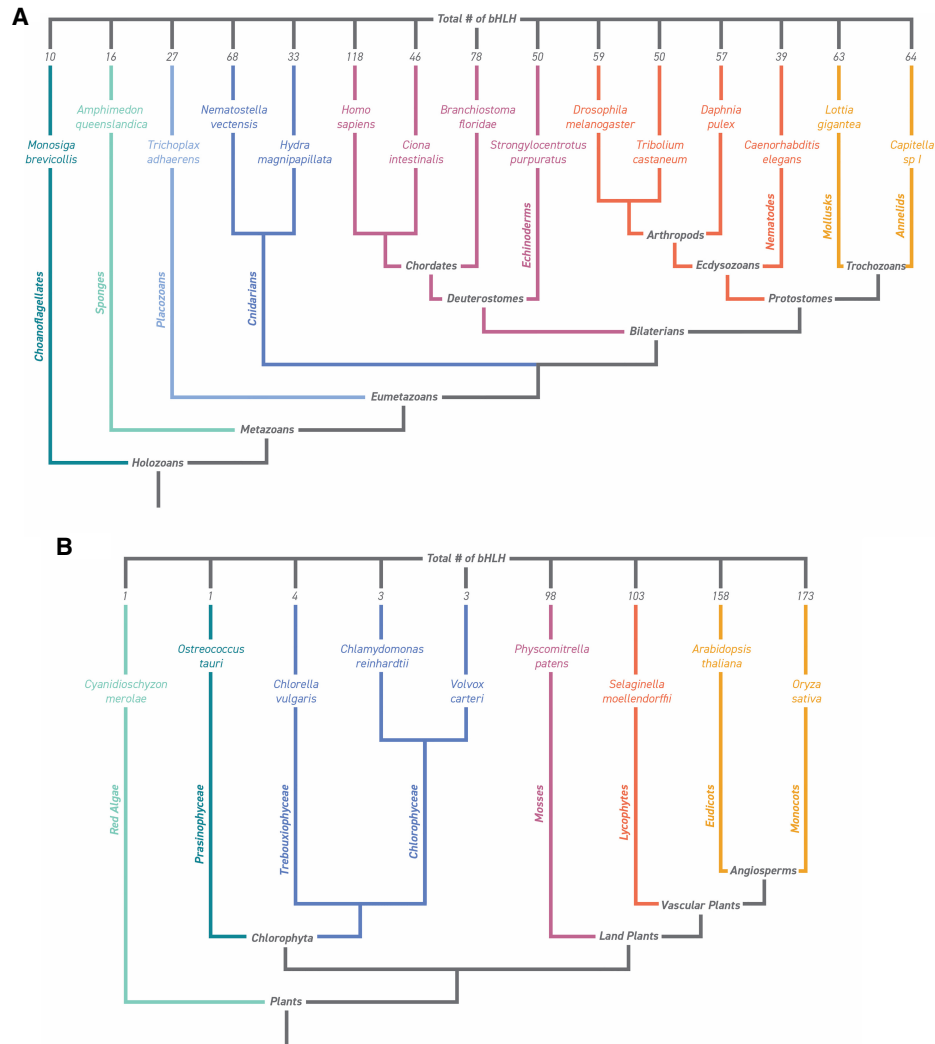
Figure 2. Schematic diagram of HLH genes encoded within the genomes of multicellular organisms. (*A*) Diagram depicting the number of HLH genes associated with holozoans; i.e., animals and closest unicellular relatives but excluding fungi (Simionato et al. 2007). (*B*) Diagram indicating the number of HLH genes associated with the plant kingdom (Pires and Dolan 2010a). Phylogenetic relationships are indicated. Note that the lengths of the connectors in the diagram do not correspond to differences in geological time.

bilateria evolution, which is closely associated with increasing tissue and lineage complexity. The diversification of HLH genes in metazoans and plants may be directly related to the acquisition of multicellular life forms. Locus duplication from an ancestral gene allows genes to be regulated by novel ensembles of regulatory and architectural elements to establish new patterns of gene expression. Combined with the remarkable ability of HLH genes to program and reprogram patterns of gene expression, gene duplication and diversification from ancestral genes may underpin the mechanism by which, at least in part, cellular diversity was generated.

E proteins

E12 and E47 are members of a subset of HLH proteins, branded as E proteins. E proteins also include HEB, E2-2,

daughterless, and HLH-2. E proteins have been characterized predominantly in the context of immune cell development (Bain and Murre 1998). However, they are also involved in other developmental pathways, including *D. melanogaster* sex determination and neurogenesis and gonadogenesis in *C. elegans* (Caudy et al. 1988; Sallee et al. 2017). During embryogenesis, E47 promotes cortical plate neural development and, in adult neural precursor cells, represses an astrocyte-specific gene program. In contrast, the E-protein antagonist Id3 orchestrates BMP2-induced astrocyte differentiation (Pfurr et al. 2017). Upon cortical brain injury, Id3 is induced by BMP2 to neutralize E47 activity and promote astrocyte differentiation (Bohrer et al. 2015). The E2A proteins also function as tumor suppressors and are associated with a wide variety of chromosomal translocations associated with childhood leukemias (Bain et al. 1997; Aspland et al. 2001). Notably, HEB has the ability to functionally replace E2A in supporting B-

cell development (Zhuang et al. 1998). These observations support the notion that gene duplication from ancestral genes may have occurred in part to permit genes being placed under novel regulatory elements rather than providing de novo biochemical activities.

E-proteins bind as homodimers or heterodimers to DNA. They are not lineage-restricted, but their mRNA levels vary among different cell types (Murre et al. 1989a). E-protein levels are regulated by multiple mechanisms, such as dimerization-induced degradation (Sallee and Greenwald 2015). E proteins contain three highly conserved regions, named AD1, AD2, and AD3 (Aronheim et al. 1993; Massari et al. 1996; Zhang et al. 2004; Chen et al. 2013). The AD1 domain is present in all E proteins, folds into a helical structure, and contains a four-amino-acid motif named LDFS (Massari et al. 1996). The LDFS motif is also present in HLH proteins, including Rtg3, a nutrient sensor in *S. cerevisiae*. The AD1 domain regulates E-protein function by recruiting the histone acetyltransferase (HAT) P300/CBP to acetylate lysine residues across the histone tails (Bradney et al. 2003; Zhang et al. 2004). Acetylated lysine residues at histone tails recruit BRD4, a member of the bromodomain and extraterminal (BET) chromatin reader proteins (Nguyen et al. 2014). H3K27ac-marked histones interact with chromatin remodelers such as BRG1 to evict nucleosomes across E-protein-bound enhancer elements and provide DNA accessibility to cooperating transcription factors (Bossen et al. 2015). The AD1 domain also has the ability to repress downstream target gene expression by recruiting members of the ETO family (Zhang et al. 2004). Three closely related ETO members (ETO, ETO-2, and MTGR1) interact with histone deacetylases (HDACs), including HDAC1 and HDAC3. Finally, as mentioned above, E proteins selectively recruit CBP/P300 and/or ETO members (HATs vs. HDACs) to enhancer regions to acetylate and/or deacetylate lysine residues. E-protein occupancy is also closely associated with demethylation of CpG residues across the enhancer landscape (Benner et al. 2015; Lio et al. 2016; Orlanski et al. 2016). In fact, compelling studies have demonstrated that the E2A proteins promote demethylation by direct recruitment of the ten-eleven translocation (TET) proteins (Lio et al. 2016).

How does E-protein occupancy at the mechanistic level activate lineage-specific programs of gene expression? Recent studies suggest that E proteins orchestrate enhancer-promoter communication at multiple levels, including geometric confinement (loop domains) and phase separation. E-protein-induced geometric confinement may involve noncoding transcription-induced loop extrusion (Isoda et al. 2017). E2A protein occupancy is closely associated with the activation of noncoding transcription, which in turn leads to recruitment of cohesion (Lin et al. 2012; Isoda et al. 2017). Once loaded, the cohesin complex activates the loop extrusion process until two CTCF sites are reached, sequestering superenhancers and promoter regions into a single loop domain (Isoda et al. 2017). E proteins may also act to orchestrate phase separation across superenhancers. It is well established that the E proteins recruit P300/CBP at both the biochemical and genome-wide levels (Bradney et al. 2003; Zhang et al. 2004; Lin

et al. 2012). P300 acetylates lysine residues at H3 and H4 histone tails. The bromodomain-containing protein BRD4 binds multiple acetylated lysine residues on H3 and H4 tails (Nguyen et al. 2014). BRD4 also contains large intrinsically disordered regions (Sabari et al. 2018). These two properties may allow BRD4 to promote phase separation across loop domains associated with E-protein and P300 occupancy. Finally, E proteins themselves may self-organize into phase-separated droplets, as demonstrated recently for other HLH proteins such as c-Myc (Boija et al. 2018). Indeed, they are present in at least two distinct physical states and accumulate into droplets during developmental progression (Quong et al. 1999). E proteins interact with a wide spectrum of transcriptional regulators, coactivators, chromatin remodelers, TET proteins, and components of the splicing machinery (Teachenor et al. 2012; Lio et al. 2016). How such a large ensemble of factors interacts with the E2A proteins is an enigma but likely involves a spectrum of weak electrostatic and hydrophobic interactions that link disordered domains in a seemingly promiscuous fashion that collectively promotes a gel-like state (Boija et al. 2018). Thus, E proteins may act as a scaffold to recruit BRD4 to clusters of enhancers, concentrating an onset of biochemical reactions and ultimately activating lineage-specific gene programs. In sum, the key roles for E proteins in orchestrating gene expression involve (1) the activation of noncoding transcription and (2) the establishment of a phase-separated state across loop domains to facilitate long-range enhancer-promoter communication.

Id proteins

The DNA-binding activities of E proteins are counteracted by the Id gene products (Benezra et al. 1990). Four Id proteins, named Id1–4, have been identified in mammalian genomes. The Id proteins contain an HLH domain but lack a basic region and, upon heterodimerization, neutralize the DNA-binding activities of bHLH proteins. That Id proteins are involved primarily in targeting E proteins was revealed in studies showing that a decline in E-protein activity overcomes the need for Id expression to promote developmental progression (Yan et al. 1997; Rivera et al. 2000; Boos et al. 2007; Miyazaki et al. 2011, 2017; Zook et al. 2018).

The Id proteins serve in a wide spectrum of developmental pathways in both health and disease (Lasorella et al. 2014). They function predominantly by modulating cell cycle progression, developmental progression, and tumor suppression (Lyden et al. 1999; Yokota et al. 1999; Schmitz et al. 2012; Lasorella et al. 2014; Miyazaki et al. 2015). Genes that are regulated by the E–Id module relating to cellular expansion and cell growth involve many, including p15, p16, p27, p27, p57, and cdk6 as well as von Hippel-Lindau (VHL) and c-myc (Prabhu et al. 1997; Pagliuca et al. 2000; Schwartz et al. 2006; Lasorella et al. 2014; Lee et al. 2016). In immune cell development, the E–Id protein axis regulates the expression of an armamentarium of genes that differ between cell types and developmental

stages. They include genes encoding for transcription factors, components of signaling pathways, antigen receptors, chemokine receptors, DNA repair factors, enzymes involved in somatic recombination, and so on.

Id transcription is regulated by a wide ensemble of receptors, including the T-cell receptor (TCR), B-cell receptor (BCR), TGFβ, BMP, FGF, and cytokine receptors (Bain et al. 2001; Kee et al. 2001; Ying et al. 2003; Yang et al. 2011). In embryonic stem cells, Id1 expression is activated by BMP involving the SMAD pathway to promote self-renewal and in the vasculature (Ying et al. 2003; Yang et al. 2014). The regulation of Id2 and Id3 expression in developing T cells is dynamic and complex. During thymocyte selection, Id3 expression is activated by pre-TCR and TCR signaling involving the ERK–MAPK–EGR1 pathway (Bain et al. 2001). In CD8 T cells, upon stimulation of the TCR, Id2 and Id3 expression rapidly declines, while E2A levels increase. Id2 levels are elevated again at the height of infection in short-lived memory and effector CD8 T cells to orchestrate survival and terminal differentiation (Omilusik et al. 2018). Id3 abundance, on the other hand, gradually increases as cells differentiate toward the long-lived memory CD8 T-cell pool (Yang et al. 2011; Omilusik et al. 2013). Regulatory T (Treg) cells express high levels of Id3 that rapidly decline upon activation, whereas Id2 levels rise (Miyazaki et al. 2014). Likewise, in naïve B cells, Id3 levels decline upon triggering of the BCR to initiate a germinal center-specific program of gene expression (Chen et al. 2016; Gloury et al. 2016). Thus, a common feature that arises is that, in adaptive immune cells, antigen receptor-mediated signaling activates or suppresses Id expression to modulate E-protein DNA-binding activity to orchestrate developmental progression beyond the checkpoint.

