



Review

# Mechanisms of Binding Specificity among bHLH Transcription Factors

Xabier de Martin <sup>1</sup>, Reza Sodaei <sup>1</sup> and Gabriel Santpere <sup>1,2,\*</sup> 

<sup>1</sup> Neurogenomics Group, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), DCEXS, Universitat Pompeu Fabra, 08003 Barcelona, Spain; fxabierd@gmail.com (X.d.M.); reza.sodaei@gmail.com (R.S.)

<sup>2</sup> Department of Neuroscience, Yale University School of Medicine, New Haven, CT 06510, USA

\* Correspondence: gabrielsantperebaro@gmail.com

**Abstract:** The transcriptome of every cell is orchestrated by the complex network of interaction between transcription factors (TFs) and their binding sites on DNA. Disruption of this network can result in many forms of organism malfunction but also can be the substrate of positive natural selection. However, understanding the specific determinants of each of these individual TF-DNA interactions is a challenging task as it requires integrating the multiple possible mechanisms by which a given TF ends up interacting with a specific genomic region. These mechanisms include DNA motif preferences, which can be determined by nucleotide sequence but also by DNA's shape; post-translational modifications of the TF, such as phosphorylation; and dimerization partners and co-factors, which can mediate multiple forms of direct or indirect cooperative binding. Binding can also be affected by epigenetic modifications of putative target regions, including DNA methylation and nucleosome occupancy. In this review, we describe how all these mechanisms have a role and crosstalk in one specific family of TFs, the basic helix-loop-helix (bHLH), with a very conserved DNA binding domain and a similar DNA preferred motif, the E-box. Here, we compile and discuss a rich catalog of strategies used by bHLH to acquire TF-specific genome-wide landscapes of binding sites.

**Keywords:** transcription factor binding sites; E-box; bHLH; co-factors; ChIP-seq; pioneer factors; dimerization



**Citation:** de Martin, X.; Sodaei, R.; Santpere, G. Mechanisms of Binding Specificity among bHLH Transcription Factors. *Int. J. Mol. Sci.* **2021**, *22*, 9150. <https://doi.org/10.3390/ijms22179150>

Academic Editor:  
Maurizio Margaglione

Received: 27 July 2021  
Accepted: 18 August 2021  
Published: 24 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Gene expression is primarily regulated by transcription factors binding and acting on regions of the DNA which, precisely because they host this activity, become what is known as cis-regulatory elements. A recent manually curated census of transcription factors in the human genome identified 1639 of these molecules, classified in around 100 types based on their DNA binding domains (DBD) [1]. Those DBD largely, but not completely, determine the DNA sequence preferentially bound by each TF and with that the ability to influence expression on effectively close target genes. The complex and dynamic regulatory network orchestrated in a given cell is thus ultimately controlled by the interaction of TF with their targets, and the abnormal modification of such interactions can have dire consequences for the proper development and maintenance of the organism but also, be the substrate for evolutionary innovation. However, the analysis and identification of disease or evolutionarily relevant genetic mutations disrupting links in the regulatory network is challenging and refractory to comprehensive and automatable genome-wide scans analogous to those applied to protein-coding genes.

Nevertheless, the application of automated genome-wide approaches has yielded significant insights on many aspects. For example, early leverage of ChIP-seq data from multiple TFs and genome-wide chromatin maps revealed that the vast majority of transcription factor binding sites (TFBS) fell on accessible chromatin [2], with the exception of those binding sites associated with chromatin repressors or pioneer TFs. Counting canonical

motifs matches or binding events, determined *in silico* and experimentally, respectively, has allowed calculating enrichments of TFBS on specific groups of sequences. These methods often provide meaningful global observations on broad dynamics of the regulatory landscape in development, tissue-specific functions, evolution and disease. For example, this approach guided the search for master regulators on differentiation processes, pathological events and a plethora of other biological processes; and also, under an evolutionary perspective, enlightened the putative impact on the regulatory network of gain, losses and functionally repurposed accessible chromatin regions [3]. The analysis of aggregated data on TFBS has also allowed the study of global conservation patterns using within-population segregating sites and substitutions across species' phylogenetic trees. Highly informative positions in a motif accumulate fewer polymorphisms than flanking or more degenerate positions, indicating the ability to detect the effect of purifying selection acting on a group of binding sites [4]. Moreover, a fraction of genomic variants falling on TFBS are associated with an allelic imbalance of chromatin accessibility [4,5] and with changes in TF binding and gene expression [6] and are particularly enriched in GWAS signal [7]. Similar evidence for purifying selection can be observed when leveraging polymorphism data with fixed substitutions between species on TFBS using a MacDonald–Kreitman framework [8], which, in addition, has the potential to reveal the fraction of adaptive substitutions, *i.e.*, driven by positive selection, in groups of binding sites.

## 2. Variability and Complexity of Transcription Factor Regulatory Activity

To identify the relevant disrupting events in the regulatory network directly from the scrutiny of the genome and aggregated datasets from multiple TFs is a herculean task. At least four major types of reasons can explain these difficulties in building generalizable approaches. First, most TF present variable but often notorious discrepancies between the *in silico* predicted TFBS motifs versus those experimentally determined, for example using ChIP-seq. Analyses of TF ChIP-seq data, which presents its own set of biases and caveats, forced to abandon simplistic models of TF-motif binding and revealed multiple nuances, including a large proportion of binding events where no predicted motif can be identified, diversity on the motif itself including departure from canonical k-mers, and variation of the binding site landscapes across developmental times and cell types. Moreover, most TFs can recognize DNA “shape motifs” based on preferred DNA local physical characteristics along the genome, which may or may not contain their canonical sequence motif [9].

Second, an additional source of variability involves the composition, “grammar”, of the entire cis-regulatory element, which determines the fact that TFBSs can undergo different modes of selection depending on the specific regulatory region, TF and motif. Classification of enhancer organization initially defined two extreme models, which can coexist in many instances [10]. The two models are: (A) *Enhanceosome*: In this model, TFBSs need a precise order and spacing in a sequence and therefore work synergistically as a unit. Disrupting a piece triggers loss of function. (B) *Billboard*: This model allows for more flexibility, since TFBSs spacing and order are not relevant and removing one chunk can have little or no evolutionary/deleterious effect. A third model, called the Collective mode, adds the cooperative dimension of certain TFs, which can be recruited into enhancers by cooperative protein-protein interactions on sequences with a lax grammar of TFBS.

A third type of argument refers to the diversity of co-factors and dimerization partners of each TF in each cellular context. TF can bind to DNA in monomeric, homo- or hetero-oligomeric forms, and the choice of partner has consequences on the specific motif recognized by the complex. Our current uncertainties on the dynamics of TF partnerships, and our incomplete catalog of combinations between TFs and co-factors, hamper comprehensively meaningful scans. Cataloging TF-TF spatiotemporal interactions is a tremendous endeavor itself, since multiple modes of TF-TF cooperation are possible, generating a huge combinatorial potential with thousands of possible interactions. For example, two TF might directly interact to increase DNA binding affinity. On this, structural analysis revealed instances where the oligomerization is performed before the DNA binding occurs, while

other instances require the DNA molecule to allow the formation of the complex. In this direct interaction modality, particular TF pairs can bind to composite motifs, or suboptimal motifs, that differ from those preferred predicted motifs of their individual components. In addition, TF cooperation can be also indirect, for example occurring when the binding of one TF relaxes the energy requirements for another TF to bind nearby.

Another form of indirect cooperation occurs when the effect of one TF on the state of the chromatin may benefit the binding of other TFs. This connects with the fourth layer of variability and complexity, which is that TFs can affect transcription by multiple mechanisms. These include the direct recruitment of RNA Pol II, recruitment of histone modifiers, nucleosome displacement, recruitment of modifiers of DNA methylation, and binding steric competition with other TFs. Many of the mechanisms of TF binding are affected by the concentration of the TF of interest and/or its TF partners and co-factors [11]. Concentration of a TF influences the degree of specificity of its DNA binding sites, with higher concentrations enabling lower affinity binding. It has been suggested that this mechanism could be manifested in particular genomic regions by increasing TF concentration in specific nuclear subdomains (reviewed in Kribelbauer et al. 2019 [12]). Moreover, TF concentration can alter the passive competition that exists between certain TFs and nucleosomes or DNA methylation [2]. Measuring the concentration of the TF is not entirely predictive of the binding landscape per se, as the activity of many TFs is further modified by dynamic post-translational modifications such as phosphorylation, which can affect their subcellular localization and dimerizing partners.

In this review, we focused on variability and complexity between members of one particular family of TFs, the basic helix-loop-helix (bHLH) to describe how this multimodal diversity determining the modes of action of TFs and their DNA binding specificities can be found within one structural family of TFs with a highly conserved DBD. This detailed exploration of bHLH particularities will illustrate the need for individual TF in-depth experimental studies disclosing motif variability, tissue-specificity, choice of partners and co-factors, post-translational modifications and effects on chromatin states and gene expression.

### 3. The bHLH Family of Transcription Factors

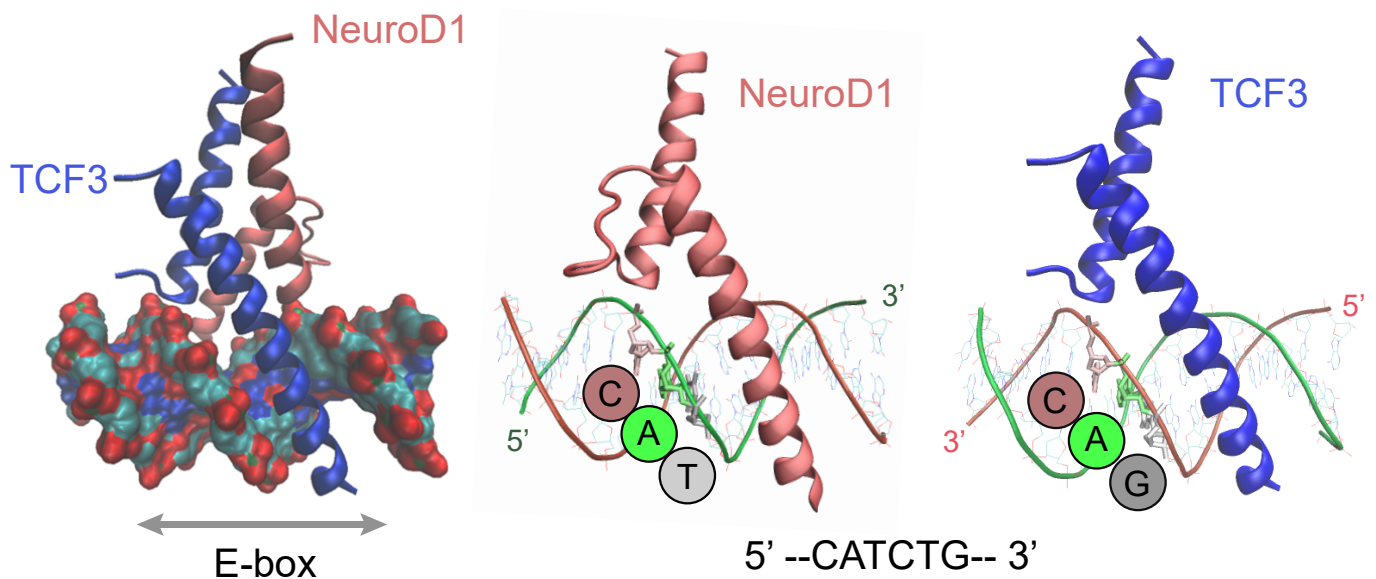
The basic helix-loop-helix (bHLH) transcription factors represent the second most populated family of transcription factors in the human genome, [1] presenting a bit over 100 members. The definition of this class is based on a common motif in the 3D structure of the DNA binding domain: an alpha-helix with a basic domain in the N-terminal end, which interacts with DNA, followed by a loop and a second alpha-helix. The two alpha-helices after the basic domain confer the platform for the formation of the bHLH dimers. This common configuration is partially modified in a subset of bHLH that contains a leucine zipper domain carboxy-terminal to the second alpha-helix (e.g., MAX), a Per-Arnt-Sim (PAS) domain (e.g., NPAS4) or an Orange domain (e.g., HES1). bHLH molecules were initially broadly classified by Murre et al. using a mixture of qualitative criteria in six classes [13]. Class I factors show expression among multiple tissues and dimerize with their lineage-restricted class II partners. Class III is composed of MYC proteins, and class IV of MYC interacting proteins. Class V was defined by HLH lacking the basic DNA binding domain which form inactive dimers, and finally, class VI represented a group of transcriptional repressors containing proline in their basic region. In a subsequent revision of this classification, a seventh class was incorporated to group those TFs presenting a PAS domain [14]. Alternative classification systems have been proposed based on aminoacid sequence alignments, forming classes A, B, C, D and E [15,16]. The study of phylogenetic relationships among bHLH additionally suggested that bHLH of classes III and IV in Murre et al. [13] classification, class B in Atchley and Fitch were the most probable ancestral bHLH classes of the family [17]. Moreover, orthologous comparisons in multiple organisms including plants, yeast and metazoan, revealed families of bHLH differentially represented among groups of organisms indicating a different time of appearances of

bHLH gene subfamilies, while no such families included genes observed in plant and animals, indicating independent radiation of bHLH genes in the two kingdoms [16]. Of the 44 subfamilies identified in metazoans, 43 were represented in the common ancestor of all bilaterians, indicating an old origin for most bHLH subfamilies accompanied by multiple lineage-specific differences in the individual bHLH repertoire [16,17]. Additionally, bHLH phylostratigraphy further supported the idea that class B was at the root of opisthokonta (including metazoan and fungi) bHLH radiation [17], while studies in plants also indicated that plant bHLHs evolved from class B members present in all eukaryotes [18].

bHLH factors have been shown to regulate a rich plethora of biological processes, such as neurogenesis (reviewed in: [19]), myogenesis [20–25], hematopoiesis [26,27], response to environmental and physiological signals [28,29], including the genetic control of circadian rhythms [30–32], and cell cycle/proliferation (reviewed in: [33]). While presenting these variegated roles and cell-types expression patterns, bHLH members recognize a short degenerate CANNTG motif known as Ephrussi-Box or, most commonly by its shorter name, E-box. As can be argued for many TF families, the high degree of in vitro derived motif overlap among TFs of the same family has raised the question of how target specificity of individual transcription factors is achieved in vivo [34]. In the case of the bHLH family, 110 factors could theoretically compete for binding to E-boxes which, in the case of the human genome, occurs as frequently as ~15M times, if we aggregate over all possible CANNTG hexanucleotides occurrences counted by Liu et al. [35]. Although regional and cell-type restricted expression largely contributes to avoiding collisions in E-box usage, still many bHLH proteins tend to be co-expressed in one cell type at a time. Therefore, some underlying mechanisms must exist where each factor acquires the ability to regulate its own specific targets involved in a particular biological process. The aim of this review is to shed light upon those mechanisms.

#### 4. DNA-Motif Preferences

A detailed understanding of how bHLH factors recognize DNA has to come from structural analyses. Studies of bHLH protein structure during the early 90's, particularly in MAX, TCF3, USF or MyoD, revealed key common aspects of bHLH dimerization, binding to DNA and the preference of bHLH proteins to specific half-sites of the E-boxes [36–39]. The conserved basic helix-loop-helix domain of these proteins consists of two alpha helices connected by a loop. bHLH factors dimerize through this domain and contact DNA with a region rich in basic amino acids located in the N-terminal end of the first helix, termed the basic region. Each monomer of this dimeric structure contacts half of the E-box CANNTG sequence, but they do it in opposing strands, resulting in each monomer recognizing a "CAN" half site (Figure 1). As we will see in detail, since half-site sequences are informative of the specific proteins and dimer configurations of bHLH factors, E-boxes could be readily described by their strand-oriented half-sites, (i.e., CAC-CAC, CAG-CAT, CAG-CAG, etc.).



**Figure 1.** Structure of Tcf3-NeuroD1 heterodimer bound to a CATCTG E-box. The bHLH domains of NeuroD1 and Tcf3 are shown in red and blue, respectively. NeuroD1 binds the CAT half-site of the E-box in the forward strand (pink) and Tcf3 binds the CAG half-site in the reverse strand (green). This representation has been produced with VMD v1.9.4 from the published X-ray crystal PDB:2ql2 from Longo et al. [40].

Early *in vitro* electrophoretic studies revealed that different bHLH factors have different nucleotide preferences in the central, flanking, or even the core positions of the E-box [41–59]. Later high-throughput *in vitro* studies, namely protein binding microarrays and HT-SELEX, permitted quantitative affinity assessment of multiple factors towards a large number of sequences, from where a comprehensive catalog of DNA binding motifs was derived [60–62].

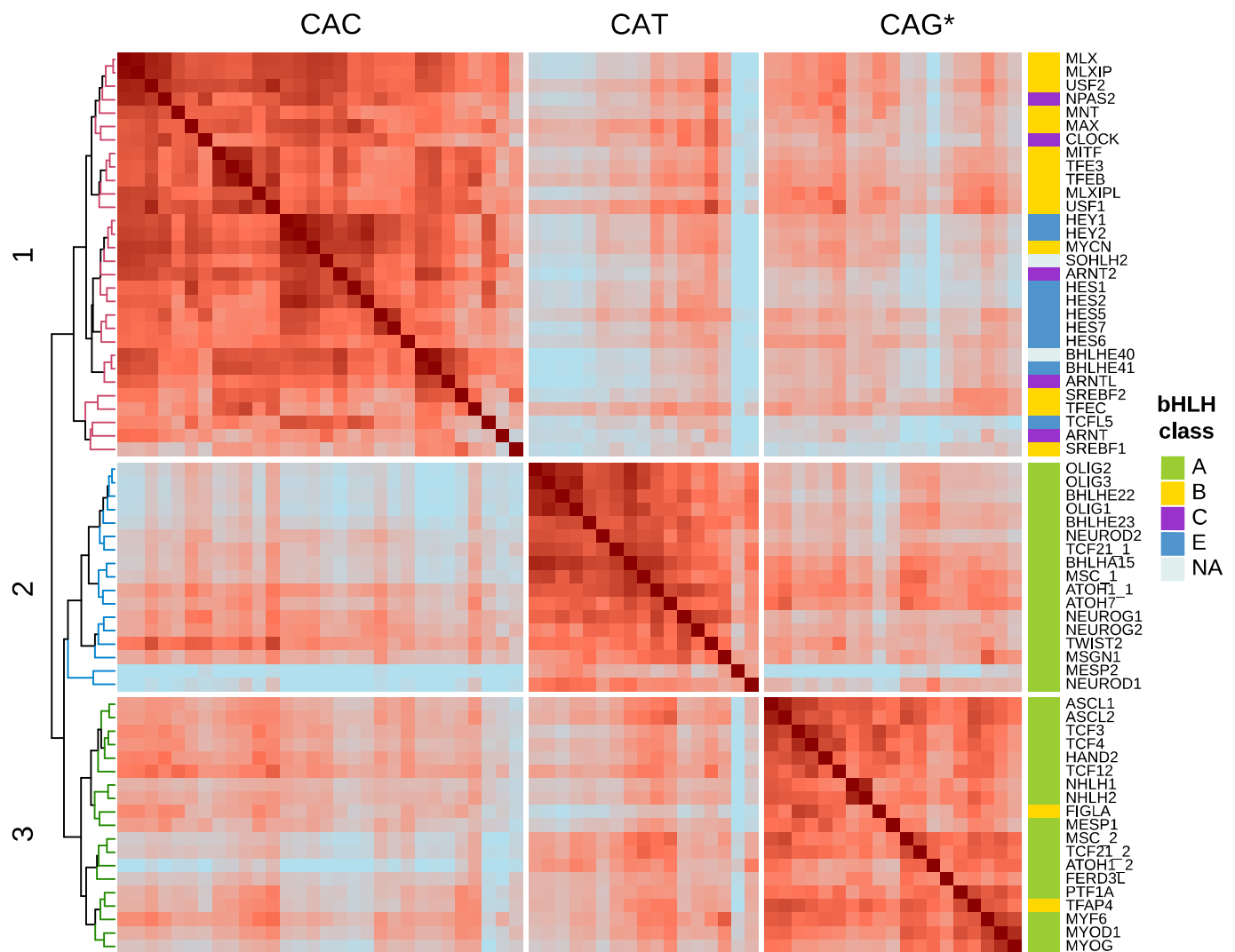
If we cluster bHLH factors by similarity of their preferred motif, derived from *in vitro* high-throughput homodimer assays [1,60,61] (Figure 2), three main clusters can be readily identified. Cluster 1 is composed of bHLH recognizing CAC half-sites, cluster 2 TFs recognize CAT half-sites and cluster 3 members bind to E-boxes containing at least one CAG site. These three clusters have correspondence with bHLH classification systems based both on phylogenetic relationships and qualitative criteria [13,15,16,63], and also specifically reflect amino acid variation in the DNA binding domain, as we have represented in Figure 3.

