



**UNIVERSIDAD DE CANTABRIA/
EUSKAL HERRIKO UNIBERTSITATEA**

MASTER IN MOLECULAR BIOLOGY AND BIOMEDICINE

MASTER'S THESIS

**Analysis of the role of CpG islands in
the re-establishment of enhancer-
dependent gene expression patterns
after mitotic exit.**

Performed by: Miguel Sánchez Pérez

Supervised by: Álvaro Rada Iglesias / Helena Gómez Asenjo

**Transcriptional Regulation in Development and Congenital Diseases /
Department of Cellular and Molecular Signalling / IBBTEC**

SANTANDER 2025

Abstract

All cells within an organism typically share the same genetic material. Therefore, the differentiation of individual cell lineages is driven by a specific epigenetic and transcriptional profile, which needs to be tightly regulated to maintain cellular identity. Various genetic and epigenetic elements—such as histones modifications, transcription factors, promoters and enhancers—play a key role in guiding and preserving cell fate. In proliferative cells, transcription is completely halted during mitosis and reinitiated upon entry into G1. To maintain cellular identity, gene expression programs must be re-established after mitosis, including the reactivation of transcription factor genes that play crucial roles in cell identity regulation. These genes are typically controlled by distal enhancers.

Here, we hypothesize that CpG-rich enhancers (i.e. enhancers associated with CpG islands) might be particularly important to ensure the timely and precise reactivation of regulatory genes as cells transition from mitosis to G1, thus restoring cell identity. To test this hypothesis, we selected two enhancers that are active in mouse embryonic stem cells (mESC) and created different constructs combining them with or without a CpG island. These constructs were inserted at 100 kb from the *Gata6* promoter in mESCs, generating transgenic lines. By evaluating *Gata6* expression via RT-qPCR in asynchronous cells, we observed that while the CpG island alone has a negative effect on *Gata6* expression, the combination of an active enhancer with a CpG island increases its expression. Although these results are preliminary, the establishment of these cell lines will enable future studies in a mitotic context.

Background

During the cell cycle, cells undergo a profound reorganization of nuclear architecture, particularly during mitosis, when global transcription is silenced, chromatin becomes highly compacted, and most transcription factors disengage from their binding sites. This transient but dramatic shutdown raises a central question in cell biology: How can cells faithfully maintain their transcriptional identity across successive rounds of division? The answer lies in the capacity of cells to preserve epigenetic and regulatory information, which ensures that transcriptional programs characteristic of a given cell type is rapidly reestablished upon mitotic exit.

Disruption of this tightly controlled process has been increasingly linked to developmental disorders and oncogenesis. In embryonic development, failure to reactivate enhancers or promoters in a timely and cell type-specific manner can lead to defective lineage specification and congenital anomalies (Pachano et al., 2021; Angeloni & Bogdanović, 2021).

In cancer, aberrant bookmarking or enhancer misregulation contributes to oncogenic transcriptional programs: bookmarking factors such as MYC and RUNX1 can drive uncontrolled proliferation when deregulated, while mutations in enhancer-associated cofactors like BRD4 promote transcriptional addiction in tumour cells (Pelham-Webb et al., 2021). A key regulator of differential gene expression programs are the enhancers, that govern spatiotemporal and quantitative expression dynamics of target genes. Enhancers are widely believed to contact the target promoters to effect transcriptional activation via transcription factor requirement (Contreras & Perea-Resa, 2024). Within this framework, active enhancers have emerged as critical epigenetic memory platforms, functioning not merely as passive regulatory elements but as dynamic hubs that safeguard transcriptional competence during mitosis

Active enhancers are distal cis-regulatory elements characterized by specific chromatin signatures, including the enrichment of histone modifications such as H3K27ac and H3K4me1, as well as the recruitment of transcriptional coactivators such as p300/CBP. In contrast, inactive or poised enhancers display either repressive marks (e.g., H3K27me3) or bivalent signatures that maintain them in a primed state. Active enhancers confer a unique predisposition to rapidly resume regulatory activity following transcriptional arrest (Proudhon et al., 2016). Genome-wide studies employing ChIP-seq and ATAC-seq have shown that a subset of active enhancers remains marked during mitosis and, in some instances, continues to be bound by specific transcription factors and it's been studied that disruption of H3K27 acetylation on enhancers can lead to aberrant gene activation, contributing to developmental defects and oncogenic transcriptional programs (Creyghton et al.2010). These sites act as mitotic bookmarks, facilitating the rapid reactivation of gene expression programs in the early G1 phase (Teves et al., 2016; Festuccia et al., 2019).

The concept of mitotic bookmarking encompasses the retention of regulatory proteins (including transcription factors, chromatin remodelers, and epigenetic cofactors) on selected genomic loci during mitosis. While chromatin is globally condensed and transcriptional activity is silenced, bookmarking ensures that key regulatory information is not lost. This mechanism is exemplified by the persistence of histone modifications such as H3K4me3 at promoters, which serve as stable epigenetic cues. Conversely, marks such as H3K27ac may be partially erased during mitosis, requiring their reestablishment by coactivators like p300 upon mitotic exit. Bookmarking transcription factors—including GATA1, FOXA1, OCT4, SOX2, and c-MYC—are capable of binding condensed chromatin, thereby acting as functional “molecular placeholders” that prime

the genome for rapid transcriptional reactivation (Kadauke & Blobel, 2013; Teves et al., 2016).

Beyond ensuring continuity of transcriptional programs, bookmarking has profound implications for cell fate decisions, developmental plasticity, and disease. In stem cells, bookmarking mechanisms underpin pluripotency by enabling the swift reactivation of genes central to stem cell identity. For example, in human embryonic stem cells, active enhancers retain partial accessibility and remain associated with cofactors such as BRD4 and p300, thereby preserving regulatory competence even during mitosis (Zhu et al., 2023). The cooperative interplay between pioneer factors like OCT4 and SOX2 and chromatin remodelers highlights a multilayered regulatory logic by which enhancer bookmarking stabilizes cell identity. Conversely, in oncogenesis, disruption of bookmarking mechanisms can lead to inappropriate enhancer activation or silencing, contributing to aberrant transcriptional circuits that fuel tumorigenesis (Raccaud & Suter, 2018).

An additional layer of complexity arises from the interplay between enhancers and CpG islands (CGIs). These genomic regions, typically hypomethylated and enriched in CpG dinucleotides, are strongly associated with promoters and transcription initiation sites across vertebrates (Angeloni & Bogdanović, 2021). However, not all CGIs are promoter-associated. So-called orphan CGIs function as enhancer-associated regulatory modules. Their activity is not dictated by linear genome proximity but by higher order three dimensional chromatin folding, which enables physical contact between distal enhancers and their target promoters. Strikingly, orphan CGIs have been demonstrated to potentiate the activity of poised enhancers, amplifying their regulatory impact and modulating gene responsiveness (Bell & Vertino, 2017; Pachano et al, 2021). This enhancer–CGI cooperation underscores the evolutionary and functional versatility of CGIs beyond their canonical role at promoters.

Recent high-resolution chromatin conformation analyses and single cell epigenomic approaches have provided direct evidence that subsets of enhancers retain structural and functional hallmarks during mitosis. These enhancers act as epigenetic beacons, ensuring that lineage-specific genes are promptly reactivated after mitotic exit (Pelham-Webb et al., 2021; Zhu et al., 2023). Thus, enhancer bookmarking is now recognized as an integral component of cellular memory systems, bridging the apparent gap between transcriptional silencing during mitosis and the continuity of identity programs.

Collectively, the studies reviewed here converge on the idea that enhancers are emerging as central players in the maintenance of transcriptional identity across cell

division. From histone modifications and bookmarking factors to the cooperation with orphan CpG islands, enhancers provide both structural and functional continuity in gene regulation. Their ability to retain regulatory features during mitosis ensures that daughter cells faithfully inherit transcriptional states, thereby linking epigenetic memory to developmental fidelity. As such, enhancers are no longer viewed solely as distal regulatory elements but as active guardians of cellular identity, a concept that is gaining increasing prominence in developmental biology, stem cell research, and cancer epigenetics (Rada-Iglesias et al. 2011; Calo & Wysocka 2013).

A deeper understanding of the role of active enhancers and CpG islands (CGIs), maintaining the transcriptional profile after mitosis, requires advances in several fundamental areas of research and their subsequent integration, enabling more targeted investigations. Progress in genetics, epigenetics, and cell biology applied to mitosis has shed new light on this process, while also highlighting unresolved questions in which regulatory elements such as enhancers and CGIs may play a decisive role (Ito & Zaret, 2022). Complementary studies focusing on the functional mechanisms of enhancers and CGIs have revealed the intricate network of interactions underlying key biological processes, including the cell cycle, differentiation, and the maintenance of cellular identity (Panigrahi & O'Malley, 2021). Moreover, advances in transcriptomics, particularly the development of techniques such as single-cell RNA sequencing (scRNA-seq), have opened new horizons by enabling high-resolution, cell-by-cell analyses of transcriptional landscapes (Kulkarni et al., 2019).

The project presented here arises from this context as an initial approach to studying the potential role of CpG islands in the reactivation of genes after mitosis in conjunction with active enhancers. To achieve this goal, mouse embryonic stem cells (mESCs) were selected as the *in vitro* model system. The *Gata6* gene is transcriptionally inactive in mESCs, which made it an appropriate choice for exposure to different genetic constructs of regulatory elements, thereby enabling a more meaningful assessment of its transcriptional activation (Coux et al., 2024). For the design of these regulatory elements, two active enhancers (one associated with the *Klf2* gene and another with the *Tbx3* gene) were selected, together with the CpG island linked to the *Sox1* gene, which had been previously employed by the research group (Pachano et al., 2021). These elements were used to generate distinct constructs that will be inserted approximately 100 kb upstream of *Gata6*, followed by transcriptional analyses via RT-qPCR in asynchronous cells.

This strategy aims to establish the foundational bases, regulatory elements, and experimental protocols required to facilitate future investigations within a mitotic context.