The Id proteins are small proteins that are highly conserved and remarkably similar. Interfering with the expression of two or more Id genes leads to embryonic lethality (Lyden et al. 1999). They display similar if not identical dimerization specificities (Prabhu et al. 1997). Id proteins are short-lived, with half-lives <20 min. Their protein levels are regulated by the ubiquitin–proteasome pathway. Both E3 ligases and deubiquitylases dictate Id abundance. Id1, Id2, and Id4 all carry a D box in their C-terminal domain that is targeted by the anaphase-promoting complex/cyclosome (APC/C–CDH1) E3 ubiquitin ligase (Lasorella et al. 2006). The APC/C–CDH1 complex targets these Id proteins for degradation to promote cell cycle arrest, whereas the ubiquitin-specific protease 1 (USP1) modulates Id protein abundance to control cell growth (Williams et al. 2011). The stability of Id proteins is also critically dependent on partner choice. The *Drosophila* Id homolog Emc has a short half-life unless it forms heterodimers with daughterless (Li and Baker 2018). Pairing daughterless with proneural gene products rather than with Emc leads to rapid depletion of the Emc pool. Although Emc transcript levels are uniform across the progenitor cell population, Emc protein levels are dictated by competition for heterodimerization with the proneural bHLH proteins. Thus, Id expression is regulated at multiple levels, involving both transcriptional and post-transcriptional inputs to modulate cell growth and developmental progression.

The Myc network

The Myc proteins were among the first HLH proteins to be characterized in molecular detail. The founding member of this family is v-Myc, originally identified in retroviruses associated with the development of animal cancers and acquired from a cellular locus named c-Myc (Sheiness et al. 1978; Roussel et al. 1979; Sheiness and Bishop 1979). c-Myc controls cell growth, differentiation, metabolism, and death but, when aberrantly expressed, readily promotes the development of a wide spectrum of malignancies. (Carroll et al. 2018). Myc binds DNA as a heterodimer with its partners, Max and Mxd (Blackwood and Eisenman 1991; Ayer et al. 1993). Myc interacts with a coactivator complex named TRRAP to recruit HATs and chromatin remodelers to activate gene expression, whereas the Mxd–Max heterodimer recruits mSin3, which serves as a scaffold for HDACs (HDAC1 and HDAC2) to silence transcription (Ayer et al. 1995; McMahon et al. 1998). The main function of the Myc–Max heterodimers is to promote cell proliferation, metabolism, and size. Myc orchestrates proliferation by activating the expression of genes encoding for cell cycle regulators. Myc regulates cellular metabolism by regulating the expression of an armamentarium of metabolic enzymes and transporters affecting glutamine and glucose metabolism to support lipid, amino acid, and nucleotide biosynthesis (Zhang et al. 2007; Wang et al. 2011). Indirectly, c-Myc influences chromatin topology at a global scale by regulating ATP levels to modulate loop extrusion (Kieffer-Kwon et al. 2017). Although most of the genes regulated by c-Myc are transcribed by RNA polymerase II, c-Myc also activates the expression of genes transcribed by RNA polymerases I and III. Prominent among these are the transfer RNA (tRNA) and 5S RNA genes (Gomez-Roman et al. 2003) and ribosomal genes RNA polymerase I (genes encoding ribosomal RNA) (Arabi et al. 2005; Grewal et al. 2005). Most recently, elegant studies demonstrated that the c-myc proteins have the ability to form phase-separated droplets in conjunction with the MED1 subunit of the Mediator complex (Boija et al. 2018). How gelation of the Myc proteins and related proteins modulates cell growth and transformation will be an area of intense investigation for years to come.

Myogenic HLH proteins

MyoD was the first protein identified that programs transdifferentiation (Davis et al. 1987). Since its initial discovery, three MyoD-related HLH proteins have been identified, including Myf5, Mrf4, and myogenin. MyoD, Myf5, and Mrf4 expression acts coordinately to specify muscle cell fate, whereas myogenin regulates the terminal differentiation of myoblasts (Hasty et al. 1993; Nabeshima et al. 1993; Rudnicki et al. 1993; Kassar-Duchossoy et al. 2004). Interestingly, unlike MyoD/Myf-5 double-deficient

mice, the *D. melanogaster* MyoD homolog nautilus orchestrates the development of only a subset of muscle progenitor cells (Enriquez et al. 2012). Thus, while the structures, expression patterns, and functions of many HLH proteins have been remarkably well conserved throughout evolution, the role of some bHLH proteins in cell type specification has changed.

E-box sites are widespread. There are literally thousands of E-box sites that span the genome. This raises the question of how subsets of E-box-binding sites are selected. For MyoD and NeuroD2, functional specificity is dictated by differences in binding site specificities. Both factors share common binding sites but also are associated with unique binding site specificities, and it is these sites that are more closely associated with lineage-specific transcription signatures (Fong et al. 2012). This subset of MyoD-binding sites was associated with composite binding sites for the homeodomain-containing proteins PBX and MEIS (Fong et al. 2015). Notably, it was shown that MyoD can be converted to a neuronal differentiation factor by preventing its ability to interact with PBX1, demonstrating that binding site selection can dictate functional specificity (Fong et al. 2015).

Once bound to DNA, MyoD recruits P300/CBP to promote (akin to E proteins) acetylation of lysine residues at the tails of H3 and H4 (Puri et al. 1997). MyoD-induced H3K27ac in turn recruits the SWI/SNF chromatin remodeling machinery to orchestrate nucleosome depletion and promote chromatin accessibility (Forcales et al. 2012). MyoD also has the ability to bind E-box sites associated with a closed chromatin environment but only upon interacting with the homeobox-containing protein PBX (Berkes et al. 2004). MyoD and myogenin exhibit unique regulatory roles at similar ensembles of target genes (Cao et al. 2006). At immediate targets, MyoD activity is sufficient to activate a full program of gene expression (Cao et al. 2006). However, at later genes, MyoD expression is not sufficient for activating downstream target gene transcription. Rather, it requires myogenin for full activation (Cao et al. 2006). These data demonstrate that closely related bHLH proteins perform unique roles, and not all of them are associated with pioneering activities. Distinctive motifs outside but adjacent to the bHLH domain seem responsible for this feature. Likewise, Myf-5 and MyoD bind overlapping sites yet have distinct functional features. Whereas Myf-5 induces histone acetylation at H4 in the absence of polymerase II recruitment, MyoD promotes histone acetylation at H4 and recruitment of RNA polymerase II (Conerly et al. 2016). Thus, MyoD and Myf5 are associated with the same DNA-binding site preference but have diverged at overlapping binding sites by segregating distinct steps in gene activation. Collectively, these as well as other observations indicate that Myf5 is predominantly a chromatin remodeler, Myogenin is primarily a transcriptional activator, and MyoD performs both functions. Consequently, Myf5 might act, at least in part, by enabling subsequent activation of a myogenic gene program by the combined activities of myogenin and MyoD (Conerly et al. 2016).

The neuronal HLH proteins

Prominent among the bHLH proteins that promote neurogenesis and specify neural cell fate are the proneural proteins (Cubas et al. 1991). In *Drosophila*, the proneural HLH proteins are first activated in a group of cells named the proneural cluster. Amid the proneural cluster, the cell with the highest abundance of proneural bHLH protein establishes neural identity (Doe and Goodman 1985; Cabrera 1990; Hartenstein and Posakony 1990). Paradoxically, upon reaching its maximum abundance, proneural gene expression is silenced. The decline in proneural protein abundance is a key step, since persistent abundance of these factors severely perturbs terminal differentiation (White and Jarman 2000). How is proneural abundance regulated upon reaching its highest levels of expression? An evolutionarily conserved post-translation mechanism that readily switches proneural activity on and off in progenitor cells may be key. At peak levels, proneural gene expression is switched off by phosphorylation of a single amino acid (serine or threonine) residue located in the atonal, scute, and Neurogenin-2 HLH domains (Quan et al. 2016). Phosphorylation of residues located in the proneural bHLH domains suppresses the activity of the proneural bHLH proteins to permit terminal differentiation. In a process named lateral inhibition, a single cell will differentiate, while neighboring cells are prevented from also adopting the neural cell fate. Lateral inhibition is mediated by the Notch signaling pathway, which, upon ligand engagement, activates the expression of bHLH genes located in the Enhancer of split complex (Bailey and Posakony 1995; Lai et al. 2000). Enhancer of split complex gene products in turn divert cells to adopt the epidermal cell fate by suppressing a neural-specific transcription signature. These well-designed studies show how proneural bHLH proteins dictate the epidermal versus neuronal cell fate choice among neighboring cells.

In murine neural progenitors, the proneural proteins also play essential roles. Hes1 and Hes5 maintain the neural stem cell pool by repressing proneural gene expression (Nakamura et al. 2000). The proneural bHLH proteins Ascl1 and Neurogenin-2 activate a neuronal cell-specific transcription signature and antagonize the expression of an astrocytic-specific gene program, whereas Olig2 specifies the oligodendrocyte cell fate (Sun et al. 2001; Lu et al. 2002). All three genes are expressed in neural progenitors, raising the question of how distinct cell fates emerge amid competing pathways. Recent studies have been revealing. The self-renewal activities of Ascl1, neurogenin, and Olig2 were closely associated with oscillatory patterns of gene expression (Imayoshi et al. 2013). However, upon differentiation into neurons, Asc1 expression was sustained rather than oscillatory, and the expression of neurogenin and Olig2 was suppressed. Olig2 and Ascl1 expression was persistent under conditions that promote oligodendrocyte or astrocytic cell fate. Furthermore, optogenetic modulation of Ascl1 expression revealed that oscillatory Ascl1 expression was essential to promote self-renewal, whereas persistent Ascl1 expression orchestrated a neuronal gene program (Imayoshi et al. 2013). These

elegant studies show how the dynamics of bHLH gene expression, rather than simple abundance, dictates developmental choice.

HLH gene expression in morphogenesis

The key morphogenetic events in the early embryo, including gastrulation, germ layer formation, and somitogenesis, are all executed by HLH proteins that act coordinately with other transcriptional regulators. For example, in *D. melanogaster*, Twist activates a mesodermal-specific program of gene expression, whereas the zinc finger-containing protein Snail suppresses the transcription of genes associated with ectodermal cell fate. In vertebrates, Twist proteins promote neural crest tube closure by modulating cell migration and differentiation of neural crest and mesenchymal progenitor cells. Intriguing studies have demonstrated that the Twist proteins may impact metastasis. Twist expression is particularly abundant in cells that are metastasizing, and loss of Twist expression was shown to suppress the ability of tumor cells to intravasate into the blood stream (Yang et al. 2004). More recent studies provided a new twist to Twist function. A series of elegant experiments showed that Twist1 activates a cancer stem cell-specific program of gene expression in both skin and mammary cells that is independent of its role in orchestrating epithelial–mesenchymal transition and tumor infiltration (Beck et al. 2015).

The Twist proteins Hand1 and Hand2 also orchestrate cardiac myocyte development (Conway et al. 2010). Hand proteins are typical HLH proteins that form heterodimers with the Twist proteins, are partially complementarily expressed in the developing heart, and are coexpressed in the cardiac outflow tract. Mice deficient for both Hand 1 and Hand2 display severe cardiac defects (George and Firulli 2018). How the Twist proteins modulate seemingly unrelated pathways such as the epithelial–mesenchymal transition, tumor stemness, and cardiomyocyte identity deserves further scrutiny.

Somitogenesis is also controlled and enforced by the bHLH proteins. Segmentation occurs very early during embryogenesis through the formation of somites that give rise to skeletal muscle and the vertebrae. During somitogenesis, the unsegmented presomitic or paraxial mesoderm progressively generates epithelial somites in an anterior-to-posterior direction. The process of somite segmentation in mice is repeated every 2 h and is again dictated by HLH proteins (Bessho et al. 2001). Hes7 expression is initiated from the posterior and then progresses into the anterior regions of the presomitic mesoderm. In presomitic cells, Hes7 transcription oscillates, with each cycle of expression giving rise to a pair of somites. In the absence of Hes7, somites, vertebrae, and ribs fail to segregate (Bessho et al. 2001). Likewise, sustained Hes7 expression readily leads to fused somites, consistent with the idea that oscillating patterns of Hes7 expression serve to orchestrate somitogenesis (Fig. 3). The oscillating pattern of Hes7 expression is controlled by a feedback mechanism. Interference with the feed-
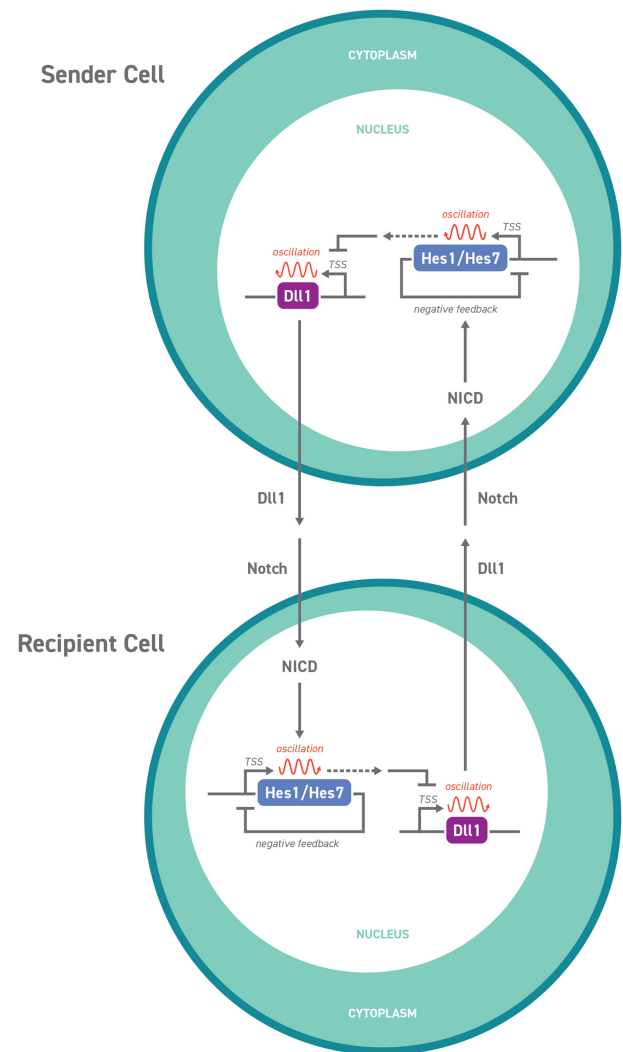
Figure 3. A molecular oscillator instructs somitogenesis. Oscillating patterns of HES 1/7 gene expression are transmitted between sender and recipient cells by the Notch–Dll1 pathway. Upon triggering by binding, the Dll1 the Notch intracellular domain (NICD) is released to activate Hes1/7 expression. Hes1/7 expression is regulated by a negative feedback circuitry that results in an oscillating pattern of gene expression. Hes1/7 expression in turn suppresses Dll1 expression to orchestrate a bidirection oscillating pattern of Dll1 expression between sender and recipient cells (Kageyama et al. 2018).

back mechanism causes sustained Hes7 expression and fusion of the somites. A remarkable series of experiments has explored how neighboring cells in the somites coordinate oscillating patterns of Hes7 gene expression (Fig. 3; Shimojo et al. 2016). Briefly, oscillation was transmitted between neighboring cells by Notch–DL1-mediated signaling. Dampening the expression of the Notch ligand DLl1 interfered with Hes7 periodic transcription and blocked somitogenesis. Thus, the Notch signaling module coordinates oscillating patterns of proneural bHLH gene expression between sender and recipient cells to generate a segmented body plan (Fig. 3).

