The FASEB Journal

FASEB
Federation of American Societies for Experimental Biology

RESEARCH ARTICLE OPEN ACCESS

The Direct and Individual Transcriptional Function of the Human Homodimeric and Heterodimeric Basic Helix–Loop–Helix Transcription Factors E47 and Scleraxis

Karla Miranda López-Delucio[1,2] | Tania Dessire Sandoval-Gurubel[1] | Carina Becerril[1] | Carlos Campero-Basaldua[3] | Alicia González[3] | Ana Lilia Torres-Machorro[1]

[1]Laboratorio de Biología Celular, Departamento de Investigación en Fibrosis Pulmonar, Instituto Nacional de Enfermedades Respiratorias "Ismael Cosío Villegas", Mexico City, Mexico | [2]Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Mexico City, Mexico | [3]Departamento de Bioquímica y Biología Estructural, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City, Mexico

Correspondence: Ana Lilia Torres-Machorro (ana.torres@iner.gob.mx)

Received: 20 December 2024 | **Revised:** 13 May 2025 | **Accepted:** 2 June 2025

Funding: This research was funded by Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) through the "Fondo Sectorial de Investigación para la Educación: Ciencia Básica" (Grant CB2017-2018 A1-S-8737) to A.L.T.-M.

Keywords: *COL1A2* | core promoter | E47 | *ENO1* | *Saccharomyces cerevisiae* | Scleraxis (SCX) | TCF3 | *TGFB1*

ABSTRACT

The basic helix–loop–helix (bHLH) transcription factors (TFs) are essential in development and disease. Their function is regulated at multiple levels, including the structuring of homo- or heterodimeric forms among members of the family. Because most bHLH TFs have numerous dimerization partners, the commonly used overexpression or deletion experimental approaches in humans often generate results influenced by the altered regulatory balance of the TF network. To study the direct transcriptional role of two human bHLH TFs, we expressed them in an isolated system (yeast) with no additional tissue-specific bHLH TFs. The transcriptional effect was measured utilizing a GFP reporter controlled by human regulatory sequences containing different amounts of the bHLH TF consensus binding sites, the E-boxes. The individual transcriptional contributions of heterodimeric SCX-E47 or homodimeric E47 were compared over two human regulatory regions implicated in fibrosis: *COL1A2* and *TGFB1*. Briefly, the heterodimeric SCX-E47 was the best activating form. The *COL1A2* regulatory region showed the most significant transcriptional changes despite having fewer E-boxes (five) than the *TGFB1* region (13). Finally, the context of the nearby TF binding sites and the core promoter was also relevant for the final individual transcriptional effect of the bHLH TFs tested.

1 | Introduction

The bHLH TFs constitute the second-largest TF family [1]. They have essential roles in developmental processes and cell differentiation [2, 3]. These TFs are highly expressed early in development, with a later reduction in expression and restriction to specific tissues or growth conditions. Part of the complexity of bHLH TFs regulation is set through covalent post-translational modification and the establishment of many possible homodimeric and heterodimeric interactions among members (and non-members) of the bHLH TF family [3, 4]. These TFs have been studied in detail in mammalian systems through artificial overexpression or deletion of the bHLH of interest [5–8]. Nevertheless, the results obtained are not necessarily a direct transcriptional effect of the bHLH TF on the DNA, as the TF manipulation can affect the transcription of

more bHLH TFs [9, 10] and generate, block, or alter protein-protein interactions with other TFs that end up changing the delicate regulatory balance of the cellular bHLH TF network [11–13].

As an example, the highly studied bHLH TF MYOD can form homodimers, homotetramers, and heterodimers with the bHLH TFs E12, E47, TWIST1, BHLHE41, ASCL3, ID1, ID2, ID3, HEY1, and HES1 [3]. Although not all possible interactors are always expressed simultaneously, MYOD overexpression or deletion affects transcription directly and indirectly due to distorted protein–protein interactions. It is also worth mentioning that most bHLH TF studies do not assess the dimeric form involved but study the TFs as isolated entities [3]. Thus, the results are usually interpreted without considering the interactions established with endogenous TFs.

The dysregulation of bHLH TFs has relevant roles in multiple pathologies, including cancer, schizophrenia, and fibrosing diseases [5, 14–16]. The E47 protein, encoded by the *TCF3* gene, is ubiquitously expressed and involved in the B-lymphocyte and thymocyte development [17–19]. It can function as a homodimer or as a heterodimer. Furthermore, E47 interacts with most type II tissue-specific bHLH TFs, including Scleraxis (SCX), MYOD, TCF21, NEUROD1, HAND1, and ATOH1 [3]. SCX dysregulation is associated with fibrosing diseases due to its involvement in chondrogenesis and mesoderm development [20, 21]. Except for one in vitro work [22], most studies suggest that SCX functionality depends on its interaction with E47 [5, 20, 23].

Contrastingly, a considerable amount of the basic knowledge about transcription is derived from studies in yeast, which is an excellent genetic model [24, 25]. This unicellular eukaryote seems far from humans in the evolutionary tree. However, mutant complementation with human genes [24] demonstrates that yeast conserves most eukaryotic basal transcriptional and epigenetic machinery.

Therefore, to assess the direct transcriptional role of human bHLH TFs in an isolated environment, free from other possible interfering tissue-specific human bHLH TFs, we induced the expression of two naturally interacting human bHLH TFs in *Saccharomyces cerevisiae*: SCX and E47. Then, we utilized a GFP reporter gene to analyze the transcriptional effect of E47 homodimers or SCX-E47 heterodimers over a human regulatory region containing E-boxes, the bHLH TF binding sites [26]. The regulatory sequences were further manipulated to have contrasting basal transcription levels and E-box content, resulting in differential effects mediated by the TFs tested.

2 | Materials and Methods

2.1 | Construction of Reporter Genes

The *COL1A2* regulatory sequence is a natural target of SCX-E47 [27] and was a synthetic gene (Twist Bioscience) corresponding to the endogenous regulatory sequence of the human *COL1A2* gene (positions −990 to +10). The sequence was slightly modified to include restriction sites for subcloning on either side of the sequence: *Kpn*I on the 5′ side and *Ppu*MI and *Sph*I on the 3′ side. The *COL1A2* promoter (positions −114 to +10) was surrounded by the *Nhe*I and *Ppu*MI/*Sph*I restriction sites to allow the swapping of core promoter sequences. The original *Apa*I, *Bgl*II, and *Ppu*MI restriction sites within the sequence were removed with single nucleotide changes for cloning. The *TGFB1* regulatory sequence containing 13 E-boxes was also a synthetic gene corresponding to the endogenous regulatory region from positions −1077 to −137 relative to the transcription start point of the *TGFB1_1* promoter. The *Kpn*I and *Nhe*I restriction sites surrounded the sequence for subcloning purposes.

The core promoter sequences were also synthetic. The *ENO1* core promoter was identical to the yeast endogenous regulatory sequence from position −118 to +10, surrounded by the *Nhe*I and *Ppu*MI restriction sites. The original negative regulatory sequence of *ENO1* [28], located at the endogenous positions −126 to −166, was not part of the core promoter subcloned next to the E-box containing upstream regulatory sequences. The *ADO1* core promoter was the *S. cerevisiae* endogenous regulatory sequence from position −130 to +16, surrounded by the *Nhe*I and *Ppu*MI restriction sites. The complete *ENO1* regulatory sequence, positions −1460 to +10, was also used as a control, including the 5′ *Kpn*I and 3′ *Ppu*MI and *Sph*I restriction sites (pAT197). The original *Sph*I site at position 791 was mutated to remove the site (G to C). The short *ENO1* reporter was constructed by cutting pAT197 with *EcoRI* to extract the *ENO1* regulatory sequence from position −685 to +10. The *EcoRI* fragment also contained the complete *GFP* coding sequence and downstream *ADH1* regulatory sequences. This insert was subsequently subcloned into pRS412 to obtain the final pAT224 plasmid. Detailed maps of the *COL1A2*, *TGFB1*, and *ENO1* regulatory sequences indicating the exact localization of E-boxes, additional transcription factor finding sites, and other features are in Figure S1.

The *GFP-S65T* reporter gene was subcloned from the original pRH3003 plasmid (donated by Randy Hampton, UCSD) by PCR. This GFP variant is brighter and has improved stability and faster maturation times than the wild-type [29, 30]. The primes were KpnISphIGFP-F (CCAAAAGGTACCAAAAGCATG CCTATGGGTAAAGGAGAAGAACT) and EcoRI-ADH1-R (CCAACTGAAT TCCCGGTAGAGGTGTGGTCA) and contained the *Kpn*I and *Eco*RI restriction sites for subcloning into pRS412 (Cen plasmid, *ADE2*; kindly donated by Lorraine Pillus, UCSD) to generate plasmid pAT195 that was used as a no-signal control due to the absence of a promoter sequence. The *COL1A2* regulatory sequence was subcloned into pAT195 and localized upstream of the GFP coding sequence using KpnI and SphI restriction sites (pAT196). The *COL1A2* promoter was swapped for the *ENO1* (pAT199) or *ADO1* (pAT198) core promoters using the *Nhe*I and *Ppu*MI restriction sites. The *COL1A2* regulatory sequence was replaced with the *TGFB1* E-box regulatory sequence in the pAT199 and pAT196 plasmids using the *Kpn*I and *Nhe*I restriction sites to construct pAT201 and pAT202, respectively. All constructions were confirmed by restriction digestion or sequencing.

For the GFP protein to be used as a model of promoter activity, three assumptions were met in this system: the growth rate was

constant (Figure S2), the culture was in the exponential growth phase, and the protein levels were at a steady state [31].

2.2 | Human *SCX* and *E47* (*TCF3*) Gene Sequence Optimization for *S. cerevisiae*

The sequences of the human *E47* and *SCX* transcription factors were optimized for their expression in yeast using the GenScript OptimumGene codon optimization tool. The sequences were further modified to include restriction sites for subcloning in the pESC-*URA*3 2μ vector (Agilent). The optimized human *SCX* gene included the *Xho*I (and *Kpn*I) restriction sites upstream of the start codon, whereas the *Xba*I, *Pme*I, and *Nhe*I restriction sites were included downstream of the stop codon. The optimized *E47* human gene included the *Bgl*II restriction sites upstream of the start codon. The *Pac*I site was added to the 3′ extreme of the sequence. Restriction sequences identical to those added on the extremes were removed from the coding region to avoid unwanted cuts during cloning. Optimization changed the codon adaptation index (CAI) of the *SCX* gene from 0.4 to 0.95. In contrast, the average GC content decreased from 75% to 42% after optimization. For the *E47* gene, the optimization resulted in a change from a CAI of 0.48 to 0.94, while the average GC content decreased from 67% to 41%. Twist Bioscience then synthesized the optimized genes.