By integrating the 8 bHLH-DNA structures available to the date of their study, De Masi et al. [64] identified residues at positions 1, 2, 5, 6, 8, 9, 12 and 13 of the basic domain as the ones making base-specific contacts with different positions of the half site and its surrounding bases. Importantly, not only residues in the basic region but also some amino acids in the loop and second helix have also been shown to contact DNA [36]. Additionally, residues in positions 1, 2, 4, 6, 8, 10, 12–14, 17, 47–51 interact non-specifically with the phosphate backbone, both within the E-box and with the flanking positions [64]. In regard to these contacts, it has been proposed that bHLH factors scan DNA with an unfolded basic domain that makes non-specific contacts with the phosphate backbone until they find their preferred E-box. When that occurs, the alpha-helical conformation of the basic domain is stabilized and the specific contacts with the nucleotide bases are made [65,66].

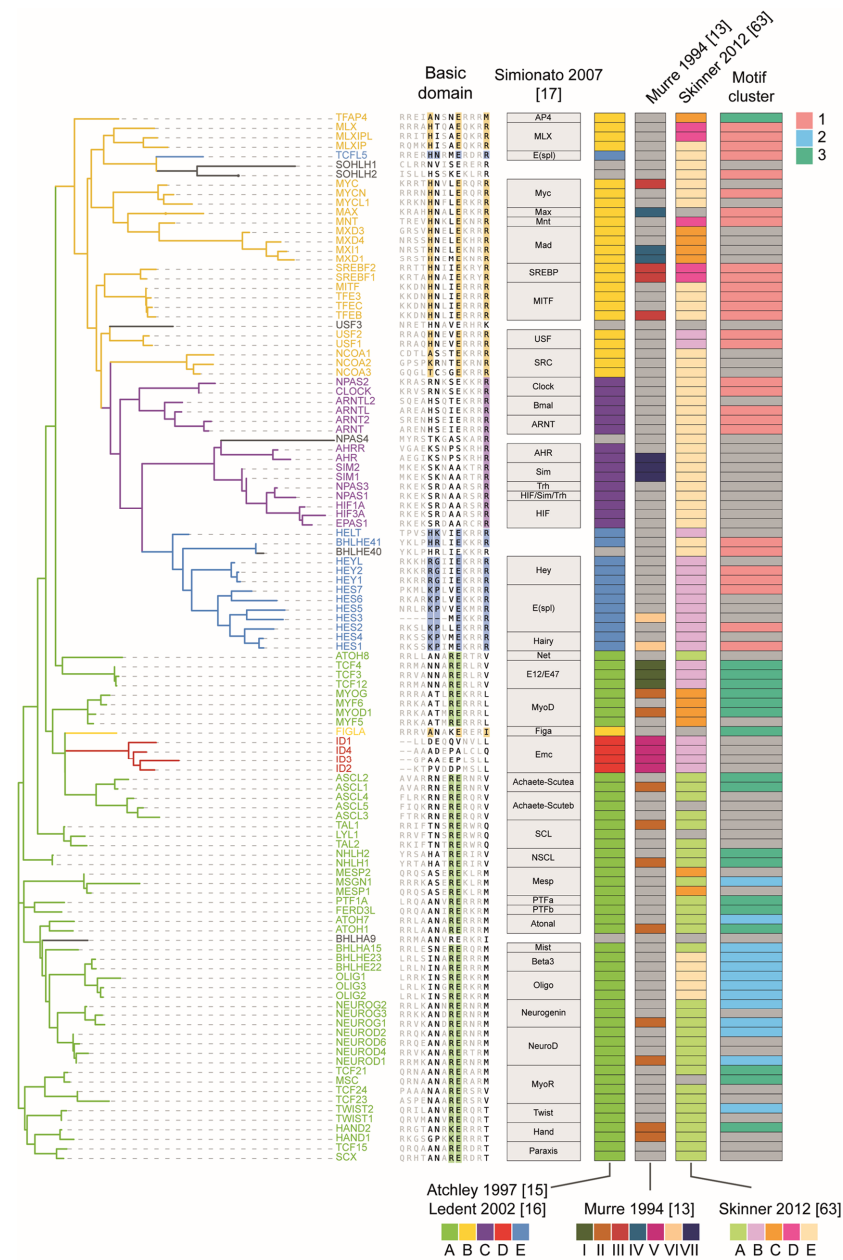
Among the amino acids in the basic region that contact bases of the E-box, the glutamic acid in position 9 undergoes the strongest contact with DNA [64]. This position is also largely conserved, as all the bHLH factors that bind to the “CA” core dinucleotide contain it, as we have illustrated in the multiple sequence alignment of the basic domain of human bHLH TFs in Figure 3. As we will discuss below, class C bHLH-PAS TF constitutes an



exception of this, with some members showing Serine or Alanine at position 9, and also class D ID proteins, which lack the basic region (Figure 3). The other DNA-contacting amino acids, which establish weaker bonds with nucleotide bases, give binding specificity to bases surrounding the “CA” core dinucleotide of the half-site and show less conservation across bHLH classes, although they are rather more conserved at the intra-class level. This, thus, constitutes the most direct mechanisms by which bHLH factors acquire target specificity: sequence preferences at central and flanking positions of the CANN TG E-box recognized by amino acids in the basic region. As we will see, in some cases, the “CA” canonical nucleotide in the core can be subject to variations too.



**Figure 2.** Heatmap representing motif similarity values among bHLH TFs calculated in Lambert et al. [1] from data derived from high-throughput in vitro assays of bHLH homodimers. Hierarchical clustering identifies three groups with a preference for E-boxes containing CAC (cluster 1, red), CAT (cluster 2, blue) and CAG half-sites (cluster 3, green). Homodimers of clusters 1 and 2 bind symmetrical CAT-CAT or CAC-CAC E-boxes, while members of cluster 3 require the presence of CAG in at least one of the half-sites (this difference is indicated by \*). A TF can appear in multiple clusters if it is represented by multiple annotated motifs, but when all of them belong to the same cluster, the TF is only shown once. bHLH classes determined by Atchley and Fitch [15] and Ledent et al. [16] are shown on the right column.



**Figure 3.** Representation of the phylogenetic relationships, alignment of the basic domain, and different classification systems of bHLH factors. The phylogenetic tree and the alignment were downloaded from the online database provided by Lambert et al. (<http://humanfbs.ccb.utoronto.ca/dbdsTable.php?dbd=bHLH>, accessed on 15 April 2021). The tree was inferred from the alignment of the whole bHLH domain, but here we only represent the basic domain as it contains the most relevant positions with respect to binding. Importantly, the tree does not imply true ancestral phylogenetic relationships among bHLH classes. The amino acids in the five positions that better separate the phylogenetic classes are colored, taking as a reference amino acids described by Atchley and Zhao [66], although we find some minor differences in those diagnostic amino acids, because they used bHLH sequences from multiple species, while we focused in human bHLH factors. In the right, different classification systems are displayed: the subfamily as annotated by Simionato et al. [17], the phylogenetic classes by Atchley et al. and Ledent et al. [15,16], the Murre classes based on both structural and functional criteria [13], the phylogenetic classes by Skinner et al. [63] inferred from the sequence of the whole protein, and finally, our clusters derived from in vitro binding affinity experiments. The boxes are colored in gray when no information about the classification was available for the corresponding gene in the corresponding original study.

Given this relationship between amino acids in the basic domain and DNA motif preferences, it is not surprising that the phylogenetic tree of bHLH factors inferred from the bHLH domain was found largely aligned with three previously determined groups (A, B, C) based on similarity in binding affinities [15,67,68] (Figure 3). Two more groups, D [15], and E [16], were added based on both phylogeny and binding preferences, forming the phylogenetic classification of five classes that we will use throughout the text. The study of the alignment of the basic domain yielded 5 positions (5, 6, 8, 9, 13) that better classified those five groups [69] (shown in Figure 3). Additionally, multivariate discriminant analysis applied on multiple amino acid features, including polarity, hydrophobicity and secondary structure, among others, highlighted positions 8, 9, 10, 12, all in the basic domain, and position 49, in the loop between the two helices, to collectively explain 86% of inter-group bHLH variation [69].

Following that phylogenetic classification, group A factors, which include tissue-restricted differentiation factors such as Neurogenins or Myogenins (class II in Murre et al. [13]) and their ubiquitously expressed partners (class I in Murre et al. [13]), the E proteins, are mainly characterized by an arginine (R) at position 8 of the basic domain and a hydrophobic or polar (M, L, V, T) residue at position 13 [38,39,69]. While some members of group A prefer CAG half-sites (e.g., FIGLA, MSC, ASCL2, TCF21), other subfamilies of this group (e.g., Twist, Mist and proneural Beta3, Oligo, Neurogenin and NeuroD) prefer a CAT half-site [60,61,64,65,70,71]. Group B proteins (bHLH-LZ), which include the cell proliferation regulator Myc, Max and Mad subfamilies among others, prefer a CAC half-site and present an invariant arginine in position 13 and frequently a histidine at position 5 of the basic domain [15,37,44,51,64,67,69,72,73]. This classification is supported additionally by both structural and electrophoretic studies that tested the effect of point mutations in the basic region, showing that arginine in position 13 specifies the central cytosine, through interaction with the guanine in the opposite strand [37,38,44,51,64,67,74,75]. Class A factors do not present an arginine in that position, and their hydrophobic or polar amino acids do not contact the central nucleotides, thus being unable to specify the CAC half-site, and preferring CAG or CAT instead [39,40].

Both class C and E represent lineages derived from class B, and as such, they possess an arginine at position 13 of the basic domain and overall prefer a CAC half-site. Class C (bHLH-PAS) members, which can respond to physiological/environmental signals such as hypoxia (HIF subfamily) [29] and xenobiotics (AHR subfamily) [28] and regulate circadian rhythms (Clock subfamily) [30,31,76], show no consistent pattern of amino acids at the critical positions, apart from the arginine at position 13. Some of them, for example, ARNT, ARNTL and CLOCK possess a glutamic acid in position 9 of the basic region and preferentially bind to the CAC half-site, whereas others do not have that critical amino acid and bind to non-canonical half-sites, such as HIF1A and SIM proteins, containing an Alanine at position 9 and binding to (A/G)C and GT(A/G)C E-boxes respectively [30,31,60,61,64,77,78]. Factors of class E, which include Hes, Hey and Dec repressors, are unique in that they contain a proline (a glycine in the case of Hey factors) in their basic region, typically but not exclusively at position 6, which is predicted to destabilize the binding to DNA. Factors of this class usually form asymmetric homodimers, where one factor binds to a CAC and the other to non-canonical CTN or CGC half-sites [43,46,58,69,73,79–82]. The class A factor HAND1, analogously to class E factors, contains a proline in position 6 of its basic region, and can also bind to degenerate half-sites, such as CGT [83,84]. Finally, class D is formed by ID proteins, which lack the basic region, and thus, are not able to bind DNA but can dimerize with other bHLH antagonizing their activity [15,85,86].

It is important to consider that, even if certain positions possess a high classificatory potential, they do not necessarily represent a recognition code with independent one-to-one correspondences between the residues and the nucleotides (i.e., a scenario where one residue specifies one base 100% of the times) [64]. Amino acids, both in the basic and HLH regions, can affect the spatial positioning of each other and thus the specific ways in which DNA contacts are made. For example, the alanine at position 5 and threonine



at position 6 of the basic domain of MyoD influence how the helix interacts with DNA and thus indirectly establish its sequence preferences [72]. Similarly, the divergence in binding specificity of *Drosophila* Scute and Atonal factors is mediated by residues that face away from the DNA-contacting surface, that likely affect the conformation of the helix of the basic domain [87,88]. In consequence, in order to accurately predict DNA sequence preference from amino acid composition, higher order structural models that take into account combinations instead of individual amino acids must be employed [64,69].

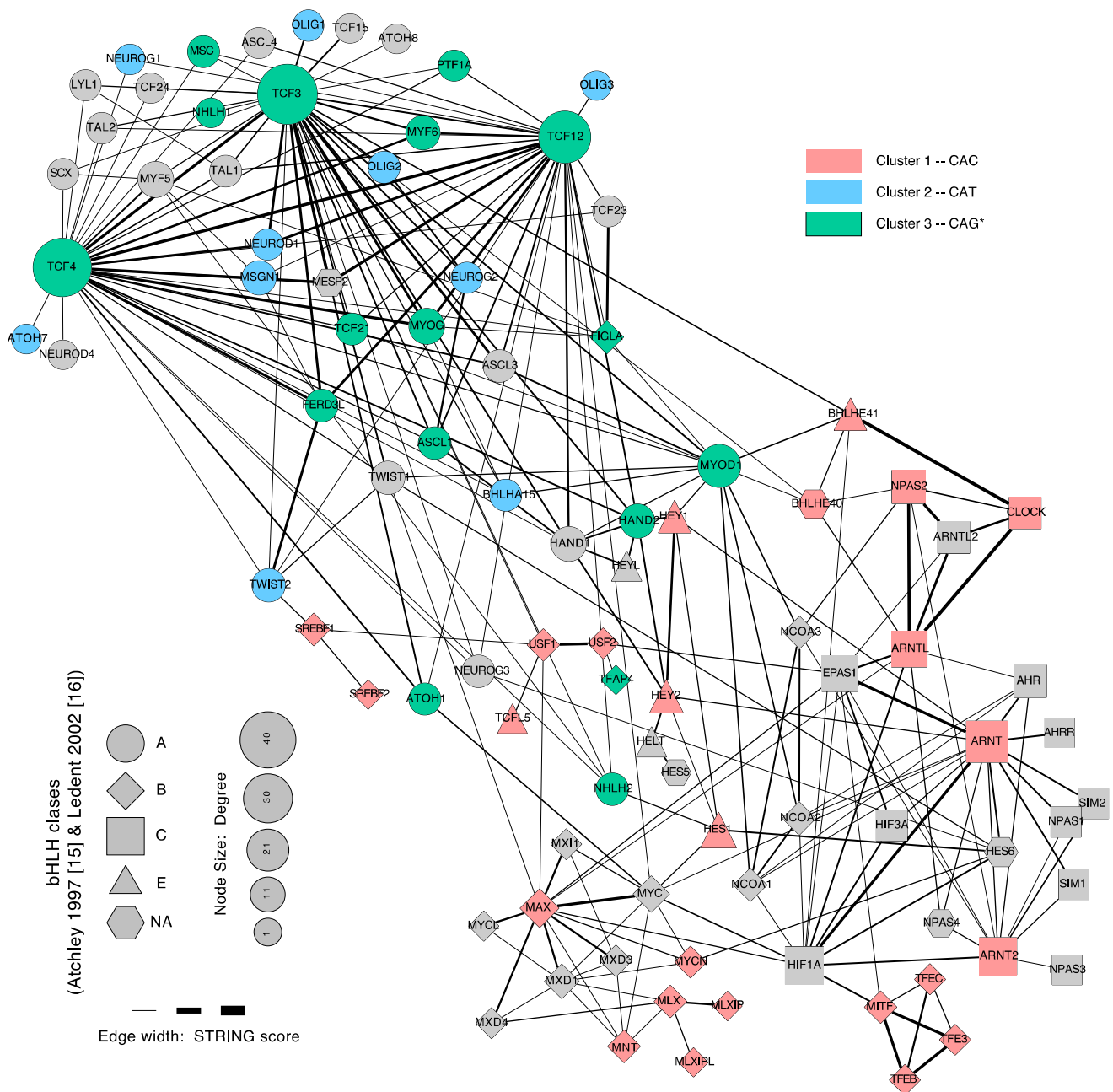
Contacts between amino acids and flanking bases of the E-box are generally not as strong as those with the central nucleotide, resulting in subtle nucleotide preferences, but sometimes they critically affect bHLH binding, and provide bHLH factors additional specificity over other members of their same class that recognize the same central dinucleotide [36,38,42,47,51,53,57,64,74,89–96]. For example, early *in vitro* studies proved that a thymine 5' adjacently flanking the E-box could differentiate the binding of TCF3 vs. MyoD [41], Myc vs. Max [53,57] and yeast PHO4 vs. Cbf1 [57]. As shown more recently by ChIP-seq studies, a 5' flanking GT dinucleotide is preferentially bound by USF but not by MYC-MAX [95], MSC and ASCL1 prefer a 3' flanking G whereas MyoD does not [97,98], and ASCL1/2 factors prefer a 5' G while MYOD1 prefers a 5' A in a subset of binding sites [99], for example. Furthermore, some bHLH factors can interact with additional motifs upstream or downstream of the E-box, thanks to other protein domains. For example, NEUROD1 recognizes an extra AT-rich sequence 3bp away from its half-site thanks to a AT-hook domain [60], and HIF factors interact with extra flanking motifs too, via their PAS-A domain [77,100].

Finally, certain post-translational modifications can also affect DNA binding of bHLH factors. Phosphorylation of Serines or Threonines in the basic (e.g., HAND, Hes1, NeuroM (NeuroD4) and myogenic factors) [101–104], HLH (e.g., Twist1, HAND, and Neurogenin/Achaete-Scute/Atonal subfamilies) [70,101,105,106] and other (e.g., Max) [107] domains can impede DNA binding of the mentioned factors.

## 5. Dimerization

In the sections above, we have mostly considered variability of bHLH sequence preference at the individual level. However, as previously stated, bHLH factors bind DNA as dimers, and the partner choice can have profound effects on the resulting preferred DNA binding motif. Dimerization is mediated by the helix-loop-helix-domain and, only in certain families, by an additional adjacent domain: the leucine zipper (LZ) in the class B factors [94], the Per-Arn-Sim (PAS) domain in class C factors [100,108] and the Orange domain in class E factors [109]. These class-specific domains favor dimerization of factors of the same class, although cross-dimerization events across classes are also possible (Figure 4). Classes A and D can easily cross-dimerize, as both depend uniquely on the HLH domain for dimerization.