These studies are provocative, raising the possibility that similar mechanisms operate in other tissues, such as the lymphoid organs, where a wide ensemble of cell types interacts to transmit information between sender and recipient cells to orchestrate developmental progression and generate an effective immune response.

HLH–PAS proteins

The dominant players in the circadian clock are the HLH and PAS domain-containing proteins CLOCK and BMAL1 (Bargiello and Young 1984; Reddy et al. 1984, Zehring et al. 1984). The HLH and PAS domains of circadian clock proteins are intertwined to promote heterodimerization (Fig. 1; Huang et al. 2012). Their binding sites contain ACGTG or GCGTG core sequences. CLOCK and BMAL form heterodimers to induce the expression of their transcriptional repressors: Per1, Per2, Cry1, and Cry2 (Fig. 4). During the early afternoon and late evening, CLOCK/ BMAL1 heterodimers bind in a rhythmic fashion to E-box target sites, where they evict nucleosomes to activate Per1 and Cry1 transcription (Allada et al. 1998; Darlington et al. 1998; Rutila et al. 1998). PER and CRY proteins accumulate in the cytoplasm, which then form heterodimers that enter the nucleus (Fig. 4; Vosshall et al. 1994). During late night and early morning, PER/CRY heterodimers form a higher-order complex with CLOCK/ BMAL heterodimers to suppress downstream target gene expression. Next, PER undergoes a series of phosphorylation modifications that promotes its degradation and release from the CLOCK/BMAL activator (Fig. 4). The CLOCK/BMAL heterodimer then binds its cognate E-box-binding sites to repeat the process of evicting nucleo-some at its targets. Thus, the core transcriptional HLH activators CLOCK and BMAL1 drive the expression of their own suppressors (Per1–3 and Cry1–2) to generate a remarkably well-conserved negative feedback loop (Reppert and Weaver 2001).

HLH proteins that contain PAS domains also serve as oxygen sensors. Specifically, the key players in the cellular response to hypoxia are the HIF proteins (Wang et al. 1995). There are three HIF proteins (named HIF-1α, HIF-2α, and HIF-3α) that readily form heterodimers with the ARNT (HIF-1β) proteins to bind to a distinct class of E-box-binding sites (Semenza 2012). HIF-1α and HIF-1β serve as oxygen sensors, whereas the role of HIF-3α is less well understood (Keith et al. 2012). Under normal oxygen levels, two proline residues in the HIF-1α and HIF-2α oxygen-dependent degradation domain are hydroxylated by prolyl hydroxylase domain (PHD)-containing proteins (Bruick and McKnight 2001). The activity of PHD proteins requires access to both oxygen and α-ketoglutarate. Once proline residues are hydroxylated, the HIF-1α and HIF-2α proteins are targeted to the ubiquitin–proteasome pathway for degradation mediated by VHL tumor suppressor protein-dependent ubiquitination (Zhang et al. 2007; Kaelin and Ratcliffe 2008). Specifically, VHL recruits an E3 ubiquitin–protein ligase complex that catalyzes a covalent interaction of ubiquitin to HIF-1α residues, serving as a signal for degradation. However, under hypoxic conditions, the HIF proteins avoid proteolysis and readily dimerize with the ARNT proteins, activating programs of gene expression linked with angiogenesis and erythropoiesis (Semenza 2012). HIF proteins also activate the expression of genes associated with glucose transport and glycolysis to provide metabolic needs for cells growing in hypoxic conditions (Girgis et al. 2012).
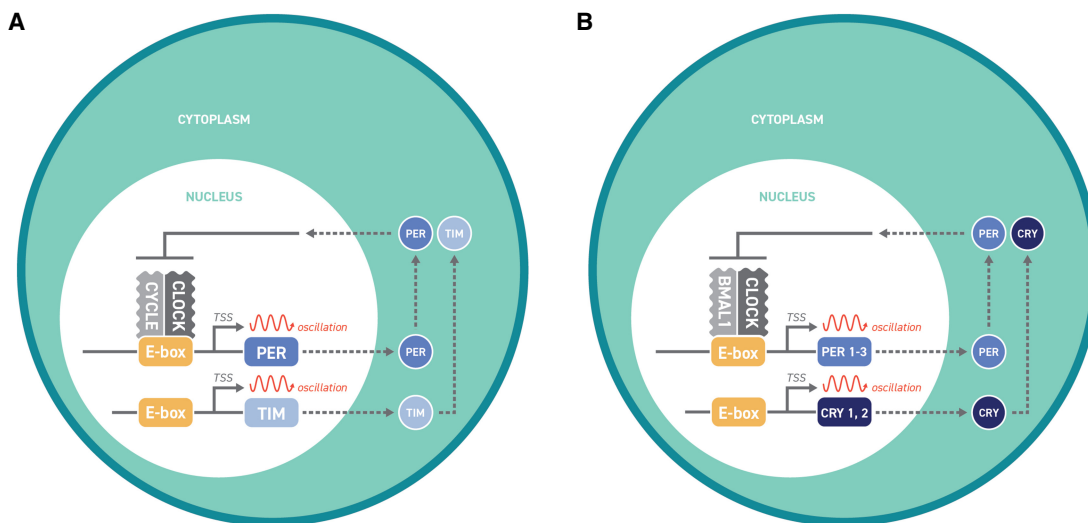

Figure 4. Role of HLH proteins in orchestrating circadian gene expression. (*A*) The *D. melanogaster* HLH proteins CLOCK and CYCLE activate the expression of PER and TIM. PER and TIM transcripts are exported to the cytoplasm for translation and heterodimer formation. They enter the nucleus to suppress CLOCK:CYCLE-mediated transactivation, generating a negative feedback loop with oscillating patterns of gene expression. (*B*) The mammalian HLH proteins CLOCK and BMAL activate the expression of PER and CRY. PER and CRY transcripts are exported to the cytoplasm for translation and heterodimer formation. They enter the nucleus to suppress CLOCK:BAML-mediated transactivation, generating a negative feedback loop with oscillating patterns of gene expression (Lowrey and Takahashi 2004).

HLH proteins in hematopoiesis

Together with other transcriptional regulators, HLH proteins specify the fates of virtually all immune cell types. HLH protein activity begins in hematopoietic stem cells (HSCs), which give rise to all other blood cells. At least four HLH proteins—SCL, LYL1, E2A, and Id1—are involved at this stage. SCL specifies the HSC fate, whereas SCL, LYL1, E2A, and Id1 maintain the HSC compartment (Fig. 5; Shivdasani et al. 1995; Semerad et al. 2009; Souroullas et al. 2009; Singh et al. 2018). Prominent in orchestrating early erythropoiesis is a transcription factor complex containing the HLH proteins E2A and SCL as well as an adaptor protein named LMO2 that interacts with SCL (Fig. 5; Soler et al. 2010). Recently, the structure of the E2A:SCL:LMO2 ternary complex was resolved (Omari et al. 2013). It was an eye-opener. It showed how adaptor proteins have the ability to increase DNA-binding site specificity of HLH heterodimers that, by themselves, display little DNA-binding selectivity. Specifically, upon interacting with SCL, LMO2 released new hydrogen bonds in the SCL:E47 heterodimer that strengthened heterodimerization but also induced rotation in E47 (Omari et al. 2013). The rotation in E47 altered the binding site preference for the ternary complex to such a degree that much of the complex site preference was contributed by another factor, named GATA-1 (Omari et al. 2013; Hewitt et al. 2016). This is a cardinal finding, since it shows at the atomic level how HLH proteins have the ability to select their target sites with great precision.

HLH proteins also promote the development of intermediate progenitors affecting virtually the entire spectrum of early hematopoiesis. Specifically, LYL1 promotes the development of T-cell progenitors, whereas the E2A proteins, together with Ikaros and PU.1, help direct the developmental progression of the lymphoid-primed multipotent progenitors (LMPPs) and their differentiated progeny, which include macrophage–dendritic progenitor cells (MDPs), granulocyte–macrophage progenitors (GMPs), and common lymphoid progenitors (CLPs) (Dias et al. 2008; Semerad et al. 2009; Zohren et al. 2012). MDPs give rise to macrophages and two distinct types of dendritic cells: antigen-presenting classical dendritic cells (cDCs) and interferon-producing plasmacytoid dendritic cells (pDCs). pDC development is controlled by E2-2, whereas cDC maturation is orchestrated in part by Id2 (Fig. 5; Cisse et al. 2008). In cDCs, Id2 neutralizes E2-2 activity and may also modulate dendritic cell development via other pathways. E2-2 directly activates a dendritic-specific program of gene expression by recruiting P300/CBP and, in pDCs, acts in concert with the transcriptional corepressor ETO (MTG16) to suppress the cDC-specific gene program. Id2 expression in pDCs is directly suppressed by the combined activities of E2-2 and MTG16 (Fig. 5; Grajkowska et al. 2017).

CLPs give rise to innate lymphoid cells (ILCs), B-lineage cells, and T-lineage cells. ILC development does not require E2A, E2-2, or HEB activity, although E-protein expression in the thymus is essential to suppress ILC2 development (Fig. 5; Miyazaki et al. 2017; Wang et al.
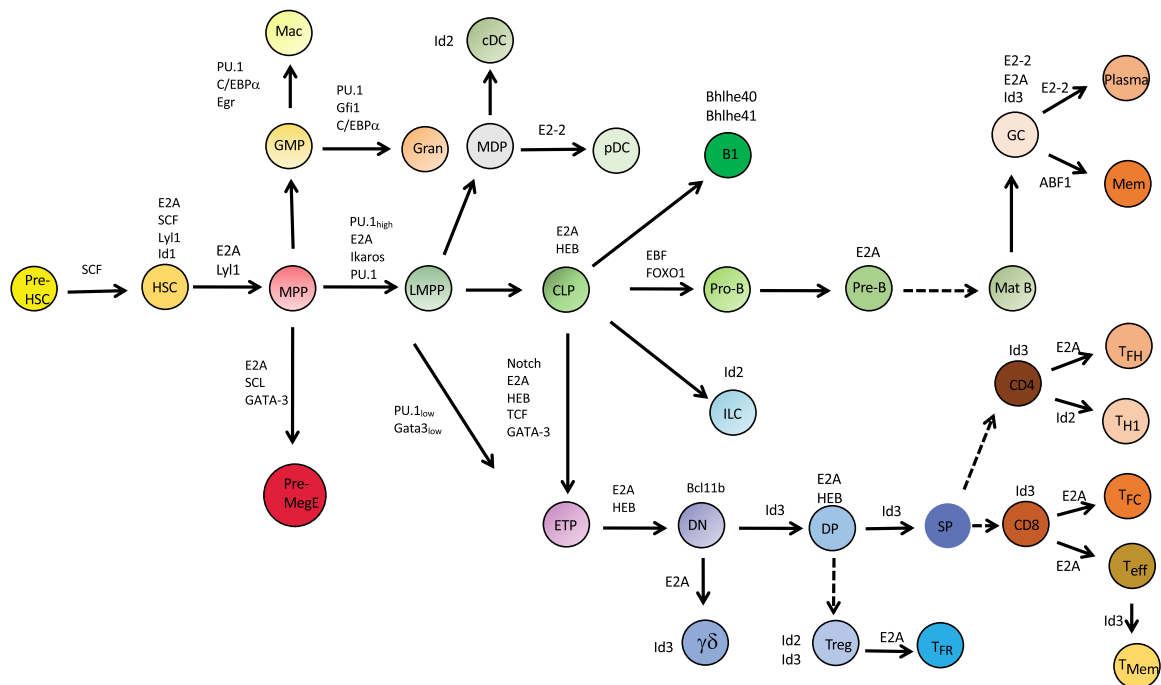
Figure 5. HLH proteins and the generation of immune cell diversity. The role of HLH proteins in hematopoiesis and immune cell development is indicated. Dashed lines indicate developmental transitions likely involving multiple intermediates and transitions between primary and secondary lymphoid compartments.