2.3 | Subcloning of Optimized Human Genes Into Yeast Inducible Plasmid

The synthetic genes were subcloned into the yeast-inducible plasmid pESC-*URA3* (Agilent). SCX and E47 were also subcloned for co-expression on either side of the bidirectional promoter *pGAL1, GAL10* of pESC-*URA3*. The restriction enzymes used for subcloning SCX were *Xho*I and *Nhe*I (pAT75). E47 was subcloned into pESC-*URA3* with *Bgl*II and *Pac*I (pAT100). Finally, E47 was also subcloned into pAT75 using the same restriction enzymes: pAT119.

2.4 | Yeast Strain

An *S. cerevisiae* strain null for the lysosomal peptidase *PEP4* was constructed in the W303 genetic background. A haploid *pep4Δ::kanMX* strain was derived from the BY diploid heterozygous KO collection (kindly donated by L. Pillus) by tetrad dissection. Its gDNA was used as a template for PCR amplification with the primers PEP4-KO-F (CTTGAGCTCCTCAATTGTAT) and PEP4-KO-R (GAATTAGAGGTGCTCCGTA). Using the homologous recombination standard method, the PCR product obtained was used for the genomic replacement of the endogenous gene in the wild-type W303-1a strain (LPY5, donated by L. Pillus). The resulting G418-resistant strain was verified by molecular genotyping with the PEP4-KO-F and KANMX-R primers (CAGCCATTACGCTCGTCATC) and backcrossed to the wild-type strain. Tetrad dissection generated the strain ATY38 (MAT@ *ade2-1 can1-100 GAL his3-11 leu2,3112 trp1-1 ura3-1 pep4Δ::KanMX*). The *PEP4* haploid null strains exhibit a chronological lifespan defect [32] and cannot grow after 1 or 2 years in the freezer. Therefore, during this work, the *PEP4* null strain had to be dissected fresh from the diploid (ATY32), generating identical strains to ATY38: ATY187, ATY188, and ATY231.

To avoid high autofluorescence in red W303 cells, they were used at low densities (<1 OD600) and transformed with an *ADE2* plasmid (reporter plasmids) that produced white cells (pAT195, 196, 197, 199, 201, 202, or 224) [33].

2.5 | Transformation and Expression of Human bHLH TFs in Yeast

The reporter and inducible plasmids were co-transformed into yeast *pep4Δ* strains using the quick transformation method previously reported [34]. The inducible plasmids coding for the bHLH TFs were selected in dropout media lacking uracil, whereas the reporter genes were selected in dropout media lacking adenine. The double selection was achieved with dropout media lacking both uracil and adenine.

Co-transformed single colonies were used for starter 3 mL cultures in YM -ura- ade- raffinose media and grown for 3 days at 30°C to achieve saturation and synchronization. Then, 45–60 mL cultures were seeded at 0.05 A600 to have them grow to log phase overnight. Cultures between 0.5 and 0.9 A600 were used for induction. A time zero (T0), sample was removed as a no-induction control before adding galactose at the final 2% concentration to induce the expression of the TFs of interest.

The viability of the strains transformed with each plasmid pair, before and after bHLH TF induction by galactose, was confirmed by dilution growth assays [35] (Figure S3).

2.6 | Immunoblots and Co-Immunoprecipitation

To verify the induction of expression of the human TFs, proteins were extracted from 10 mL fractions of the culture, as reported previously [35]. Briefly, washed cell pellets were lysed using glass beads beating in PBS supplemented with the protease inhibitors Calbiochem V cocktail and 1 mM PMSF. Samples containing 1× loading buffer were boiled for 10 min and loaded into 12% SDS-PAGEs. Proteins separated in gels were transferred to 0.2 μm PDVF membranes. After blocking with 5% milk, the anti-SCX (PA523943, Thermo Scientific) and anti-TCF3 (PA520900, Thermo Scientific) antibodies were used to verify the induction of the gene of interest at 1:1500 dilutions. The anti-beta-actin C4 antibody (SC-47778, Santa Cruz) was used as a loading control with a 1:1000 dilution for all immunoblots. The anti-GFP antibody (IGFP63, Biolegend) was used at a 1:1000 dilution. For the co-immunoprecipitations, 50 mL cultures were grown to A600 = 0.8 and were induced for SCX-E47 expression for 2 h. After centrifugation and washing, pellets were lysed by bead beating in 100 μL of cold IP lysis buffer (50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 0.25% nonidet-P40, 10% glycerol, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.05% β-mercaptoethanol, 1× Calbiochem V protease inhibitors cocktail, and 1 mM PMSF). The cleared lysate was incubated overnight with a 1:60 dilution of anti-Myc (9E10.2, Biolegend). The IP mixture was incubated at 4°C with protein A-sepharose beads for 1 h. Afterward, beads were washed with IP lysis buffer

and wash buffer (50 mM HEPES-KOH pH 7.5, 250 mM NaCl, and 1 mM EDTA) to finish eluting with 25 μL of 2.5× sample loading buffer.

2.7 | Cell Preparation for Flow Cytometry

About 10^7 cells were pelleted, washed with PBS, and fixed with 8% paraformaldehyde (PFA) for 10 min, followed by an overnight incubation at 4°C in 4% PFA. Fixed cells were washed with FACS solution (0.2 M Tris–HCL pH 7.5, 20 mM EDTA) and stored in PBS until use. The cells were diluted to 2×10^6 cells/ml and sonicated for 10 s in water using a VWR ultrasonic cleaner just before reading.

2.8 | Cytometer Data Acquisition, Analyses, and Statistics

Thirty thousand events were recorded for each sample using the BD flow cytometer FACS ARIA II (P07800079) at Laboratorio Nacional Conahcyt de Investigación y Diagnóstico por Inmunocitofluorometría (LANCIDI). The voltage parameters were set to 460 for SSC and 150 for FSC. The GFP fluorescence parameter was automatically set during equipment calibration. GFP was read on the logarithmic scale. All reads were captured using the DIVA software version 8.0.2. All data were stored as .fcs files and analyzed with the FlowJo 10.8.1 software. Data were graphed as dot plots with FSC-H and SSC-H axes to select for singles. Then, singles were graphed for GFP on the *X*-axis relative to SSC-A on the *Y*-axis. The GFP data were also graphed relative to FSC-A, obtaining similar results (not shown). The data derived from the non-signal control (Q1 quadrant) was removed from the quantification by setting the Q3 quadrant, which contained all GFP-positive cells. Finally, the GFP-positive varying population (vp) was measured in a quadrant at 50k SSC-A (Figure S4). The median fluorescence intensity (MED) and percentage of GFP-positive cells were measured for each quadrant for at least three independent experiments and compared relative to time zero. The normality of the data was confirmed with the Shapiro–Wilk test, followed by an ANOVA group comparison test. The differences between specific samples were assessed using the Student's *t*-test. Data were normalized to time zero for comparisons among different experiments, and statistics were evaluated using unpaired *t*-tests.

2.9 | Nucleosome Scanning Assay (NuSA)

Performed as previously described [36, 37]. Twenty-five milliliters of yeast cultures co-transformed with pAT119 (inducible SCX-E47) and either pAT196 (*COL1A2*) or pAT202 (*TGFB1-COL1A2*) reporter genes were grown in drop-out media with 2% raffinose until logarithmic phase (T0). The nucleosome positioning was analyzed at time zero and after 90 min of SCX-E47 induction (galactose) over both regulatory regions. Quantitative PCR analysis was made using a LightCycler 480 Instrument II (Roche Molecular Systems) with 10 μL reactions containing five μL of the 2× master mix Maxima SYBR Green/fluorescein (Thermo Scientific), 0.5 μL of each primer [10 μM], and 1 ng of undigested or MNase digested DNA (25 min). Triplicate reactions were run for 40 cycles at 95°C for 15 s, 58°C for 15 s, and 60°C for 45 s. Relative protection was calculated as a ratio to the control positioned nucleosome of *VCX1* [37]. The graphics were made with GraphPad Prism 6 (San Diego, CA, USA). Primer sequences are in Table S1.

3 | Results

E47 has over 40 possible dimerization partners, whereas SCX has at least five [3]. E47 and SCX function as heterodimers [20, 38]; E47 can also function as a homodimer [39], whereas the role of SCX as a homodimer remains unknown. Due to the multiple possible dimerization partners for both factors, we analyzed their direct and precise role over a known human target (*COL1A2*) [27]. For this, we used *Saccharomyces cerevisiae* as an artificial eukaryotic in vivo system, where the interactions with other tissue-specific bHLH TFs were prevented because these bHLH TFs are not expressed in yeast [40]. The transcriptional effects were observed utilizing a plasmid GFP reporter, where the chromatin structure was assumed to be constant [41, 42].

3.1 | System Construction

To build a system to analyze the regulatory role of human bHLH TFs over target genes, we used two types of plasmids for expression in yeast: one coding for the human TFs and the second containing the reporter gene. The first plasmid type was multicopy, coding for the optimized human bHLH TFs SCX and E47 under the inducible, bidirectional promoter *pGAL1, GAL10* (pESC, Agilent). The second plasmid type was a CEN plasmid containing a GFP reporter gene under the control of the human regulatory region for the *COL1A2* gene. This gene was selected because it is a known SCX target [27] and contains 5 E-boxes, the bHLH TF binding sites. Furthermore, the sequences of two E-boxes had reported binding sites for SCX [20, 27]. The *COL1A2* core promoter contained motifs like the TATA box and the Inr. Thinking that the reporter gene's basal expression could influence the bHLH TFs' regulatory effect, it was modulated by changing the core promoter sequence.