Although grouping of bHLH factors according to sequence specificity reflects multiple shared binding properties within each group, as we have mentioned above, intra-group differences also exist in the manifested preferences towards flanking, central and core ("CA") positions of the E-box (Figure 3), and thus dimerization with different partners confers the ability of one given bHLH to attain binding to variegated sets of sequences [47,70,105,110–113]. For example, the TWIST1 (CAT preference) homodimer binds more effectively to the CATATG (CAT-CAT E-box) sequence, but when heterodimerizing with HAND2 and TCF3 (both with a CAG preference) it binds to the CATCTG sequence (CAT-CAG E-boxes) [70]. Regarding the flanking preferences, as previously mentioned, Max tolerates a T flanking its half-site much better than Myc, and thus, Max homodimers and Myc-Max heterodimers target different sequences [47].



**Figure 4.** Network representation of protein-protein interactions among bHLH TF catalogued in the STRING database. Only experimental/biochemical score was taken into account, from experiments of human and mouse proteins. When experimental data was available for both species, the highest score was considered. STRING score is reflected in the width of edges in the network, while the size of each node represents its degree centrality. The shape of each node indicates the bHLH classification by Atchley and Fitch [15] and Ledent et al. [16]. Nodes are colored by the motif similarity cluster derived from Figure 2. (The difference of CAG is indicated by \*).

However, the sequence specificity of a dimer is not necessarily predicted as a sum of the half-site preferences of its monomers. The monomers alter the structural conformation of their partners, usually forming non-symmetric dimer structures where one monomer binds with high affinity to its preferred half-site in a ‘specific conformation’ and the other, with a lower affinity to a non-preferred half-site in a ‘non-specific’ conformation [36,40,64,72,74]. Moreover, a monomer makes extensive contacts with the half-site that corresponds to its partner; with the DNA backbone via multiple amino acids, and

with nucleotide bases via the residue in position 13 of the basic domain [36,40,64]. Consequently, a monomer can bind different half-sites when dimerizing with different partners. For example, as a homodimer TCF3 selects CACCTG (CAC-CAG E-boxes), binding CAC specifically, and CAG non-specifically [36,42]. TCF3 binds to the CAC half-site too when heterodimerized with MyoD and to the CAT half-site when heterodimerized with Twist and Neurod1 [40,42,72]. *Drosophila* hairy and E(spl) factors and their mammalian homologs of the class E bind as homodimers to sequences containing a CAC and a CAC/CGC/CTN half-sites [43,46,58,73,79,80,82,91,110,114], thus presumably forming an asymmetric structure where one monomer binds specifically to a CAC and the other to a variable half-site.

Class C proteins also form asymmetric dimer structures upon DNA binding [100,108,115]. This way, ARNT (which has a glutamate at position 9 of the basic domain) binds to CAC half-sites while partnering with AHR (Isoleucine at position 9), HIF-1a (Alanine at position 9) or SIM proteins (Alanine at position 9) for example, that bind to (T/G)NGC, (A/G)C and GT(A/G)C half-sites respectively [64,116]. Furthermore, ARNT changes its flanking sequence preferences depending on the dimerization partners, for example, binding to a GTTCTCAC half-site upon heterodimerizing with AHR and to a CAGCAC when by heterodimerizing with HLF [117]. Class B factors generally form symmetric dimers that bind to symmetric sequences, but some exceptions exist. For example, SREBP proteins, bind as symmetric TCACGTGA sequence, but can also adopt an asymmetric structure and target a ATCACnCCAC, making contacts with ATCAC and GTGG half sites [51,118,119]. This dual mode of binding is conferred by the presence of an atypical Tyrosine at position 12, instead of the conserved Arginine [51].

In Figure 4 we have represented the protein-protein interaction network of bHLH factors obtained from filtered interactions deposited in STRING database [120]. A myriad of possible dimeric combinations can potentially occur between bHLH factors, both within and, to a lesser extent, between subfamilies. It is important to establish that a number of these STRING interactions do not necessarily occur *in vivo*, since the proposed partners may not be colocalizing spatially or temporally in the organism. Fast advances in single-cell technology make it possible to conceive the completion of a bHLH expression atlas across cell types and developmental times to add and remove edges in this network. Dimer composition depends on the relative concentration of the factors, plus the relative dimerization affinities they have between them [105]. Moreover, phosphorylation of the HLH domain of some factors influences their choice of partner [19,70,105,121–123]. For example, it impairs HIF1A association with ARNT [123], homodimerization of chicken MyoD [121] and promotes Neurog2 heterodimerization with Tcf3 [122] and Olig2 homodimerization [124].

As the network of *in vitro* possible dimers is highly connected (Figure 4), multiple factors can potentially compete *in vivo* for a common dimerization partner. Thus, a bHLH protein can deprive dimerization partners of another protein, and thus act as an indirect inhibitor [26,125,126].

However, not all the bHLH dimers are able to bind DNA. Factors of class D, ID proteins in vertebrates, lack a basic domain and consequently form non-DNA binding heterodimers with other factors [15,85,86]. Factors of this group dimerize preferentially with factors of class A, and within this class, more strongly with the E proteins (TCF3, TCF4, TCF12) [127–129]. By sequestering them in the form of inactive heterodimers, ID proteins reduce the concentration of available E proteins, thus indirectly impeding the E protein-dependent DNA binding of other factors of group A [86]. This way, Id proteins are capable of inhibiting biological processes directed by class A factors, such as myogenesis [85,86] and B-cell differentiation [130,131]. Class E factors, which typically homodimerize, as mentioned, or heterodimerize via their bHLH-Orange domains and repress transcription [110], have also been shown to heterodimerize with class A and class C factors, impairing the binding to their cognate sequences, and thus acting in a manner similar to Id proteins [43,50,58,76,110,132–134]. However, it is not certain that these interactions occur at physiological expression levels [132], and other possible mechanisms of repression in class E are more relevant *in vivo* [110], as we will describe below. Class A

factors, such as Twist, Mist1, Atoh8 and Hand1 also have been shown to be able to form inactive heterodimers [84,126,135,136].

Dimer-dependent repression can be achieved among bHLH by yet another mechanism, in which the “repressor” partner drives recognition of the same sequences but is transcriptionally inert, so indirectly represses transcription by competing for binding with the transcriptionally active dimer [47,137–140]. Further, some bHLH factors can dimerize both with activator and repressor partners [94,141,142]. For example, MAX can dimerize with MYC transcriptional activator, with MAD, MGA and MNT repressors, and with itself as a transcriptionally inert homodimer, and all those dimers can compete for binding to a common set of sequences [94,138,142]. Similarly, ARNT2 represses transcription as a homodimer, and turns into an activator when heterodimerizing with NPAS4 [141]. BHLH factors can also dimerize with members of other transcription factor families, which most of the time results in the formation of dimers unable to bind DNA [143–149].

## 6. Cooperative Binding with Other Transcription-Factors

Protein-protein interactions of bHLH factors are not restricted to their bHLH dimerization partners and as generally occurs in all families of transcription factors, additional cofactors are needed to make effective the activation or repression of their target genes. This includes several effector molecules such as chromatin remodelers, mediator complex, histone modifiers and enzymes regulating DNA methylation, but also other transcription factors. Indeed, certain bHLH members can interact with other transcription factors and cooperatively bind DNA. Under this mode of cooperation, TFs bind jointly to DNA through protein-protein interactions (sometimes through bridging cofactors), enhancing both their affinity to DNA and their ability to recruit transcriptional machinery.

This type of interaction can be mediated by amino acids scattered through the bHLH domain, and sometimes outside of it. For several bHLH factors, amino acids in the basic region that are not essential to DNA binding have been proved critical to their function and/or binding preferences [25,52,88,150–154]. Some of these residues may influence sequence binding preferences through subtle contacts with DNA, as shown more recently by de Masi et al. [64], while others can directly affect the conformation of the main DNA contacting residues, as discussed above. Moreover, these residues that have an influence but are not essential for DNA binding have been implicated in protein-protein contacts with additional transcription factors. For example, all myogenic factors (MYOG, MYOD1, MYF5 and MYF6) contain an Alanine in position 5 and a Threonine in position 6 of the basic domain, known for some as the myogenic code (Figure 3), which has proven to be essential in the formation of cooperatively binding complexes with Mef2 and Pbx/Meis transcription factors [155,156]. Amino acids in the HLH domain but facing away from the dimerization surface, as well as some located outside the bHLH domain, have also been suggested to influence binding and target specificity through interactions with additional transcription factors [21,157–162]. In some cases, well-defined additional domains outside the bHLH mediate the interaction, as in the case of Ptf1a or Neurod1 [157,163–165].

Multiple instances of direct cooperative binding between bHLH and other TFs have been identified. For example: HES1-c-Myb [166], c-Myc-TFII-I [167], c-Myc-USF [168], MYC-YY1 [168], Ptf1a-Rbpj [157,163,165], USF1-Ets1 [169], Neurod1-PDX1 [164], Twist1-PRC1/2 [170] and yeast Cbf1-Met4-Met28 (as a complex) [118,171–173]. This cooperative binding can constitute a mechanism of binding specificity of individual bHLH predicted to bind similar E-boxes, so as only the bHLH with the capacity to interact with another factor that binds to an adjacent motif will activate (or repress) target gene expression [87]. Conversely, it is also possible that in situations involving multiple interacting factors, only one of the factors can recognize specifically the E-box. For instance, this scenario has been described in one enhancer of the Notch ligand Delta1, where both Ascl1 and Neurog2 can interact with the Brn1/2 POU factors but only Ascl1 recognizes the particular E-box [87,174].



Furthermore, as the cooperative complex stabilizes protein-DNA interactions, it allows transcription factors to bind suboptimal sequences, which could not bind individually [118]. For example, it has been shown that in the myogenin promoter, DNA-bound Pbx1A-Meis1 dimers recruit MyoD to non-canonical E-box sequences [155,160]. Analogously, Myc is recruited by resident chromatin proteins and by proteins of the transcriptional machinery to promoters, which allows binding to less preferred E-box, and even to random sequences [89,175]. In the case of PTF1-RBPJ1, both factors allow variations in their cognate sequences [163]. Interactions with cooperating factors can also modify the structure of the bHLH dimer, altering its DNA recognition. For example, in hematopoietic cells, TAL1-E47 (a splice variant of TCF3) dimers, bind to the bridging cofactor LMO2, which in turn interacts with different transcription factors that determine the complex' target specificity: Sp1 in hematopoietic progenitors [176], GATA1/2 in erythroid cells [125,177–179] and RUNX1, ETS1 and GATA3 in leukemogenic T-cells [125,180]. Interaction with LMO2 modifies the bHLH dimer in such a way, that in most cases, only E47 and not TAL1 binds DNA, to a TG dinucleotide 7-9bps upstream the sequence motif of the cooperating factor [125,177,181,182].

A bHLH can also cooperatively interact with itself forming homotetramers that bind to tandem E-boxes: TWIST1 [71,183], MYOD1 [25,184,185], NEUROD2 [184], yeast Cbf1p [186], C-Myc-Max dimers [187] and MLXIPL-MLX dimers [188] for example can accomplish this. It has been shown that homotypic clustering of multiple binding sites of a bHLH factor strongly enhances binding to DNA and transcriptional response [70,185,189,190], which is understood as a pervasive mechanism across TFs in general [191]. In addition to cooperative binding through homotypic complexes, cooperative binding independent of physical interactions between the transcription factors [192] and cooperative recruitment of transcriptional cofactors can help explain the enhanced transcriptional response associated with these clusters.

Indirect cooperativity among TFs can manifest as cooperative transcriptional activation via independent binding to DNA. This type of cooperation can require a complex motif grammar along the associated genomic region. For example, in the mouse ventral neural tube, chicken NeuroM or Neurog2 can bind DNA in the HB9 promoter and then form a complex with adjacently bound LIM-homeodomain (LIM-HD) factors, through the LIM adapter Ldb1 (NLI), which act as a bridge. Two Lhx3 factors bound at both sides of the bHLH factors are sufficient for V2 interneuron generation, whereas in the formation of motor neurons, those sites are occupied by Isl1 factors, and the Lhx3 factors are located some nucleotides further away [157,193,194]. It has been shown that phosphorylation of Neurog2 facilitates the interaction with the NLI adaptors in the generation of the motor neurons [195]. Other bHLH factors present in the system, such as Ascl1, can also bind to the same E-boxes, but cannot interact with NLI, so the formation or not of the entire complex is what drives regulatory specificity in this case [193].

Adjacently bound factors do not always cooperate to activate transcription; in some cases, co-factor interaction mediates repression. For example, in the IgH enhancer, an unknown factor that binds to an E-box inactivates the rather distantly bound MyoD or TFE3, but not other bHLH factors [154]. Similarly, another unknown E-box binding factor specifically represses transcriptional activity of TFE3 in the prothymosin- $\alpha$  intron enhancer [196,197].

Understanding the three-dimensional architecture of the genome is critical to examine events of cooperation between bHLH factors and other TFs involving regions of the genome not adjacent in terms of DNA coordinates, when the factors are brought into contact by DNA looping. Pitx1-Neurod1 [198], MyoD-MEF2 [151], Myc-Max bivalent homotetramers [94], USF bivalent homotetramers [38] and *Drosophila* Achaete/Scute with Pannier, through the bridging cofactor Chip [199,200], for example, have been reported to interact in this fashion.

A particular modality of bHLH activity through co-factors can implicate no direct binding of the bHLH to DNA and can mediate transcriptional repression or acti-



vation. Tcf21 [201], TWIST1 [128,202,203], Hand1 [84], Hey1 [204], Hes and Hey proteins [110,148,205], and Dec proteins (BHLHE40 and BHLHE41) [76,133,206–208] can act by repressing the activity of other previously bound TFs, as corepressors, while some other factors, such as MyoD, HAND2 and HES1, can act as coactivators [156,209–211]. As a general principle, TFs that regulate targets as part of an enhanceosome, can also be recruited by other factors without binding DNA by themselves and coactivate transcription, as in the case of the yeast Tye7p factor [186].

Furthermore, Twist has been shown to inhibit two general transcriptional coactivators: p300 and PCAF [14]. And Hes/Hey proteins are able to influence transcription through yet additional mechanisms: by binding and blocking the basal transcriptional machinery, by promoting the degradation of TFs, and by promoting complex formation with other factors and kinases, facilitating phosphorylation and activation [110,212].

## 7. Chromatin Accessibility and Pioneer Factors

Chromatin accessibility is a major determinant of *in vivo* transcription factor binding, and an additional source by which they acquire binding specificity. Some factors can only bind open chromatin, while others, termed pioneer factors, can access highly packed, closed chromatin, and promote its remodeling. Differences in chromatin accessibility among cell types can contribute to explain different observed E-box occupancy among bHLH with similar motif preferences. Experiments conducting comparisons of binding sites after inducing ectopic expression of certain TFs in other cell types have helped to establish to which degree E-box binding is determined by accessible chromatin landscapes. These observations are influenced by the pioneer ability of the compared bHLH member to bind nucleosome-packed chromatin.

Several bHLH factors, such as USF1/2 [213], TCF3 (in B cells) [213], yeast Pho4 [214], HIF1A-ARNT [215], and MYC [138,142,216] among others, have shown to preferentially target accessible chromatin. This preference for open chromatin is particularly evident when comparing binding of TFs across different cell types. For example, when comparing MYOD1 and NEUROD2 binding in P19 cells vs. fibroblasts [184] and MYOD1 also in myotubes vs. rhabdomyosarcoma cells [92].

Multiple bHLH members of the A class, which generally participate in cell differentiation processes, can act as pioneer factors, albeit with differences among them in terms of supporting evidence, cell types or additional co-factor requirements for this pioneer activity. Ascl1 has been robustly shown to act as a pioneer factor in fibroblasts [217], in glioblastoma cells [218], in neural progenitors [219], but not in keratinocytes [217]. Neurod1 binds to silenced chromatin of regulatory elements of neuronal genes during neurogenesis in neural progenitors [220]. In pericytes, however, Neurod1 pioneer activity requires co-expression with Sox2 (another pioneer TF) to target inaccessible DNA [221]. Pioneer activity has also been suggested for Neurog2, presenting a neurogenic role in fibroblasts [222], and for MYOD1, when ectopically expressed in mouse embryonic stem cells [99].

Soufi et al. [98] associated the pioneer characteristics of some transcription factors to the length of the basic helix-1 and to their ability to bind centrally degenerate motifs in the surface of the nucleosome [98]. For example, the pioneer factor ASCL1 has a short basic helix 1 and thus contacts only the “CA” core dinucleotide, leaving the central dinucleotide free for nucleosome binding, which is reflected on the centrally degenerate E-boxes bound by ASCL1 in nucleosome-rich targets. They also found that MYC, which preferentially targets open chromatin, binding to an invariant CACGTG motif, also can target closed chromatin, through a centrally degenerate E-box, presumably binding only to the core “CA” through a partially folded basic helix-1 [98]. MYC co-binds with other factors when targeting inaccessible chromatin, and this interaction probably stabilizes the weak binding of the partially unfolded MYC basic region to the centrally degenerate E-box [98].

Even if some TF have a long basic-helix-1 motif, such as MyoD [98] and were consequently predicted to not bind nucleosome-rich sites, MYOD1, contrary to the notion derived from Fong et al. [184] and MacQuarrie et al. [92], presented similar ability to bind inaccessi-

ble chromatin when ectopically expressed in mouse embryonic stem cells as ASCL1/2 [99], and both TFs majorly bound the same sites when ectopically expressed in mouse embryonic fibroblasts compared to their native cell types, neural progenitors and myotubes, respectively [162]. Of note, sites preferentially bound by Ascl1 in fibroblasts were enriched in centrally degenerate E-boxes, as observed by Soufi et al. [98], whereas Myod1-preferred sites harbored E-boxes with a fixed central GC [162]. Contrary to Soufi et al. [98], predictions, Casey et al. [99] did not report centrally degenerate E-boxes in closed chromatin bound by ASCL1, ASCL2, and MYOD1, but only a slight preference for the GG central dinucleotide [99] which suggest the existence of other mechanisms of pioneer bHLH/E-box interactions. For example, ASCL1, ASCL2 and MYOD1, but not NEUROD1 or TCF21, binding sites in closed chromatin revealed a spatially reiterated pattern of E-boxes separated by ~10–15bp [99]. This constrained pattern led to the suspicion that such E-boxes could be accessible in the surface of the nucleosome and allow tetrameric complexes to bind in those loci.

Another possible mechanism that can help explain the apparent pioneer activity of MYOD1 involves the action of co-factors. In Q. Y. Lee et al. [162], experiments, canonical E-boxes were absent in about half of the Myod1-enriched sites, and in that case, Myod1 binding regions were enriched for additional motifs, including Homeobox (Pbx/Meis), MADS and REST, whereas Ascl1-preferred sites were more enriched in E-boxes and depleted in additional motifs. This fits with previous evidence showing that Myod1 can form tetrameric complexes with Pbx/Meis factors to bind non-canonical E-boxes in the nucleosome-rich myogenin promoter [155,223–225] and suggests that in this case Myod1 pioneering activity is facilitated or mediated by cooperative binding with other pioneer factors, as was proposed for Myc by Soufi et al. [98]. Of note, Casey et al. [99] also found Pbx/Meis motifs enriched in Myod1-bound sites, however, those were not specifically enriched in closed chromatin.