Objectives

In this project, we hypothesize that enhancers associated with CpG islands play a pivotal role in the reactivation of genes upon mitotic exit, thereby preserving the cellular identity of daughter cells. To test this, we will expose an inactive gene (*Gata6*) to different combinations of regulatory elements, expecting, according to our hypothesis, that the combination of an active enhancer linked to a CpG island will exhibit the strongest transcriptional activation response. The specific objectives of the project are:

1. **Construction of regulatory element vectors**

Two constructs were designed for each of the selected enhancers (Klf2 TF and Tbx3 TF). The primary construct will consist of the corresponding enhancer sequence fused to the CpG island (CGI) of the *Sox1* gene. In parallel, a construct harbouring four tandem repeats of the corresponding enhancer sequence (Multi Transcription Factor Binding Site, MTFBS) will be generated. As an additional control, the CGI alone will be cloned. All constructs will be flanked by homology arms to facilitate CRISPR-Cas9-mediated insertion into target cell lines.

2. **Generation of transgenic cell lines**

As the baseline, we will use a previously generated transgenic mESC line (*Gata6*ko11.1 Hemizygous). This line is hemizygous for the *Gata6* gene, so our goal is to integrate the designed constructs into the remaining allele using CRISPR-Cas9, targeting an insertion site located ~100 kb upstream of the *Gata6* promoter.

3. **Preliminary analysis of *Gata6* expression in asynchronous cells**

Following the selection of successfully engineered cell lines carrying the regulatory constructs, all lines will be cultured in parallel and subjected to transcriptional analysis by RT-qPCR under asynchronous conditions. This will establish a baseline expression profile to inform subsequent studies in mitotic cells.

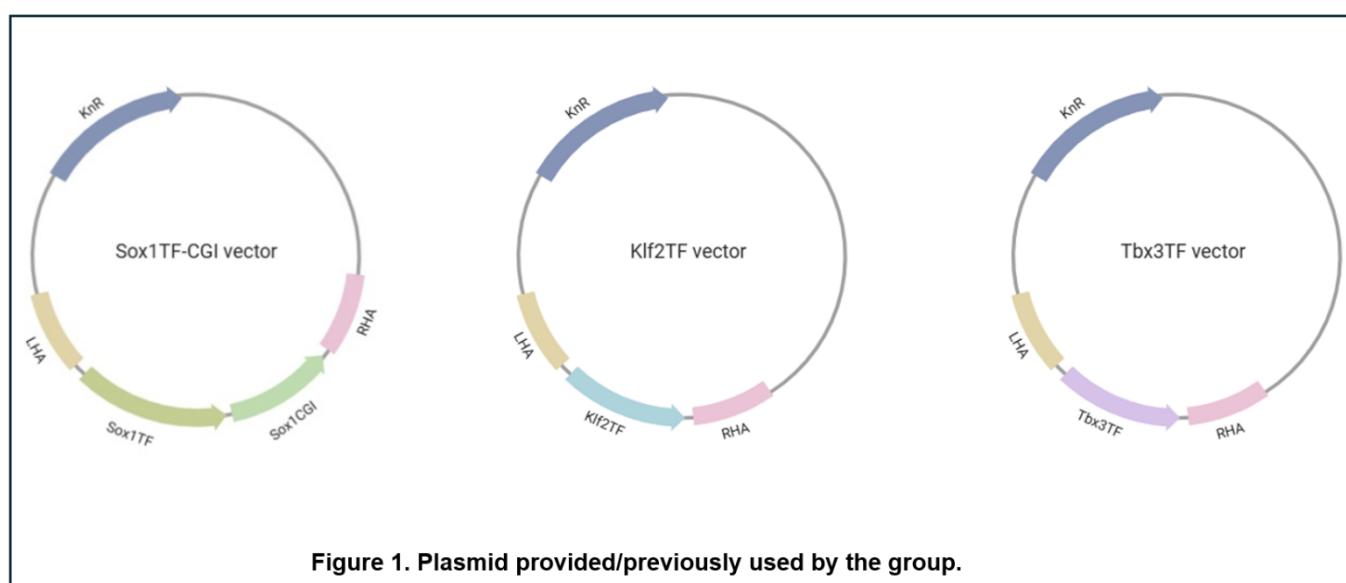
Methodology

1) Construction of regulatory element vectors.

1.1 Insert amplification and purification.

The inserts required for the construction of our vectors were amplified from plasmids previously generated and used by our laboratory team (Fig. 1). Amplification was carried out by PCR using the corresponding primer pairs for each case (Fig. 2). Reactions were performed with Kapa HiFi DNA polymerase (Fig. 3), following the manufacturer's protocol, with <1 µg of plasmid DNA as template. The annealing temperature and extension time were optimized individually for each primer set, which had been evaluated in silico using OligoEvaluator (Thermo Fisher Scientific). The final reaction volume was 25µL. For the construction of the MTFBS vectors, each insert was amplified three times using different primer combinations designed to enable subsequent fragment assembly.

The size and integrity of the amplified inserts were verified by agarose gel electrophoresis (1.5% agarose, 100 V, 30 min). From this point forward, all agarose gel electrophoresis analyses will utilize NZYDNA Ladder III (LdrIII; 200–10,000 bp) or NZYDNA Ladder VII (LdrVII; 100–3,000 bp) as molecular weight markers. To prepare the fragments for restriction digestion, PCR products were purified using the PCR clean up kit from Biotools (#21.202). The final washing step of the protocol was performed twice to maximize purity of the inserts.



1.2 Digestion and purification of vectors and inserts.

Digestion reactions were performed using different restriction enzymes (Fig. 3). Since all digestions involved double-enzyme combinations, enzymes were selected based on buffer and incubation temperature compatibility, using the Thermo Scientific Double Digest Calculator. Ultimately, all digestions were carried out in CutSmart 10× buffer (Thermo Fisher Scientific), with overnight incubations at 37 °C. Enzyme volumes were adjusted to provide a total of 10 U per reaction. The final reaction volume was 50 µL.

For digested vectors, linearized products were resolved by electrophoresis on a 1% agarose gel, and the corresponding bands were excised and purified using the PCR clean up kit from Biotools (#21.202). Digested inserts, in contrast, were directly purified with the same kit without requiring gel verification.

1.3 Ligation and transformation into *E. coli* DH5α.

Ligation reactions were performed using T4 ligase (Fig. 3). Each reaction had a final volume of 20 µL and was incubated for 8 h at 21 °C, followed by overnight incubation at 12 °C. A 5:1 vector-to-insert molar ratio was employed, and calculations for each reaction were verified using the NEBioCalculator Ligation Tool (New England Biolabs).

Following ligation, the recombinant plasmids were introduced into *E. coli* DH5α competent cells. Transformation was carried out using a standard heat-shock protocol. Briefly, competent cells stored at –80 °C were thawed on ice and mixed with 1–5 µL of ligation product per 50 µL of competent bacteria. After 30 min on ice, cells were heat shocked at 42 °C for 45 s and immediately transferred to ice for 2 min. Subsequently, 950 µL of LB medium (without antibiotic) pre-warmed to room temperature was added, and the suspension was incubated at 37 °C for 30 min with shaking at 250 rpm. Transformed cells were plated onto LB-agar plates containing kanamycin as selective antibiotic and incubated overnight at 37 °C (~16 h)

1.4 Colony PCR and minipreps for sequencing.

From each transformation, five isolated colonies grown on kanamycin-containing plates were selected. Individual colonies were picked with sterile pipette tips, resuspended in 50 µL of Milli-Q water, and used as template for colony PCR. Several primer combinations were employed depending on the construct to be verified (Fig. X). Colony PCR reactions were performed with NZYTech DNA polymerase in a total volume of 25 µL, and PCR products were analyzed by agarose gel electrophoresis (1.5% agarose) to confirm the expected fragment sizes.

Positive colonies were subsequently cultured by inoculating 20 µL of the bacterial suspension into 5 mL of LB medium supplemented with kanamycin. Cultures were incubated overnight at 37 °C with agitation at 200 rpm. From each overnight culture, 0.5 mL was mixed with 0.5 mL of glycerol and stored at –80 °C as a glycerol stock, while the remaining 4.5 mL were processed for plasmid purification using the NZYTech Miniprep Kit (#MB01002). Purified plasmids were subjected to Sanger sequencing (<https://eurofinsgenomics.eu/>) to confirm the correct assembly of the constructs.