The original human *COL1A2* core promoter was replaced with one of two yeast TATA-box-containing core promoters: the strong promoter *ENO1* [28, 43] or the core promoter of a poorly expressed gene, *ADO1* [43] (Figure 1). Thus, some resulting reporters contained human-yeast hybrid sequences combining the upstream human *COL1A2* regulatory region with the yeast core promoters (Figure 1). The GFP reporter basal expression levels were compared by immunoblot and flow cytometry using a plasmid coding for GFP with no regulatory region as a no-signal control. The immunoblot showed that the most efficient regulatory region was *COL1A2*, followed by *COL1A2-ENO1*, whereas *COL1A2-ADO1* showed no GFP expression in the western blot (Figure 1B). The flow cytometry data were comparable to the immunoblots (Figure 1C–F) and indicated that the median fluorescence intensity for *COL1A2* was the highest, while the *COL2A2-ENO1* and *COL1A2-ADO1* intensities were similar (Figure 1G). Nevertheless, differences were more evident when analyzing the percentage of GFP-expressing cells, where almost 90% of the *COL1A2* reporter cells were positive for GFP, followed by 48% for the *COL1A2-ENO1*

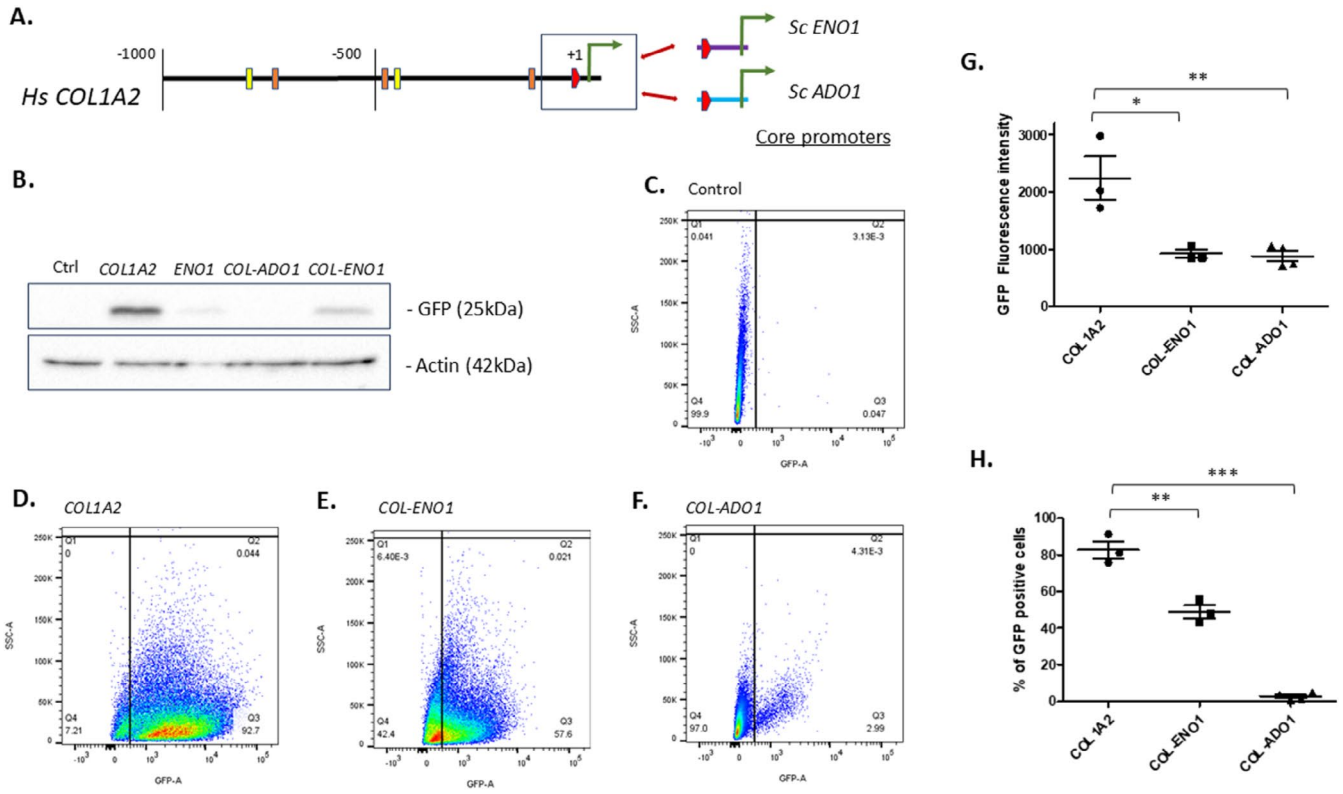

FIGURE 1 | System construction and basal reporter expression levels. (A) Schematic representation of the 1 kb human *COL1A2* regulatory region upstream of the reporter gene GFP. The sequence contained 5 E-box sequences (orange), with two SCX binding sites (yellow). The construction also included the original human core promoter sequence with consensus TATA and Inr sequences. To modulate the *COL1A2* transcription levels, two constructions replaced the *COL1A2* core promoter sequence with the *S. cerevisiae ENO1* or *ADO1* core promoter sequences. (B) Basal reporter GFP expression levels for the different constructs by immunoblot. The control construction (Ctrl) included the GFP coding sequence with no upstream regulatory regions. The original *ENO1* regulatory sequence (1.5 kb upstream of the transcription start point) was used as a control for an original yeast regulatory sequence (*ENO1*). (C–F) Examples of flow cytometry dot plots produced by different GFP-reporter constructions include the control (C), *COL1A2* (D), *COL1A2-ENO1* (E), and *COL1A2-ADO1* (F). The control dot plot was used to define the gate that excluded the cells that did not express GFP. The same gate was applied to plots D–F. The pseudo colors indicate cell densities. (G) Comparison of reporter's median fluorescence intensities measured by flow cytometry. (H) Comparison of the percentage of GFP-positive cells measured by flow cytometry for tested reported constructs. Statistical differences in the normal data were assessed initially with one-way ANOVA, followed by unpaired *t*-tests. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

reporter. In contrast, the *COL1A2-ADO1* reporter exhibited only 2% GFP-positive cells (Figure 1H).

The experimental strategy is summarized in Figure 2A. The human *SCX* and *E47* sequences were optimized for expression in *S. cerevisiae* and expressed from a *pESC-URA3* plasmid, where a bidirectional galactose-inducible promoter controlled the expression of either or both genes. Therefore, we tested the regulatory role of individual SCX, homodimeric E47, and heterodimeric SCX-E47. Co-immunoprecipitation experiments confirmed the heterodimeric interactions between SCX and E47 (Figure S5). The basal GFP expression level of the reporter genes was compared to the GFP expression after 30 and 90 min of the induction of the TF expression (Figure 2A). The 30-min induction time point was considered an early response in the presence of low amounts of the TF of interest. The 90-min time point was considered a later response in the presence of higher amounts of the TF of interest (Figure 2). The yeast cells containing reporter and inducible plasmids were grown to logarithmic phase with the sugar raffinose as a carbon source due to its neutral sugar properties on the *GAL1* and *GAL10* promoters. Therefore, adding galactose to the growth media ensured a rapid and coordinated induction of the TF expression (Figure 2B–D). We verified the induction of the expression for SCX, E47, and SCX-E47 for all independent experiments (Figure 2B–D) before the GFP expression analyses by flow cytometry were made. Representative dot plots before and after 30 and 90 min of SCX-E47 inductions are shown in Figure 2E.

3.2 | The SCX-E47 Heterodimer Had Positive Transcriptional Effects Over COL1A2 Regulatory Regions

Flow cytometry evaluated the GFP reporter expression after the induction of heterodimeric SCX-E47 TFs. The no-induction crude parameters of median GFP intensity and percentage of GFP-expressing cells (T0) were compared with the two induction time points for all three reporter genes described above. As shown in Figure 3A, the median GFP intensity of the three reporter genes tested did not change upon SCX-E47 induction. Nevertheless, statistically significant differences were observed when the percentage of positive GFP-expressing cells was

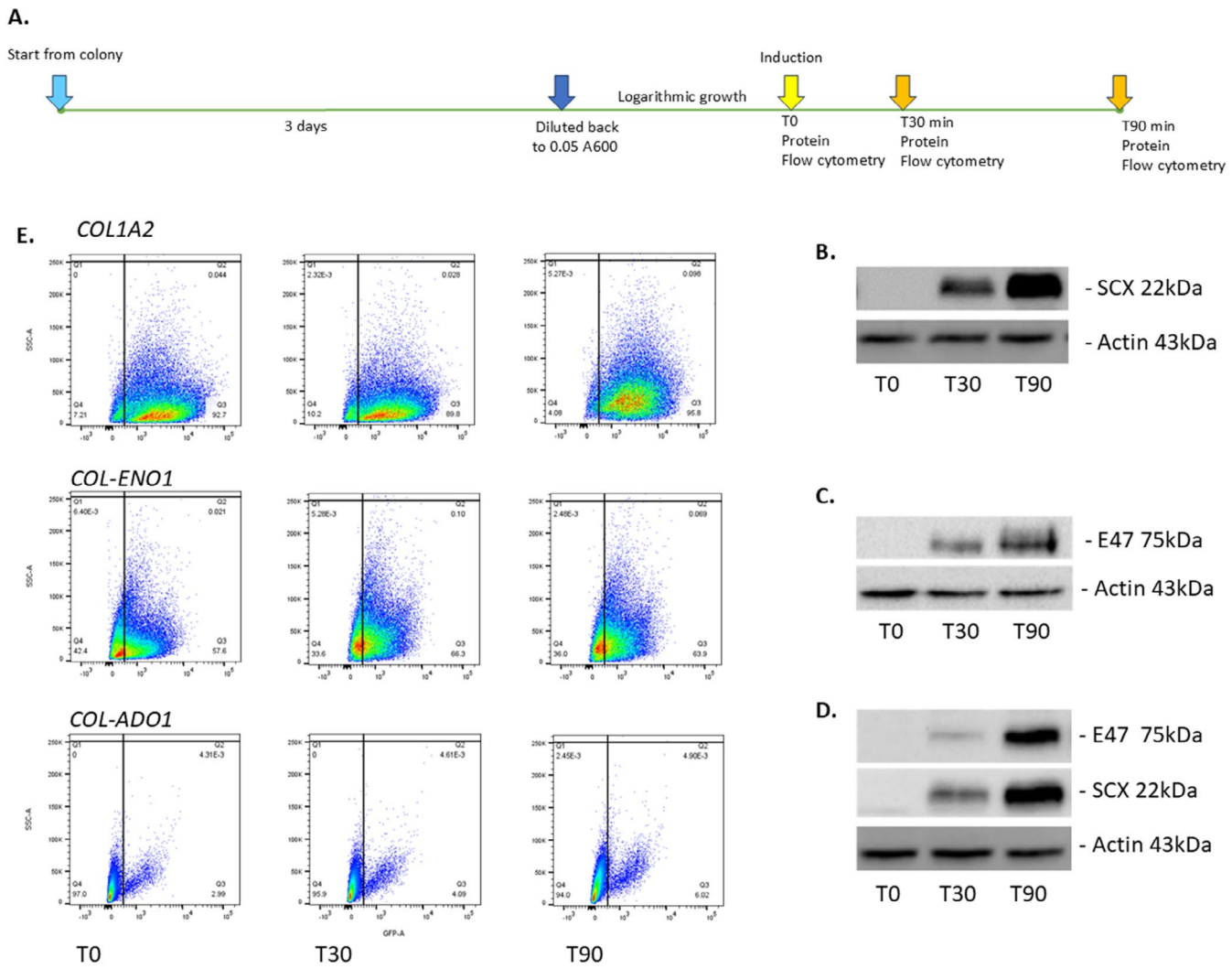

FIGURE 2 | Induction of human bHLH TF expression in yeast and flow cytometry dot plots after SCX-E47 inductions. (A) Diagram of experimental strategy. (B–D) Representative Immunoblots verifying induction of bHLH TFs expression after 30 and 90 min of growth in galactose. (B) SCX induction. (C) E47 induction. (D) SCX-E47 synchronic induction. (E) Representative flow-cytometry GFP-SSC-A dot plots after SCX-E47 inductions for the *COL1A2*, *COL1A2-ENO1*, and *COL1A2-ADO1* reporter genes. The gates and statistics are as in Figure 1.

evaluated. The *COL1A2-ADO1* regulatory region was unresponsive to SCX-E47 induction, perhaps due to low to null basal expression levels (Figure 3B). However, despite the high percentage of GFP-positive cells from the *COL1A2* original human regulatory region, a significant 5.5% increase in GFP-positive cells was observed at T90 relative to T30 (Figure 3B). In the case of the hybrid regulatory region *COL1A2-ENO1*, the increase was broader and observed earlier (T30), with an almost 9% increase, representing an 18% total difference relative to the T0 value (Figure 3B). This increase was also observed at T90, remaining a statistically significant effect.