2017). Id2 defines ILCs and is essential for ILC lineage progression, suppression of a stem cell gene program, and maintenance of long-term identity (Huang et al. 2017). Mature natural killer (NK) cells, a subset of ILCs, readily developed in the absence of Id2 but activated the expression of genes associated with a naïve CD8 gene program and consequently failed to mature into cytotoxic effector cells (Zook et al. 2018). Notably, the depletion of Id2 abundance was compensated for by the induction of Id3 expression, acting to neutralize E-protein activity (Zook et al. 2018). These data as well as other observations showed how an intricate circuitry involving carefully balanced E-protein and Id-protein abundance dictates the progression of maturing immune cells (Boos et al. 2007; Miyazaki et al. 2011, 2017; Zook et al. 2018).

HLH proteins and B-cell diversity

Early B-cell development relies on tightly regulated degrees of E-protein activity. In a subset of CLPs, E47 abundance is elevated (through mechanisms yet to be determined) to initiate development toward the B-cell fate (Fig. 5; Bain et al. 1994; Zhuang et al. 1994, 1996; Jaleco et al. 1999; Beck et al. 2009). Next, E2A and HEB activate the expression of FOXO1, which together induce EBF1 expression (Inlay et al. 2009; Mercer et al. 2011; Welinder et al. 2011; Miyai et al. 2018). EBF1 and FOXO1 then act in a feed-forward loop to orchestrate a B-lineage-specific gene program and suppress the expression of genes associated with alternative cell lineages (Mansson et al. 2012; Li et al. 2018). E2A proteins directly regulate critical B-cell targets, including genes encoding for components of the pre-BCR such as λ5, VpreB1-3, CD79a, and CD79b and the recombination-activating genes *Rag1* and *Rag2* (Kwon et al. 2008; Lin et al. 2010; Mercer et al. 2011).

E proteins also shape the topology of the immunoglobulin heavy and light chain loci of early B cells. E proteins bind to a spectrum of sites across the Igh and Igκ loci, which then sequester CBP/p300 and BRG1 to remodel the local chromatin landscape (Bradney et al. 2003; Bossen et al. 2015). These partnerships ultimately induce elaborate genomic interactions across E2A-bound sites, plausibly mediated by BRD4 by interacting with P300-induced acetylation of lysine residues at H3 and H4 tails (Lin et al. 2012). Upon pre-BCR expression, E47 abundance temporarily declines to promote cellular expansion but rises again in small pre-B cells to establish a network of cross-links that spans the entire Igκ V region cluster, ultimately activating Igk VJ rearrangement (Fig. 5; Romanow et al. 2000; Goebel et al. 2001; Inlay et al. 2004; Quong et al. 2004; Lin et al. 2012; Bossen et al. 2015). Later, in immature B cells, the E2A proteins enforce another important developmental checkpoint. In the presence of autoreactivity, E2A abundance remains high to permit continued Igκ and Igλ rearrangement (Quong et al. 2004; Beck et al. 2009). However, upon completion of an innocuous immunoglobulin receptor, E2A abundance drops, preventing continued rearrangement (Fig. 5). Collectively, these data reveal a carefully tuned developmental process

in which E2A proteins play a central role: High E2A abundance establishes a network of cross-links across the Igκ locus that not only initiates VJ rearrangement but also permits continued receptor editing until an innocuous receptor has been generated (Beck et al. 2009).

E proteins have equally important and nuanced roles in mature B cells. In the spleen and lymph nodes, follicular B cells express high levels of E2A. In contrast, marginal zone B cells require high levels of Id3 for their development (Quong et al. 2004). In naïve mature B cells, E2A abundance is low, but, upon activation, E2A levels readily accumulate to promote germinal center B-cell development, induce AID expression, and promote class switch recombination (Fig. 5; Quong et al. 1999; Sayegh et al. 2003; Chen et al. 2016; Gloury et al. 2016; Wöhner et al. 2016). Likewise, the Id proteins modulate the germinal center reaction. In activated B cells, Id3 abundance declines, releasing E proteins from their inhibitors to induce the expression of genes associated with BCR and cytokine receptor-mediated signaling (Chen et al. 2016; Gloury et al. 2016). In maturing B cells, E2A as well as E2-2 continue their activities to promote plasma cell differentiation, whereas the HLH protein ABF1 instructs germinal center cells to adopt the memory B-cell fate (Fig. 5; Massari et al. 1998; Chiu et al. 2014; Gloury et al. 2016; Wöhner et al. 2016). Finally, two additional HLH proteins—Bhlhe40 (Dec2) and Bhlhe41 (Dec1)—orchestrate the self-renewal activity and development of B1 cells (Fig. 5; Kreslavsky et al. 2017). The remarkably ability of E proteins to orchestrate B-cell development is a consequence of their ability to activate or silence hundreds of target genes that differ from each other at each developmental step—a daunting task that must involve many other factors such as EBF1, PAX5, FOXO1, IRF4, IRF8, PU.1, and others yet to be identified.

HLH proteins and the generation of T-cell diversity

E proteins also specify T-cell identity. E2A and HEB in early T-cell progenitors activate the expression of Notch1 (Fig. 5; Bain and Murre 1998; Schotte et al. 2010; Miyazaki et al. 2017). Notch signaling then coordinately acts with E proteins to specify T-cell fate. How precisely Notch signaling and E proteins establish T-cell identity remains to be determined, but recent work points to a role for noncoding transcription. Specifically, E-protein-binding sites have been identified across the Bcl11b intergenic region, which contains a noncoding transcript named ThymoD (Longabaugh et al. 2017). Activation of ThymoD transcription repositions the Bcl11b superenhancer complex from the lamina to the nuclear interior and juxtaposes the Bcl11b enhancer and promoter regions into a single loop domain (Isoda et al. 2017; Hu et al. 2018). Bcl11b expression orchestrates a T-lineage-specific gene program and suppresses the expression of genes associated with alternative cell lineages in part by silencing Id2 expression (Ikawa et al. 2010; Li et al. 2010; Longabaugh et al. 2017). In a parallel pathway, the E proteins activate the expression of *Rag1* and *Rag2* and orchestrate accessibility and recombination

across the TCRβ, TCRγ, and TCRδ loci (Agata et al. 2007; Miyazaki et al. 2017). Upon assembly of a pre-TCR complex, a signaling cascade involving the ERK–MAPK pathway lowers E47 levels but increases Id3 abundance to promote cellular expansion and developmental progression (Fig. 5; Engel and Murre 2004). Interestingly, γδ TCR-mediated signaling elevates Id3 abundance more than pre-TCR signaling. The difference in Id3 abundance may instruct cells to adopt either β or γδ T-lineage identity (Lauritsen et al. 2009).

Id3 expression not only dictates the β-lineage and γδ-lineage fates but also orchestrates γδ effector, γδ NKT, αβ NKT, and invariant innate follicular helper (T_{FH}) cellular expansion and cell fate (Ueda-Hayakawa et al. 2009; Verykokakis et al. 2010; D'Cruz et al. 2014; Miyazaki et al. 2015). How does the E–Id protein axis modulate innate γδ, NKT, and invariant T_{FH} NKT cell fates? Recent studies demonstrated that the E–Id protein module intersects with yet another regulatory circuitry, the PI3K–FOXO–mTOR pathway, to control expansion, self-renewal, and differentiation of innate-like cells (Miyazaki et al. 2015). It is tempting to speculate that these modules not only regulate the self-renewal, expansion, and differentiation of invariant innate T_{FH} cells but also act on other immune cells.

Prior to thymocyte selection, HEB levels accumulate to orchestrate TCRα locus rearrangement (Jones and Zhuang 2007; D'Cruz et al. 2010). Instructed by the expressed TCR, the thymus selects useful clones and destroys harmful or useless ones (von Boehmer et al. 1989). This selection process is at least in part enforced by the E–Id protein axis (Fig. 5; Bain et al. 1999; Rivera et al. 2000; Jones and Zhuang 2007; Jones-Mason et al. 2012; Miyazaki et al. 2015). High E2A abundance interferes with positive selection, whereas high Id3 abundance instructed by TCR-mediated ERK–MAPK signaling promotes developmental progression beyond the TCR checkpoint to promote differentiation into either CD4 or CD8 expressors and exit from the thymus (Fig. 5; Bain et al. 1997). More recent studies extended these observations involving not only Id3 but also Id2 (Miyazaki et al. 2015). Specifically, Id proteins orchestrate positive selection in two steps. The first involves the activation of Id3 expression by TCR signaling. The second step involves the activation of Id2 expression in DP thymocytes that already have received a TCR signal. In the absence of both Id2 and Id3 expression, thymocytes fail to mature into CD4 or CD8 SP cells, with the exception of a slowly expanding population of innate T_{FH}-like cells. These data suggest that differences in the strength and/or timing of Id expression, instructed by different signals involving distinct sender cells (epithelial vs. DP cells), dictate developing thymocytes to adopt either the adaptive or innate immune cell fate (Miyazaki et al. 2015). While these observations indicate heavy involvement of the E–Id protein axis in seemingly unrelated pathways, there is a common pattern: E and Id proteins orchestrate assembly of the TCR loci and enforce the pre-TCR and γδ and αβ TCR checkpoints to ensure that only cells that have generated a productive antigen receptor progress beyond the barriers.

HLH proteins and homing

Upon entry into the peripheral lymphoid organs, the vast majority of T-lineage cells remain in a naïve cell state enforced in part by Id3 (Miyazaki et al. 2011). Once exposed to invading pathogens, T cells initiate a multistep transcriptional program that instructs their differentiation from naïve to effector and/or memory-like states. Similar to B cells, the T-cell differentiation program hinges on the E–Id protein axis. While, in naïve T cells, high levels of Id3 enforce the naïve state, upon triggering the TCR, Id3 abundance declines to permit E-protein occupancy across the enhancer landscape to activate a T_{FH} gene program (Fig. 5; Miyazaki et al. 2011). T_{FH} cells not only down-regulate Id3 transcription but also need to silence Id2 expression. This is achieved by a transcriptional regulator named Bcl6 that is closely associated with T_{FH} cell development (Shaw et al. 2016). Id2 in other peripheral immune cell types acts to silence a T_{FH} gene program. For example, T_H1 cells express high levels of Id2 to suppress a T_{FH}-specific transcription signature (Shaw et al. 2016). Different patterns of Id2 versus Id3 expression have been observed in CD8 effector and memory cells. While Id3 levels initially decline during a viral infection, Id2 abundance is elevated to neutralize E-protein DNA binding to promote and maintain a terminal differentiation state for effector CD8 cells (Omilusik et al. 2018). In contrast to Id2, Id3 levels decline when a viral infection reaches maximum levels but rise again in the long-lived memory compartment to generate a functional memory compartment (Fig. 5; Yang et al. 2011). The rise and fall of Id gene expression is intriguing and indicative of a complex circuitry built to respond rapidly to cues generated by a continuously changing viral environment. Intriguing studies have recently revealed a prominent role for E and Id proteins in yet another CD8 cell type: so-called follicular cytotoxic T cells (T_{FC}). T_{FC} cells home to the B-cell follicles to eradicate viral infected T_{FH} cells as well as B cells (Im et al. 2016). T_{FC} cells share a transcription signature with T_{FH} cells that is again orchestrated by the E and Id proteins (Im et al. 2016; Leong et al. 2016). Finally, the E–Id module also controls the homing of follicular Treg cells (T_{FR}). In Treg cells, Id2 and Id3 expression enforces the T_{FR} checkpoint to prevent premature maturation by silencing a T_{FH}-like specific gene program (Fig. 5; Miyazaki et al. 2014). In later stages, declining gradients of Id2 and Id3 abundance promote the developmental progression toward a more mature T_{FR} phenotype by activating the expression of the chemokine receptor CXCR5. Thus, a common regulatory gene network with heavy involvement of the E and Id proteins has evolved to orchestrate the homing of T_{FH}, T_{FC}, and T_{FR} cells during a viral infection to the B-cell follicles.

HLH proteins and programmed transdifferentiation

Illuminating studies by Lassar and Weintraub (Davis et al. 1987) revealed that the expression of a single bHLH protein, MyoD, was sufficient to reprogram nonmuscle cells

into skeletal muscle. Demonstration of MyoD-driven reprogramming in a wide ensemble of cell types soon followed. Cell types reprogrammed by MyoD include chondrocytes, adipocytes, and retinal epithelial cells (Weintraub et al. 1989). HLH proteins also have the ability to genetically transdifferentiate nonneuronal cells into neurons. Specifically, forced Ascl1 and Neurogenin-2 expression reprograms astrocytes into terminally differentiated neurons (Berninger et al. 2007; Heinrich et al. 2010). Human cortical neurons are most efficiently generated from human embryonic stem cells by forced expression of Neurogenin-2. By combining the expression of three factors (Ascl1, Brn2, and Myt1l), mouse embryonic fibroblasts readily converted into functional neurons (Vierbuchen et al. 2010). The bHLH region of Ascl1 is rather small and does not interact preferentially with the two central E-box nucleotides, indicative of occupancy to a single side of the DNA surface. Consequently, Ascl1 is a pioneer factor; i.e., it occupies the majority of binding sites in fibroblasts without help from either Brn2 and Myt1l, allowing it to activate silent genes that are sequestered in closed chromatin. Most recently, human adult peripheral T cells were genetically programmed into neuronal-like cells that displayed all of the key features associated with neurons (Tanabe et al. 2018). Finally, forced expression of the bHLH protein Neurogenin-3 in combination with PDX1 and MAFA efficiently transdifferentiated pancreatic acinar cells into β-like cells in vivo (Zhou et al. 2008). Remarkably, the programmed β-like cells were maintained for >1 year and reversed diabetes (Zhou and Melton 2018).