Name	Sequence 5'-3'	Lenght	Tm (°C)	Enzyme restriction site added?
18-Klf2-enh-FW-1	ATTATTACTAGTCCAACATGTTCTAAAGTGGTC	35	63,3	Yes (SpeI)
24-Tbx3-enh-FW-2	ATTATTACTAGTGTTCATTCTAGGCCAGAGTCG	34	62,7	Yes (SpeI)
149-Klf2-Clal-RV-1	ATTATTATCGATGTTTGTGGGATAGGGGAC	31	63	Yes (Clal)
150-Tbx3-Clal-RV-1	ATTATTATCGATGAAATCTGTAATCCAGTACCACC	32	63,4	Yes (Clal)
26-LHA-100kb-FW	CTTGATGAGGAGGAGACCTTG	21	62,8	No
27-RHA-100kb-RV	CCCTTGATACTGAATCCATCTC	22	61,3	No
137-Amp-repair-Gata6-F	CAATGTCCTTGCCAAATCCT	20	63,8	No
138-Amp-repair-Gata6-R	AATCCCCTGTGGCTCGTTTCT	22	68,7	No
139-Klf2-RV	CGAGGCTTGACCACCTTTTAGG	21	64,7	No
141-Tbx3-RV	CGACTCTGGCCTAGAAAGAAC	21	62,7	No
145-Klf2-FW-2	GAAGAATCTGCACTTAACCCG	21	62,5	No
147-Tbx3-FW-2	CCAATGAATCCATCAGCC	19	66,3	No
151-Klf2-MTF-Clal-F	ACAAACATCGATACGGCAATC	21	63,6	No
152-Klf2-MTF-BamHI-F	CACAAACAGTTTGGATCGGTG	21	65,3	No
153-Klf2-MTF-XbaI-F	CACAAACGGTGTGAGAGTTG	20	65,2	No
154-Klf2-MTF-MluI-F	ACAAACGCAATATGTCATATCCG	23	64,4	No
155-Klf2-R	TGCTTTGTGCTCACTGCAC	23	65,1	No
156-Klf2-F	CCAACATGTTCTAAAGTGGTC	23	63,3	No
157-K/T-MTF-MluI-R	ACGCGTCGGATGACATATTG	22	65,5	No
158-Klf2-sp-Clal-F	ATTATTATCGATACGGCAATCAGTC	50	63,3	Yes (Clal)
159-Klf2-sp-BamHI-MluI-R	ATTATTACGCGTATTATTGGATCCAGCCACCGATCCAAAC	60	63	Yes (BamHI+MluI)
160-Klf2-BamHI-F	ATTATTGGATCCCAACATGTTCTAAAGTGGTC	35	63,3	Yes (BamHI)
161-Klf2-sp-XbaI-MluI-R	ATTATTACGCGTATTATTCTAGAGCGACCAACTCTCAACACC	62	63	Yes (XbaI+MluI)
162-Klf2-XbaI-F	ATTATTCTAGACCAACATGTTCTAAAGTGGTC	35	63,3	Yes (XbaI)
163-Klf2-MluI-R	ATTATTACGCGTCGGATATGACATATTGCG	48	63	Yes (MluI)
164-Tbx3-MTF-Clal-F	GATTTTCATCGATACGGCAATC	21	63,1	No
165-Tbx3-MTF-BamHI-F	CAGATTTCAGTTTGGATCGGTG	22	65,4	No
166-Tbx3-MTF-XbaI-F	CAGATTTCGGTGTGAGAGTTG	22	63,5	No
167-Tbx3-MTF-MluI-F	GATTTCCGAATATGTCATATCCG	23	63,9	No
168-Tbx3-F	GGTCAGCCATGAATTCAG	20	65,5	No
169-Tbx3-R	GACTACGGAGTTAGTTCAAGGCC	23	63,9	No
170-Tbx3-sp-Clal-F	ATTATTATCGATACGGCAATCAGTC	48	62,7	Yes (Clal)
171-Tbx3-sp-BamHI-R	ATTATTACGCGTATTATTGGATCCAGCCACCGATCCAAAC	65	63,4	Yes (BamHI+MluI)
172-Tbx3-BamHI-F	ATTATTGGATCCGTTCACTTAGGCCAGAGTCG	33	62,7	Yes (BamHI)
173-Tbx3-sp-XbaI-MluI-R	ATTATTACGCGTATTATTCTAGAGCGACCAACTCTCAACACC	67	63,4	Yes (XbaI+MluI)
174-Tbx3-XbaI-F	ATTATTCTAGAGTTCACTTAGGCCAGAGTCG	33	62,7	Yes (XbaI)
175-Tbx3-MluI-R	ATTATTACGCGTCGGATATGACATATTGCG	53	63,4	Yes (MluI)
176-Klf2-MTF-XbaI-R	AGGAACATGTTGGTCTAGAGCG	22	64,3	No
177-Tbx3-MTF-XbaI-R	ATGAACCTAGAGCCGACCACTC	23	62,6	No
178-P60-CGI-Endika	ACGAGTAACGCCAGCCTAGA	20	63,8	No
179-P59-CGI-Endika	GGAGAAGGAGGTAGCGAAG	20	64	No
180-MluI-CGI-R	ATTATTACGCGTAGACAAAGAGCGTCGCG	40	63,2	Yes (MluI)
181-SpeI-CGI-F	ATTATTACTAGTATTATCCGCCCTATTGAGC	31	60,5	Yes (SpeI)
182-Klf2-MTF-Clal-R	GATTGCCGTATCGATGTTGT	21	63,6	No
183-Klf2-MTF-BamHI-R	CACCGATCCAAACTGTTTGTG	21	65,3	No
184-Tbx3-MTF-Clal-R	GATTGCCGTATCGATGAAATC	21	63,1	No
185-Tbx3-MTF-BamHI-R	CACCGATCCAAACTGAAATCTG	22	65,4	No

Figure 2. Primers used during cloning and genotyping workflows. Primers 18 to 147 were provided by the laboratory, while primers 147 and below were specifically designed for this project.

Ezyme	Reference
SpeI-HF	#R3133S
MluI-HF	#R3198L
Clal	#R0197S
XBA1	#R0145L
bamh1-HF	#R36386
T4 Ligase	#EL0016
NZYTaQ II 2x Green Master Mix	#MB358
Kapa HiFi HotStart ReadyMix	#KK2601
GoTaq® DNA Polymerase	#M3001

Figure 3. Enzymes employed during cloning and genotyping workflows

2) Generation of transgenic cell lines

2.1 Parental cell line and cell culture conditions

A transgenic mouse embryonic stem cell (mESC) line hemizygous for the *Gata6* gene had been previously generated in our laboratory. In the remaining allele, a premature stop codon was introduced. This line, referred to as E14 *Gata6*KO hemizygous 11.1, was used as the parental line for the present study. Consequently, all modifications associated with *Gata6* required targeting of only one allele.

The parental line and all subsequently generated transgenic lines were cultured on gelatine-coated plates in KnockOut DMEM (Life Technologies, 10829018) supplemented with 15% fetal bovine serum (FBS; Life Technologies, 10082147) and leukemia inhibitory factor (LIF).

2.2 CRISPR-Cas9 Vector and Donor Amplification for Transfection

To introduce our constructs into the *Gata6*^{KO} hemizygous 11.1 cell line, the CRISPR-Cas9 system was employed. The components of this system, including the guides required for insertion 100 kb upstream of the *Gata6* promoter, are contained within the PX330A vector (guides 25+26) (Fig.5), previously constructed in the laboratory (Pachano et al, 2021).

The different constructs cloned into plasmids were used as donor templates for transfection. Each donor carries the sequence to be inserted, flanked by homology arms (LH/RH) required for homology-directed knock-in. These donors were amplified by PCR using Kapa High-Fidelity polymerase (Fig. 3).

For MTFBS donors, due to the presence of multiple repetitive sequences, PCR amplification was avoided; instead, the donor was excised directly from the vector. In all cases, donors were purified using the QIAquick PCR Purification Kit, leaving them ready for transfection.

2.3 Transfection of Mouse Embryonic Stem Cells *Gata6*KO 11.1 hemizygous

Mouse embryonic stem cells (mESCs *Gata6*KO 11.1 hemizygous) were transfected with the sgRNA–Cas9-expressing vector. Together with the knock-in donor using Lipofectamine (Thermo Scientific, L3000001). The protocol was previously optimized in the laboratory and proceeded as follows:

At least one hour prior to transfection, culture medium was aspirated and replaced with fresh medium. Sufficient wells of 12-well plates coated with gelatine were prepared in advance for transfection. Next, proceed to prepare the mixes required for transfection.

Mix A: 50 μ L of Opti-MEM containing <1 μ g DNA, 250ng PX330A vector and 2 μ L of Lipofectamine 3000 reagent (per transfection).

Mix B: 50 μ L of Opti-MEM containing 1 μ L of Lipofectamine L3000001 (per transfection).

Mix A was added to Mix B gently and incubated for 15 minutes at room temperature to allow complex formation. Cells were washed, trypsinized, and counted. 150,000 cells per well were plated in approximately 100 μ L of medium in each well of a 12-well plate. The transfection mixture from step 3 was added dropwise to the plated cells. Additional Opti-MEM was added to reach a total volume of 500 μ L per well. Cells were incubated with the mixture for 1–2 hours.

After incubation, 1 mL of standard stem cell medium (S+L medium) was added to each well, and cells were allowed to recover overnight. Different vector:donor ratios were tested in multiple replicates to optimize transfection efficiency, including 250:250, 250:100, and 250:50 ng per well. The following day, transfection efficiency was evaluated by fluorescence microscopy. Puromycin (selection antibiotic; the vector carries both the CRISPR-Cas9 machinery and puromycin resistance gene) was added at 1.8 μ L/mL of medium. Cells were incubated for 24 hours in the presence of puromycin. If live cells were observed in the negative control (non-transfected cells), selection was repeated for an additional 24 hours to ensure complete elimination of non-transfected cells.

2.4 Genotyping of Populations and Single-Cell Clones

Genotyping was performed by PCR using various primer combinations designed to confirm the correct insertion of different regions of our construct (Fig. 4). In most cases, NZYTech Green polymerase was used; however, for amplification of CpG island regions, GoTaq polymerase was employed due to its higher efficiency in GC-rich sequences.