3.3 | The Solo Expression of E47 or SCX Had Little Effect on the COL1A2 Reporter Genes

The effect of expressing E47 or SCX in isolation was only assessed for the *COL1A2* and *COL1A2-ENO1* reporters. The induction of unaccompanied E47 to analyze its role as a homodimer showed no change in the median fluorescence intensity and the percentage of GFP-positive cells driven by the *COL1A2* regulatory region (Figure 4A). Similarly, no changes were seen for either fluorescence parameter for the *COL1A2-ENO1* regulation by the E47 homodimer (Figure 4A).

In agreement with previous reports, finding no regulatory role for SCX as a homodimer [5], the induction of the TF SCX alone did not affect the median GFP intensity or percentage of GFP-positive cells for both *COL1A2* and *COL1A2-ENO1* reporters (Figure 4B).

3.4 | A Changing Population Analysis Uncovers Key SCX-E47 Regulatory Effects

Although we assumed that the yeast strain expressed a "constant" population of TFs that could affect the reporter genes tested, we still expected small transcriptional changes when a single pair of TFs was induced. This assumption was based on the fact that the most potent TF regulatory effects typically result from synergistic interactions with other proteins, some of which may not be expressed in the yeast context.

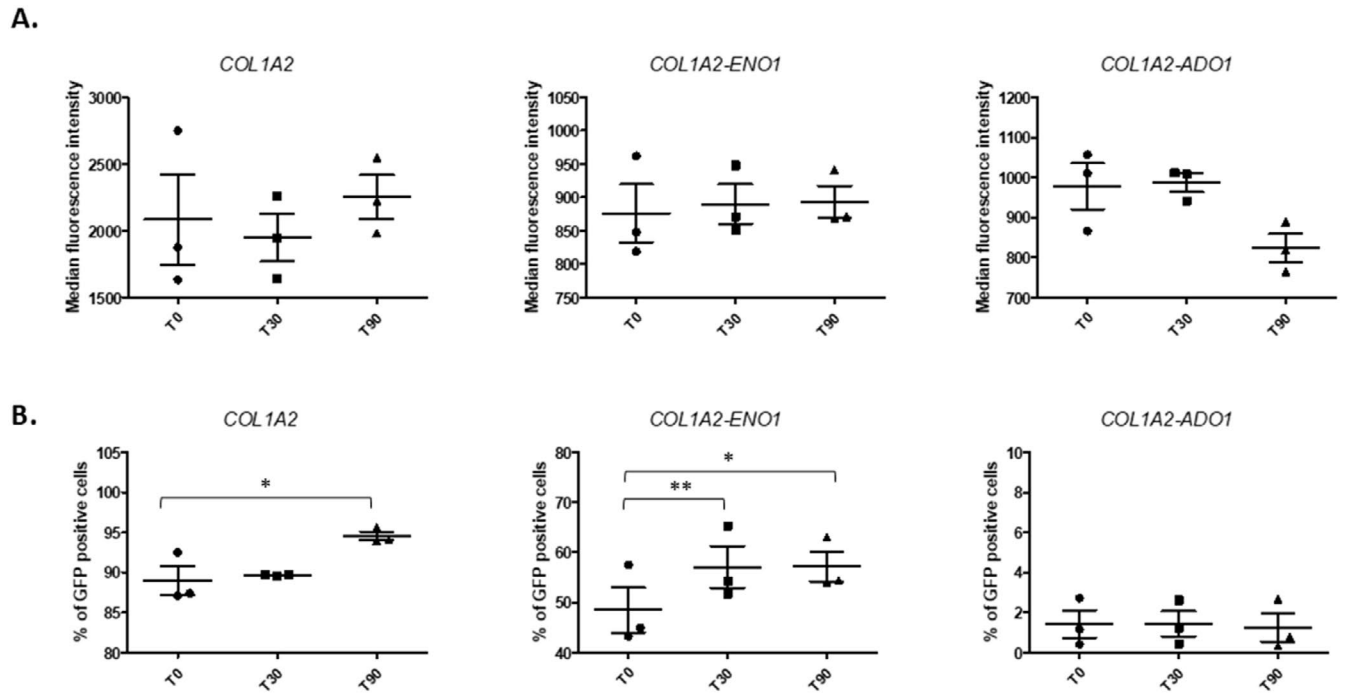

FIGURE 3 | The SCX-E47 heterodimer exerts a positive transcriptional effect on transcriptionally active *COL1A2* regulatory sequences. (A) The Median Fluorescence Intensity levels from the *COL1A2*, *COL1A2-ENO1*, and *COL1A2-ADO1* regulatory regions did not change after 30 (T30) and 90 min (T90) of induction of SCX-E47. (B) The percentage of GFP-positive cells changed significantly after 90 min of SCX-E47 induction for the *COL1A2* regulatory region and after 30 and 90 min for the *COL1A2-ENO1* regulatory region. No changes were observed for the *COL1A2-ADO1* regulatory region. At least three independent experiments were performed for each construction. Statistical differences were measured with the Repeated Measures ANOVA, the Tukey post-test, and the paired *t*-tests. $*p < 0.05$, $**p < 0.01$.

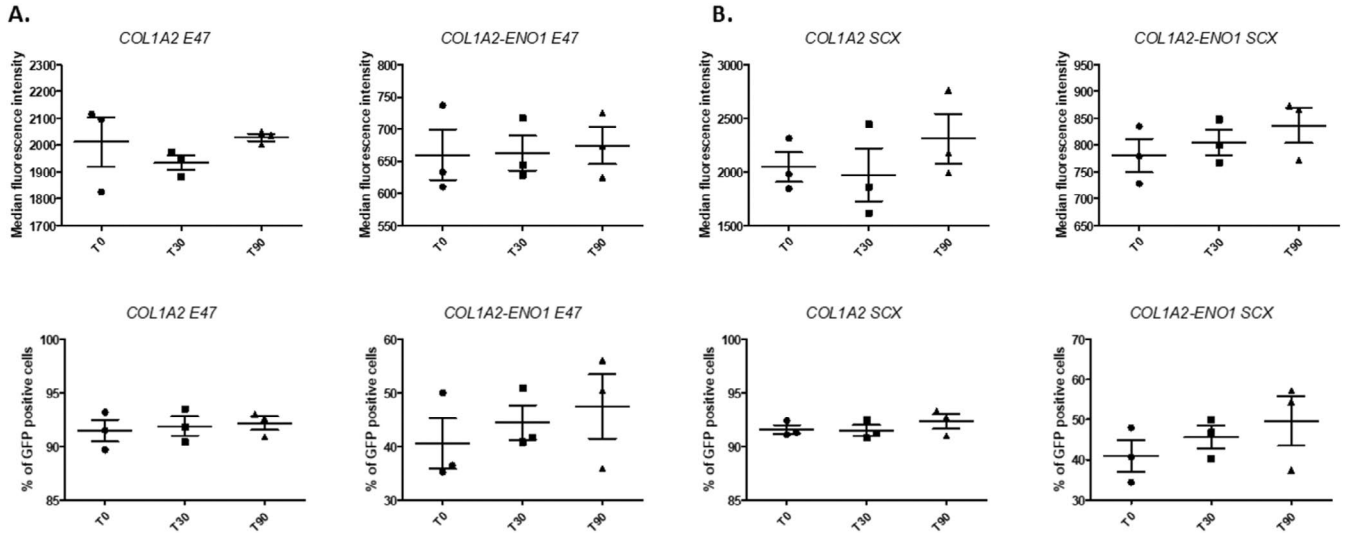

FIGURE 4 | The isolated induction of E47 or SCX did not affect the GFP expression from *COL1A2* regulatory sequences. (A) The Median Fluorescence Intensity levels and percentage of GFP-positive cells from the *COL1A2* and *COL1A2-ENO1* regulatory regions did not change after 30 (T30) and 90 min (T90) of E47 induction. (B) The Median Fluorescence Intensity levels and percentage of GFP-positive cells from the *COL1A2* and *COL1A2-ENO1* regulatory regions did not change after 30 (T30) and 90 min (T90) of SCX induction. No statistical differences were observed for any graph in the figure. At least three independent experiments were performed for each construction.

Reasoning that the bHLH transcriptional effects could be diluted within the constantly expressing GFP population, only the data of the changing population were analyzed (see Methods and Figure S4) and normalized to make comparisons among groups. Thus, we compared the T30 and T90 induction time points among reporter genes and TF combinations tested.

When the SCX-E47 effect was compared for *COL1A2* and *COL1A2-ENO1*, it was found that the median GFP fluorescence was affected only in *COL1A2* by 40%. Contrastingly, the percentage of positive GFP-expressing cells augmented similarly at T30 and T90 for both reporters (Figure 5A), almost doubling the number of positive cells by T90. This suggests that neither the