Despite these striking results, translating transdifferentiation strategies into the clinic requires additional achievements to (1) improve the efficiency of reprogramming and (2) develop stable and functional transplantation procedures. Recent studies have shed light on how to optimize in vivo programming efficiency. The ability of HLH proteins to genetically program transdifferentiation varies greatly among recipient cells. While MyoD readily converts nonmuscle cell lines into skeletal muscle cells, forced expression of MyoD in *C. elegans* failed to efficiently program transdifferentiation (Fukushige and Krause 2005). Thus, the ability of MyoD to genetically program transdifferentiation is limited by cellular context. Recent elegant experiments have provided mechanistic insight into how cellular context dictates the ability to genetically program transdifferentiation (Fong et al. 2012). Two recipient cell types—embryonic fibroblasts and embryonic carcinoma cell lines (P19)—were examined for their ability to transdifferentiate into muscle or neurons upon MyoD or NeuroD2 expression. Embryonic fibroblasts readily converted into muscle upon forced MyoD expression, while NeuroD2 expression did not promote transdifferentiation. Conversely, P19 cells efficiently transdifferentiated into neurons when NeuroD2 was overexpressed, while forced MyoD expression displayed only a limited ability to orchestrate a skeletal muscle-specific program of gene expression. Interestingly, NeuroD2 and MyoD share thousands of binding sites in cells overexpressing each factor. However, a subset of binding sites was specific for each fac-

tor, and this subset was associated with the induction of a myogenic versus neuronal-specific gene program. Additionally, for each cell type, NeuroD2 or MyoD occupancy was highly enriched for E-box sites located in accessible chromatin (Fong et al. 2015). While the induction of lineage-specific gene programs by forced expression of lineage-specific bHLH proteins is dictated by the chromatin landscape, it is not the complete story. Successful programming also depends on the presence of other factors that either prevent or permit reprogramming, leaving many details to be uncovered.

Conclusion

The HLH domain is an ancient DNA-binding domain that, in unicellular organisms, became involved in regulating the biosynthesis of phospholipids and amino acids. In multicellular organisms, the structure of the HLH domain closely resembled those expressed in unicellular organisms. However, HLH proteins evolved to greater complexity by acquiring additional domains, such as the leucine zipper, the PAS domains, and highly conserved transactivation/repression domains. Despite these permutations in structure, the ability of HLH to activate lineage-specific programs of gene expression across species is universal. The most remarkable example of this deep conservation of function involves HLH proteins that control sexual dimorphism. Specifically, the human ortholog (E2A) can replace HLH-2 to orchestrate *C. elegans* gonadogenesis (Sallee and Greenwald 2015).

As the 30 years have gone by, many questions remain: For example, what mechanisms control the ability of HLH proteins to activate lineage-specific gene programs, reprogram differentiated cell types, and faithfully orchestrate antigen receptor assembly. Numerous studies involving all classes of HLH proteins indicate that HLH proteins predominantly bind enhancer elements, recruiting HATs, chromatin remodelers, and coactivators (McMahon et al. 1998; Bradney et al. 2003; Zhang et al. 2004; Forcales et al. 2012; Lin et al. 2012; Teachenor et al. 2012; Bossen et al. 2015; Fong et al. 2015; Grajkowska et al. 2017). Prominent among the factors recruited by HLH proteins is the bromodomain protein BRD4. BRD4 binds acetylated H3 and H4 lysine residues and is associated with large intrinsically disordered domains that may establish a phase-separated state across superenhancers (Hnisz et al. 2017; Sabari et al. 2018). An HLH protein-induced phase-separated state may serve to compartmentalize enhancers and promoters or antigen receptor variable, diversity, and joining elements during the somatic recombination process. These observations lead to the cardinal conclusion that bHLH proteins primarily serve to facilitate the compartmentalization of transcriptional components at lineage-specific genes.

Genes encoding for HLH proteins arose in unicellular organisms >600 million years ago, duplicated from ancestral genes, diversified, and are closely associated with the establishment of multicellularity. The notion that duplication and diversification of HLH genes from ancestral

genes are closely associated with the generation of multi-cellular life is perhaps not surprising. Duplication from an ancestral locus allows HLH genes to be placed under the control of novel enhancers and insulators. These reconfigurations facilitate new spatial and temporal programs of gene expression, generating ever-increasing cellular diversity.

Acknowledgments

References

Agata Y, Tamaki N, Sakamot S, Ikawa T, Masudo K, Kawamoto H, Murre C. 2007. Regulation of T cell receptor β gene rearrangements and allelic exclusion by the helix–loop–helix protein, E47. *Immunity* **27:** 871–884. doi:10.1016/j.immuni.2007.11.015

Allada R, White NE, So WV, Hall JC, Rosbash M. 1998. A mutant *Drosophila* homolog of mammalian *Clock* disrupts circadian rhythms and transcription of period and timeless. *Cell* **93:** 791–804. doi:10.1016/S0092-8674(00)81440-3

Arabi A, Wu S, Shiue C, Ridderstrale K, Larsson LG, Wright APH. 2005. c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription. *Nat Cell Biol* **7:** 303–310. doi:10.1038/ncb1225

Aronheim A, Shiran R, Rosen A, Walker MD. 1993. The E2A gene product contains two separable and functionally distinct transcription activation domains. *Proc Natl Acad Sci* **90:** 8063–8067. doi:10.1073/pnas.90.17.8063

Aspland SE, Bendall HH, Murre C. 2001. The role of E2A–PBX1 in leukemogenesis. *Oncogene* **20:** 5708–5717. doi:10.1038/sj.onc.1204592

Ayer DE, Kretzner L, Eisenman RN. 1993. Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell* **72:** 211–222. doi:10.1016/0092-8674(93)90661-9

Ayer DE, Lawrence QA, Eisenman RN. 1995. Mad–Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* **80:** 767–776. doi:10.1016/0092-8674(95)90355-0

Bailey AM, Posakony W. 1995. Suppressor of Hairless directly activates transcription of *Enhancer of split* Complex genes in response to Notch receptor activity. *Genes Dev* **9:** 2609–2622. doi:10.1101/gad.9.21.2609

Bain G, Murre C. 1998. The role of E-proteins in B- and T-lymphocyte development. *Semin Immunol* **10:** 143–153. doi:10.1006/smim.1998.0116

Bain G, Maandag ECR, Izon DJ, Amsen D, Kruisbeek AM, Weintraub BC, Krop I, Schlissel MS, Feeney A, van Roon M, et al. 1994. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* **79:** 885–892. doi:10.1016/0092-8674(94)90077-9

Bain G, Engel I, Robanus Maandag EC, te Riele HP, Voland JR, Sharp LL, Chun J, Huey B, Pinkel D, Murre C. 1997. E2A deficiency leads to abnormalities in αβ T-cell development and to rapid development of T-cell lymphomas. *Mol Cell Biol* **17:** 4782–4791. doi:10.1128/MCB.17.8.4782

Bain G, Quong MW, Soloff RS, Hedrick SM, Murre C. 1999. Thymocyte maturation is regulated by the activity of the helix–loop–helix protein, E47. *J Exp Med* **190:** 1605–1616. doi:10.1084/jem.190.11.1605

Bain G, Cravatt CB, Loomans C, Alberola-Ila J, Hedrick SM, Murre C. 2001. Regulation of the helix–loop–helix proteins, E2A and Id3, by the Ras–Erk MAPK cascade. *Nat Immunol* **2:** 165–171. doi:10.1038/84273

Bao Y, Xu F, Shimeld SM. 2017. Phylogenetics of lophotrochozoan bHLH genes and the evolution of lineage-specific gene duplicates. *Genome Biol Evol* **9:** 869–886. doi:10.1093/gbe/evx047

Bargiello TA, Young MW. 1984. Molecular genetics of a biological clock in *Drosophila*. *Proc Natl Acad Sci* **81:** 2142–2146. doi:10.1073/pnas.81.7.2142

Beck K, Peak MM, Ota T, Nemazee D, Murre C. 2009. Distinct roles for E12 and E47 in B cell specification and the sequential rearrangement of immunoglobulin light chain loci. *J Exp Med* **206:** 2271–2284. doi:10.1084/jem.20090756

Beck B, Lapouge G, Rorive S, Drogat B, Desaedelaere K, Delafaille S, Dubois C, Salmon I, Willekens K, Marine JC, et al. 2015. Different levels of Twist1 regulate skin tumor initiation, stemness, and progression. *Cell Stem Cell* **16:** 67–79. doi:10.1016/j.stem.2014.12.002

Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H. 1990. The protein Id: a negative regulator of helix–loop–helix DNA binding proteins. *Cell* **61:** 49–59. doi:10.1016/0092-8674(90)90214-Y

Benner C, Isoda T, Murre C. 2015. New roles for DNA cytosine modifications, eRNA, anchors and superanchors in developing B cell progenitors. *Proc Natl Acad Sci* **112:** 12776–12781. doi:10.1073/pnas.1512995112

Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, Tapscott SJ. 2004. Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol Cell* **14:** 465–477. doi:10.1016/S1097-2765(04)00260-6

Berninger B, Costa MR, Koch U, Schroeder T, Sutor B, Grothe B, Gotz M. 2007. Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. *J Neurosci* **27:** 8654–8664. doi:10.1523/JNEUROSCI.1615-07.2007

Bessho Y, Sakata R, Komatsu S, Shiota K, Yamada S, Kageyama R. 2001. Dynamic expression and essential functions of Hes7 in somite segmentation. *Genes Dev* **15:** 2642–2647. doi:10.1101/gad.930601

Blackwood EM, Eisenman RN. 1991. Max: a helix–loop–helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* **251:** 1211–1217. doi:10.1126/science.2006410

Bohrer C, Pfurr S, Mammadzada K, Schildge S, Plappert L, Hils M, Pous L, Rauch KS, Dumit VI, Pfeifer D, et al. 2015. The balance of Id3 and E47 determines neural stem/precursor cell differentiation into astrocytes. *EMBO J* **34:** 2804–2819. doi:10.15252/embj.201591118

Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, et al. 2018. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175:** 1–14. doi:10.1016/j.cell.2018.10.042

Boos MD, Yokota Y, Eberl G, Kee BL. 2007. Mature natural killer cell and lymphoid tissue-inducing cell development requires Id2-mediated suppression of E protein activity. *J Exp Med* **204:** 1119–1130. doi:10.1084/jem.20061959

Bossen C, Murre CS, Chang AN, Mansson R, Rodewald HR, Murre C. 2015. The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth. *Nat Immunol* **16:** 775–784. doi:10.1038/ni.3170

Bradney C, Hjelmeland M, Komatsu Y, Yoshida M, Yao TP, Zhuang Y. 2003. Regulation of E2A activities by histone acetyltransferases in B lymphocyte development. *J Biol Chem* **278:** 2370–2376. doi:10.1074/jbc.M211464200

Bruick RK, McKnight SL. 2001. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294:** 1337–1340. doi:10.1126/science.1066373

Cabrera CV. 1990. Lateral inhibition and cell fate during neurogenesis in *Drosophila:* the interactions between *scute, Notch* and *Delta. Development* **109:** 733–742.

Cao Y, Kumar RM, Penn BH, Berkes CA, Kooperberg C, Boyer LA, Young RA, Tapscott SJ. 2006. Global and gene-specific analyses show distinct roles for MyoD and Myog at a common set of promoters. *EMBO J* **25:** 502–511. doi:10.1038/sj.emboj.7600958

Carroll PA, Freie BW, Mathsyaraja H, Eisenman RN. 2018. The myc transcription factor network: balancing metabolism, proliferation and oncogenesis. *Front Med* **12:** 412–425. doi:10.1007/s11684-018-0650-z

Caudy M, Vässin H, Brand M, Tuma R, Jan LY, Jan YN. 1988. Daughterless, a *Drosophila* gene essential for both neurogenesis and sex determination, has sequence similarities to myc and the achaete–scute complex. *Cell* **55:** 1061–1067. doi:10.1016/0092-8674(88)90250-4

Chen WY, Zhang J, Geng H, Du Z, Nakadai T, Roeder RG. 2013. A TAF4 coactivator function for E proteins that involves enhanced TFIID binding. *Genes Dev* **27:** 1596–1609. doi:10.1101/gad.216192.113

Chen S, Miyazaki M, Chandra V, Fisch KM, Chang AN, Murre C. 2016. Id3 orchestrates germinal center B cell development. *Mol Cell Biol* **36:** 2543–2552. doi:10.1128/MCB.00150-16

Chiu Y, Lin I, Su S, Wang K, Yang S, Tsai D, Hsieh Y, Lin K. 2014. Transcription factor ABF-1 suppresses plasma cell differentiation but facilitates memory B cell formation. *J Immunol* **193:** 2207–2217. doi:10.4049/jimmunol.1400411

Church GM, Ephrussi A, Gilbert W, Tonegawa S. 1985. Cell-type-specific contacts to immunoglobulin enhancers in nuclei. *Nature* **313:** 798–801. doi:10.1038/313798a0

Cisse B, Caton ML, Lehner M, Maeda T, Scheu S, Locksley R, Holmberg D, Zweier C, den Hollander NS, Kant SG, et al. 2008. Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell* **135:** 37–48. doi:10.1016/j.cell.2008.09.016