Once bound, pioneer factors remodel chromatin to leave DNA accessible for other transcription factors. For example, Myod1 binding to non-canonical E-boxes via Pbx/Meis promotes chromatin remodeling of the myogenin promoter, which makes previously hidden E-boxes accessible for Myod1 binding [155,223]. Moreover, the ability to remodel chromatin can crosstalk with post-translational modifications of certain bHLH pioneer factors. Neurog2 and Ascl1 can be phosphorylated on multiple Serine-Proline sites, with increasing phosphorylated sites implying decreasing affinity to DNA [226,227]. Therefore, promoters that are epigenetically available are largely insensitive to Neurog2 and Ascl1 phospho-status, while those that require substantial remodeling quantitatively respond to Neurog2 and Ascl1 phospho-status [226,227].

Finally, pioneering activities can yet derive from another mode of cooperative binding, independent of physical interactions. For example, Ptf1a co-binds with Fox and GATA factors in the pancreas and with Sox and Hox in the neural tube [228,229]. Those factors are pioneers in their respective tissues, opening chromatin and thus allowing Ptf1a binding.

## 8. DNA Modifications

As an additional source of binding specificity, bHLH factors can also differentially recognize chemical modifications of DNA bases. Cytosines in CpG sites frequently present a methyl group bound to their 5th position, which can be progressively oxidized to 5-hydroxymethylcytosine (5hmC), then 5-formylcytosine (5fC), and then be subsequently transformed into 5-carboxylcytosine (5caC) [230]. Symmetrical methylation of the central CpG of E-boxes has shown to prevent Myc-Max, Max-Max and HIF1A-ARNT binding to DNA [215,231–233]. In the case of Max homodimers, and oxidation to a 5caC restores the affinity to the level of the unmodified cytosine [233]. This recognition of the centrally modified cytosines is mediated by the Arginine at position 13 of Max, and conservation of this amino acid in all class B factors (Figure 2) suggests they all interact equally with that modified base [234]. In vitro methylation interference assays on the guanines in the central

dinucleotide also disrupt the binding of MYC-MAX in canonical or non-canonical E-boxes containing central CG or TG dinucleotides [44].

Conversely, modification of the central CpG has very little effect on TCF4 binding, whereas any type of C modification of the core CA (5mC, 5hmC, 5fC or 5CaC) has a negative impact on binding affinity [235]. However, a 5caC in the CpG immediately flanking the E-box enhances the binding of Tcf4, Tcf3, Tcf12 and Ascl1 [235,236]. The crystal structure of Tcf3 shows that Arginines of positions 1 and 2 of the basic region make these contacts [236], which are conserved in factors of the subfamilies Net, E12/E47, MyoD, Atonal, Mist, Neurogenin, NeuroD and MyoD (Figure 2).

## 9. Shape

As discussed above, bHLH factors do not only bind DNA through specific contacts with nucleotide bases, but through non-specific interactions with the phosphodiester backbone too [64]. Sometimes, the latter mode of binding prevails over the former, and allows the factors to sense the 3D shape of DNA. While it is true that DNA shape ultimately depends on the nucleotide sequence, different nucleotide combinations can result in the same shape. Thus, a bHLH protein that heavily relies on shape recognition can bind to different sequences, including non-E-box motifs [9].

Samee et al. [9] developed an algorithm to detect shape motifs from DNA sequence, and when applying it to ChIP-seq data of 7 bHLH factors, found that 5 of them, USF1, MAX, MXI1, TAL1 and BHLHE40, recognized specific shape motifs. This mode of binding has been proposed to account for the large divergence on *in vivo* binding landscapes of MYC-MAX heterodimers vs. MAX homodimers [9,237]. Recognition of DNA shape in positions distally flanking the E-box has also been proposed to drive target specificity of Ascl1 vs. Neurog2 [238] and yeast Tye7 vs. Cbf1 vs. Pho4 [239].

## 10. Binding to Non-B DNA

Some evidence suggests that binding of bHLH factors to DNA can imply DNA structure other than the classical Watson-Crick double helix or B-form. For example, MyoD and Myf6 homodimers were shown to bind four-stranded structures, called G-quadruplex, that are formed in guanine-rich tracts, which are enriched in promoters of human genes [240–242], for example in promoters of muscle-specific genes [243–245]. Such bHLH homodimers bound more tightly the quadruplex structures than the E-box containing B-form DNA, whereas heterodimers with E-proteins, or homodimers composed by the bHLH domain alone, preferred E-boxes over the quadruplex [243,244,246]. In contrast, homodimers of Myog, another myogenic regulatory factor, bound weakly the tetraplex structure [247]. However, it remains to be elucidated how this modality of binding affects *in vivo* expression of target genes. One hypothesis states that these G-quadruplex might sequester transcriptionally inert MyoD and Myf6 homodimers, this way promoting activation of muscle genes, as they no longer compete with their heterodimeric form with E-proteins which display a higher affinity for E-boxes [243,246,248].

## 11. Expression Levels

An obvious natural mechanism that can restrict collisions of bHLH on the same E-box is the spatiotemporal expression confinement of certain bHLH members. Multiple experiments inducing ectopic expression or over-expression of bHLH indicate that many collisions on E-boxes could be possible if certain bHLH were ever to share space and time or were expressed at higher levels. However, we already know that many bHLH with similarly preferred motifs are indeed co-expressed *in vivo*, which results in an allowed, if not required, set of interactions among them. For instance, when the factors that compete for the binding sites are all activators of transcription, this redundancy may result in enhanced transactivation accompanied by an increased robustness to mutations, which occurs likely in processes such as neuronal differentiation [249]. Different bHLH factors can also target the same E-box sequentially, carrying out complementary functions. Myf5

and MyoD bind to the same sites, but Myf5 binds first inducing histone acetylation and the subsequent binding of MyoD results in the recruitment of Pol II and thus activation of gene expression [250].

In contrast, bHLH factors that repress transcription, such as MNT [142], Dece (BHLHE40 and BHLHE41) [208,251], Bhlha15 [136], HES1 [140], yeast Cbf1 [214] and Msc [23] for example, and also non-bHLH repressors, such as like Snai1/2 [143,152,252], Myt1l [253], and ZEB [197,254] antagonize activator bHLH factors when competing for the same E-box sites. These E-box competing repressors can impede activator bHLH binding to a subset of E-boxes related to a particular biological function and impose thresholds on activator concentrations to trigger transcription.

Myc is the most studied factor regarding overexpression. When overexpressed, typically in tumor cells, in addition to its preferred CACGTG motif, it binds low-affinity sequences that show no resemblance to the E-box, such as AACGTT, thus broadly occupying the euchromatic cis-regulatory landscape of the cell [138,142,255]. Tal1 and Olig2 have also been shown to bind to degenerate E-boxes when overexpressed in cancer [98], and Atoh7 also binds to non-preferred motif sequences when overexpressed in vitro [140].

## 12. ChIP-Seq

The different modes of binding discussed above explain why in vivo genome-wide binding sites can hardly be inferred solely from the presence of TF binding motifs determined in vitro. The divergence from the in-silico prediction is repeatedly shown by ChIP-seq studies, the most widely used technique for in vivo binding assessment. In this section, we will discuss how ChIP-seq can inform us about the mechanisms of bHLH binding specificity that we have explained above, and we will describe the current status of the accumulated body of data of bHLH ChIP-seqs, across cell types and bHLH families.

By applying de novo motif discovery algorithms to the regions determined to be bound by a bHLH ChIP-seq, E-boxes typically appear as the most enriched motifs while often also indicating central and flanking sequence preferences. Top-enriched E-box motifs determined from ChIP-seq experiments can sometimes occlude a more nuanced scenario of motifs preferences. For example, Neurod2 or MyoD can bind motifs with different central dinucleotides, and while GC dinucleotide is associated with common targets between both TF, GA and GG are more associated with neuronal and myogenic genes, respectively [184]. The stratification of TF binding regions according to multiple functional and biological criteria can reveal subgroups of enriched E-boxes. Additional motifs of other transcription factors can also be enriched in the bound regions, indicating putative cooperative binding [95,98,162,184,189,256–261]. Finding fixed spacing patterns between the motifs [182] and/or leveraging ChIP-seq data of the co-enriched transcription factors and finding overlapping binding sites provides further validation of these inferred co-binding events [95,256].

Comparing ChIP-seq binding landscapes of bHLH factors with chromatin accessibility maps (as determined for instance by ATAC-seq, MNase-seq or H3K27ac) before the onset of the TF expression can be used to analyze to which extent the factor can bind to closed chromatin and thus act as a pioneer factor [92,99,184,216,260]. Conversely, when assessing chromatin accessibility posterior to TF expression, its remodeling capability can be determined [162,238,260]. Interestingly, Lee et al. [162] found that binding strength, measured by the intensity of the ChIP-seq signal, rather than mere binding, correlated with the extent of subsequent chromatin modification and transcriptional activation. These interesting findings suggest a qualitative promiscuity of binding of bHLH TFs with similar motif preference, resolved only when assessing binding quantitatively. Finally, ChIP-seq results can also be combined with genome-wide assessments of other epigenetic modifications such as DNA methylation, and measure correlations with bHLH binding. This is the case of Neurod2, whose binding sites are associated with regions undergoing hypomethylation during neuronal development, due to the interaction between Neurod2 and TET2 [262].



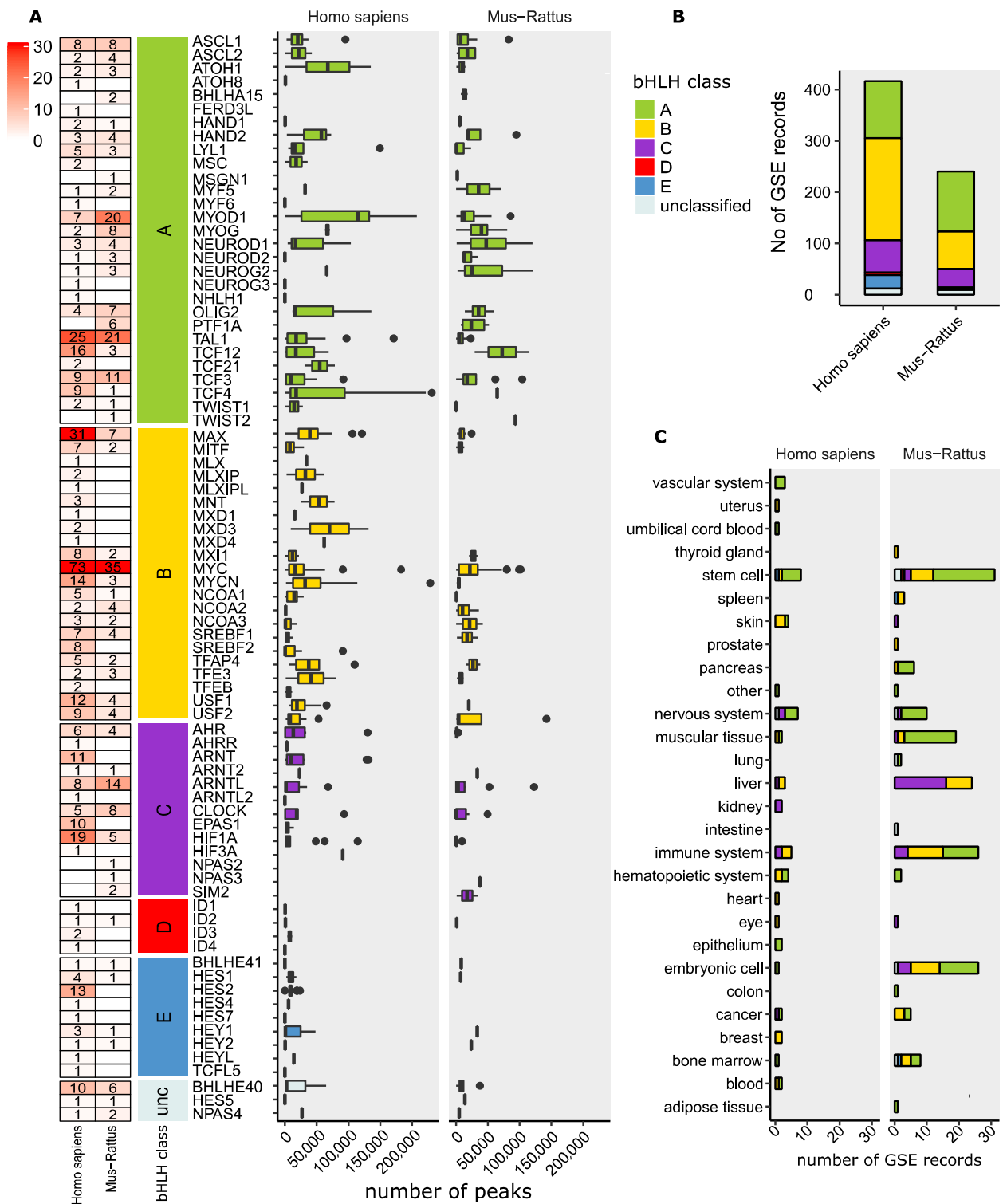
Not surprisingly, large differences arise when comparing binding landscapes of a factor in different cell types [184,260]. Therefore, to answer the question of how bHLH proteins acquire binding specificity over the other members of their class, binding of two or more factors has to be tested in the same cell type at the same time. This way, multiple studies have compared ChIP-seq binding landscapes between bHLH factors, finding different degrees of overlap for the binding sites, and attributing the differential binding to sequence preferences on central or flanking nucleotides of the E-box [90,95,97,99,184,214,238,261], cooperative binding with other factors [90,99,162,184,186,261], the ability to target closed chromatin [99] and DNA shape recognition [9,238,239].

As we have explained through this review, multiple mechanisms can explain why sequence motif preferences escape prediction in *in vivo* systems. For example, in *Drosophila* embryo ChIP-seq assays, Twist binds the TA central dinucleotide 7% of the times, whereas in the *in vitro* SELEX assay 35.6% of the times [71]. One possible explanation for this is the choice of dimerization partners. *In vitro* binding assays are performed with a single dimer of the factor, typically the homodimer, whereas *in vivo*, the factor can potentially be dimerizing with multiple partners with different sequence preferences, and which can affect its own half-site preference. Complex formation with other transcription factors can modify the structure of the bHLH dimer and thus its sequence preference, as in the case of TCF3-TAL1 binding to only a half-site of the E-box when forming a complex with LMO2-GATA/ETs1/Runx1 [177,181,182,263]. Additionally, as exemplified by MYC, a bHLH factor can be recruited by chromatin-bound proteins and thus target a wide variety of sequence combinations, including those with low affinity with the dimer [89,175,216]. When targeting inaccessible chromatin, bHLH factors can also bind to a modified E-box, concretely a centrally modified one, as proposed by Soufi et al. [98]. Further, and as we have seen above, sometimes recruited bHLH factors do not even interact with DNA [156,186,211]. As a result of this, a fraction of ChIP-bound regions by a factor may not contain its preferred sequence, or directly no E-box at all, and instead is the preferred motif of the recruiting factor. Moreover, when a factor heavily relies on DNA shape for binding, many sites may not contain an E-box, nor any additional factor motifs. Whereas classical PWM models that treat nucleotides independently fail to identify these shape binding sites, models that consider higher-order interactions between nucleotides or that explicitly use DNA shape characteristics can be more accurate [9,118].

Despite the aforementioned shortcomings, ChIP-seq experiments represent fundamental steps to build TF regulatory networks and to compare those networks across species and biological conditions. We have analyzed the current status of aggregated bHLH ChIP-seq data using the Gene Transcription Regulation Database (GTRD) [264] which integrates and re-process, using a standardized protocol, TF ChIP-seq datasets deposited in ENCODE and Short Read Archive (Figure 5 and Table S1) and performed on human and rodents.

There is great disparity in the number of studies dedicated to each bHLH TF (Figure 5A). We found that 32 bHLH members have never been interrogated by ChIP-seq (Table S1), while some other members have received particular attention. For example, MYC is the most surveyed TF in the database and includes 73 human and 35 mouse records, followed by MAX, an MYC dimerization partner with 35 records in humans and 7 records in rodents. MYC ChIP-seq experiments have been conducted in a wide variety of tissues and cell types, including immune system cells, liver, embryonic stem cells and bone marrow in rodents, whereas in humans, the majority of studies used cell lines, and a few skin and breast tissues (Table S1). Because of MYC's growth/oncogenic activity [33], it is not surprising that most of these ChIP-seq records were gathered in the context of cancer biology. Moreover, if we attend to specific bHLH classes, class E and D have rarely been studied by ChIP-seq in rodents or humans. In the majority of cases, these classes show one record per gene (Figure 5A). On aggregate, human studies almost double studies performed on rodents, albeit with some differences in the proportion of bHLH classes (Figure 5B).





**Figure 5.** Summary measures of ChIP-seq experiments conducted on bHLH TF using data collected from GTRD (<https://gtrd.biouml.org>, accessed on 17 May 2021) (A) Number of GSE records for each available bHLH transcription factor (32 had no available data in GTRD, see Table S1). Boxplot shows the number of peaks detected by MACS2 for each transcription factor averaging over the sum of experiments with the same GSE record ID, variability of numbers could reflect technical or biological differences among conditions and replicates. (B) The number of studies conducted in each class of bHLH in humans and rodents. (C) Barplot showing the distribution of tissues evaluated by ChIP-seq experiments stratified by species and representing the contribution of each bHLH class.

We also observed large variability in the average number of peaks among TFs and also within TFs across their various studies (Figure 5A). Using again MYC as an example, the average number of peaks ranged from 6 in the rodent's embryonic fibroblast (GSE67694) to 101596 in the thyroid gland (GSE85648), and in humans from 40 in fetal lung cell lines (GSE81899) to 182929 in breast carcinoma (GSE1006866). This variability is reflecting many sources of variation ranging from biological conditions to technical aspects such as specific antibodies, induced TF expression, etc. (Table S1).

Finally, the distribution by tissue and cell types shows inter-class differences. For example, class C with many known members of core clock genes such as ARNTL and CLOCK has been extensively studied in the liver (Figure 5C), while studies in stem cells are largely dominated by ChIP-seq experiments on class A bHLH TFs.

### 13. Conclusions and Future Directions

We have presented here a plethora of molecular mechanisms that influence and make possible the establishment of specific regulatory networks among different bHLH transcription factors. We have seen bHLH sequence differences determining motif preferences, alongside multiple types of such motif preferences: E-box central and flanking nucleotides, as well as flanking motifs, motif spacing and other forms of complex motif grammar; together with an intricate set of spatiotemporally regulated co-factor interactions also affecting the DNA binding landscape. In particular, the case of MYOD1 and ASCL1 represents an example rich with nuances to illustrate the intermingling aspects of pioneering activity, structural differences, sequence preference and co-factor requirements, and also the value of leveraging ChIP-seq experiments derived from multiple cell types. As more ChIP-seq data accumulates, the degree to which a given E-box can be or cannot be bound by many bHLH and which are the reasons for that binding sharedness or discrepancy will gain more detail, and what now constitute examples will potentially become generalizable principles. This will also benefit from the completion of full organisms' cell-type gene expression atlases, at different developmental stages, in the line of initiatives such as the Human Cell Atlas, which will allow the exact description of which bHLH, and bHLH co-factors, are actually co-expressed, and thus can potentially collaborate or collide for DNA binding. In addition, ChIP-seq experiments conducting comparative and quantitative binding landscapes on ectopically expressed TF, combined with protein domain shuffling, will continue to be a valuable tool to dissect binding specificity mechanisms among phylogenetically close bHLH with very similar motif preferences. In that regard, recent findings suggest that certain bHLH regulating very different differentiation programs can actually bind with high promiscuity on a similar set of E-boxes and, under certain conditions independent of that bHLH binding or expression, drive the differentiation program of the other bHLH [162]. This kind of analysis is fundamental and reminds us to avoid single TF-centered reductionist approaches to understand a regulatory network. Finally, the development and standardization of high-throughput TF-ChIP-seq techniques will allow the implementation of more comprehensive experimental designs that will aid us to understand the mechanisms shaped by natural selection that allowed and accommodated the radiation and functional specializations of all bHLHs.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijms22179150/s1>.