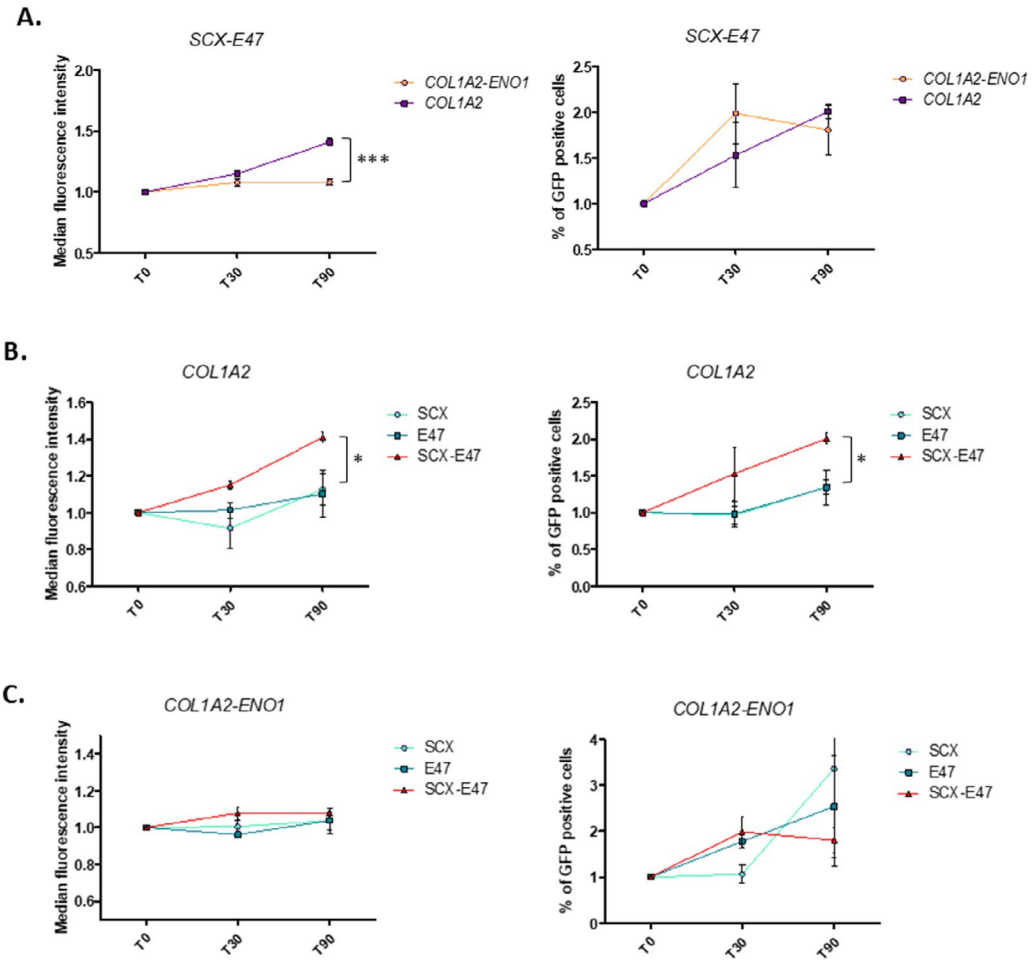

FIGURE 5 | Data comparison for specific bHLH TFs over the *COL1A2* and *COL1A2-ENO1* regulatory regions detailed the SCX-E47 transcriptional effects. The compared data were derived after gating the flow cytometry data to visualize the cells that altered the cloud's shape. (A) After data normalization, the median fluorescence intensity and percentage of GFP-positive cells for the *COL1A2* and *COL1A2-ENO1* regulatory regions were graphed when the expression of SCX-E47 was induced. (B) Comparison graph as in A. but for the *COL1A2* regulatory region after SCX-E47 induction relative to individual *SCX* or *E47* inductions. (C) Comparison graph as in B. for the *COL1A2-ENO1* regulatory region. The statistical differences were assessed with a two-way ANOVA test and Bonferroni post-tests for detailed individual differences. $*p < 0.05$, $***p < 0.001$.

basal transcription nor the core promoter composition significantly altered the SCX-E47 positive transcriptional effect.

When analyzing only the *COL1A2* reporter gene for the three TFs combinations tested, we confirmed the relevant role of the SCX-E47 heterodimer over the homodimers through statistically significant changes in median GFP intensity and the percentage of GFP-positive cells (Figure 5B). Using the same analysis for the *COL1A2-ENO1* reporter, the E47 effect at the percentage of GFP-positive cells did not differ from the SCX-E47 effect at 30 min, suggesting that the E47 homodimer may also promote an early transcriptional effect (Figure 5C).

3.5 | Role of SCX-E47 Heterodimers Over a Dense E-Box Regulatory Region (*TGFB1*)

Other potential SCX-E47 target genes involved in mesenchymal gene expression also contained multiple E-boxes within 2 kb upstream of the transcription start point (Figure S6). Among them, the *TGFB1* regulatory region [44] encompassed five SCX binding sites within a sequence containing 13 E-boxes (Figure 6A,B). Therefore, we constructed two reporter genes, including the *TGFB1* E-box regulatory region and the initially tested *COL1A2* and *ENO1* core promoters (Figure 6A–C). Unexpectedly, the expression of GFP originating from the *TGFB1* regulatory domain was markedly diminished compared to that from *COL1A2* (Figure 6C). Nonetheless, we postulated that the transcriptional modulation mediated by SCX-E47 over this sequence would likely yield a more pronounced impact than that of *COL1A2*.

Surprisingly, no significant gene expression changes were observed for SCX, E47, or SCX-E47 over either *TGFB1* reporters when comparing T0 to T30 and 90 (Figure S7). Thus, we analyzed the changing population data and compared it to the *COL1A2* data. Figure 6D shows that SCX-E47 did not induce the GFP expression from the *TGFB-COL* regulatory region, whereas expression from the *COL1A2* doubled at T90 when measured as a percentage. Furthermore, SCX-E47 also induced a statistically significant change in the fluorescence median intensity from the *COL1A2* regulatory region relative to the *TGFB-COL* sequence (Figure 6E).

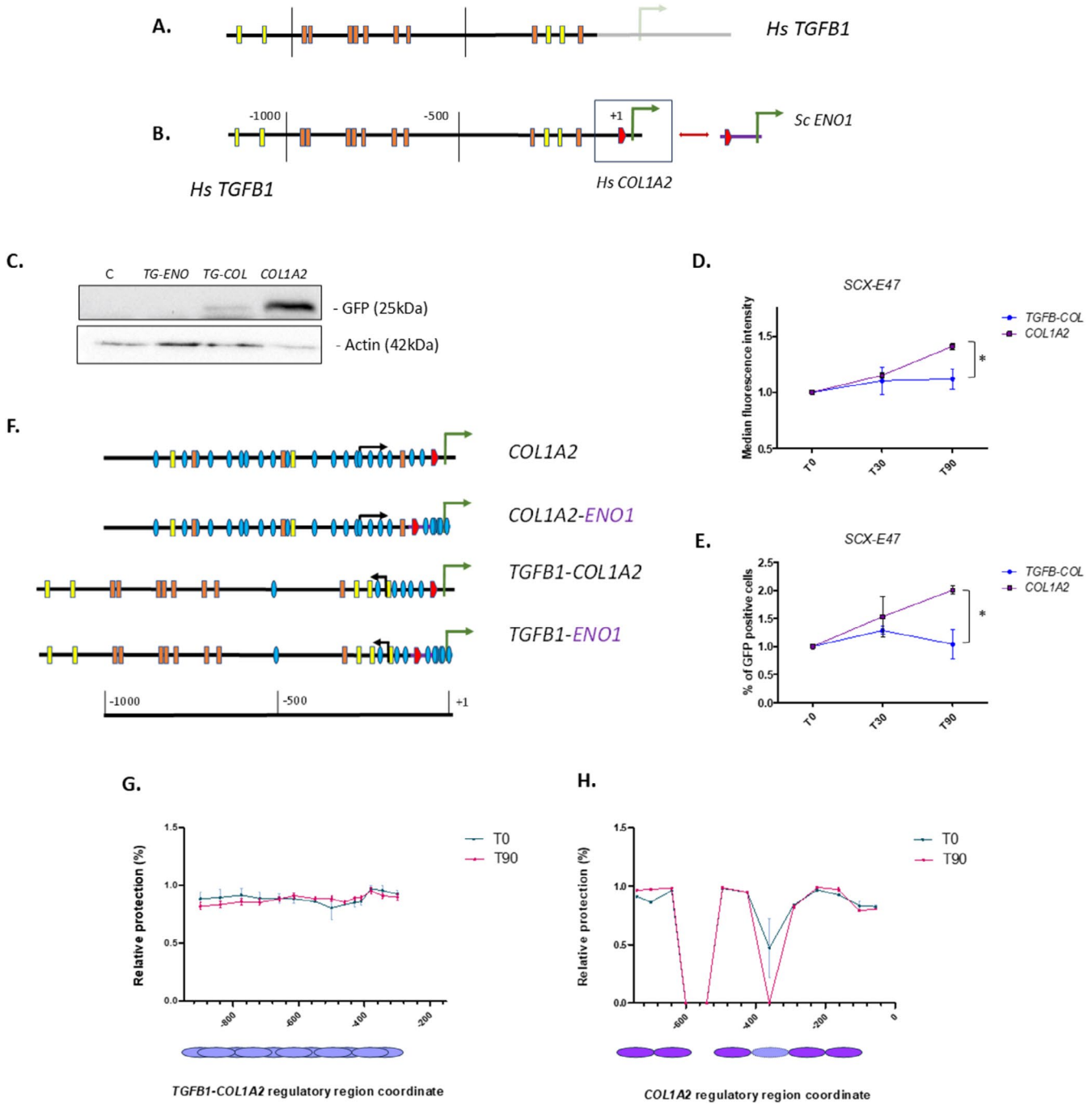

FIGURE 6 | The *TGFB1* E-box regulatory region was unresponsive to SCX/E47 bHLH TFs. (A) Schematic representation of the E-box regulatory region of the *TGFB1* regulatory region. The tsp. shown is within the *TGFB1_1* promoter sequence, located about 500 bps upstream of the main *TGFB1_2* promoter in the human genome. (B) Schematic representation of the *TGFB1-COL1A2* construct containing the *TGFB1* E-box region and the *COL1A2* core promoter, structuring a human-human hybrid regulatory sequence (*TG-COL*). The *COL1A2* promoter sequence was swapped with the yeast *ENO1* promoter to structure the *TGFB1-ENO1* human-yeast hybrid regulatory sequence (*TG-ENO*). The *TGFB1* sequence contained 13 E-boxes (orange) with five SCX binding sites (yellow). (C) Basal reporter GFP expression levels for the different constructs by immunoblot. The blot was stripped and probed for Actin as a loading control. (D) Data comparison for SCX-E47 over the *COL1A2* and *TGFB-COL1A2* regulatory regions. (E) Median Fluorescent Intensity and percentage of GFP-positive cells graphs comparing levels between *TGFB-COL1A2* and *COL1A2* regulatory sequences upon SCX-E47 induction. (F) The transcription factor binding sites for the regulatory regions tested were assessed at Swissregulon and schematically localized as blue ovals in the E-box maps of the sequences studied. Orange and yellow rectangles indicate E-boxes, whereas the purple promoters correspond to yeast sequences. Shorter arrows depict additional promoters reported present within the genomic sequence. The statistical differences were assessed with a two-way ANOVA test and Bonferroni post-tests for detailed individual differences. $*p < 0.05$. (G, H) NuSA for *TGFB1* (G) and *COL1A2* (H) regulatory regions under basal expression conditions (T0) and after SCX-E47 induction (T90). Three independent experiments were analyzed by qPCR and graphed. Below each graph, a schematic representation of nucleosome positioning is included. Nucleosomes were represented as purple ovals.