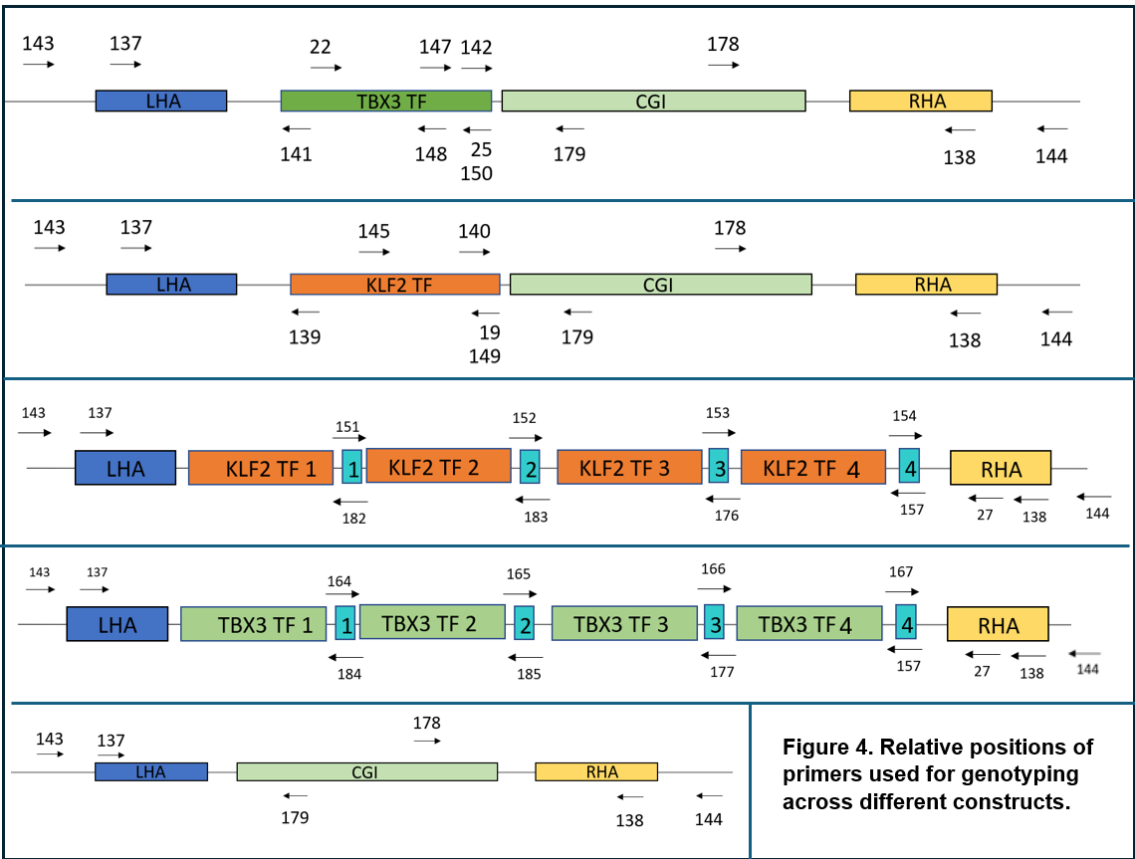
For preliminary genotyping of total transfected populations, DNA was extracted using Proteinase K according to the following protocol (Q&D protocol):

Detach cells from the plate and centrifuge. Remove the supernatant and add to the pellet: 50 μ L Milli-Q water, 50 μ L 2 $\times$ lysis buffer, and 2 μ L Proteinase K. Last, incubate at 65°C for 10 min, followed by 98°C for 3 min.

The resulting DNA was used to confirm the presence of positive cells in the populations. Single-cell seeding was then performed in 96-well plates for subsequent individual genotyping. Parallel seeding with 10 cells per well was also carried out to allow genotyping of small populations in cases where single-cell clones did not yield positive results; positive 10-cell populations were then expanded to single cells in 96-well plates.

After single-cell seeding, wells exhibiting more than one colony or no colonies after 10 days were discarded. Once monoclonal populations were selected, DNA was extracted either following the previously described protocol or using the NZY Tissue gDNA Isolation Kit to obtain higher purity samples.

For genotyping of clones, the same protocols applied to the populations were used, along with additional primer combinations to comprehensively characterize the quality of the insertion, including potential inversions, duplications, or other rearrangements (Fig. 4).



Name	Sequence 5'-3'
25-Gata6-guide	CACCCTCATAACTCACGACGCTCC
26-Gata6-guide	AAACGGAGCGTCGTGAGTTATGAG

Figure 5. Guide RNAs cloned into the PX330A vector, which encodes the CRISPR-Cas9 system and a puromycin resistance gene for selection (Pachano et al, 2021)

3) Preliminary analysis of *Gata6* expression in asynchronous cells

To evaluate the impact of introducing various genetic constructs into our cell lines, we performed quantitative PCR (qPCR) to measure the expression levels of the *Gata6* gene. The analysis included newly generated cell lines alongside a negative control (*Gata6*ko11.1 Hemizygous), previously characterized and lacking regulatory element integration. Additionally, we included laboratory-established lines containing only enhancer elements (Tbx3 TF and Klf2 TF) without CpG island association.

All cell lines were cultured in parallel until reaching confluence in a 12-well plate. Upon achieving optimal growth, total RNA was extracted using the NZY Total RNA Isolation Kit (#MB13402) and resuspended in 50 µL of nuclease-free water. RNA concentration was quantified using a NanoDrop spectrophotometer. To eliminate potential genomic DNA contamination, samples were treated with DNaseI using the TURBO DNA-Free Kit (AM1907).

Complementary DNA (cDNA) synthesis was performed using the NZYtech First-Strand cDNA Synthesis Kit (MB12502), following the manufacturer's protocol. All procedures were conducted on ice using nuclease-free tubes to preserve RNA integrity.

qPCR was conducted in a 96-well plate format using specific primers targeting *Gata6*, along with primers for two housekeeping genes (*Eef1a* and *Hprt*), the expression level of this housekeeping genes will be used to normalize the expression levels of *Gata6*. Two primers were targeting each gene, giving a total of six (Fig.6). To validate expression levels across clones. each clone was analysed in triplicate with each primer to calculate average expression values and to exclude samples with aberrant melting temperatures.

Following data filtration, expression levels of *Gata6* were compared across the different cell lines, highlighting differential gene expression patterns associated with the specific construct present in each clone.

Name	Sequence 5'-3'
Eef1a1-F	AGCGTAGCCAGCACTGATTT
Eef1a1-R	TAGACGAGGCAATGTTGCTG
Hprt1_F	CAAGGGCATATCCAACAACA
Hprt1_R	GCCCCAAAATGGTTAAGGTT
Gata6-F	CTACACAAGCGACCACCTCA
Gata6-R	TGTAGAGGCCGTCTTGACCT

Figure 6. qPCR primers used in the expression analysis of *Gata6*

Results

1) Construction of regulatory element vectors.

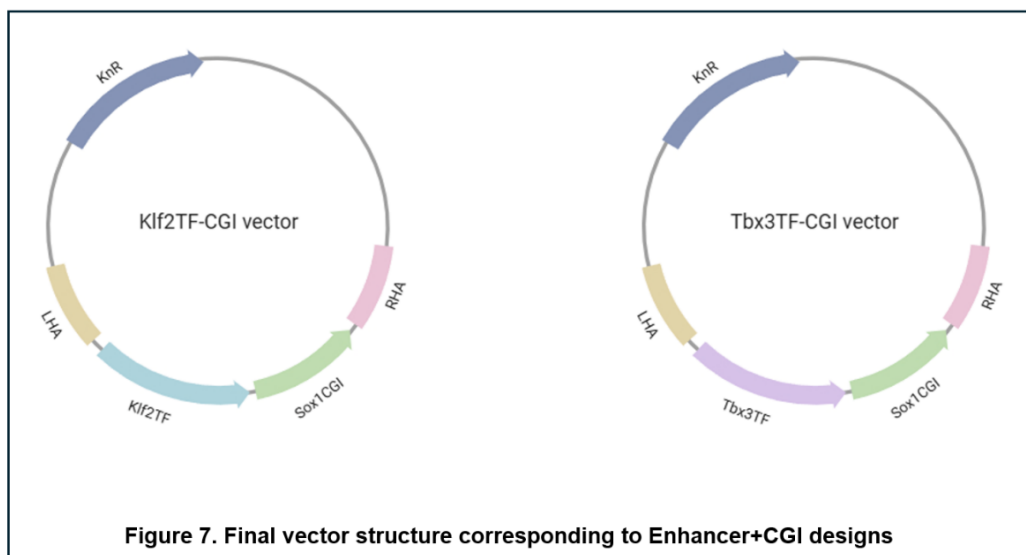
1.1 Vectors containing Enhancers and CGI regions

The enhancer sequences associated with the transcription factors Klf2 and Tbx3 were amplified by PCR using primer pairs 18+149 and 24+150, respectively (Fig.2). The templates used for amplification were the plasmids Klf2TFvector and Tbx3TFvector (Fig.1).

The resulting amplicons, along with the Sox1TF+CGI vector (Fig.1), were digested with the restriction enzymes *SpeI* and *Clal* (Fig.3). This digestion step served two purposes: it prepared the inserts for ligation and simultaneously removed the Sox1-associated enhancer fragment from the Sox1TF+CGI backbone.

Consequent digested products were ligated using T4 DNA ligase (Fig.3) at a molar insert-to-vector ratio of 5:1, yielding the desired constructs containing either the Klf2 or Tbx3 enhancer sequences in combination with the CGI region. The ligation products were transformed into *E. coli* DH5 α competent cells via heat shock, and transformants were selected on kanamycin-containing agar plates.

For colony screening, PCR was performed using primer pairs 26+149 for Klf2 and 26+150 for Tbx3 (Fig2, Fig.4), with NZYtaq (Fig.3). Positive colonies were cultured, and plasmid DNA was extracted. The integrity of the constructs was confirmed by Sanger sequencing. Sequencing results validated the correct assembly of the enhancer-CGI constructs. Glycerol stocks of the confirmed recombinant *E. coli* strains were prepared and stored for future use.



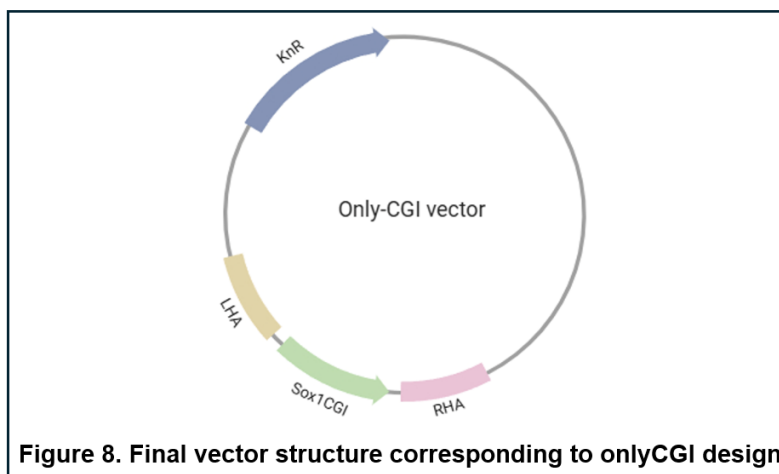
1.2 Vector containing only the CGI region

The CpG island (CGI) of Sox1 was amplified by PCR using the previously constructed Klf2+CGI vector as template (Fig.7). Primer pair 180+181 (Fig.2) was used to specifically amplify the CGI fragment.