Conerly ML, Yao Z, Zhong JW, Groudine M, Tapscott SJ. 2016. Distinct activities of Myf5 and MyoD indicate separate roles in skeletal muscle lineage specification and differentiation. *Dev Cell* **36:** 375–385. doi:10.1016/j.devcel.2016.01.021

Conway SJ, Firulli B, Firulli AB. 2010. A bHLH code for cardiac morphogenesis. *Pediatric Cardiol* **31:** 318–324. doi:10.1007/s00246-009-9608-x

Cubas P, de Celis J-F, Campuzano S, Modolell J. 1991. Proneural clusters of *achaete-scute* expression and the generation of sensory organs in the *Drosophila* imaginal wing disc. *Genes Dev* **5:** 996–1008. doi:10.1101/gad.5.6.996

Darlington TK, Wager-Smith K, Ceriani MF, Staknis D, Gekakis N, Steeves TDL, Weitz CJ, Takahashi JS, Kay SA. 1998. Closing the circadian loop: CLOCK-induced transcription of its own inhibitors, *per* and *tim. Science* **280:** 1599–1603. doi:10.1126/science.280.5369.1599

Davis RL, Weintraub H, Lassar AB. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51:** 987–1000. doi:10.1016/0092-8674(87)90585-X

Davis RL, Cheng PF, Lassar AB, Weintraub H. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60:** 733–746. doi:10.1016/0092-8674(90)90088-V

D'Cruz LM, Knell J, Fujimoto JK, Goldrath AW. 2010. An essential role for the transcription factor HEB in thymocyte survival, Tcra rearrangement and the development of natural killer T cells. *Nat Immunol* **11:** 240–249. doi:10.1038/ni.1845

D'Cruz LM, Stradner MH, Yang CY, Goldrath AW. 2014. E and Id proteins influence invariant NKT cell sublineage differentiation and proliferation. *J Immunol* **192:** 2227–2236. doi:10.4049/jimmunol.1302904

Degnan BM, Vervoort M, Larroux C, Richards GS. 2009. Early evolution of metazoan transcription factors. *Curr Opin Genet Dev* **19:** 591–599. doi:10.1016/j.gde.2009.09.008

Dias S, Månsson R, Gurbuxani S, Sigvardsson M, Kee BL. 2008. E2A proteins promote development of lymphoid-primed multipotent progenitors. *Immunity* **29:** 217–227. doi:10.1016/j.immuni.2008.05.015

Dobzhansky T. 1964. Biology, molecular and organismic. *Am Zool* **4:** 443–452.

Doe CQ, Goodman CS. 1985. Early events in insect neurogenesis. II. The role of cell interactions and cell lineage in the determination of neuronal precursor cells. *Dev Biol* **111:** 206–219. doi:10.1016/0012-1606(85)90446-4

Ellenberger T, Fass D, Arnaud M, Harrison SC. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix–loop–helix dimer. *Genes Dev* **8:** 970–980. doi:10.1101/gad.8.8.970

Engel I, Murre C. 2004. E2A proteins enforce a proliferation checkpoint in developing thymocytes. *EMBO J* **23:** 202–211. doi:10.1038/sj.emboj.7600017

Enriquez J, de Taffin M, Crozatier M, Vincent A, Dubois L. 2012. Combinatorial coding of *Drosophila* muscle shape by Collier and Nautilus. *Dev Biol* **363:** 27–39. doi:10.1016/j.ydbio.2011.12.018

Ephrussi A, Church GM, Tonegawa S, Gilbert W. 1985. B lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. *Science* **227:** 134–140. doi:10.1126/science.3917574

Fong AP, Yao Z, Zhong JW, Cao Y, Ruzzo WL, Gentleman RC, Tapscott SJ. 2012. Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev Cell* **22:** 721–735. doi:10.1016/j.devcel.2012.01.015

Fong AP, Yao Z, Zhong JW, Johnson NM, Farr GH, Maves L, Tapscott SJ. 2015. Conversion of MyoD to a neurogenic factor: binding site specificity determines lineage. *Cell Rep* **10:** 1937–1946. doi:10.1016/j.celrep.2015.02.055

Forcales SV, Albini S, Giordani L, Malecova B, Cignolo L, Chernov A, Coutinho P, Saccone V, Consalvi S, Williams R, et al. 2012. Signal-dependent incorporation of MyoD-BAF60c into Brg1-based SWI/SNF chromatin-remodelling complex. *EMBO J* **31:** 301–316. doi:10.1038/emboj.2011.391

Fukushige T, Krause M. 2005. The myogenic potency of HLH-1 reveals wide-spread developmental plasticity in early *C. elegans* embryos. *Development* **132:** 1795–1805. doi:10.1242/dev.01774

George RM, Firulli AB. 2018. Hand factors in cardiac development. *Anat Rec* doi:10.1002/ar.23910

Girgis CM, Cheng K, Scott CH, Gunton JE. 2012. Novel links between HIFs, type 2 diabetes, and metabolic syndrome. *Trends Endocrinol Metab* **23:** 372–380. doi:10.1016/j.tem.2012.05.003

Goebel P, Janney N, Valenzuela JR, Romanow WR, Murre C, Feeney AJ. 2001. Localized gene-specific induction of accessibility to V(D)J recombination induced by E2A and early B cell factor in nonlymphoid cells. *J Exp Med* **194:** 645–656. doi:10.1084/jem.194.5.645

Gloury R, Zotos D, Zuidscherwoude M, Masson F, Liao Y, Hasbold J, Corcoran LM, Hodgkin PD, Belz GT, Shi W, et al. 2016. Dynamic changes in Id3 and E-protein activity orchestrate germinal center and plasma cell development. *J Exp Med* **213:** 1095–1111. doi:10.1084/jem.20152003

Gomez-Roman N, Grandori C, Eisenman RN, White RJ. 2003. Direct activation of RNA polymerase III transcription by c-Myc. *Nature* **421:** 290–294. doi:10.1038/nature01327

Grajkowska LT, Ceribelli M, Lau CM, Warren ME, Tiniakou I, Nakandakari S, Bunin A, Haecker H, Mirny LA, Staudt LM, et al. 2017. Isoform-specific expression and feedback regulation of E protein TCF4 control dendritic cell lineage specification. *Immunity* **46:** 65–77. doi:10.1016/j.immuni.2016.11.006

Grewal SS, Li L, Orian A, Eisenman RN, Edgar BA. 2005. Myc-dependent regulation of ribosomal RNA synthesis during *Drosophila* development. *Nat Cell Biol* **7:** 295–302. doi:10.1038/ncb1223

Hartenstein V, Posakony JW. 1990. A dual function of the *Notch* gene in *Drosophila* sensillum development. *Dev Biol* **142:** 13–30. doi:10.1016/0012-1606(90)90147-B

Hasty P, Bradley A, Morris JH, Edmondson DG, Venuti JM, Olson EN, Klein WH. 1993. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* **364:** 501–506. doi:10.1038/364501a0

Heinrich C, Blum R, Gascón S, Masserdotti G, Tripathi P, Sánchez R, Tiedt S, Schroeder T, Götz M, Berninger B. 2010. Directing astroglia from the cerebral cortex into subtype specific functional neurons. *PLoS Biol* **8:** e1000373. doi:10.1371/journal.pbio.1000373

Hewitt KJ, Johnson KD, Gao X, Keles S, Bresnick EH. 2016. The hematopoietic stem and progenitor cell cistrome: GATA factor-dependent *cis*-regulatory mechanisms. *Curr Top Dev Biol* **118:** 45–76. doi:10.1016/bs.ctdb.2016.01.002

Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. 2017. A phase separation model for transcriptional control. *Cell* **169:** 13–23. doi:10.1016/j.cell.2017.02.007

Hu G, Cui K, Fang D, Hirose S, Wang X, Wangsa D, Jin W, Ried T, Liu P, Zhu J, et al. 2018. Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. *Immunity* **48:** 227–242.e8. doi:10.1016/j.immuni.2018.01.013

Huang N, Chelliah Y, Shan Y, Taylor CA, Yoo SH, Partch C, Green CB, Zhang H, Takahashi JS. 2012. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* **337:** 189–194. doi:10.1126/science.1222804

Huang Q, Seillet C, Belz GT. 2017. Shaping innate lymphoid cell diversity. *Front Immunol* **8:** 1569–1578. doi:10.3389/fimmu.2017.01569

Ikawa T, Hirose S, Masuda K, Kakugawa K, Satoh R, Shibano-Satoh A. 2010. An essential developmental checkpoint for production of the T cell lineage. *Science* **329:** 93–96. doi:10.1126/science.1188995

Im SJ, Hashimoto M, Gerner MY, Lee J, Kissick HT, Burger MC, Shan Q, Hale JS, Lee J, Nasti TH, et al. 2016. Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537:** 417–421. doi:10.1038/nature19330

Imayoshi I, Isomura A, Harima Y, Kawaguchi K, Kori H, Miyachi H, Fujiwara T, Ishidate F, Kageyama R. 2013. Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. *Science* **342:** 1203–1208. doi:10.1126/science.1242366

Inlay M, Tian H, Lin T, Xu Y. 2004. Important roles for E protein binding sites within the immunoglobulin κ chain intronic enhancer in activating $V_\kappa J_\kappa$ rearrangement. *J Exp Med* **200:** 1205–1211. doi:10.1084/jem.20041135

Inlay MA, Bhattacharya D, Sahoo D, Serwold T, Seita J, Karsunky H, Plevritis SK, Dill DL, Weissman IL. 2009. Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev* **23:** 2376–2381. doi:10.1101/gad.1836009

Isoda T, Moore A, He Z, Chandra V, Aida M, Denholtz M, van Hamburg JP, Fisch K, Chang AN, Fahl S, et al. 2017. Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. *Cell* **171:** 103–119.e18. doi:10.1016/j.cell.2017.09.001

Jaleco AC, Stegmann AP, Heemskerk MH, Couwenberg F, Bakker AQ, Weijer K, Spits H. 1999. Genetic modification of human B cell development: B cell development is inhibited by the dominant-negative helix–loop–helix factor Id3. *Blood* **94:** 2637–2646.

Jones ME, Zhuang Y. 2007. Acquisition of a functional T cell receptor during T lymphocyte development is enforced by HEB and E2A transcription factors. *Immunity* **27:** 860–870. doi:10.1016/j.immuni.2007.10.014

Jones-Mason ME, Zhao X, Kappes D, Lasorella A, Iavarone A, Zhuang Y. 2012. E protein transcription factors are required for the development of CD4+ lineage T cells. *Immunity* **36:** 348–61. doi:10.1016/j.immuni.2012.02.010

Kaelin WG, Ratcliffe PJ. 2008. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol Cell* **30:** 393–402. doi:10.1016/j.molcel.2008.04.009

Kageyama R, Shimojo S, Isomura A. 2018. Oscillatory control of notch signaling in development. *Adv Exp Med Biol* **1066:** 265–277. doi:10.1007/978-3-319-89512-3_13

Kassar-Duchossoy L, Gayraud-Morel B, Gomès D, Rocancourt D, Buckingham M, Shinin V, Tajbakhsh S. 2004. Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature* **431:** 466–471. doi:10.1038/nature02876

Kee BL, Rivera RR, Murre C. 2001. Id3 inhibits B lymphocyte progenitor growth and survival in response to TGF-β. *Nat Immunol* **2:** 242–247. doi:10.1038/85303

Keith B, Johnson RS, Simon MC. 2012. HIF1α and HIF2α: sibling rivalry in hypoxic tumour growth and progression. *Nat Rev Cancer* **12:** 9–22. doi:10.1038/nrc3183

Kieffer-Kwon KR, Nimura K, Rao SSP, Xu J, Jung S, Pekowska A, Dose M, Stevens E, Mathe E, Dong P, et al. 2017. Myc regulates chromatin decompaction and nuclear architecture during B cell activation. *Mol Cell* **67:** 566–578.e10. doi:10.1016/j.molcel.2017.07.013

Kreslavsky T, Vilagos B, Tagoh H, Poliakova DK, Schwickert TA, Wöhner M, Jaritz M, Weiss S, Taneja R, Rossner MJ, et al. 2017. Essential role for the transcription factor Bhlhe41 in regulating the development, self-renewal and BCR repertoire of B1a cells. *Nat Immunol* **18:** 442–455. doi:10.1038/ni.3694

Kwon K, Hutter C, Sun Q, Bilic I, Cobaleda C, Malin S, Busslinger M. 2008. Instructive role of the transcription factor E2A in early B lymphopoiesis and germinal center B cell development. *Immunity* **28:** 751–762. doi:10.1016/j.immuni.2008.04.014

Lai EC, Bodner R, Posakony JW. 2000. The *Enhancer of split* Complex of *Drosophila* includes four Notch-regulated members of the bearded gene family. *Development* **127:** 3441–3455.