To understand the differential effect of SCX-E47 over the *COL1A2* and *TGFB1* reporters, we further analyzed the sequence content and nucleosome positioning of the regulatory regions tested. Figure 6F shows that the *COL1A2* E-boxes were contained within a region very dense in TF binding sites, whereas the *TGFB1* E-boxes were localized further upstream of the closest region dense in TF binding sites. Utilizing nucleosome scanning assays, we found that the *TGFB1* regulatory region was wrapped entirely in a random continuous nucleosome distribution (Figure 6G). Contrastingly, the *COL1A2* regulatory region was organized as other transcription regulatory regions, characterized by well-positioned nucleosomes surrounding a nucleosome-free region (Figure 6H). Additionally, upon the SCX-E47 induction, the nucleosome positioning was further improved (Figure 6H).

3.6 | SCX-E47 Also Modulated an Independent TF-Dense Regulatory Region

To confirm the relevance of contiguous TF binding sites for regulation by bHLH TFs, we tested, as above, the regulatory role of SCX-E47 over the yeast *ENO1* regulatory sequence, which contained 12 E-boxes embedded within a region densely packed with other TF binding sites (Figure 7A). A shorter *ENO1* regulatory sequence was also examined, lacking other transcription factor binding sites upstream of the SCX-binding site (Figure 7B). These two reporter genes tested also differed in basal transcription levels (Figure 7C). The induction of SCX-E47 did not affect the median fluorescence intensity from the varying population of either regulatory sequence (Figure 7D). Contrastingly, the percentage of GFP-positive cells increased 90 min after the SCX-E47 induction only for the *ENO1* regulatory sequence and not for the short *ENO1* reporter (Figure 7E). Additionally, the transcriptional induction was similar to that driven by the *COL1A2* regulatory region (Figure 7E).

4 | Discussion

4.1 | Yeast as a System to Study Tissue-Specific TFs

In contrast to the diversity and extent of bHLH TFs in humans, yeast has few reported examples, including at least 8 TFs [40] and one protein involved in DNA replication [45]. In humans, most type II bHLH TFs have a role in development and cell differentiation, whereas in yeast, these TFs mostly have metabolic roles [40]. Yeast offered an in vivo system containing the basal RNA polymerase II transcriptional machinery and numerous sequence-specific TFs that establish a "basic" transcriptional environment where the effect of "external" TFs can be analyzed. The absence of tissue-specific bHLH TFs was warranted; additionally, many yeast bHLH TFs were assumed to be absent or expressed at low levels because they are functional or expressed only under specific metabolic conditions not met in our experiments. This would apply, for example, to *PHO4*, *INO2*, and *INO4*, as the media used contained adequate amounts of phosphate and inositol. Additionally, we assumed that *RTG1* would sequester *RTG3* in the cytoplasm due to the presence of functional mitochondria [46, 47].

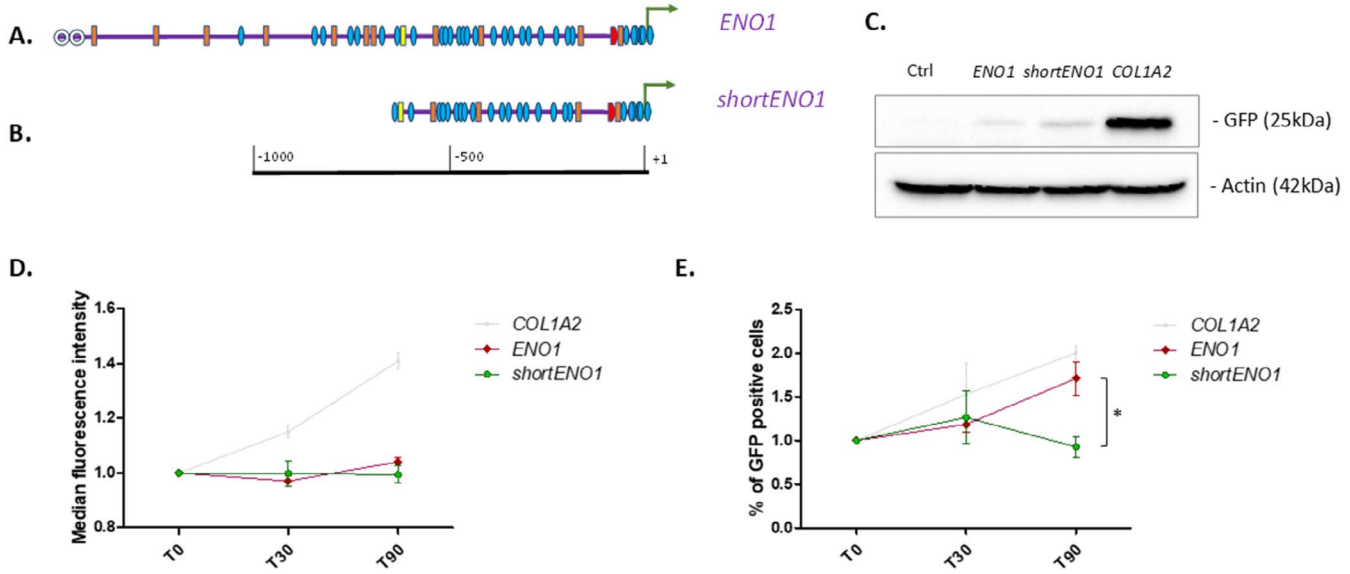

FIGURE 7 | The SCX/E47 bHLH TFs also modulated the *ENO1* regulatory region. (A) Schematic representation of the E-box regulatory region of the *Saccharomyces cerevisiae's ENO1* regulatory region. The *ENO1* sequence contained 12 E-boxes (orange) with one SCX binding site (yellow). The transcription factor binding sites for the regulatory regions tested were schematically localized as blue ovals. The sequence also contained two 5′ replication origins (represented by white circles). (B) Schematic representation of the *shortENO1* construct lacking the transcription factor binding sites localized upstream of the SCX E-box (yellow). (C) Basal reporter GFP expression levels for the different constructs by immunoblot. The *shortENO1* regulatory region exhibited higher basal expression levels, possibly due to removing replication origins. The blot was stripped and probed for Actin as a loading control. (D, E) Data comparison for SCX-E47 over the *ENO1* and *shortENO1* regulatory regions. Varying population Median Fluorescent Intensity (D) and percentage of GFP-positive cells graphs (E) comparing levels between *ENO1* and *shortENO1* regulatory sequences upon SCX-E47 induction. The *COL1A2* data were included for comparison in both graphs as gray lines. The statistical differences were assessed using a two-way ANOVA test and Bonferroni post hoc tests for detailed individual comparisons. $^*p < 0.05$.

A similar system in yeast was previously used to extend our understanding of mammalian bHLH TFs [48]. The system only tested the viability of yeast dependent on the function of a human bHLH TF. Opposingly, we optimized the expression of human genes in yeast, utilized an inducible bHLH TF expression system, and measured the reporter expression in single cells. This allowed a detailed analysis of the transcriptional dynamics.

Despite the advantages mentioned above, including the possibility of using the system to analyze other bHLH TF heterodimers, the system has two main drawbacks to consider when interpreting the data: the inherent differences between yeast and mammals and the use of a plasmid reporter. The chromatin structure and transcriptional cofactors do differ between mammalian cells and yeast. Furthermore, yeast does not express many TFs that potentially bind to the regulatory regions tested. Additionally, the chromatin context differs between chromosomal and plasmid DNA. Thus, this study would certainly complement a parallel study performed within the human chromosomal context.

4.2 | Reporter Gene and Data Interpretation

The GFP reporter was measured by flow cytometry to analyze and differentiate cell-to-cell data, allowing discrimination of the non-GFP-expressing population from the positive population. The quantitative parameters measured by the flow cytometer and analyzed in this study were the percentage of GFP-positive cells and the median fluorescence intensity. We reasoned that the values obtained provided different information. A change in the median intensity after the induction of the TFs could mean that the basal GFP expression was affected when the bHLH TFs induced boosted or repressed the ongoing transcriptional activity [49]. On the other hand, a change in the percentage of GFP-positive cells could mean that more cells started expressing GFP upon TF induction or that the expression level reached the threshold required for detection by the cytometer. Thus, the percentage changes could be related to having very low or non-GFP-expressing cells start expressing it when the bHLH transcription factors SCX and E47 were present. The flow cytometry analysis also allowed better visualization of the transcriptional effect by removing the unchanging population. This process was also related to a cell morphology change for unknown reasons.

The induction of SCX-E47 expression primarily altered the number of cells expressing GFP for two reporter genes. The increase was higher for the *COL1A2-ENO1* relative to the *COL1A2* gene, perhaps because the basal expression level of the *COL1A2* reporter gene was very high (90%). The observed changes in the percentage of GFP-positive cells were in total agreement with the known regulatory sequences (enhancers) role in increasing the probability of transcription [50, 51]. The only observed change in the median fluorescence intensity was for SCX-E47 over the *COL1A2* regulatory region compared to the *COL1A2-ENO1* sequence, suggesting that SCX-E47 could potentiate transcriptional active units over this particular sequence.

The results with the *ADO1* core promoter showed null expression of GFP in western blots and very few GFP-positive cells by cytometry. This would suggest that the addition of the *ADO1* promoter resulted in non-transcribed constructs despite a "potent" upstream regulatory region (human *COL1A2*). This result reinforces the conclusion of previous reports, which found that upstream regulatory regions could not function without a functional promoter [52].

4.3 | The Heterodimeric SCX-E47 as a Regulator of the *COL1A2* Gene

It is well established from experiments in mammalian cells that E47 functions as both a homodimer and heterodimer with multiple dimerization partners [3]. Except for one in vitro report [22], the SCX function appears dependent on heterodimeric interactions with E proteins [5]. Turning to the evidence presented in this work, the function of SCX in the absence of heterodimerization partners may be scarce or nonexistent for the *COL1A2* and *TGFB1* proximal regulatory regions. E47 had a mild homodimeric role at early times for the *COL1A2-ENO1* reporter. Thus, comparing the results of homodimeric E47 with SCX-E47, the early time point may also depend on E47.

Due to the observed change in the percentage of GFP-positive cells, the most successful dimeric complex regulating the *COL1A2* gene was SCX-E47, which did not affect the *TGFB1* gene.

4.4 | Significance of the Enhancer-Promoter Specificity

The data with core promoter differences could also be related to the known enhancer's preference for specific promoters [53, 54], which have been primarily studied by comparing canonical promoter motifs such as TATA or DPE. However, because all the tested core promoters contained a TATA box, the different transcriptional efficiencies could relate to the known enhancer-specific promoter preferences or the distinct localization of the TATA box in yeast and human promoters (−25 and −100, respectively).

4.5 | Relevance of the Sequence and Genetic Context for bHLH TF Function

As an isolated system, we expected that the transcriptional changes resulting from overexpression of the TFs of interest would result in small changes because the context of other human tissue-specific transcription factors would be absent. Nevertheless, yeast expresses several transcription factors conserved in humans, which also structure a genomic context of DNA-bound regulators that could interact with the E-box-binding transcription factors studied. These additional DNA-bound proteins had a role in our results.