Following purification, the PCR product and the Klf2+CGI vector backbone were digested with MluI and SpeI to enable directional ligation. This digestion step also removes the entire Klf2+CGI fragment from the plasmid, leaving the vector backbone empty and suitable for the insertion of the Sox1 CpG island as a standalone regulatory element. The digested insert and vector were ligated and subsequently transformed into *Escherichia coli* DH5 α .

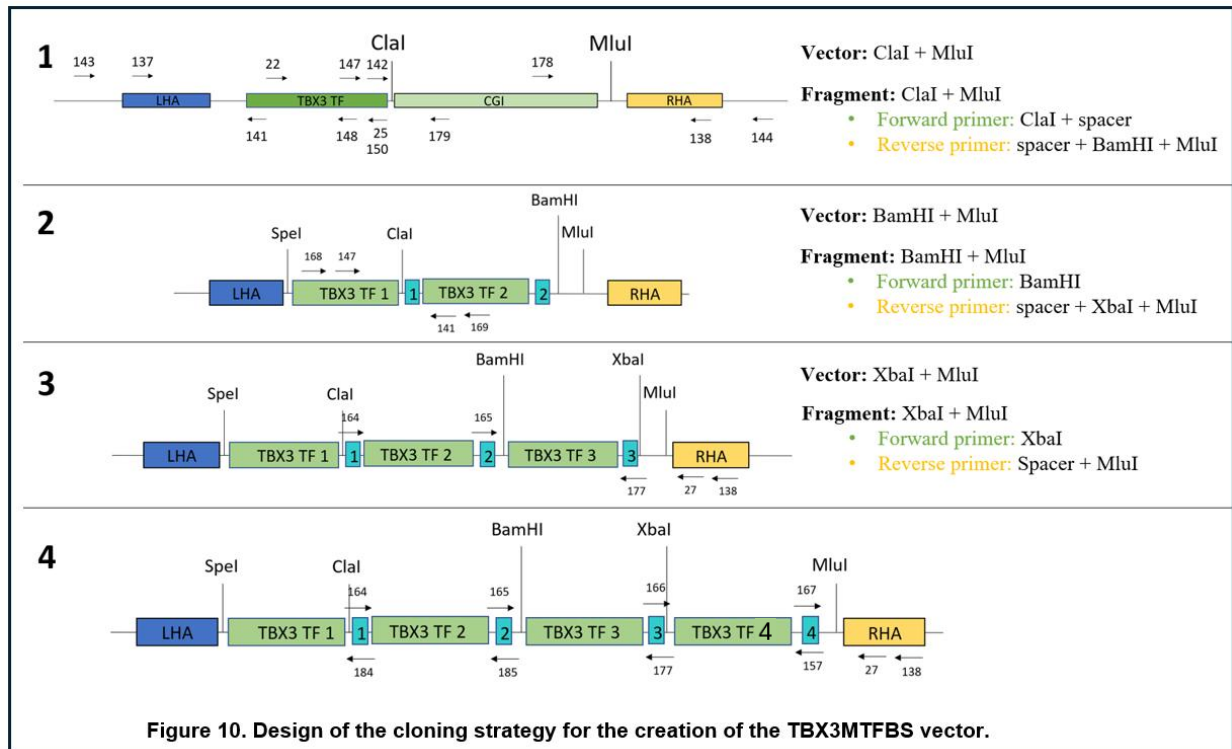
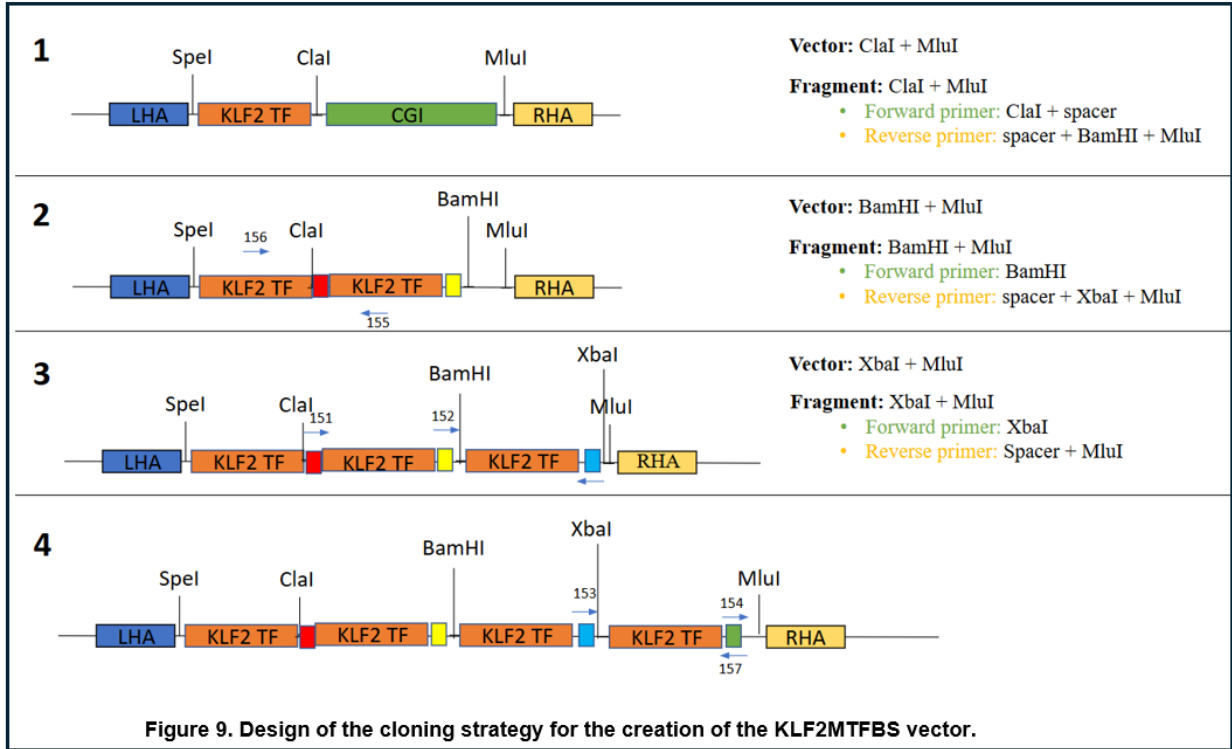
Colony screening was performed by PCR using primers 178+138(Fig.2, Fig.4) and NZYtech Green Master Mix(Fig.3). Positive colonies were cultured, and plasmid DNA was extracted.

The integrity of the construct was confirmed by Sanger sequencing, which validated the correct assembly of the Vector OnlyCGI. Glycerol stocks of the transformed bacterial clones were prepared and stored as part of the plasmid library.



1.3 Vectors containing four tandem repeats of the enhancer region (MTFBS).

To construct the MultiTFBS-KLF2 and MultiTFBS-TBX3 vectors, specific primers were designed to amplify compatible fragments for sequential assembly (Fig.2). This strategy aimed to generate vectors containing four tandem repeats of each corresponding enhancer sequence (Fig.9, Fig.10)



The fragments were amplified from the vectors generated in the initial step (VectorKLF2+CGI and VectorTBX3+CGI) using the following primer pairs.

- KLF2 inserts:

k2 (insert 2): primers 158 + 159

k3 (insert 3): primers 160 + 161

k4 (insert 4): primers 162 + 163

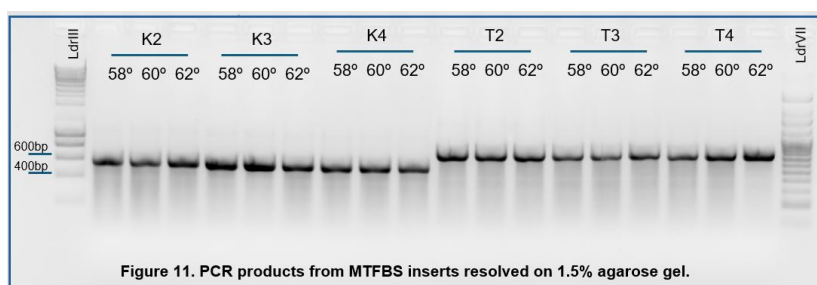
- TBX3 inserts:

t2 (insert 2): primers 170 + 171

t3 (insert 3): primers 172 + 173

t4 (insert 4): primers 174 + 175

Since these primers had not been previously validated, PCR reactions were performed using three different annealing temperatures, using Kapa polymerase. Amplification was successful under all tested conditions. The expected band sizes are 513 bp for the KLF2 inserts and 651 bp for the TBX3 inserts, not accounting for the slight variation introduced by the specific primer overhangs, which depend on the restriction enzyme recognition sites intended to be added to the amplicons.

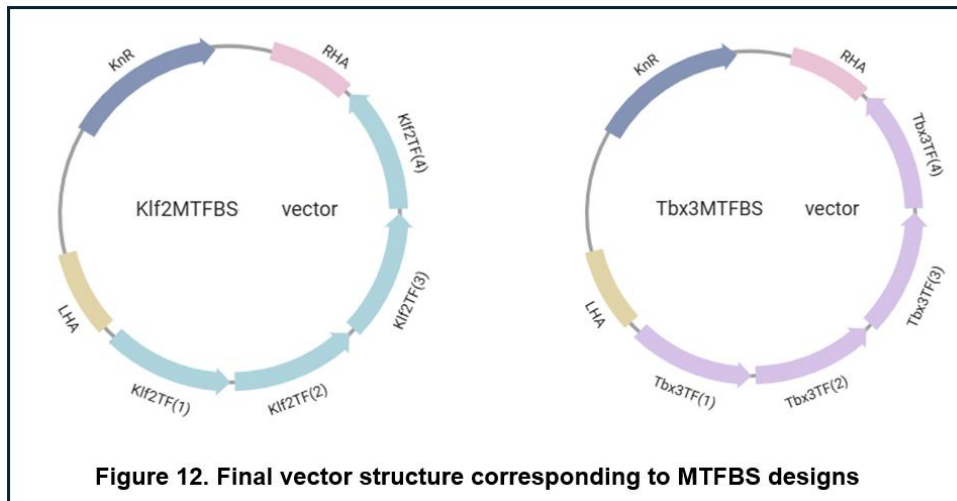


The PCR products were purified and used as inserts for the assembly of the MultiTFBS vectors. Each primer pair was designed to introduce specific restriction sites required for the sequential cloning strategy (see Fig. 9 and Fig. 10). After each assembly step, the resulting construct was transformed into *E. coli* DH5α and verified by Sanger sequencing prior to proceeding with the next fragment.