**Author Contributions:** All authors contributed to the collection of information and writing of this review. All authors have read and agreed to the published version of the manuscript.

**Funding:** G.S. received the support of a fellowship from "la Caixa" Foundation (ID 100010434). The fellowship code is LCF/BQ/PI19/11690010. G.S. is also supported by Ministerio de Ciencia e Innovación, Spain (PID2019-104700GA-I00) and by the NIH grant R01HG010898-01.

**Acknowledgments:** We would like to thank Semyon Kolmykov for providing help and data from GTRD.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2018**, *172*, 650–665. [[CrossRef](#)]
- Thurman, R.E.; Rynes, E.; Humbert, R.; Vierstra, J.; Maurano, M.T.; Haugen, E.; Sheffield, N.C.; Stergachis, A.B.; Wang, H.; Vernot, B.; et al. The accessible chromatin landscape of the human genome. *Nature* **2012**, *489*, 75–82. [[CrossRef](#)] [[PubMed](#)]
- Vierstra, J.; Rynes, E.; Sandstrom, R.; Zhang, M.; Canfield, T.; Scott Hansen, R.; Stehling-Sun, S.; Sabo, P.J.; Byron, R.; Humbert, R.; et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **2014**, *346*, 1007–1012. [[CrossRef](#)] [[PubMed](#)]
- Maurano, M.T.; Haugen, E.; Sandstrom, R.; Vierstra, J.; Shafer, A.; Kaul, R.; Stamatoyannopoulos, J.A. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **2015**, *47*, 1393–1401. [[CrossRef](#)] [[PubMed](#)]
- Atak, Z.K.; Taskiran, I.I.; Demeulemeester, J.; Flerin, C.; Mauduit, D.; Minnoye, L.; Hulselmans, G.; Christiaens, V.; Ghanem, G.E.; Wouters, J.; et al. Interpretation of allele-specific chromatin accessibility using cell state-aware deep learning. *Genome Res.* **2021**, *31*, 1082–1096. [[CrossRef](#)] [[PubMed](#)]
- Kasowski, M.; Grubert, F.; Heffelfinger, C.; Hariharan, M.; Asabere, A.; Waszak, S.M.; Habegger, L.; Rozowsky, J.; Shi, M.; Urban, A.E.; et al. Variation in transcription factor binding among humans. *Science* **2010**, *328*, 232–235. [[CrossRef](#)] [[PubMed](#)]
- Yan, J.; Qiu, Y.; Ribeiro dos Santos, A.M.; Yin, Y.; Li, Y.E.; Vinckier, N.; Nariai, N.; Benaglio, P.; Raman, A.; Li, X.; et al. Systematic analysis of binding of transcription factors to noncoding variants. *Nature* **2021**, *591*, 147–151. [[CrossRef](#)]
- Gronau, I.; Arbiza, L.; Mohammed, J.; Siepel, A. Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. *Mol. Biol. Evol.* **2013**, *30*, 1159–1171. [[CrossRef](#)] [[PubMed](#)]
- Samee, M.A.H.; Bruneau, B.G.; Pollard, K.S. A De Novo Shape Motif Discovery Algorithm Reveals Preferences of Transcription Factors for DNA Shape Beyond Sequence Motifs. *Cell Syst.* **2019**, *8*, 27–42.e6. [[CrossRef](#)]
- Rubinstein, M.; Souza, F.S.J. de Evolution of Transcriptional Enhancers and Animal Diversity. *Philos. Trans. Royal Soc. B Biol. Sci.* **2013**, *368*, 20130017. [[CrossRef](#)]
- Klemm, S.L.; Shipony, Z.; Greenleaf, W.J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **2019**, *20*, 207–220. [[CrossRef](#)]
- Kribelbauer, J.F.; Rastogi, C.; Bussemaker, H.J.; Mann, R.S. Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. *Annu Rev Cell Dev Bi* **2019**, *35*, 1–23. [[CrossRef](#)]
- Murre, C.; Bain, G.; van Dijk, M.A.; Engel, I.; Furnari, B.A.; Massari, M.E.; Matthews, J.R.; Quong, M.W.; Rivera, R.R.; Stuiver, M.H. Structure and function of helix-loop-helix proteins. *BBA Gene Struct. Expr.* **1994**, *1218*, 129–135. [[CrossRef](#)]
- Massari, M.E.; Murre, C. Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. *Mol. Cell. Biol.* **2000**, *20*, 429–440. [[CrossRef](#)]
- Atchley, W.R.; Fitch, W.M. A natural classification of the basic helix-loop-helix class of transcription factors. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 5172–5176. [[CrossRef](#)]
- Ledent, V.; Paquet, O.; Vervoort, M. Phylogenetic analysis of the human basic helix-loop-helix proteins. *Genome Biol.* **2002**, *3*, 1–18. [[CrossRef](#)] [[PubMed](#)]
- Simionato, E.; Ledent, V.; Richards, G.; Thomas-Chollier, M.; Kerner, P.; Coornaert, D.; Degnan, B.M.; Vervoort, M. Origin and diversification of the basic helix-loop-helix gene family in metazoans: Insights from comparative genomics. *BMC Evol. Biol.* **2007**, *7*, 1–18. [[CrossRef](#)] [[PubMed](#)]
- Heim, M.A.; Jakoby, M.; Werber, M.; Martin, C.; Weisshaar, B.; Bailey, P.C. The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* **2003**, *20*, 735–747. [[CrossRef](#)]
- Dennis, D.J.; Han, S.; Schuurmans, C. bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res.* **2019**, *1705*, 48–65. [[CrossRef](#)]
- Block, N.E.; Miller, J.B. Expression of MRF4, a myogenic helix-loop-helix protein, produces multiple changes in the myogenic program of BC3H-1 cells. *Mol. Cell. Biol.* **1992**, *12*, 2484–2492. [[CrossRef](#)] [[PubMed](#)]
- Dezan, C.; Meierhans, D.; Kunne, A.G.E.; Allemann, R.K. Acquisition of myogenic specificity through replacement of one amino acid of MASH-1 and introduction of an additional  $\alpha$ -helical turn. *Biol. Chem.* **1999**, *380*, 705–710. [[CrossRef](#)]
- Fujisawa-Sehara, A.; Nabeshima, Y.; Komiyama, T.; Uetsuki, T.; Asakura, A.; Nabeshima, Y.I. Differential trans-activation of muscle-specific regulatory elements including the myosin light chain box by chicken MyoD, myogenin, and MRF4. *J. Biol. Chem.* **1992**, *267*, 10031–10038. [[CrossRef](#)]
- Lu, J.; Webb, R.; Richardson, J.A.; Olson, E.N. MyoR: A muscle-restricted basic helix-loop-helix transcription factor that antagonizes the actions of MyoD. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 552–557. [[CrossRef](#)] [[PubMed](#)]
- Penn, B.H.; Bergstrom, D.A.; Dilworth, F.J.; Bengal, E.; Tapscott, S.J. A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev.* **2004**, *18*, 2348–2353. [[CrossRef](#)] [[PubMed](#)]

25. Weintraub, H.; Dwarki, V.J.; Verma, I.; Davis, R.; Hollenberg, S.; Snider, L.; Lassar, A.; Tapscott, S.J. Muscle-specific transcriptional activation by MyoD. *Genes Dev.* **1991**, *5*, 1377–1386. [[CrossRef](#)] [[PubMed](#)]
26. Porcher, C.; Liao, E.C.; Fujiwara, Y.; Zon, L.I.; Orkin, S.H. Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. *Development* **1999**, *126*, 4603–4615. [[CrossRef](#)]
27. Shivdasani, R.A.; Mayer, E.L.; Orkin, S.H. Absence of blood formation in mice lacking the T-cell leukaemia oncogene tal-1/SCL. *Nature* **1995**, *373*, 432–434. [[CrossRef](#)] [[PubMed](#)]
28. Larigot, L.; Juricek, L.; Dairou, J.; Coumoul, X. AhR signaling pathways and regulatory functions. *Biochim. Open* **2018**, *7*, 1–9. [[CrossRef](#)] [[PubMed](#)]
29. Nakayama, K.; Kataoka, N. Regulation of gene expression under hypoxic conditions. *Int. J. Mol. Sci.* **2019**, *20*, 3278. [[CrossRef](#)]
30. McDonald, M.J.; Rosbash, M.; Emery, P. Wild-Type Circadian Rhythmicity Is Dependent on Closely Spaced E Boxes in the *Drosophila* timeless Promoter. *Mol. Cell. Biol.* **2001**, *21*, 1207–1217. [[CrossRef](#)]
31. Nakahata, Y.; Yoshida, M.; Takano, A.; Soma, H.; Yamamoto, T.; Yasuda, A.; Nakatsu, T.; Takumi, T. A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. *BMC Mol. Biol.* **2008**, *9*, 1–11. [[CrossRef](#)] [[PubMed](#)]
32. Sato, F.; Kawamoto, T.; Fujimoto, K.; Noshiro, M.; Honda, K.K.; Honma, S.; Honma, K.I.; Kato, Y. Functional analysis of the basic helix-loop-helix transcription factor DEC1 in circadian regulation: Interaction with BMAL1. *Eur. J. Biochem.* **2004**, *271*, 4409–4419. [[CrossRef](#)] [[PubMed](#)]
33. Carroll, P.A.; Freie, B.W.; Mathsyaraja, H.; Eisenman, R.N. The MYC transcription factor network: Balancing metabolism, proliferation and oncogenesis. *Front. Med.* **2018**, *12*, 412–425. [[CrossRef](#)] [[PubMed](#)]
34. Aguilar-Rodríguez, J.; Peel, L.; Stella, M.; Wagner, A.; Payne, J.L. The architecture of an empirical genotype-phenotype map. *Evolution (N. Y.)* **2018**, *72*, 1242–1260. [[CrossRef](#)]
35. Liu, Z.; Venkatesh, S.S.; Maley, C.C. Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples. *BMC Genom.* **2008**, *9*, 1–17. [[CrossRef](#)] [[PubMed](#)]
36. Ellenberger, T.; Fass, D.; Arnaud, M.; Harrison, S.C. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* **1994**, *8*, 970–980. [[CrossRef](#)] [[PubMed](#)]
37. Ferré-D'Amaré, A.R.; Prendergast, G.C.; Ziff, E.B.; Burley, S.K. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **1993**, *363*, 38–45. [[CrossRef](#)] [[PubMed](#)]
38. Ferré-D'Amaré, A.R.; Pognonec, P.; Roeder, R.G.; Burley, S.K. Structure and function of the b/HLH/Z domain of USF. *EMBO J.* **1994**, *13*, 180–189. [[CrossRef](#)]
39. Ma, P.C.M.; Rould, M.A.; Weintraub, H.; Pabo, C.O. Crystal structure of MyoD bHLH domain-DNA complex: Perspectives on DNA recognition and implications for transcriptional activation. *Cell* **1994**, *77*, 451–459. [[CrossRef](#)]
40. Longo, A.; Guanga, G.P.; Rose, R.B. Crystal structure of E47-NeuroD1/Beta2 bHLH domain-DNA complex: Heterodimer selectivity and DNA recognition. *Biochemistry* **2008**, *47*, 218–229. [[CrossRef](#)]
41. Murre, C.; McCaw, P.S.; Vaessin, H.; Caudy, M.; Jan, L.Y.; Jan, Y.N.; Cabrera, C.V.; Buskin, J.N.; Hauschka, S.D.; Lassar, A.B.; et al. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **1989**, *58*, 537–544. [[CrossRef](#)]
42. Blackwell, T.K.; Weintraub, H. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **1990**, *250*, 1104–1110. [[CrossRef](#)] [[PubMed](#)]
43. Akazawa, C.; Sasai, Y.; Nakanishi, S.; Kageyama, R. Molecular characterization of a rat negative regulator with a basic helix-loop-helix structure predominantly expressed in the developing nervous system. *J. Biol. Chem.* **1992**, *267*, 21879–21885. [[CrossRef](#)]
44. Blackwell, T.K.; Huang, J.; Ma, A.; Kretzner, L.; Alt, F.W.; Eisenman, R.N.; Weintraub, H. Binding of myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell. Biol.* **1993**, *13*, 5216–5224. [[CrossRef](#)]
45. Yokoyama, C.; Wang, X.; Briggs, M.R.; Admon, A.; Wu, J.; Hua, X.; Goldstein, J.L.; Brown, M.S. SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene. *Cell* **1993**, *75*, 187–197. [[CrossRef](#)]
46. Ishibashi, M.; Sasai, Y.; Nakanishi, S.; Kageyama, R. Molecular characterization of HES-2, a mammalian helix-loop-helix factor structurally related to *Drosophila* hairy and Enhancer of split. *Eur. J. Biochem.* **1993**, *215*, 645–652. [[CrossRef](#)]
47. Fisher, F.; Crouch, D.H.; Jayaraman, P.S.; Clark, W.; Gillespie, D.A.F.; Goding, C.R. Transcription activation by Myc and Max: Flanking sequences target activation to a subset of CACGTG motifs in vivo. *EMBO J.* **1993**, *12*, 5075–5082. [[CrossRef](#)] [[PubMed](#)]
48. Whitelaw, M.; Pongratz, I.; Wilhelmsson, A.; Gustafsson, J.A.; Poellinger, L. Ligand-dependent recruitment of the Arnt coregulator determines DNA recognition by the dioxin receptor. *Mol. Cell. Biol.* **1993**, *13*, 2504–2514. [[CrossRef](#)]
49. Tontonoz, P.; Kim, J.B.; Graves, R.A.; Spiegelman, B.M. ADD1: A novel helix-loop-helix transcription factor associated with adipocyte determination and differentiation. *Mol. Cell. Biol.* **1993**, *13*, 4753–4759. [[CrossRef](#)]
50. Takebayashi, K.; Sasai, Y.; Sakai, Y.; Watanabe, T.; Nakanishi, S.; Kageyama, R. Structure, chromosomal locus, and promoter analysis of the gene encoding the mouse helix-loop-helix factor HES-1. Negative autoregulation through the multiple N box elements. *J. Biol. Chem.* **1994**, *269*, 5150–5156. [[CrossRef](#)]
51. Kim, J.B.; Spotts, G.D.; Halvorsen, Y.D.; Shih, H.M.; Ellenberger, T.; Towle, H.C.; Spiegelman, B.M. Dual DNA binding specificity of ADD1/SREBP1 controlled by a single amino acid in the basic helix-loop-helix domain. *Mol. Cell. Biol.* **1995**, *15*, 2582–2588. [[CrossRef](#)]



52. Robert, L.D.; Pei-Feng, C.; Lassar, B.A.; Harold, W. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **1990**, *60*, 733–746.
53. Halazonetis, T.D. Determination of the c-MYC DNA-binding site. *Proc. Natl. Acad. Sci.* **1991**, *88*, 6162–6166. [[CrossRef](#)] [[PubMed](#)]
54. Blackwood, E.M.; Eisenman, R.N. Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* **1991**, *251*, 1211–1217. [[CrossRef](#)] [[PubMed](#)]
55. Kerkhoff, E.; Bister, K.; Klempnauer, K.H. Sequence-specific DNA binding by Myc proteins. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4323–4327. [[CrossRef](#)] [[PubMed](#)]
56. Wright, W.E.; Binder, M.; Funk, W. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol. Cell. Biol.* **1991**, *11*, 4104–4110. [[CrossRef](#)] [[PubMed](#)]
57. Fisher, F.; Goding, C.R. Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. *EMBO J.* **1992**, *11*, 4103–4109. [[CrossRef](#)]
58. Sasai, Y.; Kageyama, R.; Tagawa, Y.; Shigemoto, R.; Nakanishi, S. Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. *Genes Dev.* **1992**, *6*, 2620–2634. [[CrossRef](#)]
59. Reyes, H.; Reisz-Porszasz, S.; Hankinson, O. Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science* **1992**, *256*, 1193–1195. [[CrossRef](#)]
60. Yin, Y.; Morgunova, E.; Jolma, A.; Kaasinen, E.; Sahu, B.; Khund-Sayeed, S.; Das, P.K.; Kivioja, T.; Dave, K.; Zhong, F.; et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **2017**, *356*. [[CrossRef](#)]
61. Jolma, A.; Yan, J.; Whittington, T.; Toivonen, J.; Nitta, K.R.; Rastas, P.; Morgunova, E.; Enge, M.; Taipale, M.; Wei, G.; et al. DNA-binding specificities of human transcription factors. *Cell* **2013**, *152*, 327–339. [[CrossRef](#)] [[PubMed](#)]
62. Grove, C.A.; De Masi, F.; Barrasa, M.I.; Newburger, D.E.; Alkema, M.J.; Bulyk, M.L.; Walkout, A.J.M. A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors. *Cell* **2009**, *138*, 314–327. [[CrossRef](#)]
63. Skinner, M.K.; Rawls, A.; Wilson-Rawls, J.; Roalson, E.H. Basic helix-loop-helix transcription factor gene family. *Differentiation* **2011**, *80*, 1–8. [[CrossRef](#)] [[PubMed](#)]
64. De Masi, F.; Grove, C.A.; Vedenko, A.; Alibés, A.; Gisselbrecht, S.S.; Serrano, L.; Bulyk, M.L.; Walkout, A.J.M. Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Res.* **2011**, *39*, 4553–4563. [[CrossRef](#)] [[PubMed](#)]
65. Bouard, C.; Terreux, R.; Honorat, M.; Manship, B.; Ansieau, S.; Vigneron, A.M.; Puisieux, A.; Payen, L. Deciphering the molecular mechanisms underlying the binding of the TWIST1/E12 complex to regulatory E-box sequences. *Nucleic Acids Res.* **2016**, *44*, 5470–5489. [[CrossRef](#)]
66. Pellanda, P.; Dalsass, M.; Filipuzzi, M.; Loffreda, A.; Verrecchia, A.; Castillo Cano, V.; Thabussot, H.; Doni, M.; Morelli, M.J.; Soucek, L.; et al. Integrated requirement of non-specific and sequence-specific DNA binding in Myc-driven transcription. *EMBO J.* **2021**, *40*, 1–17. [[CrossRef](#)]
67. Dang, C.V.; Dolde, C.; Gillison, M.L.; Kato, G.J. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins (DNA-protein interaction/transcription). *Genetics* **1992**, *89*, 599–602.
68. Swanson, H.I.; Chan, W.K.; Bradfield, C.A. DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *J. Biol. Chem.* **1995**, *270*, 26292–26302. [[CrossRef](#)]
69. Atchley, W.R.; Zhao, J. Molecular architecture of the DNA-binding region and its relationship to classification of basic helix-loop-helix proteins. *Mol. Biol. Evol.* **2007**, *24*, 192–202. [[CrossRef](#)]
70. Firulli, B.A.; Redick, B.A.; Conway, S.J.; Firulli, A.B. Mutations within helix I of twist1 result in distinct limb defects and variation of DNA binding affinities. *J. Biol. Chem.* **2007**, *282*, 27536–27546. [[CrossRef](#)]
71. Ozdemir, A.; Fisher-Aylor, K.I.; Pepke, S.; Samanta, M.; Dunipace, L.; McCue, K.; Zeng, L.; Ogawa, N.; Wold, B.J.; Stathopoulos, A. High resolution mapping of Twist to DNA in Drosophila embryos: Efficient functional analysis and evolutionary conservation. *Genome Res.* **2011**, *21*, 566–577. [[CrossRef](#)]
72. Kophengnavong, T.; Michnowicz, J.E.; Blackwell, T.K. Establishment of Distinct MyoD, E2A, and Twist DNA Binding Specificities by Different Basic Region-DNA Conformations. *Mol. Cell. Biol.* **2000**, *20*, 261–272. [[CrossRef](#)] [[PubMed](#)]
73. Van Doren, M.; Bailey, A.M.; Esnayra, J.; Ede, K.; Posakony, J.W. Negative regulation of proneural gene activity: Hairy is a direct transcriptional repressor of achaete. *Genes Dev.* **1994**, *8*, 2729–2742. [[CrossRef](#)] [[PubMed](#)]
74. Shimizu, T.; Toumoto, A.; Ihara, K.; Shimizu, M.; Kyogoku, Y.; Ogawa, N.; Oshima, Y.; Hakoshima, T. Crystal structure of PHO4 bHLH domain-DNA complex: Flanking base recognition. *EMBO J.* **1997**, *16*, 4689–4697. [[CrossRef](#)] [[PubMed](#)]
75. Brownlie, P.; Ceska, T.A.; Lamers, M.; Romier, C.; Stier, G.; Teo, H.; Suck, D. The crystal structure of an intact human Max-DNA complex: New insights into mechanisms of transcriptional control. *Structure* **1997**, *5*, 509–520. [[CrossRef](#)]
76. Sato, F.; Bhawal, U.K.; Kawamoto, T.; Fujimoto, K.; Imaizumi, T.; Imanaka, T.; Kondo, J.; Koyanagi, S.; Noshiro, M.; Yoshida, H.; et al. Basic-helix-loop-helix (bHLH) transcription factor DEC2 negatively regulates vascular endothelial growth factor expression. *Genes Cells* **2008**, *13*, 131–144. [[CrossRef](#)] [[PubMed](#)]
77. Kimura, H.; Weisz, A.; Ogura, T.; Hitomi, Y.; Kurashima, Y.; Hashimoto, K.; D’Acquisto, F.; Makuuchi, M.; Esumi, H. Identification of hypoxia-inducible factor 1 ancillary sequence and its function in vascular endothelial growth factor gene induction by hypoxia and nitric oxide. *J. Biol. Chem.* **2001**, *276*, 2292–2298. [[CrossRef](#)] [[PubMed](#)]
78. Murre, C. Helix-loop-helix proteins and the advent of cellular diversity: 30 years of discovery. *Genes Dev.* **2019**, *33*, 6–25. [[CrossRef](#)]