To maintain a constant histone composition and chromatin structure, we used plasmid reporters with high expression levels and restricted long-distance interactions with other motifs [41, 42]. A possible disadvantage of "floating reporter plasmids" is that they could lead to random unwanted interactions with other chromosomal domains. Nevertheless, the interactions would be different in independent experiments and, therefore, diluted in the results. Additionally, because the most proximal

regulatory regions are closest to the model of the enhancer effects over a linked gene [55] and because a short distance between the core promoter and the regulatory domains tested may not require looping as part of the activation process [56], we considered episomes as an informative system where we assumed a relatively constant genomic context for each reporter. Nevertheless, when the nucleosome positioning of the regulatory regions tested was evaluated, the *COL1A2* and *TGFB1* sequences had contrasting architectures that were undoubtedly relevant for the SCX-E47 individual function.

As the yeast promoters were expected to produce the highest basal GFP levels, we were surprised to find that the original *COL1A2* upstream and core promoter regulatory sequences produced the highest basal GFP levels relative to the other regulatory regions. Therefore, we hypothesize that the high transcriptional capacity of the *COL1A2* upstream and core promoter regulatory sequences in yeast may be attributable to a specific sequence composition, nucleosome positioning, diverse transcription factor binding motifs, and the spatial arrangements of these motifs, which have been progressively honed throughout evolution.

The fact that the strongest yeast promoter tested (*ENO1*) negatively affected the transcriptional effectiveness of the *COL1A2* regulatory region could be related to a partially incompatible sequence environment between out-of-context human and yeast sequences. Nevertheless, the construction exhibited a significantly elevated basal transcriptional output, with approximately 50% of the GFP-positive cells and a corresponding median fluorescence intensity. These attenuated basal GFP levels facilitated enhanced visualization of the transcriptional impact of the assessed bHLH transcription factors.

The results observed with the *TGFB1* reporter were unexpected because the original hypothesis only focused on the E-box content. Thus, having less transcription with more E-boxes seemed contradictory. Therefore, we analyzed the reported chromosomal composition of the studied regions to understand the obtained results. The sequence analysis found a relevant difference: the E-boxes of the *COL1A2* regulatory region were embedded in an area containing numerous (more than 20) binding sites for other TFs. At the same time, the numerous *TGFB1* E-boxes were far from clustered TF binding sites. Additional contrasting sequence characteristics between the *COL1A2* and *TGFB1* regulatory regions could also affect the SCX-E47 binding. These characteristics include different sequence combinations of some of the 16 possible E-boxes and the range between the promoter and the E-boxes. Furthermore, the binding of basic bHLH TFs is contingent upon the sequence context surrounding the E-box [57, 58].

The NuSA also demonstrated contrasting chromatin landscapes when comparing the *TGFB1* and *COL1A2* regulatory regions. Randomly positioned nucleosomes wrapped the first, whereas the *COL1A2* regulatory region had well-positioned nucleosomes and a nucleosome-free region that characterize transcriptional regulatory regions [59]. Even though the nucleosome positioning did not change much upon the SCX-E47 induction, one of the nucleosomes' positioning was further improved. Other bHLH TFs have been demonstrated to bind nucleosomes [60]. In this particular case, one SCX-E47 binding site is localized at the nucleosome that becomes better positioned after the induction of its expression.

Furthermore, when analyzing the original chromosomal context of the regulatory regions tested using the UCSC genomic browser, we found that the *COL1A2* regulatory region had some of the molecular signatures of enhancers [53], including DNase I hypersensitive sites and characteristic histone modifications, such as H3K4me1 and H3K27ac. The core promoter region was reported to be a binding site for the TBP and the RNA polymerase II, including the H3K4me3 mark [61]. The RNA polymerase II binding region extended up to 80% of the regulatory region studied. The above chromosomal details and our data show that the sequence upstream of *COL1A2* is adequate to promote transcription, even within the yeast plasmid context. Therefore, the reduced transcription rates when the human core promoter was replaced with a yeast sequence could be explained by an alteration in the structure of the originally permissive *COL1A2* core promoter sequence. In agreement with our NuSA data, the 1 kb human *TGFB1* genomic context reported at UCSC was localized outside DNase I hypersensitivity areas or RNA polymerase II binding regions. Additionally, an antisense promoter was also found within this region.

To confirm that the contrasting results between *COL1A2* and *TGFB1-COL1A2* reporters were related to the concentration of TF binding sites surrounding E-boxes, we studied the effect of inducing SCX-E47 over the *ENO1* regulatory region, which also contained multiple additional TF binding sites. The resulting transcriptional activation was similar to that observed for the *COL1A2* regulatory region. Furthermore, removing TF binding sites localized upstream of the E-box resulted in the loss of the SCX-E47 activating capacity. Thus, we propose that the localization of E-boxes within a sequence rich in other TF binding sites would enable the binding of other TFs, the formation of transcriptionally permissive and cooperative functional units, and the exchange of coregulators [50, 62]. Finally, this aligns with the two-step target search process by TFs, where an initial global search using protein–protein interactions, particularly those involving intrinsically disordered regions [63], identifies the correct "protein cloud". This is followed by a subsequent local search for specific DNA motifs, utilizing the DNA-binding domain [64]. Thus, the high abundance of TF binding sites in the *COL1A2* and *ENO1* regulatory regions may serve as a recruiting protein cloud for the SCX-E47 bHLH heterodimer.

Even though this is a basic research work identifying punctual roles of bHLH TFs over specific (out-of-context) human regulatory regions, we believe that this type of analysis has a potential long-term impact on health, as the altered expression of bHLH TFs in humans leads to multiple diseases, including fibrosis, where the regulators and regulatory regions tested here are relevant.

Author Contributions

A.L.T.-M. conceived and designed the research. A.L.T.-M. and K.M.L.-D. constructed most of the plasmids and performed most of the

experimental procedures. T.D.S.-G. constructed the *TGFB1* plasmids and one of the *PEP4* null strains. C.B., C.C.-B., A.G., and A.L.T.-M. conceived and performed the NuSA assays. A.L.T.-M. made the analyses and wrote the paper.

Acknowledgments

This paper serves as fulfillment for the requirements of obtaining a Master of Science degree in the biomedicine field at Posgrado en Ciencias Biológicas, UNAM, for K.M.L.-D. K.M.L.-D. received the fellowship CVU: 1190609 from Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) during her M.Sc. studies. We thank Dr. Ana María Cevallos for the critical discussions. We acknowledge the technical support of the Laboratorio Nacional Conahcyt de Investigación y Diagnóstico por Inmunocitofluorometría (LANCIDI), Dámaris-Priscila Romero-Rodríguez, Karen Bobadilla, and Criselda Mendoza-Milla. Lorraine Pillus and Randy Hampton kindly donated strains and plasmids.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available in the materials and methods, results, and/or supporting information of this article.

References

1. S. A. Lambert, A. Jolma, L. F. Campitelli, et al., "The Human Transcription Factors," *Cell* 175 (2018): 598–599.

2. C. Murre, "Helix-Loop-Helix Proteins and the Advent of Cellular Diversity: 30 Years of Discovery," *Genes & Development* 33 (2019): 6–25.

3. A. L. Torres-Machorro, "Homodimeric and Heterodimeric Interactions Among Vertebrate Basic Helix-Loop-Helix Transcription Factors," *International Journal of Molecular Sciences* 22 (2021): 12855.

4. R. M. Barnes and A. B. Firulli, "A Twist of Insight—The Role of Twist-Family bHLH Factors in Development," *International Journal of Developmental Biology* 53 (2009): 909–924.

5. M. Ramirez-Aragon, F. Hernandez-Sanchez, T. S. Rodriguez-Reyna, et al., "The Transcription Factor SCX Is a Potential Serum Biomarker of Fibrotic Diseases," *International Journal of Molecular Sciences* 21 (2020): 5012.

6. M. H. Farah, J. M. Olson, H. B. Sucic, R. I. Hume, S. J. Tapscott, and D. L. Turner, "Generation of Neurons by Transient Expression of Neural bHLH Proteins in Mammalian Cells," *Development* 127 (2000): 693–702.

7. W. El Yakoubi, C. Borday, J. Hamdache, et al., "Hes4 Controls Proliferative Properties of Neural Stem Cells During Retinal Ontogenesis," *Stem Cells* 30 (2012): 2784–2795.

8. C. Porcher, W. Swat, K. Rockwell, Y. Fujiwara, F. W. Alt, and S. H. Orkin, "The T Cell Leukemia Oncoprotein SCL/Tal-1 Is Essential for Development of all Hematopoietic Lineages," *Cell* 86 (1996): 47–57.

9. S. J. Rhodes and S. F. Konieczny, "Identification of MRF4: A New Member of the Muscle Regulatory Factor Gene Family," *Genes & Development* 3 (1989): 2050–2061.

10. M. J. Thayer, S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar, and H. Weintraub, "Positive Autoregulation of the Myogenic Determination Gene MyoD1," *Cell* 58 (1989): 241–248.

11. M. Garber, N. Yosef, A. Goren, et al., "A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals," *Molecular Cell* 47 (2012): 810–822.

12. H. Han, J. W. Cho, S. Lee, et al., "TRRUST v2: An Expanded Reference Database of Human and Mouse Transcriptional Regulatory Interactions," *Nucleic Acids Research* 46 (2018): D380–D386.

13. M. Bergsland, D. Ramskold, C. Zaouter, S. Klum, R. Sandberg, and J. Muhr, "Sequentially Acting Sox Transcription Factors in Neural Lineage Development," *Genes & Development* 25 (2011): 2453–2464.

14. I. Belle and Y. Zhuang, "E Proteins in Lymphocyte Development and Lymphoid Diseases," *Current Topics in Developmental Biology* 110 (2014): 153–187.

15. D. C. Bersten, A. E. Sullivan, D. J. Peet, and M. L. Whitelaw, "bHLH-PAS Proteins in Cancer," *Nature Reviews Cancer* 13 (2013): 827–841.

16. L. H. Wang and N. E. Baker, "E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease," *Developmental Cell* 35 (2015): 269–280.

17. G. Bain, I. Engel, E. C. Robanus Maandag, et al., "E2A Deficiency Leads to Abnormalities in Alphabeta T-Cell Development and to Rapid Development of T-Cell Lymphomas," *Molecular and Cellular Biology* 17 (1997): 4782–4791.

18. K. Beck, M. M. Peak, T. Ota, D. Nemazee, and C. Murre, "Distinct Roles for E12 and E47 in B Cell Specification and the Sequential Rearrangement of Immunoglobulin Light Chain Loci," *Journal of Experimental Medicine* 206 (2009): 2271–2284.