The procedure was essentially the same for both selected enhancers. Starting from the Enhancer+CGI construct, the CGI region was excised via digestion with ClaI and MluI, and replaced with the corresponding fragment 2, digested with the same enzymes. For fragment 3, both the insert and the intermediate vector (containing two verified enhancer copies) were digested with BamHI and MluI.

Fragment 4 was cloned by digesting both the insert and the vector, now containing three confirmed enhancer copies, with XbaI and MluI.

This strategy resulted in final constructs lacking the CGI region but containing four tandem repeats of the enhancer sequence: KLF2MTFBS vector and TBX3MTFBS vector. Finally, the correct assembly of the final vectors was confirmed by Sanger sequencing. Bacterial glycerol stocks were prepared and stored for future use.



2) Generation of transgenic cell lines

The results presented below focus on the characterization of the transgenic lines that were ultimately selected for inclusion in the Gata6 expression analysis via RT-qPCR. Although several additional lines were generated throughout the course of the project, these were excluded from the expression study for various reasons.

A complete list of all transgenic lines generated will be provided in a summary table (Fig.27). All lines, regardless of their inclusion in the expression analysis, have been cryopreserved for potential future characterization or application in new experimental contexts.

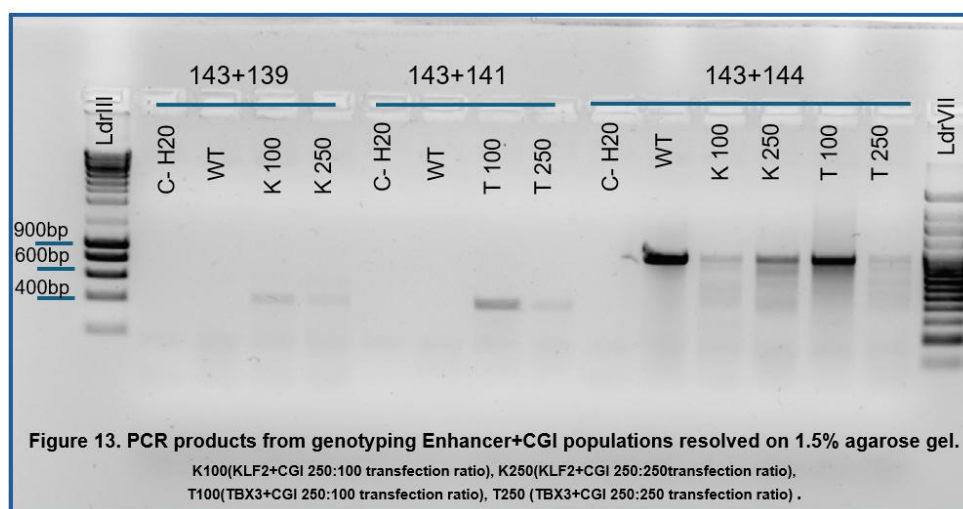
Accordingly, the genotyping PCR results shown here aim to provide the clearest possible overview of the characterization status of the selected lines, even if this includes data from lines that were eventually discarded.

2.1 Insertion of Enhancer+CGI constructs.

Genotyping of the cell lines generated following transfection with the KLF2+CGI, and TBX3+CGI donor constructs was initially performed on the entire post-transfection population. In each case, genomic DNA from the parental line (Gata6KO11.1 hemizygous, referred to as WT) was included as a control. This approach allowed us to confirm the presence of at least one correctly integrated clone within the population and to optimize the primer pairs that would be used for subsequent genotyping (see Fig. 2 and Fig. 4).

The expected band sizes were 418 bp for KLF2+CGI (primers 143+139) and 408 bp for TBX3+CGI (primers 143+141). Primer pair 143+144 was used across all samples, as these primers anneal to genomic regions flanking the insertion site. In positive clones, the expected amplicon size exceeds the amplification capacity of standard PCR conditions, resulting in no product. In contrast, WT samples (both from the control and within the transfected populations) produce a 909 bp band, enabling the exclusion of false negatives due to poor DNA quality.

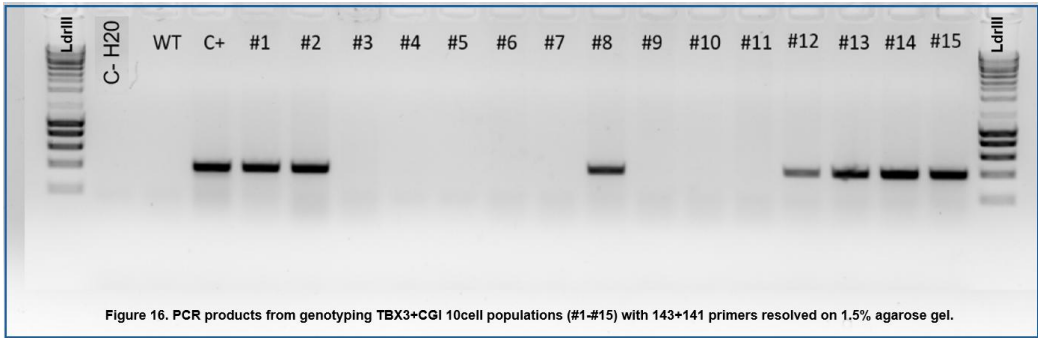
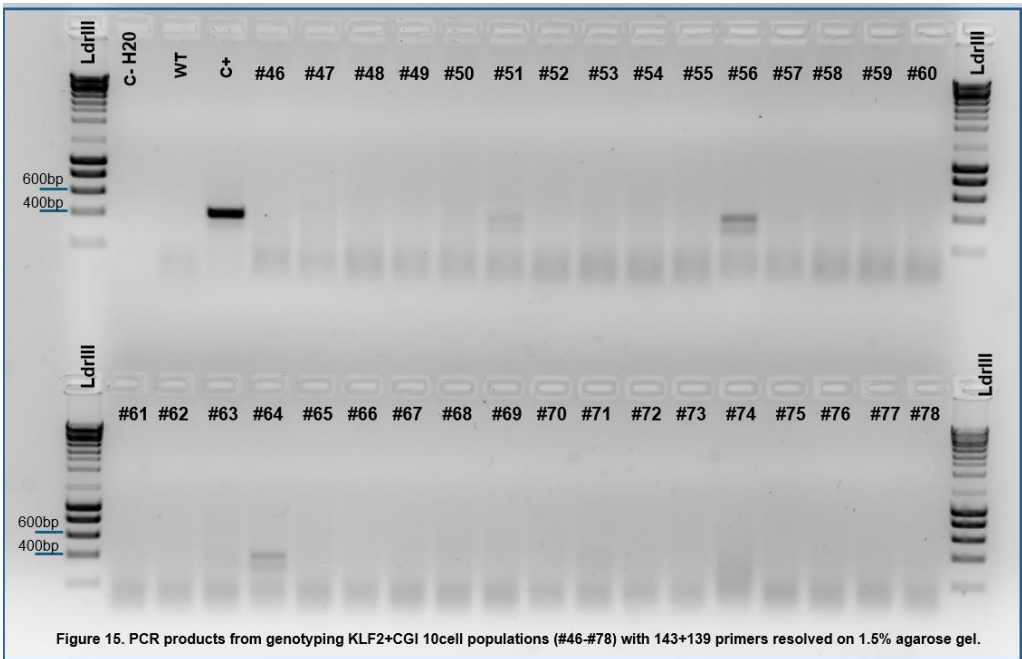
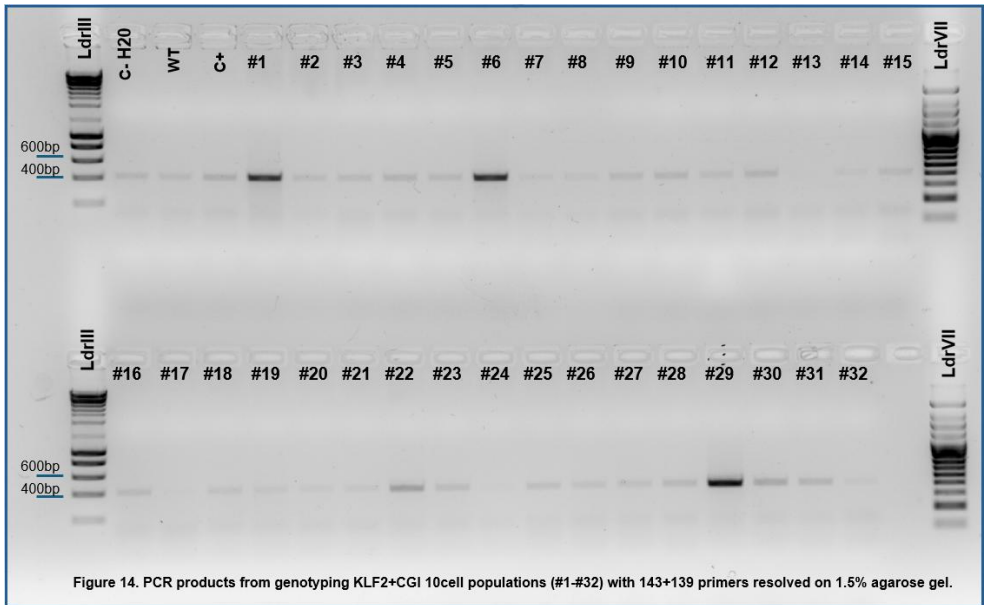
If necessary, transfections were repeated to obtain additional clones.



Following confirmation of genotyping efficiency and the presence of positively transfected cells, 10-cell and single-cell seeding was performed in 96-well plates. Initial screening focused on the 10-cell populations, while single-cell cultures were maintained as a reserve in case additional clones were needed.