Lasorella A, Stegmüller J, Guardavaccaro D, Liu G, Carro MS, Rothschild G, de la Torre-Ubieta L, Pagano M, Bonni A, Iavarone A. 2006. Degradation of Id2 by the anaphase-promoting complex couples cell cycle exit and axonal growth. *Nature* **442:** 471–474. doi:10.1038/nature04895

Lasorella A, Benezra R, Iavarone A. 2014. The ID proteins: master regulators of cancer stem cells and tumour aggressiveness. *Nat Rev Cancer* **14:** 77–91. doi:10.1038/nrc3638

Lauritsen JP, Wong GW, Lee SY, Lefebvre JM, Ciofani M, Rhodes M, Kappes DJ, Zúñiga-Pflücker JC, Wiest DL. 2009. Marked induction of the helix–loop–helix protein Id3 promotes the γδ T cell fate and renders their functional maturation Notch independent. *Immunity* **31:** 565–575. doi:10.1016/j.immuni.2009.07.010

Lee SB, Frattini V, Bansal M, Castano AM, Sherman D, Hutchinson K, Bruce JN, Califano A, Liu G, Cardozo T, et al. 2016. An Id2-dependent mechanism for VHL inactivation in cancer. *Nature* **529:** 172–177. doi:10.1038/nature16475

Leong YA, Chen Y, Ong HS, Wu D, Man K, Deleage C, Minnich M, Meckiff BJ, Wei Y, Hou Z, et al. 2016. CXCR5[+] follicular cytotoxic T cells control viral infection in B cell follicles. *Nat Immunol* **17:** 1187–1196. doi:10.1038/ni.3543

Li K, Baker NE. 2018. Regulation of the *Drosophila* ID protein Extra macrochaetae by proneural dimerization partners. *Elife* **7:** e33967. doi:10.7554/eLife.33967

Li L, Leid M, Rothenberg EV. 2010. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* **329:** 89–93. doi:10.1126/science.1188989

Li R, Cauchy P, Ramamoorthy S, Boller S, Chavez L, Grosschedl R. 2018. Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. *Genes Dev* **32:** 96–111. doi:10.1101/gad.309583.117

Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J, et al. 2010. A global network of transcription factors, involving E2A, EBF1 and FOXO1, that orchestrates B cell fate. *Nat Immunol* **11:** 635–643. doi:10.1038/ni.1891

Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, Murre C. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat Immunol* **13:** 1196–1204. doi:10.1038/ni.2432

Lio CW, Zhang J, González-Avalos E, Hogan PG, Chang X, Rao A. 2016. Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. *Elife* **5:** e18290. doi:10.7554/eLife.18290

Longabaugh WJR, Zeng W, Zhang JA, Hosokawa H, Jansen CS, Li L, Romero-Wolf M, Liu P, Kueh HY, Mortazavi A, et al. 2017. Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. *Proc Natl Acad Sci* **114:** 5800–5807. doi:10.1073/pnas.1610617114

Lowrey PL, Takahashi JS. 2004. Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu Rev Genomics Hum Genet* **5:** 407–441. doi:10.1146/annurev.genom.5.061903.175925

Lu QR, Sun T, Zhu Z, Ma N, Garcia M, Stiles CD, Rowitch DH. 2002. Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection. *Cell* **109:** 75–86. doi:10.1016/S0092-8674(02)00678-5

Lyden D, Young AZ, Zagzag D, Yan W, Gerald W, O'Reilly R, Bader BL, Hynes RO, Zhuang Y, Manova K, et al. 1999. Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* **401:** 670–677. doi:10.1038/44334

Ma PC, Rould MA, Weintraub H, Pabo CO. 1994. Crystal structure of MyoD bHLH domain–DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77:** 451–459. doi:10.1016/0092-8674(94)90159-7

Mansson R, Welinder E, Ashber J, Lin YC, Benner C, Glass CK, Sigvardsson M, Murre C. 2012. Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. *Proc Natl Acad Sci* **108:** 17402–17407.

Massari ME, Jennings PA, Murre C. 1996. The AD1 transactivation domain of E2A contains a highly conserved helix which is required for its activity in both *Saccharomyces cerevisiae* and mammalian cells. *Mol Cell Biol* **16:** 121–129. doi:10.1128/MCB.16.1.121

Massari ME, Rivera RR, Voland JR, Quong M, Breit TM, van Dongen JM, de Smit O, Murre C. 1998. Characterization of ABF-1, a novel basic helix–loop–helix transcription factor expressed in activated B lymphocytes. *Mol Cell Biol* **18:** 3130–3139. doi:10.1128/MCB.18.6.3130

McMahon SB, Van Buskirk HA, Dugan KA, Copeland TD, Cole MD. 1998. The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* **94:** 363–374. doi:10.1016/S0092-8674(00)81479-8

Mercer EM, Lin YC, Benner C, Jhunjhunwala S, Dutkowski J, Flores M, Sigvardsson M, Ideker T, Glass CK, Murre C. 2011. Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors. *Immunity* **35:** 413–425. doi:10.1016/j.immuni.2011.06.013

Miyai T, Takano J, Endo TA, Kawakami E, Agata Y, Motomura Y, Kubo M, Kashima Y, Suzuki Y, Kawamoto H, et al. 2018. Three-step transcriptional priming that drives the commitment of multipotent progenitors toward B cells. *Genes Dev* **32:** 112–126. doi:10.1101/gad.309575.117

Miyazaki M, Rivera RR, Miyazaki K, Lin YC, Agata Y, Murre C. 2011. The opposing roles of the transcription factor E2A and its antagonist Id3 that orchestrate and enforce the naive fate of T cells. *Nat Immunol* **12:** 992–1001. doi:10.1038/ni.2086

Miyazaki M, Miyazaki K, Chen SY, Itoi M, Miller M, Lu L-F, Varki N, Chang A, Broide DH, Murre C. 2014. Id2 and Id3 maintain the regulatory T cell pool to suppress inflammatory disease. *Nat Immunol* **15:** 767–776. doi:10.1038/ni.2928

Miyazaki M, Miyazaki K, Chen S, Chandra V, Agata Y, Chang A, Kawamoto H, Murre C. 2015. The E–Id protein axis modulates the activities of the PI3K–AKT–mTORC1–Hif1a and c-myc/p19Arf pathways to suppress innate variant T$_{FH}$ cell development, thymocyte expansion, and lymphomagenesis. *Genes Dev* **29:** 409–425. doi:10.1101/gad.255331.114

Miyazaki M, Miyazaki K, Chen K, Jin Y, Turner J, Moore AJ, Saito R, Yoshida K, Ogawa S, Rodewald HR, et al. 2017. The E-Id protein axis specifies adaptive lymphoid cell identity and suppresses thymic innate lymphoid cell development. *Immunity* **46:** 818–834.e4. doi:10.1016/j.immuni.2017.04.022

Murre C, McCaw PS, Baltimore D. 1989a. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56:** 777–783. doi:10.1016/0092-8674(89)90682-X

Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka S, Lassar A, et al. 1989b. Interactions between heterologous helix–loop–helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58:** 537–544. doi:10.1016/0092-8674(89)90434-0

Nabeshima Y, Hanaoka K, Hayasaka M, Esumi E, Li S, Nonaka I, Nabeshima Y. 1993. Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364:** 532–535. doi:10.1038/364532a0

Nair SK, Burley SK. 2003. X-ray structures of Myc–Max and Mad–Max recognizing DNA: molecular bases of regulation by proto-oncogenic transcription factors. *Cell* **112:** 193–205. doi:10.1016/S0092-8674(02)01284-9

Nakamura Y, Sakakibara SI, Miyata T, Ogawa M, Shimazaki T, Weiss S, Kageyama R, Okano H. 2000. The bHLH gene *hes1* as a repressor of the neuronal commitment of CNS stem cells. *J Neurosci* **20:** 283–293. doi:10.1523/JNEUROSCI.20-01-00283.2000

Nguyen UT, Bittova L, Müller MM, Fierz B, David Y, Houck-Loomis B, Feng V, Dann GP, Muir TW. 2014. Accelerated chromatin biochemistry using DNA-barcoded nucleosome libraries. *Nat Methods* **11:** 834–840. doi:10.1038/nmeth.3022

Omari EK, Hoosdally SJ, Tuladhar K, Karia D, Hall-Ponsele E, Platonova O, Vyas P, Patient R, Porcher C, Mancini EJ. 2013. Structural basis for LMO2-driven recruitment of the SCL:E47bHLH heterodimer to hematopoietic-specific transcriptional targets. *Cell Rep* **4:** 135–147. doi:10.1016/j.celrep.2013.06.008

Omilusik KD, Shaw LA, Goldrath AW. 2013. Remembering one's ID/E-ntity: E/ID protein regulation of T cell memory. *Curr Opin Immunol* **25:** 660–666. doi:10.1016/j.coi.2013.09.004

Omilusik KD, Nadjsombati MS, Shaw LA, Yu B, Milner JJ, Goldrath AW. 2018. Sustained Id2 regulation of E proteins is required for terminal differentiation of effector CD8+ T cells. *J Exp Med* **215:** 773–783. doi:10.1084/jem.20171584

Orlanski S, Labi V, Reizel Y, Spiro A, Lichtenstein M, Levin-Klein R, Koralov SB, Skversky Y, Rajewsky K, Cedar H, et al. 2016. Tissue specific DNA demethylation is required for proper B-cell differentiation and function. *Proc Natl Acad Sci* **113:** 5018–5923. doi:10.1073/pnas.1604365113

Pagliuca A, Gallo P, De Luca P, Lania L. 2000. Class A helix–loop–helix proteins are positive regulators of several cyclin-dependent kinase inhibitors promoter activity and negatively affect cell growth. *Cancer Res* **60:** 1376–1382.

Paroush Z, Finley RL, Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-Horowicz D. 1994. Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79:** 805–815. doi:10.1016/0092-8674(94)90070-1

Parraga A, Bellosell L, Ferré-D'Amaré AR, Burley SK. 1998. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* **6:** 661–672. doi:10.1016/S0969-2126(98)00067-7

Pfurr S, Chu YH, Bohrer C, Greulich F, Beattie R, Mammadzada K, Hils M, Arnold SJ, Taylor V, Schachtrup K, et al. 2017. The E2A splice variant E47 regulates the differentiation of projection neurons via p57(KIP2) during cortical development. *Development* **144:** 3917–3931. doi:10.1242/dev.145698

Pires N, Dolan L. 2010a. Origin and diversification of basic helix–loop–helix proteins in plants. *Mol Biol Evol* **27:** 862–74. doi:10.1093/molbev/msp288

Pires N, Dolan L. 2010b. Early evolution of bHLH proteins in plants. *Plant Signal Behav* **5:** 911–912. doi:10.4161/psb.5.7.12100

Prabhu S, Ignatova A, Park ST, Sun XH. 1997. Regulation of the expression of cyclin-dependent kinase inhibitor p21 by E2A and Id proteins. *Mol Cell Biol* **17:** 5888–5896. doi:10.1128/MCB.17.10.5888

Puri PL, Sartorelli V, Yang XJ, Hamamori Y, Ogryzko VV, Howard BH, Kedes L, Wang JY, Graessmann A, Nakatani Y, et al. 1997. Differential roles of p300 and PCAF acetyltransferases in muscle differentiation. *Mol Cell* **1:** 35–45. doi:10.1016/S1097-2765(00)80005-2

Quan XJ, Yuan L, Tiberi L, Claeys A, De Geest N, Yan J, van der Kant R, Xie WR, Klisch TJ, Shymkowitz J, et al. 2016. Post-translational control of the temporal dynamics of transcription factor activity regulates neurogenesis. *Cell* **164:** 460–475. doi:10.1016/j.cell.2015.12.048

Quong MW, Harris DP, Swain SL, Murre C. 1999. E2A activity is induced during B-cell activation to promote immunoglobulin class switch recombination. *EMBO J* **18:** 6307–6318. doi:10.1093/emboj/18.22.6307

Quong M, Martensson A, Langerak AW, Rivera RR, Nemazee D, Murre C. 2004. Receptor editing and marginal zone B cell development are regulated by the helix–loop–helix protein, E2A. *J Exp Med* **199:** 1101–1112. doi:10.1084/jem.20031180

Reddy P, Zehring WA, Wheeler DA, Pirrotta V, Hadfield C, Hall JC, Rosbash M. 1984. Molecular analysis of the period locus in *Drosophila melanogaster* and identification of a transcript involved in biological rhythms. *Cell* **38:** 701–710. doi:10.1016/0092-8674(84)90265-4

Reppert SM, Weaver DR. 2001. Molecular analysis of mammalian circadian rhythms. *Annu Rev Physiol* **63:** 647–676. doi:10.1146/annurev.physiol.63.1.647

Rivera RR, Johns CP, Quan J, Johnson RS, Murre C. 2000. Thymocyte selection is regulated by the helix–loop–helix inhibitor protein, Id3. *Immunity* **12:** 17–26. doi:10.1016/S1074-7613(00)80155-7

Robinson KA, Lopes JM. 2000. *Saccharomyces cerevisiae* basic helix–loop–helix proteins regulate diverse biological processes. *Nucleic Acids Res* **28:** 1499–1505. doi:10.1093/nar/28.7.1499

Romanow WR, Langerak AW, Goebel P, Wolvers-Tettero ILM, van Dongen JJM, Feeney AJ, Murre C. 2000. E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells. *Mol Cell* **5:** 343–353. doi:10.1016/S1097-2765(00)80429-3

Roussel M, Saule S, Lagrou C, Rommens C, Beug H, Graf T, Stehelin D. 1979. Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation. *Nature* **281:** 452–455. doi:10.1038/281452a0

Rudnicki MA, Schnegelsberg PN, Stead RH, Braun T, Arnold HH, Jaenisch R. 1993. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75:** 1351–1359. doi:10.1016/0092-8674(93)90621-V

Rutila JE, Suri V, Le M, So WV, Rosbash M, Hall JC. 1998. CYCLE is a second bHLH–PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila* period and timeless. *Cell* **93:** 805–814. doi:10.1016/S0092-8674(00)81441-5

Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, et al. 2018. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361:** eaar3958. doi:10.1126/science.aar3958

Sallee MD, Greenwald I. 2015. Dimerization-driven degradation of *C. elegans* and human E proteins. *Genes Dev* **29:** 1356–1361. doi:10.1101/gad.261917.115

Sallee MD, Littleford HE, Greenwald I. 2017. A bHLH code for sexually dimorphic form and function of the *C. elegans* somatic gonad. *Curr Biol* **27:** 1853–1860.e5. doi:10.1016/j.cub.2017.05.059

Sayegh CE, Quong ME, Agata Y, Murre C. 2003. E-proteins directly regulate expression of activation-induced deaminase in mature B cells. *Nat Immunol* **45:** 86–593.

Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M, Wright G, Shaffer AL, Hodson DJ, Buras E, et al. 2012. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490:** 116–120.

Schotte R, Dontje W, Nagasawa M, Yasuda Y, Bakker AQ, Spits H, Blom B. 2010. Synergy between IL-15 and Id2 promotes the expansion of human NK progenitor cells, which can be counteracted by the E protein HEB required to drive T cell development. *J Immunol* **184:** 6670–6679. doi:10.4049/jimmunol.0901508

Schwartz R, Engel I, Fallahi-Sichani M, Petrie HT, Murre C. 2006. Gene expression patterns define novel roles for E47 in cell cycle progression, cytokine mediated signaling and T-lineage development. *Proc Natl Acad Sci* **103:** 9976–9981. doi:10.1073/pnas.0603728103

Semenza GL. 2012. Hypoxia-inducible factors in physiology and medicine. *Cell* **148:** 399–408. doi:10.1016/j.cell.2012.01.021

Semerad CL, Mercer EM, Inlay MA, Weissman IL, Murre C. 2009. E2A proteins maintain the hematopoietic stem cell pool and promote the maturation of myelolymphoid and myeloerythroid progenitors. *Proc Natl Acad Sci* **106:** 1930–1935. doi:10.1073/pnas.0808866106

Shaw LA, Bélanger S, Omilusik KD, Cho S, Scott-Browne JP, Nance JP, Goulding J, Lasorella A, Lu LF, Crotty S, et al. 2016. Id2 reinforces TH1 differentiation and inhibits E2A to repress TFH differentiation. *Nat Immunol* **17:** 834–843. doi:10.1038/ni.3461

Sheiness D, Bishop JM. 1979. DNA and RNA from uninfected vertebrate cells contain nucleotide sequences related to the putative transforming gene of avian myelocytomatosis virus. *J Virol* **31:** 514–521.

Sheiness D, Fanshier L, Bishop JM. 1978. Identification of nucleotide sequences which may encode the oncogenic capacity of the avian retrovirus MC29. *J Virol* **28:** 600–610.

Shimizu T, Toumoto A, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, Oshima Y, Hakoshima T. 1997. Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. *EMBO J* **16:** 4689–4697. doi:10.1093/emboj/16.15.4689

Shimojo H, Isomura A, Ohtsuka T, Kori H, Miyachi H, Kageyama R. 2016. Oscillatory control of Delta-like1 in cell interactions regulates dynamic gene expression and tissue morphogenesis. *Genes Dev* **30:** 102–116. doi:10.1101/gad.270785.115

Shivdasani RA, Mayer EL, Orkin SH. 1995. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373:** 432–434. doi:10.1038/373432a0

Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M. 2007. Origin and diversification of the basic helix–loop–helix gene family in metazoans: insights from comparative genomics. *BMC Evol Biol* **7:** 33–39. doi:10.1186/1471-2148-7-33

Singh SK, Singh S, Gadomski S, Sun L, Pfannenstein A, Magidson V, Chen X, Kozlov S, Tessarollo L, Klarmann KD, et al. 2018. Id1 ablation protects hematopoietic stem cells from stress-induced exhaustion and aging. *Cell Stem Cell* **23:** 252–265.e8. doi:10.1016/j.stem.2018.06.001

Soler E, Andrieu-Soler C, de Boer E, Bryne JC, Thongjuea S, Stadhouders R, Palstra RJ, Stevens M, Kockx C, van IJcken W, et al. 2010. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* **24:** 277–289. doi:10.1101/gad.551810

Souroullas GP, Salmon JM, Sablitzky F, Curtis DJ, Goodell MA. 2009. Adult hematopoietic stem and progenitor cells require either Lyl-1 or Scl for survival. *Cell Stem Cell* **4:** 180–186. doi:10.1016/j.stem.2009.01.001

Sun XH, Baltimore D. 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell* **64:** 459–470. doi:10.1016/0092-8674(91)90653-G

Sun Y, Nadal-Vicens M, Misono S, Lin MZ, Zubiaga A, Hua X, Fan G, Greenberg ME. 2001. Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell* **104:** 365–376. doi:10.1016/S0092-8674(01)00224-0

Tanabe K, Ang CE, Chanda S, Olmos VH, Haag D, Levinson DF, Südhof TC, Wernig M. 2018. Transdifferentiation of human adult peripheral blood T cells into neurons. *Proc Natl Acad Sci* **115:** 6470–6475. doi:10.1073/pnas.1720273115

Tapscott SJ, Davis RL, Thayer MJ, Cheng PF, Weintraub H, Lassar AB. 1988. MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science* **242:** 405–411. doi:10.1126/science.3175662

Teachenor R, Beck K, Wright LY, Shen Z, Briggs SP, Murre C. 2012. Biochemical and phosphoproteomic analysis of the helix–loop–helix protein E47. *Mol Cell Biol* **9:** 1671–1682. doi:10.1128/MCB.06452-11

Ueda-Hayakawa I, Mahlios J, Zhuang Y. 2009. Id3 restricts the developmental potential of γδ lineage during thymopoiesis. *J Immunol* **182:** 5306–5316. doi:10.4049/jimmunol.0804249

Verykokakis M, Boos MD, Bendelac A, Kee BL. 2010. SAP protein-dependent natural killer T-like cells regulate the development of CD8[+] T cells with innate lymphocyte characteristics. *Immunity* **33:** 203–215. doi:10.1016/j.immuni.2010.07.013

Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M. 2010. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463:** 1035–1041. doi:10.1038/nature08797

von Boehmer H, Teh HS, Kisielow P. 1989. The thymus selects the useful, neglects the useless and destroys the harmful. *Immunol Today* **10:** 57–61. doi:10.1016/0167-5699(89)90307-1

Vosshall LB, Price JL, Sehgal A, Saez L, Young MW. 1994. Block in nuclear localization of period protein by a second clock mutation, timeless. *Science* **263:** 1606–1609. doi:10.1126/science.8128247

Wang GL, Jiang BH, Rue EA, Semenza GL. 1995. Hypoxia-inducible factor 1 is a basic helix–loop–helix–PAS heterodimer regulated by cellular O2 tension. *Proc Natl Acad Sci* **92:** 5510–5514. doi:10.1073/pnas.92.12.5510

Wang R, Dillon CP, Shi LZ, Milasta S, Carter R, Finkelstein D, McCormick LL, Fitzgerald P, Chi H, Munger J, et al. 2011. The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35:** 871–882. doi:10.1016/j.immuni.2011.09.021

Wang HC, Qian L, Zhao Y, Mengarelli J, Adrianto I, Montgomery CG, Urban JF Jr, Fung KM, Sun XH. 2017. Down-regulation of E protein activity augments an ILC2 differentiation program in the thymus. *J Immunol* **198:** 3149–3156. doi:10.4049/jimmunol.1602009

Weintraub H, Tapscott SJ, Davis RL, Thayer MJ, Adam MA, Lassar AB, Miller AD. 1989. Activation of muscle-specific genes in pigment, nerve, fat, liver and fibroblast cell lines by forced expression of MyoD. *Proc Natl Acad Sci* **86:** 5434–5438. doi:10.1073/pnas.86.14.5434

Welinder E, Mansson R, Mercer E, Bryder D, Sigvardsson M, Murre C. 2011. The transcription factors E2A and HEB act in concert to induce the expression of FOXO1 in the common lymphoid progenitor. *Proc Natl Acad Sci* **108:** 17402–17407. doi:10.1073/pnas.1111766108

White NM, Jarman AP. 2000. *Drosophila* Atonal controls photoreceptor R8-specific properties and modulates both receptor tyrosine kinase and hedgehog signaling. *Development* **127:** 1681–1689.

Williams SA, Maecker HL, French DM, Liu J, Gregg A, Silverstein LB, Cao TC, Carano RA, Dixit VM. 2011. USP1 deubiquitinates ID proteins to preserve a mesenchymal stem cell program in osteosarcoma. *Cell* **146:** 918–930. doi:10.1016/j.cell.2011.07.040

Wöhner M, Tagoh H, Bilic I, Jaritz M, Poliakova DK, Fischer M, Busslinger M. 2016. Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. *J Exp Med* **213:** 1201–1221. doi:10.1084/jem.20152002

Wu D, Potluri N, Lu J, Kim Y, Rastinejad F. 2015. Structural integration in hypoxia-inducible factors. *Nature* **524:** 303–308. doi:10.1038/nature14883

Yamazaki T, Liu L, Lazarev D, Al-Zain A, Fomin V, Yeung PL, Chambers SM, Lu CW, Studer L, Manley JL. 2018. TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency. *Genes Dev* **32:** 1161–1174. doi 10.1101/gad.316984.118

Yan W, Young AZ, Soares VC, Kelley R, Benezra R, Zhuang Y. 1997. High incidence of T-cell tumors in E2A-null mice and E2A/Id1 double-knockout mice. *Mol Cell Biol* **17:** 7317–7327. doi:10.1128/MCB.17.12.7317

Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C, Savagner P, Gitelman I, Richardson A, Weinberg RA. 2004. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117:** 927–939. doi:10.1016/j.cell.2004.06.006

Yang CY, Best JA, Knell J, Yang E, Sheridan AD, Jesionek AK, Li HS, Rivera RR, Lind KC, D'Cruz LM, et al. 2011. The transcriptional regulators Id2 and Id3 control the formation of distinct memory CD8+ T cell subsets. *Nat Immunol* **12:** 122–129.

Yang J, Li X, Morrell NW. 2014. Id proteins in the vasculature: from molecular biology to cardiopulmonary medicine. *Cardiovasc Res* **104:** 388–398. doi:10.1093/cvr/cvu215

Ying Q, Nichols J, Chambers I, Smith A. 2003. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115:** 281–292. doi:10.1016/S0092-8674(03)00847-X

Yokota Y, Mansouri A, Mori S, Sugawara S, Adachi S, Nishikawa S, Gruss P. 1999. Development of peripheral lymphoid organs and natural killer cells depends on the helix–loop–helix inhibitor Id2. *Nature* **397:** 702–706. doi:10.1038/17812

Zehring WA, Wheeler DA, Reddy P, Konopka RJ, Kyriacou CP, Rosbash M, Hall JC. 1984. P-element transformation with period locus DNA restores rhythmicity to mutant, arrhythmic *Drosophila melanogaster*. *Cell* **39:** 369–376. doi:10.1016/0092-8674(84)90015-1

Zhang J, Kalkum M, Yamamura S, Chait BT, Roeder RG. 2004. E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science* **305:** 1286–1289. doi:10.1126/science.1097937

Zhang H, Gao P, Fukuda R, Kumar G, Krishnamachary B, Zeller KI, Dang CV, Semenza GL. 2007. HIF-1 inhibits mitochondrial biogenesis and cellular respiration in VHL-deficient renal cell carcinoma by repression of C-MYC activity. *Cancer Cell* **11:** 407–420. doi:10.1016/j.ccr.2007.04.001

Zhou Q, Melton DA. 2018. Pancreas regeneration. *Nature* **557:** 351–358. doi:10.1038/s41586-018-0088-0

Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA. 2008. In vivo reprogramming of adult pancreatic exocrine cells to β-cells. *Nature* **455:** 627–632. doi:10.1038/nature07314

Zhuang Y, Soriano P, Weintraub HT. 1994. The helix–loop–helix gene E2A is required for B cell formation. *Cell* **79:** 875–884. doi:10.1016/0092-8674(94)90076-0

Zhuang Y, Cheng P, Weintraub H. 1996. B lymphocyte development is regulated by the combined dosage of three basic helix–loop–helix genes, E2A, E2-2, and HEB. *Mol Cell Biol* **16:** 2898–2905. doi:10.1128/MCB.16.6.2898

Zhuang Y, Barndt RJ, Pan L, Kelley R, Dai M. 1998. Functional replacement of the mouse E2A gene with a human HEB cDNA. *Mol Cell Biol* **18:** 3340–3349. doi:10.1128/MCB.18.6.3340

Zohren F, Souroullas GP, Luo M, Gerdemann U, Imperato MR, Wilson NK, Göttgens B, Lukov GL, Goodell MA. 2012. The transcription factor Lyl-1 regulates lymphoid specification and the maintenance of early T lineage progenitors. *Nat Immunol* **13:** 761–769. doi:10.1038/ni.2365

Zook EC, Li ZY, Xu Y, de Pooter RF, Verykokakis M, Beaulieu A, Lasorella A, Maienschein-Cline M, Sun JC, Sigvardsson M, et al. 2018. Transcription factor ID2 prevents E proteins from enforcing a naïve T lymphocyte gene program during NK cell development. *Sci Immunol* **3:** 10.1126/sciimmunol.aao2139.