19. C. Slattery, T. McMorrow, and M. P. Ryan, "Overexpression of E2A Proteins Induces Epithelial-Mesenchymal Transition in Human Renal Proximal Tubular Epithelial Cells, Suggesting a Potential Role in Renal Fibrosis," *FEBS Letters* 580 (2006): 4021–4030.

20. T. Furumatsu, C. Shukunami, M. Amemiya-Kudo, H. Shimano, and T. Ozaki, "Scleraxis and E47 Cooperatively Regulate the Sox9-Dependent Transcription," *International Journal of Biochemistry & Cell Biology* 42 (2010): 148–156.

21. D. Brown, D. Wagner, X. Li, J. A. Richardson, and E. N. Olson, "Dual Role of the Basic Helix-Loop-Helix Transcription Factor Scleraxis in Mesoderm Formation and Chondrogenesis During Mouse Embryogenesis," *Development* 126 (1999): 4317–4329.

22. Y. Liu, H. Watanabe, A. Nifuji, Y. Yamada, E. N. Olson, and M. Noda, "Overexpression of a Single Helix-Loop-Helix-Type Transcription Factor, Scleraxis, Enhances Aggrecan Gene Expression in Osteoblastic Osteosarcoma ROS17/2.8 Cells," *Journal of Biological Chemistry* 272 (1997): 29880–29885.

23. P. Cserjesi, D. Brown, K. L. Ligon, et al., "Scleraxis: A Basic Helix-Loop-Helix Protein That Prefigures Skeletal Formation During Mouse Embryogenesis," *Development* 121 (1995): 1099–1110.

24. B. K. Kennedy, "Mammalian Transcription Factors in Yeast: Strangers in a Familiar Land," *Nature Reviews Molecular Cell Biology* 3 (2002): 41–49.

25. M. A. Romanos, C. A. Scorer, and J. J. Clare, "Foreign Gene Expression in Yeast: A Review," *Yeast* 8 (1992): 423–488.

26. X. de Martin, R. Sodaei, and G. Santpere, "Mechanisms of Binding Specificity Among bHLH Transcription Factors," *International Journal of Molecular Sciences* 22 (2021): 9150.

27. R. A. Bagchi and M. P. Czubryt, "Synergistic Roles of Scleraxis and Smads in the Regulation of Collagen 1alpha2 Gene Expression," *Biochimica et Biophysica Acta* 1823 (2012): 1936–1944.

28. R. Cohen, T. Yokoi, J. P. Holland, A. E. Pepper, and M. J. Holland, "Transcription of the Constitutively Expressed Yeast Enolase Gene ENO1 Is Mediated by Positive and Negative Cis-Acting Regulatory Sequences," *Molecular and Cellular Biology* 7 (1987): 2753–2761.

29. G. H. Patterson, S. M. Knobel, W. D. Sharif, S. R. Kain, and D. W. Piston, "Use of the Green Fluorescent Protein and Its Mutants in Quantitative Fluorescence Microscopy," *Biophysical Journal* 73 (1997): 2782–2790.

30. P. Guerra, L. A. Vuillemenot, B. Rae, V. Ladyhina, and A. Milias-Argeitis, "Systematic in Vivo Characterization of Fluorescent Protein Maturation in Budding Yeast," *ACS Synthetic Biology* 11 (2022): 1129–1141.

31. S. Kannan, T. Sams, J. Maury, and C. T. Workman, "Reconstructing Dynamic Promoter Activity Profiles From Reporter Gene Data," *ACS Synthetic Biology* 7 (2018): 832–841.

32. A. Marek and R. Korona, "Restricted Pleiotropy Facilitates Mutational Erosion of Major Life-History Traits," *Evolution* 67 (2013): 3077–3086.

33. D. Zenklusen, A. L. Wells, J. S. Condeelis, and R. H. Singer, "Imaging Real-Time Gene Expression in Living Yeast," *Cold Spring Harbor Protocols* 2007 (2007): pdb.prot4870.

34. R. Elble, "A Simple and Efficient Procedure for Transformation of Yeasts," *BioTechniques* 13 (1992): 18–20.

35. A. L. Torres-Machorro, L. G. Clark, C. S. Chang, and L. Pillus, "The Set3 Complex Antagonizes the MYST Acetyltransferase Esa1 in the DNA Damage Response," *Molecular and Cellular Biology* 35 (2015): 3714–3725.

36. J. J. Infante, G. L. Law, and E. T. Young, "Analysis of Nucleosome Positioning Using a Nucleosome-Scanning Assay," *Methods in Molecular Biology* 833 (2012): 63–87.

37. C. Campero-Basaldua, J. Gonzalez, J. A. Garcia, et al., "Neo-Functionalization in *Saccharomyces cerevisiae*: A Novel Nrg1-Rtg3 Chimeric Transcriptional Modulator Is Essential to Maintain Mitochondrial DNA Integrity," *Royal Society Open Science* 10 (2023): 231209.

38. V. Lejard, G. Brideau, F. Blais, et al., "Scleraxis and NFATc Regulate the Expression of the Pro-alpha1(I) Collagen Gene in Tendon Fibroblasts," *Journal of Biological Chemistry* 282 (2007): 17665–17675.

39. T. Ellenberger, D. Fass, M. Arnaud, and S. C. Harrison, "Crystal Structure of Transcription Factor E47: E-Box Recognition by a Basic Region Helix-Loop-Helix Dimer," *Genes & Development* 8 (1994): 970–980.

40. K. A. Robinson, J. I. Koepke, M. Kharodawala, and J. M. Lopes, "A Network of Yeast Basic Helix-Loop-Helix Interactions," *Nucleic Acids Research* 28 (2000): 4460–4466.

41. V. Mladenova, E. Mladenov, and G. Russev, "Organization of Plasmid DNA Into Nucleosome-Like Structures After Transfection in Eukaryotic Cells," *Biotechnology & Biotechnological Equipment* 23 (2009): 1044–1047.

42. W. Tong, O. I. Kulaeva, D. J. Clark, and L. C. Lutter, "Topological Analysis of Plasmid Chromatin From Yeast and Mammalian Cells," *Journal of Molecular Biology* 361 (2006): 813–822.

43. L. David, W. Huber, M. Granovskaia, et al., "A High-Resolution Map of Transcription in the Yeast Genome," *Proceedings of the National Academy of Sciences of the United States of America* 103 (2006): 5320–5325.

44. I. Grafe, S. Alexander, J. R. Peterson, et al., "TGF-Beta Family Signaling in Mesenchymal Differentiation," *Cold Spring Harbor Perspectives in Biology* 10 (2018): a022202.

45. Y. Zhang, Z. Yu, X. Fu, and C. Liang, "Noc3p, a bHLH Protein, Plays an Integral Role in the Initiation of DNA Replication in Budding Yeast," *Cell* 109 (2002): 849–860.

46. T. Sekito, J. Thornton, and R. A. Butow, "Mitochondria-To-Nuclear Signaling Is Regulated by the Subcellular Localization of the Transcription Factors Rtg1p and Rtg3p," *Molecular Biology of the Cell* 11 (2000): 2103–2115.

47. E. D. Wong, S. R. Miyasato, S. Aleksander, et al., "Saccharomyces Genome Database Update: Server Architecture, Pan-Genome Nomenclature, and External Resources," *Genetics* 224 (2023): iyac191.

48. K. L. Mak, L. C. Longcor, S. E. Johnson, C. Lemercier, R. Q. To, and S. F. Konieczny, "Examination of Mammalian Basic Helix-Loop-Helix Transcription Factors Using a Yeast One-Hybrid System," *DNA and Cell Biology* 15 (1996): 1–8.

49. T. E. I. Taher, "Monitoring Promoter Activity by Flow Cytometry," *Methods in Molecular Biology* 1651 (2017): 65–73.

50. A. Panigrahi and B. W. O'Malley, "Mechanisms of Enhancer Action: The Known and the Unknown," *Genome Biology* 22 (2021): 108.

51. M. C. Walters, S. Fiering, J. Eidemiller, W. Magis, M. Groudine, and D. I. Martin, "Enhancers Increase the Probability but Not the Level of Gene Expression," *Proceedings of the National Academy of Sciences of the United States of America* 92 (1995): 7125–7129.

52. M. Petrascheck, D. Escher, T. Mahmoudi, C. P. Verrijzer, W. Schaffner, and A. Barberis, "DNA Looping Induced by a Transcriptional Enhancer In Vivo," *Nucleic Acids Research* 33 (2005): 3743–3750.

53. M. Bulger and M. Groudine, "Functional and Mechanistic Diversity of Distal Transcription Enhancers," *Cell* 144 (2011): 327–339.

54. J. E. Butler and J. T. Kadonaga, "Enhancer-Promoter Specificity Mediated by DPE or TATA Core Promoter Motifs," *Genes & Development* 15 (2001): 2515–2519.

55. E. Anderson, P. S. Devenney, R. E. Hill, and L. A. Lettice, "Mapping the Shh Long-Range Regulatory Domain," *Development* 141 (2014): 3934–3943.

56. K. C. Dobi and F. Winston, "Analysis of Transcriptional Activation at a Distance in *Saccharomyces cerevisiae*," *Molecular and Cellular Biology* 27 (2007): 5575–5586.

57. Z. Wang, Y. Wu, L. Li, and X. D. Su, "Intermolecular Recognition Revealed by the Complex Structure of Human CLOCK-BMAL1 Basic Helix-Loop-Helix Domains With E-Box DNA," *Cell Research* 23 (2013): 213–224.

58. R. Gordan, N. Shen, I. Dror, et al., "Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors Through DNA Shape," *Cell Reports* 3 (2013): 1093–1104.

59. W. K. M. Lai and B. F. Pugh, "Understanding Nucleosome Dynamics and Their Links to Gene Expression and DNA Replication," *Nature Reviews. Molecular Cell Biology* 18 (2017): 548–562.

60. M. Fernandez Garcia, C. D. Moore, K. N. Schulz, et al., "Structural Features of Transcription Factors Associating With Nucleosome Binding," *Molecular Cell* 75 (2019): 921–932.e926.

61. K. Hamamoto, Y. Umemura, S. Makino, and T. Fukaya, "Dynamic Interplay Between Non-Coding Enhancer Transcription and Gene Activity in Development," *Nature Communications* 14 (2023): 826.

62. C. A. Shively, J. Liu, X. Chen, K. Loell, and R. D. Mitra, "Homotypic Cooperativity and Collective Binding Are Determinants of bHLH Specificity and Function," *Proceedings of the National Academy of Sciences of the United States of America* 116 (2019): 16143–16152.

63. S. Brodsky, T. Jana, K. Mittelman, et al., "Intrinsically Disordered Regions Direct Transcription Factor in Vivo Binding Specificity," *Molecular Cell* 79 (2020): 459–471.e454.

64. M. V. Staller, "Transcription Factors Perform a 2-Step Search of the Nucleus," *Genetics* 222 (2022): iyac111.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.