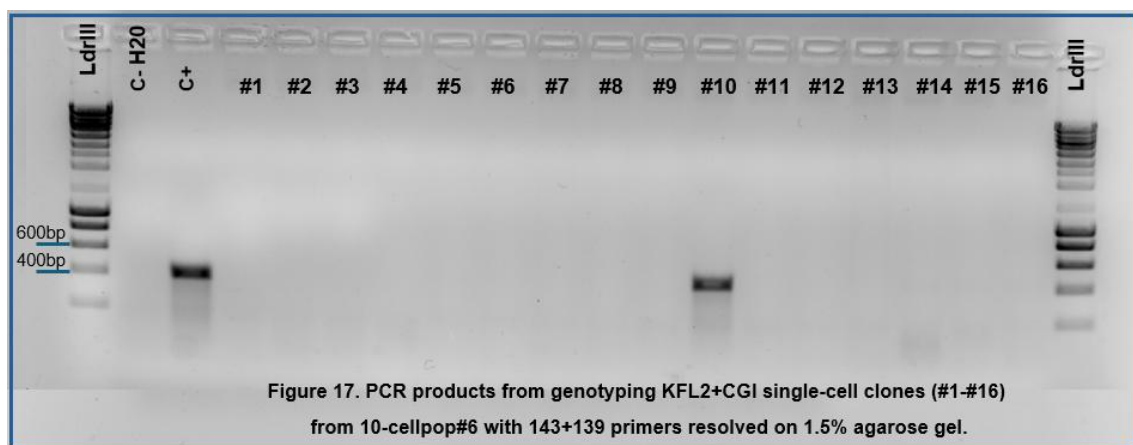
Genotyping of the 10-cell populations was carried out using previously validated primer pairs: 143+139 for KLF2+CGI and 143+141 for TBX3+CGI. These primers target the left flank of the inserted constructs, with the resulting amplicon going from upstream of the left homology arm to the corresponding enhancer sequence (Fig.4).

Band sizes expected are the same as in the total population analysis (Fig.13). Additionally, genomic DNA from previously generated cell lines containing the respective enhancer sequences was included as a positive control.



The following 10-cell populations were identified as positive: Klf2+CGI10cell-pop6, Klf2+CGI10cell-pop64, Tbx3+CGI10cell-pop2, and Tbx3+CGI10cell-pop14. These four populations were expanded until confluence in P6-well plates to proceed with clonal isolation.

Since the Klf2+CGI10cell-pop6 originated from an earlier transfection, its genotyping was performed in advance following a slightly modified protocol. After single-cell seeding, monoclonal populations were selected and subjected to genotyping PCR using the same primer pair applied to the original 10-cell population (primers 143+139). Clone #10 from this population (Klf2+CGIpop6#10) was confirmed as positive (Fig.17).



Thanks to the availability of the Kapa polymerase (previously used in the lab to amplify CG-rich sequences such as CGI) the clone Klf2+CGIpop6#10 was successfully amplified using primer pairs 145 + 138 and 143 + 149 (Fig. 4). The resulting fragments were submitted for Sanger sequencing, and subsequent alignment of the sequencing data confirmed that the construct had been correctly inserted. Therefore, this clone was directly classified as positive.

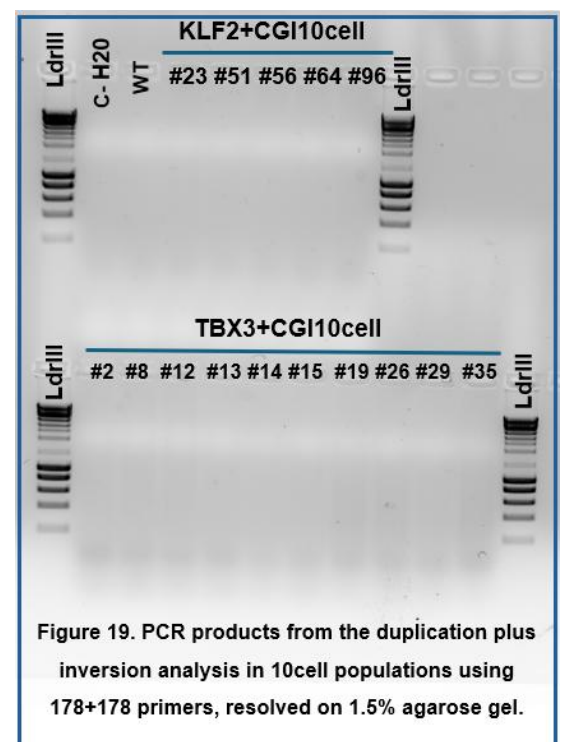
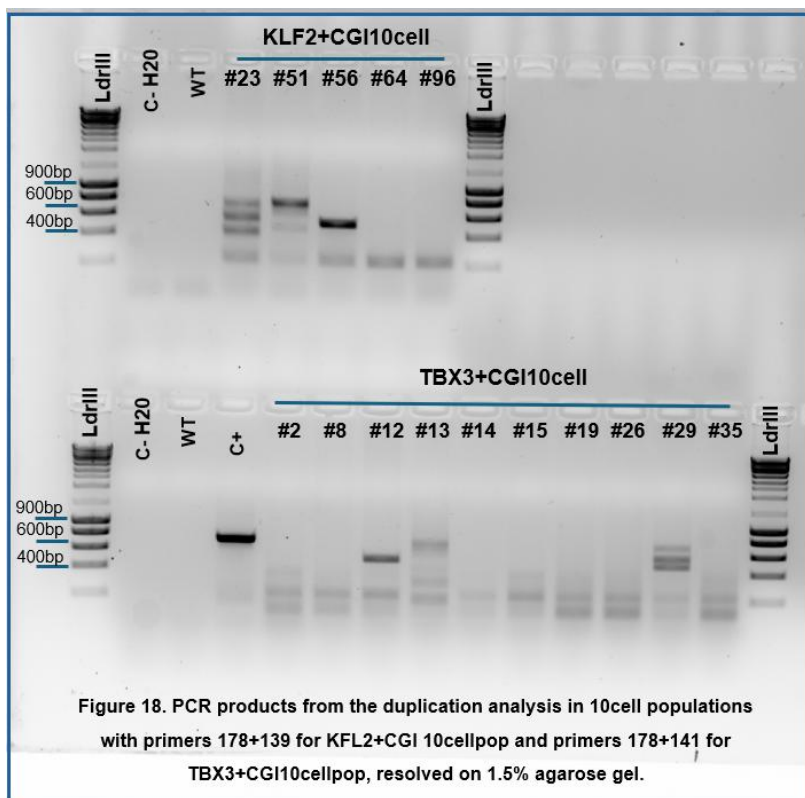
To continue genotyping Klf2+CGI10cell-pop64, Tbx3+CGI10cell-pop2 and Tbx3+CGI10cell-pop14, due to anticipating potential delays in sequencing, a preliminary screening of previously selected 10cellpop-positive clones was performed. The goal was to exclude those exhibiting duplications or duplication-inversion events prior to single-cell plating.

For this purpose, PCR was carried out using GoTaq polymerase and the following primer combinations:

- 178 + 139: Duplication analysis for Klf2+CGI
- 178 + 141: Duplication analysis for Tbx3+CGI
- 178 + 178: Detection of duplication plus inversion events in both constructs

We currently have only one available positive control for the *TBX3* constructs, consisting of genomic DNA from a cell line previously validated in the laboratory as duplication positive. In the duplication plus inversion analysis, only a single set of controls was included, as the primer pair used was identical across all samples.

Single-cell seeding in 96-well plates was performed using ten-cell populations previously screened and confirmed as negative for duplication and duplication-plus-inversion events: Klf2+CGI pop64, Tbx3+CGI pop2, and Tbx3+CGI pop14 (Fig.18, Fig.19).



The resulting single-cell populations were genotyped by PCR using the same primer pairs previously employed for the 10-cell populations: 143+139 for Klf2+CGI and 143+141 for Tbx3+CGI. These primer combinations are designed for partial genotyping on the right flank of the construct, spanning from the genomic DNA adjacent to the insertion site, through the right homology arm (RHA), and including a portion of the corresponding enhancer element (Fig.4) and The expected band sizes were 418 bp for KLF2+CGI (primers 143+139) and 408 bp for TBX3+CGI (primers 143+141) as always with this protocol.

Due to the unavailability of Kapa polymerase and the failure of multiple amplification attempts using alternative polymerases, a different approach was required. Instead of amplifying the full-length sequence of the clones for subsequent sequencing, characterization was continued via genotyping using various primer combinations. For this procedure, GoTaq polymerase was employed, as it demonstrated the ability to amplify small regions within the CpG island, enabling partial genotyping of the clones. However, this polymerase did not yield large amplicons suitable for sequencing. Consequently, a genotyping PCR targeting the right flank of the construct was performed. The resulting amplicon spanned part of the CGI and the right homology arm, using primer pair 178+144 for all clones, given that the CGI region is shared across all constructs regardless of the enhancer element. The expected band size is 708bp for all samples in this analysis. The Klf2+CGI pop6 clone #10 was included as a positive control, as its CpG island had already been sequenced. The results of both genotyping PCRs (targeting the enhancer side and the CGI side) were run together on the same gel to illustrate the final set of positive clones obtained (Fig. 20). Nevertheless, each PCR assay was independently optimized in advance.