79. Tietze, K.; Oellers, N.; Knust, E. Enhancer of splitD, a dominant mutation of *Drosophila*, and its use in the study of functional domains of a helix-loop-helix protein. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6152–6156. [[CrossRef](#)]
80. Bessho, Y.; Miyoshi, G.; Sakata, R.; Kageyama, R. Hes7: A bHLH-type repressor gene regulated by Notch and expressed in the presomitic mesoderm. *Genes Cells* **2001**, *6*, 175–185. [[CrossRef](#)] [[PubMed](#)]
81. Hwang, B.; Lee, J.H.; Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **2018**, *50*, 1–14. [[CrossRef](#)]
82. Iso, T.; Chung, G.; Hamamori, Y.; Kedes, L. HERP1 is a cell type-specific primary target of Notch. *J. Biol. Chem.* **2002**, *277*, 6598–6607. [[CrossRef](#)]
83. Firulli, A.B. A HANDful of questions: The molecular biology of the heart and neural crest derivatives (HAND)-subclass of basic helix-loop-helix transcription factors. *Gene* **2003**, *312*, 27–40. [[CrossRef](#)]
84. Knöfler, M.; Meinhardt, G.; Bauer, S.; Loregger, T.; Vasicek, R.; Bloor, D.J.; Kimber, S.J.; Husslein, P. Human Hand1 basic helix-loop-helix (bHLH) protein: Extra-embryonic expression pattern, interaction partners and identification of its transcriptional repressor domains. *Biochem. J.* **2002**, *361*, 641–651. [[CrossRef](#)] [[PubMed](#)]
85. Benzra, R.; Davis, R.L.; Lockshon, D.; Turner, D.L.; Weintraub, H. The protein Id: A negative regulator of helix-loop-helix DNA binding proteins. *Cell* **1990**, *61*, 49–59. [[CrossRef](#)]
86. Wendt, H.; Thomas, R.M.; Ellenberger, T. DNA-mediated Folding and Assembly of MyoD-E47 Heterodimers. *J. Biol. Chem.* **1998**, *273*, 5735–5743. [[CrossRef](#)]
87. Powell, L.M.; Jarman, A.P. Context dependence of proneural bHLH proteins. *Curr. Opin. Genet. Dev.* **2008**, *18*, 411–417. [[CrossRef](#)]
88. Chien, C.T.; Hsiao, C.D.; Jan, L.Y.; Jan, Y.N. Neuronal type information encoded in the basic-helix-loop-helix domain of proneural genes. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13239–13244. [[CrossRef](#)]
89. Guo, J.; Li, T.; Schipper, J.; Nilson, K.A.; Fordjour, F.K.; Cooper, J.J.; Gordân, R.; Price, D.H. Sequence specificity incompletely defines the genome-wide occupancy of Myc. *Genome Biol.* **2014**, *15*, 482. [[CrossRef](#)]
90. Hejna, M.; Moon, W.M.; Cheng, J.; Kawakami, A.; Fisher, D.E.; Song, J.S. Local genomic features predict the distinct and overlapping binding patterns of the bHLH-Zip family oncoproteins MITF and MYC-MAX. *Pigment Cell Melanoma Res.* **2019**, *32*, 500–509. [[CrossRef](#)] [[PubMed](#)]
91. Jennings, B.H.; Tyler, D.M.; Bray, S.J. Target Specificities of *Drosophila* Enhancer of split Basic Helix-Loop-Helix Proteins. *Mol. Cell. Biol.* **1999**, *19*, 4600–4610. [[CrossRef](#)] [[PubMed](#)]
92. MacQuarrie, K.L.; Yao, Z.; Fong, A.P.; Diede, S.J.; Rudzinski, E.R.; Hawkins, D.S.; Tapscott, S.J. Comparison of Genome-Wide Binding of MyoD in Normal Human Myogenic Cells and Rhabdomyosarcomas Identifies Regional and Local Suppression of Promyogenic Transcription Factors. *Mol. Cell. Biol.* **2013**, *33*, 773–784. [[CrossRef](#)]
93. Maerkl, S.J.; Quake, S.R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **2007**, *315*, 233–237. [[CrossRef](#)] [[PubMed](#)]
94. Nair, S.K.; Burley, S.K. X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA. *Cell* **2003**, *112*, 193–205. [[CrossRef](#)]
95. Wang, J.; Zhuang, J.; Iyer, S.; Lin, X.Y.; Whitfield, T.W.; Greven, M.C.; Pierce, B.G.; Dong, X.; Kundaje, A.; Cheng, Y.; et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **2012**, *22*, 1798–1812. [[CrossRef](#)] [[PubMed](#)]
96. Beltran, A.C.; Dawson, P.E.; Gottesfeld, J.M. Role of DNA sequence in the binding specificity of synthetic basic-helix-loop-helix domains. *ChemBioChem* **2005**, *6*, 104–113. [[CrossRef](#)]
97. MacQuarrie, K.L.; Yao, Z.; Fong, A.P.; Tapscott, S.J. Genome-wide binding of the basic helix-loop-helix myogenic inhibitor myocardin has substantial overlap with MyoD: Implications for buffering activity. *Skelet. Muscle* **2013**, *3*, 1–10. [[CrossRef](#)]
98. Soufi, A.; Garcia, M.F.; Jaroszewicz, A.; Osman, N.; Pellegrini, M.; Zaret, K.S. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **2015**, *161*, 555–568. [[CrossRef](#)]
99. Casey, B.H.; Kollipara, R.K.; Pozo, K.; Johnson, J.E. Intrinsic DNA binding properties demonstrated for lineage-specifying basic helix-loop-helix transcription factors. *Genome Res.* **2018**, *28*, 484–496. [[CrossRef](#)] [[PubMed](#)]
100. Wu, D.; Potluri, N.; Lu, J.; Kim, Y.; Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature* **2015**, *524*, 303–308. [[CrossRef](#)] [[PubMed](#)]
101. Murakami, M.; Kataoka, K.; Fukuhara, S.; Nakagawa, O.; Kurihara, H. Akt-dependent phosphorylation negatively regulates the transcriptional activity of dHAND by inhibiting the DNA binding activity. *Eur. J. Biochem.* **2004**, *271*, 3330–3339. [[CrossRef](#)] [[PubMed](#)]
102. Takebayashi, K.; Takahashi, S.; Yokota, C.; Tsuda, H.; Nakanishi, S.; Asashima, M.; Kageyama, R. Conversion of ectoderm into a neural fate by ATH-3, a vertebrate basic helix-loop-helix gene homologous to *Drosophila* proneural gene atonal. *EMBO J.* **1997**, *16*, 384–395. [[CrossRef](#)]
103. Ström, A.; Castella, P.; Rockwood, J.; Wagner, J.; Caudy, M. Mediation of NGF signaling by post-translational inhibition of HES-1, a basic helix-loop-helix repressor of neuronal differentiation. *Genes Dev.* **1997**, *11*, 3168–3181. [[CrossRef](#)] [[PubMed](#)]
104. Li, L.; Zhou, J.; James, G.; Heller-Harrison, R.; Czech, M.P.; Olson, E.N. FGF inactivates myogenic helix-loop-helix proteins through phosphorylation of a conserved protein kinase C site in their DNA-binding domains. *Cell* **1992**, *71*, 1181–1194. [[CrossRef](#)]
105. Fan, X.; Waardenberg, A.J.; Demuth, M.; Osteil, P.; Sun, J.Q.; Loebel, D.A.F.; Graham, M.; Tam, P.P.L. TWIST1 Homodimers and Heterodimers Orchestrate Lineage-Specific Differentiation. *Mol. Cell. Biol.* **2020**, *40*, 1–20. [[CrossRef](#)]

106. Quan, X.J.; Yuan, L.; Tiberi, L.; Claeys, A.; De Geest, N.; Yan, J.; Van Der Kant, R.; Xie, W.R.; Klisch, T.J.; Shymkowitz, J.; et al. Post-translational Control of the Temporal Dynamics of Transcription Factor Activity Regulates Neurogenesis. *Cell* **2016**, *164*, 460–475. [[CrossRef](#)]
107. Berberich, S.J.; Cole, M.D. Casein kinase II inhibits the DNA-binding activity of Max homodimers but not Myc/Max heterodimers. *Genes Dev.* **1992**, *6*, 166–176. [[CrossRef](#)] [[PubMed](#)]
108. Huang, N.; Chelliah, Y.; Shan, Y.; Taylor, C.A.; Yoo, S.H.; Partch, C.; Green, C.B.; Zhang, H.; Takahashi, J.S. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* **2012**, *337*, 189–194. [[CrossRef](#)] [[PubMed](#)]
109. Taelman, V.; Van Wayenbergh, R.; Sölter, M.; Pichon, B.; Pieler, T.; Christophe, D.; Bellefroid, E.J. Sequences downstream of the bHLH domain of the *Xenopus* hairy-related transcription factor-1 act as an extended dimerization domain that contributes to the selection of the partners. *Dev. Biol.* **2004**, *276*, 47–63. [[CrossRef](#)]
110. Fischer, A.; Gessler, M. Delta-Notch-and then? Protein interactions and proposed modes of repression by Hes and Hey bHLH factors. *Nucleic Acids Res.* **2007**, *35*, 4583–4596. [[CrossRef](#)]
111. Jones, N. Transcriptional regulation by dimerization: Two sides to an incestuous relationship. *Cell* **1990**, *61*, 9–11. [[CrossRef](#)]
112. Kadesch, T. Consequences of heteromeric interactions among helix-loop-helix proteins. *Cell Growth Differ.* **1993**, *4*, 49–55.
113. Le Dréau, G.; Escalona, R.; Fueyo, R.; Herrera, A.; Martínez, J.D.; Usieto, S.; Menendez, A.; Pons, S.; Martínez-Balbas, M.A.; Martí, E. E proteins sharpen neurogenesis by modulating proneural bHLH transcription factors' activity in an E-box-dependent manner. *Elife* **2018**, *7*, 1–29. [[CrossRef](#)]
114. Ohsako, S.; Hyer, J.; Panganiban, G.; Oliver, I.; Caudy, M. Hairy function as a DNA-binding helix-loop-helix repressor of *Drosophila* sensory organ formation. *Genes Dev.* **1994**, *8*, 2743–2755. [[CrossRef](#)]
115. Wang, Z.; Wu, Y.; Li, L.; Su, X.D. Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA. *Cell Res.* **2013**, *23*, 213–224. [[CrossRef](#)]
116. Lusska, A.; Shen, E.; Whitlock, J.P. Protein-DNA interactions at a dioxin-responsive enhancer. Analysis of six bona fide DNA-binding sites for the liganded Ah receptor. *J. Biol. Chem.* **1993**, *268*, 6575–6580. [[CrossRef](#)]
117. Kinoshita, K.; Kikuchi, Y.; Sasakura, Y.; Suzuki, M.; Fujii-Kuriyama, Y.; Sogawa, K. Altered DNA binding specificity of Arnt by selection of partner bHLH-PAS proteins. *Nucleic Acids Res.* **2004**, *32*, 3169–3179. [[CrossRef](#)]
118. Siggers, T.; Gordân, R. Protein-DNA binding: Complexities and multi-protein codes. *Nucleic Acids Res.* **2014**, *42*, 2099–2111. [[CrossRef](#)] [[PubMed](#)]
119. Párraga, A.; Bellolell, L.; Ferré-D'Amaré, A.R.; Burley, S.K. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* **1998**, *6*, 661–672. [[CrossRef](#)]
120. Jensen, L.J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; et al. STRING 8—A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **2009**, *37*, 412–416. [[CrossRef](#)] [[PubMed](#)]
121. Mitsui, K.; Shirakata, M.; Paterson, B.M. Phosphorylation inhibits the DNA-binding activity of MyoD homodimers but not MyoD-E12 heterodimers. *J. Biol. Chem.* **1993**, *268*, 24415–24420. [[CrossRef](#)]
122. Li, S.; Mattar, P.; Zinyk, D.; Singh, K.; Chaturvedi, C.P.; Kovach, C.; Dixit, R.; Kurrasch, D.M.; Ma, Y.C.; Chan, J.A.; et al. GSK3 temporally regulates Neurogenin 2 proneural activity in the neocortex. *J. Neurosci.* **2012**, *32*, 7791–7805. [[CrossRef](#)]
123. Kalousi, A.; Mylonis, I.; Politou, A.S.; Chachami, G.; Paraskeva, E.; Simos, G. Casein kinase 1 regulates human hypoxia-inducible factor HIF-1. *J. Cell Sci.* **2010**, *123*, 2976–2986. [[CrossRef](#)] [[PubMed](#)]
124. Li, H.; Paes de Faria, J.; Andrew, P.; Nitaraska, J.; Richardson, W.D. Phosphorylation Regulates OLIG2 Cofactor Choice and the Motor Neuron-Oligodendrocyte Fate Switch. *Neuron* **2011**, *69*, 918–929. [[CrossRef](#)] [[PubMed](#)]
125. El Omari, K.; Hoosdally, S.J.; Tuladhar, K.; Karia, D.; Hall-Ponselé, E.; Platonova, O.; Vyas, P.; Patient, R.; Porcher, C.; Mancini, E.J. Structural Basis for LMO2-Driven Recruitment of the SCL: E47bHLH Heterodimer to Hematopoietic-Specific Transcriptional Targets. *Cell Rep.* **2013**, *4*, 135–147. [[CrossRef](#)]
126. Spicer, D.B.; Rhee, J.; Cheung, W.L.; Lassar, A.B. Inhibition of Myogenic bHLH and MEF2 Transcription Factors by the bHLH Protein Twist. *Science* **1996**, *272*, 1476–1480. [[CrossRef](#)] [[PubMed](#)]
127. Langlands, K.; Yin, X.; Anand, G.; Prochownik, E.V. Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *J. Biol. Chem.* **1997**, *272*, 19785–19793. [[CrossRef](#)] [[PubMed](#)]
128. Jen, Y.; Weintraub, H.; Benezra, R. Overexpression of Id protein inhibits the muscle differentiation program: In vivo association of Id with E2A proteins. *Genes Dev.* **1992**, *6*, 1466–1479. [[CrossRef](#)] [[PubMed](#)]
129. Sun, X.H.; Copeland, N.G.; Jenkins, N.A.; Baltimore, D. Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. *Mol. Cell. Biol.* **1991**, *11*, 5603–5611. [[CrossRef](#)]
130. Cochrane, S.W.; Zhao, Y.; Welner, R.S.; Sun, X.H. Balance between Id and E proteins regulates myeloid-versus-lymphoid lineage decisions. *Blood* **2009**, *113*, 1016–1026. [[CrossRef](#)]
131. Rivera, R.; Murre, C. The regulation and function of the Id proteins in lymphocyte development. *Oncogene* **2001**, *20*, 8308–8316. [[CrossRef](#)] [[PubMed](#)]
132. Davis, R.L.; Turner, D.L. Vertebrate hairy and Enhancer of split related proteins: Transcriptional repressors regulating cellular differentiation and embryonic patterning. *Oncogene* **2001**, *20*, 8342–8357. [[CrossRef](#)] [[PubMed](#)]
133. Azmi, S.; Ozog, A.; Taneja, R. Sharp-1/DEC2 inhibits skeletal muscle differentiation through repression of myogenic transcription factors. *J. Biol. Chem.* **2004**, *279*, 52643–52652. [[CrossRef](#)] [[PubMed](#)]

134. Dhar, M.; Taneja, R. Cross-regulatory interaction between Stra13 and USF results in functional antagonism. *Oncogene* **2001**, *20*, 4750–4756. [[CrossRef](#)] [[PubMed](#)]
135. Ejarque, M.; Altirriba, J.; Gomis, R.; Gasa, R. Characterization of the transcriptional activity of the basic helix-loop-helix (bHLH) transcription factor Atoh8. *Biochim. Biophys. Acta Gene Regul. Mech.* **2013**, *1829*, 1175–1183. [[CrossRef](#)] [[PubMed](#)]
136. Lemerrier, C.; To, R.Q.; Carrasco, R.A.; Konieczny, S.F. The basic helix-loop-helix transcription factor Mist1 functions as a transcriptional repressor of MyoD. *EMBO J.* **1998**, *17*, 1412–1422. [[CrossRef](#)]
137. Castanon, I.; Von Stetina, S.; Kass, J.; Baylies, M.K. Dimerization partners determine the activity of the Twist bHLH protein during *Drosophila* mesoderm development. *Development* **2001**, *128*, 3145–3159. [[CrossRef](#)]
138. Allevato, M.; Bolotin, E.; Grossman, M.; Mane-padros, D.; Sladek, M.; Martinez, E. Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs. *PLoS ONE* **2017**, *7*, e0180147. [[CrossRef](#)]
139. Matter-Sadzinski, L.; Puzianowska-Kuznicka, M.; Hernandez, J.; Ballivet, M.; Matter, J.M. A bHLH transcriptional network regulating the specification of retinal ganglion cells. *Development* **2005**, *132*, 3907–3921. [[CrossRef](#)]
140. Hernandez, J.; Matter-Sadzinski, L.; Skowronska-Krawczyk, D.; Chiodini, F.; Alliod, C.; Ballivet, M.; Matter, J.M. Highly conserved sequences mediate the dynamic interplay of basic helix-loop-helix proteins regulating retinogenesis. *J. Biol. Chem.* **2007**, *282*, 37894–37905. [[CrossRef](#)]
141. Sharma, N.; Pollina, E.A.; Yap, E.L.; DiBiase, F.A.; Hrvatin, S.; Hu, L.; Lin, C.; Greenberg, M.E. ARNT2 Tunes Activity-Dependent Gene Expression through NCoR2-Mediated Repression and NPAS4-Mediated Activation. *Neuron* **2019**, *102*, 390–406.e9. [[CrossRef](#)]
142. Wolf, E.; Lin, C.Y.; Eilers, M.; Levens, D.L. Taming of the beast: Shaping Myc-dependent amplification. *Trends Cell Biol.* **2015**, *25*, 241–248. [[CrossRef](#)] [[PubMed](#)]
143. Soleimani, V.D.; Yin, H.; Jahani-Asl, A.; Ming, H.; Kockx, C.E.M.; van Ijcken, W.F.J.; Grosveld, F.; Rudnicki, M.A. Snail Regulates MyoD Binding-Site Occupancy to Direct Enhancer Switching and Differentiation-Specific Transcription in Myogenesis. *Mol. Cell* **2012**, *47*, 457–468. [[CrossRef](#)]
144. Tanoue, S.; Fujimoto, K.; Myung, J.; Hatanaka, F.; Kato, Y.; Takumi, T. DEC2-E4BP4 heterodimer represses the transcriptional enhancer activity of the EE element in the Per2 promoter. *Front. Neurol.* **2015**, *6*, 166. [[CrossRef](#)] [[PubMed](#)]
145. Pognonec, P.; Boulukos, K.E.; Aperlo, C.; Fujimoto, M.; Ariga, H.; Nomoto, A.; Kato, H. Cross-family interaction between the bHLHZip USF and bZip Fra1 proteins results in down-regulation of AP1 activity. *Oncogene* **1997**, *14*, 2091–2098. [[CrossRef](#)] [[PubMed](#)]
146. Bengal, E.; Ransone, L.; Scharfmann, R.; Dwarki, V.J.; Tapscott, S.J.; Weintraub, H.; Verma, I.M. Functional antagonism between c-Jun and MyoD proteins: A direct physical association. *Cell* **1992**, *68*, 507–519. [[CrossRef](#)]
147. Peukert, K.; Staller, P.; Schneider, A.; Carmichael, G.; Hänel, F.; Eilers, M. An alternative pathway for gene regulation by Myc. *EMBO J.* **1997**, *16*, 5672–5686. [[CrossRef](#)] [[PubMed](#)]
148. Liu, A.; Li, J.; Marin-Husstege, M.; Kageyama, R.; Fan, Y.; Gelinis, C.; Casaccia-Bonnel, P. A molecular insight of Hes5-dependent inhibition of myelin gene expression: Old partners and new players. *EMBO J.* **2006**, *25*, 4833–4842. [[CrossRef](#)] [[PubMed](#)]
149. Planque, N.; Leconte, L.; Coquelle, F.M.; Martin, P.; Saulet, S. Specific Pax-6/Microphthalmia Transcription Factor Interactions Involve Their DNA-binding Domains and Inhibit Transcriptional Properties of Both Proteins. *J. Biol. Chem.* **2001**, *276*, 29330–29337. [[CrossRef](#)]
150. Brennan, T.J.; Chakraborty, T.; Olson, E.N. Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5675–5679. [[CrossRef](#)]
151. Molkentin, J.D.; Olson, E.N. Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9366–9373.
152. Quan, X.J.; Denayer, T.; Yan, J.; Jafar-Nejad, H.; Philippi, A.; Lichtarge, O.; Vleminckx, K.; Hassan, B.A. Evolution of neural precursor selection: Functional divergence of proneural proteins. *Development* **2004**, *131*, 1679–1689. [[CrossRef](#)]
153. Skowronska-Krawczyk, D.; Ballivet, M.; Dynlacht, B.D.; Matter, J.M. Highly specific interactions between bHLH transcription factors and chromatin during retina development. *Development* **2004**, *131*, 4447–4454. [[CrossRef](#)] [[PubMed](#)]
154. Weintraub, H.; Genetta, T.; Kadesch, T. Tissue-specific gene activation by MyoD: Determination of specificity by cis-acting repression elements. *Genes Dev.* **1994**, *8*, 2203–2211. [[CrossRef](#)] [[PubMed](#)]
155. Heidt, A.B.; Rojas, A.; Harris, I.S.; Black, B.L. Determinants of Myogenic Specificity within MyoD Are Required for Noncanonical E Box Binding. *Mol. Cell. Biol.* **2007**, *27*, 5910–5920. [[CrossRef](#)]
156. Molkentin, J.D.; Black, B.L.; Martin, J.F.; Olson, E.N. Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* **1995**, *83*, 1125–1136. [[CrossRef](#)]
157. Lai, H.C.; Meredith, D.M.; Johnson, J.E. *bHLH Factors in Neurogenesis and Neuronal Subtype Specification*; Elsevier Inc.: Amsterdam, The Netherlands, 2013; ISBN 9780123972651.
158. Nakada, Y.; Hunsaker, T.L.; Henke, R.M.; Johnson, J.E. Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification. *Development* **2004**, *131*, 1319–1330. [[CrossRef](#)]
159. Ohneda, K.; Mirmira, R.G.; Wang, J.; Johnson, J.D.; German, M.S. The Homeodomain of PDX-1 Mediates Multiple Protein-Protein Interactions in the Formation of a Transcriptional Activation Complex on the Insulin Promoter. *Mol. Cell. Biol.* **2000**, *20*, 900–911. [[CrossRef](#)] [[PubMed](#)]



160. Berkes, C.A.; Bergstrom, D.A.; Penn, B.H.; Seaver, K.J.; Knoepfler, P.S.; Tapscott, S.J. Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol. Cell* **2004**, *14*, 465–477. [[CrossRef](#)]
161. Fong, A.P.; Yao, Z.; Zhong, J.W.; Johnson, N.M.; Farr, G.H.; Maves, L.; Tapscott, S.J. Conversion of MyoD to a neurogenic factor: Binding site specificity determines lineage. *Cell Rep.* **2015**, *10*, 1937–1946. [[CrossRef](#)]
162. Lee, Q.Y.; Mall, M.; Chanda, S.; Zhou, B.; Sharma, K.S.; Schaukowitz, K.; Adrian-Segarra, J.M.; Grieder, S.D.; Kareta, M.S.; Wapinski, O.L.; et al. Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes. *Nat. Cell Biol.* **2020**, *22*, 401–411. [[CrossRef](#)] [[PubMed](#)]
163. Beres, T.M.; Masui, T.; Swift, G.H.; Shi, L.; Henke, R.M.; MacDonald, R.J. PTF1 Is an Organ-Specific and Notch-Independent Basic Helix-Loop-Helix Complex Containing the Mammalian Suppressor of Hairless (RBP-J) or Its Parologue, RBP-L. *Mol. Cell. Biol.* **2006**, *26*, 117–130. [[CrossRef](#)]
164. Glick, E.; Leshkowitz, D.; Walker, M.D. Transcription Factor BETA2 Acts Cooperatively with E2A and PDX1 to Activate the Insulin Gene Promoter. *J. Biol. Chem.* **2000**, *275*, 2199–2204. [[CrossRef](#)]
165. Hori, K.; Cholewa-Waclaw, J.; Nakada, Y.; Glasgow, S.M.; Masui, T.; Henke, R.M.; Wildner, H.; Martarelli, B.; Beres, T.M.; Epstein, J.A.; et al. A nonclassical bHLH-Rbpj transcription factor complex is required for specification of GABAergic neurons independent of Notch signaling. *Genes Dev.* **2008**, *22*, 166–178. [[CrossRef](#)]
166. Allen, R.D.; Kim, H.K.; Sarafova, S.D.; Siu, G. Negative Regulation of CD4 Gene Expression by a HES-1–c-Myb Complex. *Mol. Cell. Biol.* **2001**, *21*, 3071–3082. [[CrossRef](#)]
167. Roy, A.L.; Carruthers, C.; Gutjahr, T.; Robert, B. Direct role for Myc in transcription initiation mediated by interactions with TFII-I. *Nature* **1993**, *365*, 359–361. [[CrossRef](#)]
168. Roy, A.L.; Du, H.; Gregor, P.D.; Novina, C.D.; Martinez, E.; Roeder, R.G. Cloning of an inr- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1. *EMBO J.* **1997**, *16*, 7091–7104. [[CrossRef](#)]
169. Sieweke, M.H.; Tekotte, H.; Jarosch, U.; Graf, T. Cooperative interaction of Ets-1 with USF-1 required for HIV-1 enhancer activity in T cells. *EMBO J.* **1998**, *17*, 1728–1739. [[CrossRef](#)]
170. Yang, M.H.; Hsu, D.S.S.; Wang, H.W.; Wang, H.J.; Lan, H.Y.; Yang, W.H.; Huang, C.H.; Kao, S.Y.; Tzeng, C.H.; Tai, S.K.; et al. Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition. *Nat. Cell Biol.* **2010**, *12*, 982–992. [[CrossRef](#)]
171. Blaiseau, P.L.; Thomas, D. Multiple transcriptional activation complexes tether the yeast activator Met4 to DNA. *EMBO J.* **1998**, *17*, 6327–6336. [[CrossRef](#)] [[PubMed](#)]
172. Kuras, L.; Cherest, H.; Surdin-Kerjan, Y.; Thomas, D. A heteromeric complex containing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and Met28, mediates the transcription activation of yeast sulfur metabolism. *EMBO J.* **1996**, *15*, 2519–2529. [[CrossRef](#)] [[PubMed](#)]
173. Siggers, T.; Duyzend, M.H.; Reddy, J.; Khan, S.; Bulyk, M.L. Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol. Syst. Biol.* **2011**, *7*, 1–14. [[CrossRef](#)]
174. Castro, D.S.; Skowronska-Krawczyk, D.; Armant, O.; Donaldson, I.J.; Parras, C.; Hunt, C.; Critchley, J.A.; Nguyen, L.; Gossler, A.; Göttgens, B.; et al. Proneural bHLH and Brn Proteins Coregulate a Neurogenic Program through Cooperative Binding to a Conserved DNA Motif. *Dev. Cell* **2006**, *11*, 831–844. [[CrossRef](#)]
175. Lorenzin, F.; Benary, U.; Baluapuri, A.; Walz, S.; Jung, L.A.; von Eyss, B.; Kisker, C.; Wolf, J.; Eilers, M.; Wolf, E. Different promoter affinities account for specificity in MYC-dependent gene regulation. *Elife* **2016**, *5*, 1–35. [[CrossRef](#)] [[PubMed](#)]
176. Lécuyer, E.; Herblot, S.; Saint-Denis, M.; Martin, R.; Glenn Begley, C.; Porcher, C.; Orkin, S.H.; Hoang, T. The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. *Blood* **2002**, *100*, 2430–2440. [[CrossRef](#)] [[PubMed](#)]
177. Kassouf, M.T.; Hughes, J.R.; Taylor, S.; McGowan, S.J.; Soneji, S.; Green, A.L.; Vyas, P.; Porcher, C. Genome-wide identification of TAL1's functional targets: Insights into its mechanisms of action in primary erythroid cells. *Genome Res.* **2010**, *20*, 1064–1083. [[CrossRef](#)]
178. Osada, H.; Grutz, G.; Axelson, H.; Forster, A.; Rabbitts, T.H. Association of erythroid transcription factors: Complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9585–9589. [[CrossRef](#)] [[PubMed](#)]
179. Wadman, I.A.; Osada, H.; Grütz, G.G.; Agulnick, A.D.; Westphal, H.; Forster, A.; Rabbitts, T.H. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* **1997**, *16*, 3145–3157. [[CrossRef](#)]
180. Ono, Y.; Fukuhara, N.; Yoshie, O. TAL1 and LIM-Only Proteins Synergistically Induce Retinaldehyde Dehydrogenase 2 Expression in T-Cell Acute Lymphoblastic Leukemia by Acting as Cofactors for GATA3. *Mol. Cell. Biol.* **1998**, *18*, 6939–6950. [[CrossRef](#)]
181. Han, G.C.; Vinayachandran, V.; Bataille, A.R.; Park, B.; Chan-salis, K.Y.; Keller, C.A.; Long, M.; Mahony, S.; Hardison, R.C.; Pugh, B.F. Genome-wide organization of GATA1 and TAL1 determined at high resolution. *Mol. Cell. Biology* **2016**, *36*, 157–172. [[CrossRef](#)]
182. Soler, E.; Andrieu-Soler, C.; De Boer, E.; Bryne, J.C.; Thongjuea, S.; Stadhouders, R.; Palstra, R.J.; Stevens, M.; Kockx, C.; Van Ijcken, W.; et al. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev.* **2010**, *24*, 277–289. [[CrossRef](#)]
183. Chang, A.T.; Liu, Y.; Ayyanathan, K.; Benner, C.; Jiang, Y.; Prokop, J.W.; Paz, H.; Wang, D.; Li, H.R.; Fu, X.D.; et al. An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev.* **2015**, *29*, 603–616. [[CrossRef](#)] [[PubMed](#)]

184. Fong, A.P.; Yao, Z.; Zhong, J.W.; Cao, Y.; Ruzzo, W.L.; Gentleman, R.C.; Tapscott, S.J. Genetic and Epigenetic Determinants of Neurogenesis and Myogenesis. *Dev. Cell* **2012**, *22*, 721–735. [[CrossRef](#)] [[PubMed](#)]
185. Weintraub, H.; Davis, R.; Lockshon, D.; Lassar, A. MyoD binds cooperatively to two sites in a target enhancer sequence: Occupancy of two sites is required for activation. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 5623–5627. [[CrossRef](#)]
186. Shively, C.A.; Liu, J.; Chen, X.; Loell, K.; Mitra, R.D. Homotypic cooperativity and collective binding are determinants of bHLH specificity and function. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 16143–16152. [[CrossRef](#)]
187. Walhout, A.J.M.; Gubbels, J.M.; Bernards, R.; Van Der Vliet, P.C.; Timmers, H.T.M. C-Myc/Max heterodimers bind cooperatively to the E-box sequences located in the first intron of the rat ornithine decarboxylase (ODC) gene. *Nucleic Acids Res.* **1997**, *25*, 1493–1501. [[CrossRef](#)] [[PubMed](#)]
188. Ma, L.; Sham, Y.Y.; Walters, K.J.; Towle, H.C. A critical role for the loop region of the basic helix-loop-helix/ leucine zipper protein Mlx in DNA binding and glucose-regulated transcription. *Nucleic Acids Res.* **2007**, *35*, 35–44. [[CrossRef](#)]
189. Seo, S.; Lim, J.W.; Yellajoshyula, D.; Chang, L.W.; Kroll, K.L. Neurogenin and NeuroD direct transcriptional targets and their regulatory enhancers. *EMBO J.* **2007**, *26*, 5093–5108. [[CrossRef](#)]
190. Wilkinson, G.; Dennis, D.; Schuurmans, C. Proneural genes in neocortical development. *Neuroscience* **2013**, *253*, 256–273. [[CrossRef](#)]
191. Gotea, V.; Visel, A.; Westlund, J.M.; Nobrega, M.A.; Pennacchio, L.A.; Ovcharenko, I. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.* **2010**, *20*, 565–577. [[CrossRef](#)]
192. Ezer, D.; Zabet, N.R.; Adryan, B. Homotypic clusters of transcription factor binding sites: A model system for understanding the physical mechanics of gene expression. *Comput. Struct. Biotechnol. J.* **2014**, *10*, 63–69. [[CrossRef](#)]
193. Lee, S.K.; Pfaff, S.L. Synchronization of neurogenesis and motor neuron specification by direct coupling of bHLH and homeodomain transcription factors. *Neuron* **2003**, *38*, 731–745. [[CrossRef](#)]
194. Thaler, J.P.; Lee, S.K.; Jurata, L.W.; Gill, G.N.; Pfaff, S.L. LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. *Cell* **2002**, *110*, 237–249. [[CrossRef](#)]
195. Ma, Y.C.; Song, M.R.; Park, J.P.; Henry Ho, H.Y.; Hu, L.; Kurtev, M.V.; Zieg, J.; Ma, Q.; Pfaff, S.L.; Greenberg, M.E. Regulation of Motor Neuron Specification by Phosphorylation of Neurogenin 2. *Neuron* **2008**, *58*, 65–77. [[CrossRef](#)] [[PubMed](#)]
196. Desbarats, L.; Gaubatz, S.; Eilers, M. Discrimination between different E-box-binding proteins at an endogenous target gene of c-myc. *Genes Dev.* **1996**, *10*, 447–460. [[CrossRef](#)] [[PubMed](#)]
197. Genetta, T.; Ruezinsky, D.; Kadesch, T. Displacement of an E-box-binding repressor by basic helix-loop-helix proteins: Implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. *Mol. Cell. Biol.* **1994**, *14*, 6153–6163. [[CrossRef](#)] [[PubMed](#)]
198. Poulin, G.; Lebel, M.; Chamberland, M.; Paradis, F.W.; Drouin, J. Specific Protein-Protein Interaction between Basic Helix-Loop-Helix Transcription Factors and Homeoproteins of the Pitx Family. *Mol. Cell. Biol.* **2000**, *20*, 4826–4837. [[CrossRef](#)] [[PubMed](#)]
199. Bertrand, N.; Castro, D.S.; Guillemot, F. Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **2002**, *3*, 517–530. [[CrossRef](#)] [[PubMed](#)]
200. Romain, P.; Khechumian, R.; Khechumian, K.; Arbogast, N.; Ackermann, C.; Heitzler, P. Interactions between chip and the achaete/scute-daughterless heterodimers are required for Pannier-driven proneural patterning. *Mol. Cell* **2000**, *6*, 781–790. [[CrossRef](#)]
201. Cheol, Y.H.; Gong, E.Y.; Kim, K.; Ji, H.S.; Ko, H.M.; Hyun, J.L.; Choi, H.S.; Lee, K. Modulation of the expression and transactivation of androgen receptor by the basic helix-loop-helix transcription factor pod-1 through recruitment of histone deacetylase 1. *Mol. Endocrinol.* **2005**, *19*, 2245–2257. [[CrossRef](#)]
202. Curtis, A.M.; Seo, S.B.; Westgate, E.J.; Rudic, R.D.; Smyth, E.M.; Chakravarti, D.; FitzGerald, G.A.; McNamara, P. Histone Acetyltransferase-dependent Chromatin Remodeling and the Vascular Clock. *J. Biol. Chem.* **2004**, *279*, 7091–7097. [[CrossRef](#)] [[PubMed](#)]
203. Hamamori, Y.; Wu, H.Y.; Sartorelli, V.; Kedes, L. The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, Twist. *Mol. Cell. Biol.* **1997**, *17*, 6563–6573. [[CrossRef](#)]
204. Belandia, B.; Powell, S.M.; García-Pedrero, J.M.; Walker, M.M.; Bevan, C.L.; Parker, M.G. Hey1, a Mediator of Notch Signaling, Is an Androgen Receptor Corepressor. *Mol. Cell. Biol.* **2005**, *25*, 1425–1436. [[CrossRef](#)]
205. King, I.N.; Kathiriya, I.S.; Murakami, M.; Nakagawa, M.; Gardner, K.A.; Srivastava, D.; Nakagawa, O. Hrt and Hes negatively regulate Notch signaling through interactions with RBP-J $\kappa$ . *Biochem. Biophys. Res. Commun.* **2006**, *345*, 446–452. [[CrossRef](#)]
206. Cho, Y.; Noshiro, M.; Choi, M.; Morita, K.; Kawamoto, T.; Fujimoto, K.; Kato, Y.; Makishima, M. The basic helix-loop-helix proteins differentiated embryo chondrocyte (DEC) 1 and DEC2 function as corepressors of retinoid X receptors. *Mol. Pharmacol.* **2009**, *76*, 1360–1369. [[CrossRef](#)] [[PubMed](#)]
207. Gulbagci, N.T.; Li, L.; Ling, B.; Gopinadhan, S.; Walsh, M.; Rossner, M.; Nave, K.A.; Taneja, R. SHARP1/DEC2 inhibits adipogenic differentiation by regulating the activity of C/EBP. *EMBO Rep.* **2009**, *10*, 79–86. [[CrossRef](#)] [[PubMed](#)]
208. Honma, S.; Kawamoto, T.; Takagi, Y.; Fujimoto, K.; Sato, F.; Noshiro, M.; Kato, Y.; Honma, K.I. Dec1 and Dec2 are regulators of the mammalian molecular clock. *Nature* **2002**, *419*, 841–844. [[CrossRef](#)] [[PubMed](#)]
209. Dai, Y.S.; Cserjesi, P.; Markham, B.E.; Molkenstin, J.D. The transcription factors GATA4 and dHAND physically interact to synergistically activate cardiac gene expression through a p300-dependent mechanism. *J. Biol. Chem.* **2002**, *277*, 24390–24398. [[CrossRef](#)]



210. McLarren, K.W.; Lo, R.; Grbavec, D.; Thirunavukkarasu, K.; Karsenty, G.; Stifani, S. The mammalian basic helix loop helix protein HES-1 binds to and modulates the transactivating function of the runt-related factor Cbfa1. *J. Biol. Chem.* **2000**, *275*, 530–538. [[CrossRef](#)]
211. Zang, M.X.; Li, Y.; Xue, L.X.; Jia, H.T.; Jing, H. Cooperative activation of atrial natriuretic peptide promoter by dHAND and MEF2C. *J. Cell. Biochem.* **2004**, *93*, 1255–1266. [[CrossRef](#)]
212. Kamakura, S.; Oishi, K.; Yoshimatsu, T.; Nakafuku, M.; Masuyama, N.; Gotoh, Y. Hes binding to STAT3 mediates crosstalk between Notch and JAK-STAT signalling. *Nat. Cell Biol.* **2004**, *6*, 547–554. [[CrossRef](#)]
213. Cruickshank, M.N.; Dods, J.; Taylor, R.L.; Karimi, M.; Fenwick, E.J.; Quail, E.A.; Rea, A.J.; Holers, V.M.; Abraham, L.J.; Ulgiati, D. Analysis of tandem E-box motifs within human Complement receptor 2 (CR2/CD21) promoter reveals cell specific roles for RP58, E2A, USF and localized chromatin accessibility. *Int. J. Biochem. Cell Biol.* **2015**, *64*, 107–119. [[CrossRef](#)]
214. Zhou, X.; O’Shea, E.K. Integrated Approaches Reveal Determinants of Genome-wide Binding and Function of the Transcription Factor Pho4. *Mol. Cell* **2011**, *42*, 826–836. [[CrossRef](#)]
215. Kindrick, J.D.; Mole, D.R. Hypoxic regulation of gene transcription and chromatin: Cause and effect. *Int. J. Mol. Sci.* **2020**, *21*, 8320. [[CrossRef](#)]
216. Guccione, E.; Martinato, F.; Finocchiaro, G.; Luzi, L.; Tizzoni, L.; Dall’Olio, V.; Zardo, G.; Nervi, C.; Bernard, L.; Amati, B. Myc-binding-site recognition in the human genome is determined by chromatin context. *Nat. Cell Biol.* **2006**, *8*, 764–770. [[CrossRef](#)]
217. Wapinski, O.L.; Vierbuchen, T.; Qu, K.; Lee, Q.Y.; Chanda, S.; Fuentes, D.R.; Giresi, P.G.; Ng, Y.H.; Marro, S.; Neff, N.F.; et al. XHierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **2013**, *155*, 621. [[CrossRef](#)]
218. Park, N.I.; Guilhamon, P.; Desai, K.; McAdam, R.F.; Langille, E.; O’Connor, M.; Lan, X.; Whetstone, H.; Coutinho, F.J.; Vanner, R.J.; et al. ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell Stem Cell* **2017**, *21*, 209–224.e7. [[CrossRef](#)]
219. Raposo, A.A.S.F.; Vasconcelos, F.F.; Drechsel, D.; Marie, C.; Johnston, C.; Dolle, D.; Bithell, A.; Gillotin, S.; van den Berg, D.L.C.; Ettwiller, L.; et al. Ascl1 coordinately regulates gene expression and the chromatin landscape during neurogenesis. *Cell Rep.* **2015**, *10*, 1544–1556. [[CrossRef](#)] [[PubMed](#)]
220. Pataskar, A.; Jung, J.; Smialowski, P.; Noack, F.; Calegari, F.; Straub, T.; Tiwari, V.K. NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.* **2016**, *35*, 24–45. [[CrossRef](#)] [[PubMed](#)]
221. Guillemot, F.; Hassan, B.A. Beyond proneural: Emerging functions and regulations of proneural proteins. *Curr. Opin. Neurobiol.* **2017**, *42*, 93–101. [[CrossRef](#)]
222. Smith, D.K.; Yang, J.; Liu, M.L.; Zhang, C.L. Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming. *Stem Cell Reports* **2016**, *7*, 955–969. [[CrossRef](#)] [[PubMed](#)]
223. De la Serna, I.L.; Ohkawa, Y.; Berkes, C.A.; Bergstrom, D.A.; Dacwag, C.S.; Tapscott, S.J.; Imbalzano, A.N. MyoD Targets Chromatin Remodeling Complexes to the Myogenin Locus Prior to Forming a Stable DNA-Bound Complex. *Mol. Cell. Biol.* **2005**, *25*, 3997–4009. [[CrossRef](#)]
224. Knoepfler, P.S.; Bergstrom, D.A.; Uetsuki, T.; Dac-Korytko, L.; Sun, Y.H.; Wright, W.E.; Tapscott, S.J.; Kamps, M.P. A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with Pbx-Meis1/Prep1. *Nucleic Acids Res.* **1999**, *27*, 3752–3761. [[CrossRef](#)]
225. Maves, L.; Waskiewicz, A.J.; Paul, B.; Cao, Y.; Tyler, A.; Moens, C.B.; Tapscott, S.J. Pbx homeodomain proteins direct MyoD activity to promote fast-muscle differentiation. *Development* **2007**, *134*, 3371–3382. [[CrossRef](#)]
226. Ali, F.; Hindley, C.; McDowell, G.; Deibler, R.; Jones, A.; Kirschner, M.; Guillemot, F.; Philpott, A. Cell cycle-regulated multi-site phosphorylation of neurogenin 2 coordinates cell cycling with differentiation during neurogenesis. *Development* **2011**, *138*, 4267–4277. [[CrossRef](#)] [[PubMed](#)]
227. Ali, F.R.; Cheng, K.; Kirwan, P.; Metcalfe, S.; Livesey, F.J.; Barker, R.A.; Philpott, A. The phosphorylation status of Ascl1 is a key determinant of neuronal differentiation and maturation in vivo and in vitro. *Dev.* **2014**, *141*, 2216–2224. [[CrossRef](#)]
228. Liu, Y.; MacDonald, R.J.; Swift, G.H. DNA Binding and Transcriptional Activation by a PDX1·PBX1b·MEIS2b Trimer and Cooperation with a Pancreas-specific Basic Helix-Loop-Helix Complex. *J. Biol. Chem.* **2001**, *276*, 17985–17993. [[CrossRef](#)]
229. Meredith, D.M.; Borromeo, M.D.; Deering, T.G.; Casey, B.H.; Savage, T.K.; Mayer, P.R.; Hoang, C.; Tung, K.-C.; Kumar, M.; Shen, C.; et al. Program Specificity for Ptf1a in Pancreas versus Neural Tube Development Correlates with Distinct Collaborating Cofactors and Chromatin Accessibility. *Mol. Cell. Biol.* **2013**, *33*, 3166–3179. [[CrossRef](#)] [[PubMed](#)]
230. Ito, S.; Shen, L.; Dai, Q.; Wu, S.C.; Collins, L.B.; Swenberg, J.A.; He, C.; Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011**, *333*, 1300–1303. [[CrossRef](#)]
231. Perini, G.; Diolaiti, D.; Porro, A.; Della Valle, G. In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12117–12122. [[CrossRef](#)] [[PubMed](#)]
232. Prendergast, G.C.; Lawe, D.; Ziff, E.B. Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* **1991**, *65*, 395–407. [[CrossRef](#)]
233. Wang, D.; Hashimoto, H.; Zhang, X.; Barwick, B.G.; Lonial, S.; Boise, L.H.; Vertino, P.M.; Cheng, X. MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.* **2017**, *45*, 2396–2407. [[CrossRef](#)] [[PubMed](#)]
234. Yang, J.; Zhang, X.; Blumenthal, R.M.; Cheng, X. Detection of DNA Modifications by Sequence-Specific Transcription Factors. *J. Mol. Biol.* **2020**, *432*, 1661–1673. [[CrossRef](#)] [[PubMed](#)]

235. Yang, J.; Horton, J.R.; Li, J.; Huang, Y.; Zhang, X.; Blumenthal, R.M.; Cheng, X. Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine. *Nucleic Acids Res.* **2019**, *47*, 8375–8387. [[CrossRef](#)] [[PubMed](#)]
236. Golla, J.P.; Zhao, J.; Mann, I.K.; Sayeed, S.K.; Mandal, A.; Rose, R.B.; Vinson, C. Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. *Biochem. Biophys. Res. Commun.* **2014**, *449*, 248–255. [[CrossRef](#)] [[PubMed](#)]
237. Yang, L.; Zhou, T.; Dror, I.; Mathelier, A.; Wasserman, W.W.; Gordân, R.; Rohs, R. TFBSshape: A motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res.* **2014**, *42*, 148–155. [[CrossRef](#)] [[PubMed](#)]
238. Aydin, B.; Kakumanu, A.; Rossillo, M.; Moreno-estellés, M.; Garipler, G.; Ringstad, N.; Flames, N.; Mahony, S.; Mazzoni, E.O. Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes. *Nat. Neurosci.* **2019**, *22*, 897–908. [[CrossRef](#)] [[PubMed](#)]
239. Gordân, R.; Shen, N.; Dror, I.; Zhou, T.; Horton, J.; Rohs, R.; Bulyk, M.L. Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors through DNA Shape. *Cell Rep.* **2013**, *3*, 1093–1104. [[CrossRef](#)] [[PubMed](#)]
240. Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, *35*, 406–413. [[CrossRef](#)]
241. Huppert, J.L.; Bugaut, A.; Kumari, S.; Balasubramanian, S. G-quadruplexes: The beginning and end of UTRs. *Nucleic Acids Res.* **2008**, *36*, 6260–6268. [[CrossRef](#)]
242. Huppert, J.L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005**, *33*, 2908–2916. [[CrossRef](#)]
243. Etzioni, S.; Yafe, A.; Khateb, S.; Weisman-Shomer, P.; Bengal, E.; Fry, M. Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes. *J. Biol. Chem.* **2005**, *280*, 26805–26812. [[CrossRef](#)]
244. Shklover, J.; Etzioni, S.; Weisman-Shomer, P.; Yafe, A.; Bengal, E.; Fry, M. MyoD uses overlapping but distinct elements to bind E-box and tetraplex structures of regulatory sequences of muscle-specific genes. *Nucleic Acids Res.* **2007**, *35*, 7087–7095. [[CrossRef](#)]
245. Walsh, K.; Gualberto, A. MyoD binds to the guanine tetrad nucleic acid structure. *J. Biol. Chem.* **1992**, *267*, 13714–13718. [[CrossRef](#)]
246. Yafe, A.; Etzioni, S.; Weisman-Shomer, P.; Fry, M. Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Res.* **2005**, *33*, 2887–2900. [[CrossRef](#)] [[PubMed](#)]
247. Yafe, A.; Shklover, J.; Weisman-Shomer, P.; Bengal, E.; Fry, M. Differential binding of quadruplex structures of muscle-specific genes regulatory sequences by MyoD, MRF4 and myogenin. *Nucleic Acids Res.* **2008**, *36*, 3916–3925. [[CrossRef](#)] [[PubMed](#)]
248. Shklover, J.; Weisman-Shomer, P.; Yafe, A.; Michael, F. Quadruplex structures of muscle gene promoter sequences enhance in vivo MyoD-dependent gene expression. *Nucleic Acids Res.* **2010**, *38*, 2369–2377. [[CrossRef](#)] [[PubMed](#)]
249. Bormuth, I.; Yan, K.; Yonemasu, T.; Gummert, M.; Zhang, M.; Wichert, S.; Grishina, O.; Pieper, A.; Zhang, W.; Goebbels, S.; et al. Neuronal basic helix-loop-helix proteins neurod2/6 regulate cortical commissure formation before midline interactions. *J. Neurosci.* **2013**, *33*, 641–651. [[CrossRef](#)] [[PubMed](#)]
250. Conerly, M.L.; Yao, Z.; Zhong, J.W.; Groudine, M.; Tapscott, S.J. Distinct Activities of Myf5 and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and Differentiation. *Dev. Cell* **2016**, *36*, 375–385. [[CrossRef](#)] [[PubMed](#)]
251. Li, Y.; Song, X.; Ma, Y.; Liu, J.; Yang, D.; Yan, B. DNA binding, but not interaction with Bmal1, is responsible for DEC1-mediated transcription regulation of the circadian gene mPer1. *Biochem. J.* **2004**, *382*, 895–904. [[CrossRef](#)]
252. Jiang, J.; Levine, M. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **1993**, *72*, 741–752. [[CrossRef](#)]
253. Mall, M.; Karetta, M.S.; Chanda, S.; Ahlenius, H.; Perotti, N.; Zhou, B.; Grieder, S.D.; Ge, X.; Drake, S.; Euong Ang, C.; et al. Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature* **2017**, *544*, 245–249. [[CrossRef](#)] [[PubMed](#)]
254. Postigo, A.A.; Dean, D.C. ZEB, a vertebrate homolog of Drosophila Zfh-1, is a negative regulator of muscle differentiation. *EMBO J.* **1997**, *16*, 3935–3943. [[CrossRef](#)] [[PubMed](#)]
255. Fernandez, P.C.; Frank, S.R.; Wang, L.; Schroeder, M.; Liu, S.; Greene, J.; Cocito, A.; Amati, B. Genomic targets of the human c-Myc protein. *Genes Dev.* **2003**, *17*, 1115–1129. [[CrossRef](#)] [[PubMed](#)]
256. Sessa, A.; Ciabatti, E.; Drechsel, D.; Massimino, L.; Colasante, G.; Giannelli, S.; Satoh, T.; Akira, S.; Guillemot, F.; Broccoli, V. The Tbr2 Molecular Network Controls Cortical Neuronal Differentiation Through Complementary Genetic and Epigenetic Pathways. *Cereb. Cortex* **2017**, *27*, 3378–3396. [[CrossRef](#)]
257. Huang, J.; Blackwell, T.K.; Kedes, L.; Weintraub, H. Differences between MyoD DNA binding and activation site requirements revealed by functional random sequence selection. *Mol. Cell. Biol.* **1996**, *16*, 3893–3900. [[CrossRef](#)]
258. Costa, A.; Powell, L.M.; Soufi, A.; Lowell, S.; Jarman, A.P. Atoh1 is repurposed from neuronal to hair cell determinant by Gfi1 acting as a coactivator without redistributing Atoh1's genomic binding sites. *bioRxiv* **2019**, 1–25. [[CrossRef](#)]
259. Domcke, S.; Hill, A.J.; Daza, R.M.; Cao, J.; Day, D.R.O.; Pliner, H.A.; Aldinger, K.A.; Pokholok, D.; Zhang, F.; Milbank, J.H.; et al. A human cell atlas of fetal chromatin accessibility. *Science* **2020**, *809*. [[CrossRef](#)] [[PubMed](#)]
260. Cao, Y.; Yao, Z.; Sarkar, D.; Lawrence, M.; Sanchez, G.J.; Parker, M.H.; MacQuarrie, K.L.; Davison, J.; Morgan, M.T.; Ruzzo, W.L.; et al. Genome-wide MyoD Binding in Skeletal Muscle Cells: A Potential for Broad Cellular Reprogramming. *Dev. Cell* **2010**, *18*, 662–674. [[CrossRef](#)] [[PubMed](#)]
261. Borromeo, M.D.; Meredith, D.M.; Castro, D.S.; Chang, J.C.; Tung, K.C.; Guillemot, F.; Johnson, J.E. A transcription factor network specifying inhibitory versus excitatory neurons in the dorsal spinal cord. *Dev.* **2014**, *141*, 2803–2812. [[CrossRef](#)]

- 
262. Hahn, M.A.; Jin, S.G.; Li, A.X.; Liu, J.; Huang, Z.; Wu, X.; Kim, B.W.; Johnson, J.; Bilbao, A.D.V.; Tao, S.; et al. Reprogramming of DNA methylation at NEUROD2-bound sequences during cortical neuron differentiation. *Sci. Adv.* **2019**, *5*, 1–14. [[CrossRef](#)]
  263. Pali, C.G.; Perez-Iratxeta, C.; Yao, Z.; Cao, Y.; Dai, F.; Davison, J.; Atkins, H.; Allan, D.; Dilworth, F.J.; Gentleman, R.; et al. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *EMBO J.* **2011**, *30*, 494–509. [[CrossRef](#)] [[PubMed](#)]
  264. Yevshin, I.; Sharipov, R.; Valeev, T.; Kel, A.; Kolpakov, F. GTRD: A database of transcription factor binding sites identified by ChIP-seq experiments. *Nucleic Acids Res.* **2017**, *45*, D61–D67. [[CrossRef](#)] [[PubMed](#)]