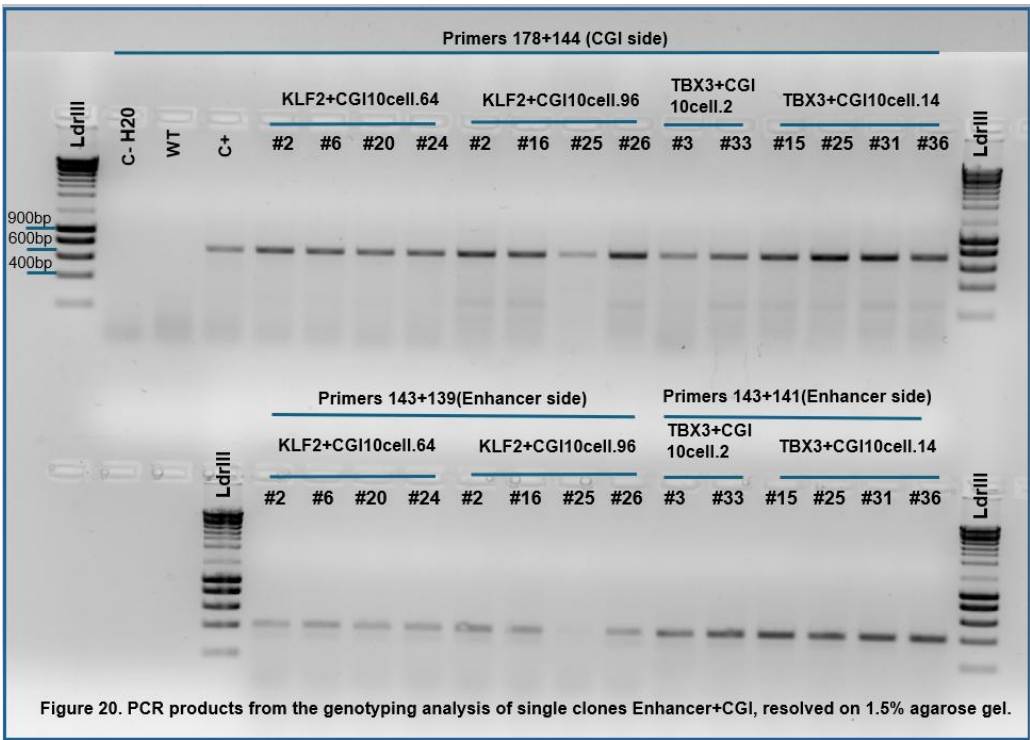


Figure 20. PCR products from the genotyping analysis of single clones Enhancer+CGI, resolved on 1.5% agarose gel.

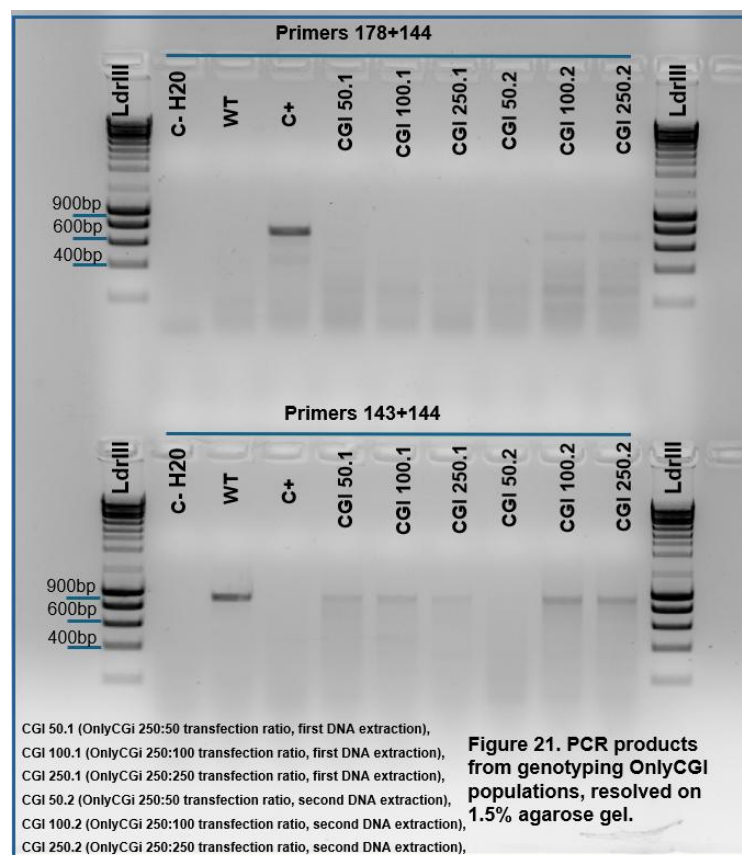
Klf2+CGI pop64 clone #6, Klf2+CGI pop64 clone #20, Tbx3+CGI pop2 clone #3, Tbx3+CGI pop14 clone #15, and Tbx3+CGI pop14 clone #31 were ultimately selected as positive clones. None of them exhibited duplication or duplication-plus-inversion events, and all tested positive in genotyping assays targeting both the enhancer side and the CGI side of the construct. Alongside Klf2+CGI pop6 clone #10, which had been previously sequenced, these clones were preserved for future analyses.

2.2 Insertion of onlyCGI construct.

To characterize cells transfected with the construct containing only the CpG island sequence of the *Sox1* gene, a strategy similar to previously employed for the enhancer+CGI constructs was followed. In this case, all PCR reactions were performed using GoTaq polymerase, due to the high GC content of the entire sequence.

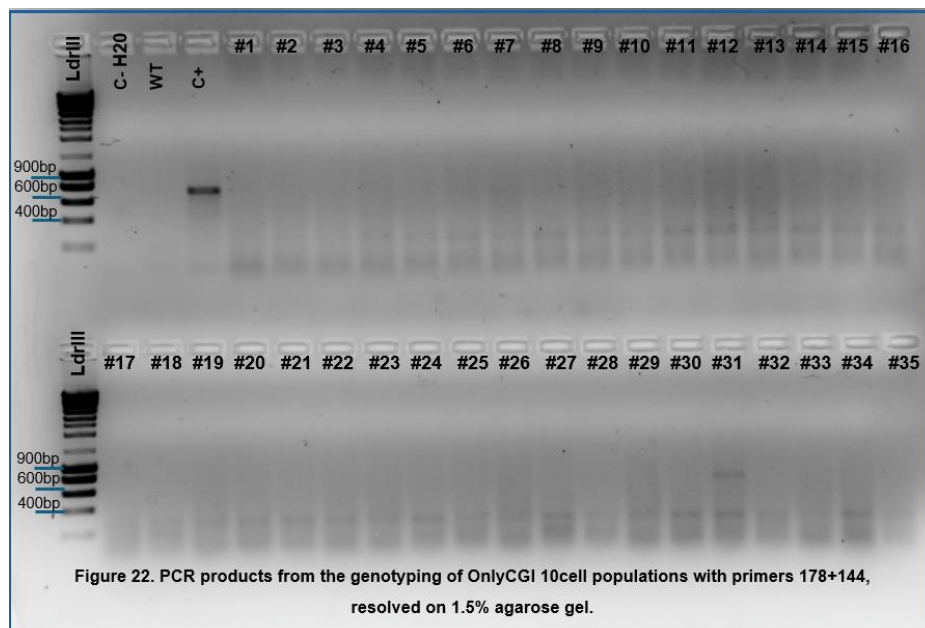
For transfection, the 50:250 ratio was also applied to improve transfection efficiency. However, as shown by the genotyping results of the total transfected populations, no improvement was observed (Fig.21). Additionally, DNA extraction from these populations had to be repeated, as initial genotyping PCR optimization attempts failed with the first DNA preparation.

As a positive control, the previously sequenced clone Klf2+CGI pop6 clone #10 was included once again. Although this clone does not contain the same construct, it does include the CpG island, making it suitable for validation purposes.

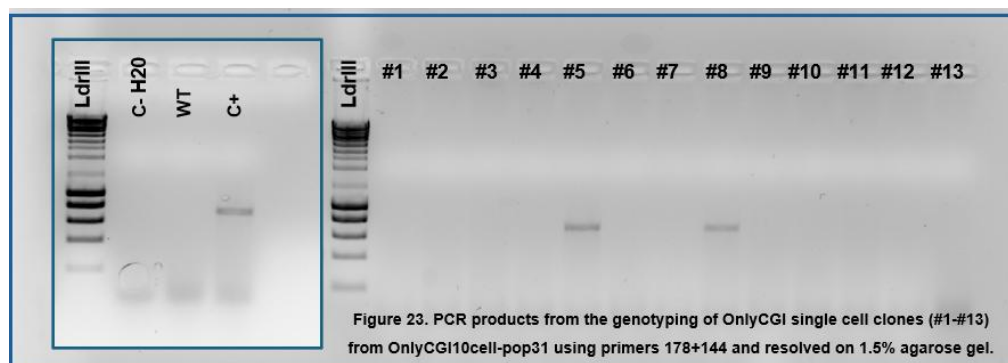


Initial analysis of the total transfected populations (Fig.21) was performed using primer pair 178+144 (Fig. 2, Fig. 4), previously used for genotyping the CGI flank of enhancer+CGI constructs, yielding an expected band size of 708 bp. Primer pair 143+144 was also included, as previously used in enhancer+CGI genotyping, to assess DNA quality. This combination produces a 901 bp band in wild-type cells, which exceeds the optimal size for efficient amplification in correctly transfected cells.

Once the presence of transfected cells was confirmed in our populations (Fig. 21), we proceeded with the seeding into ten-cell populations. These were subsequently re-analysed using primers 178 and 144. Only one population yielded a positive result, OnlyCGI10cell-pop31 (Fig. 22).



Starting from the population OnlyCGI10cell-pop31, single-cell seeding was performed to isolate monoclonal populations. The resulting clones were analysed by PCR using primer 178 and 144. The clone designated OnlyCGIpop31clone#8 tested positive (Fig. 23) and was therefore selected as the reference clone for subsequent gene expression assays.



2.3 Insertion of Multi-Transcription Factor Binding Site (MTFBS) construct

The genotyping approach for cell lines harbouring MTFBS constructs requires extensive analysis due to the large size of the insertion to be verified. Furthermore, the construct contains repetitive sequences (specifically, a tandem array of four copies of the corresponding enhancer) which complicates full-length amplification and subsequent sequencing.

To achieve optimal characterization of these lines, multiple primer combinations were designed for genotyping using NZYTaQ polymerase (Fig. 3). Initially, genotyping was performed on the total transfected populations (Fig. 24).

- For constructs containing the KLF2 enhancer (KLF2MTFBS), the following primer pairs were used:

154+144: expected amplicon size of 628 bp

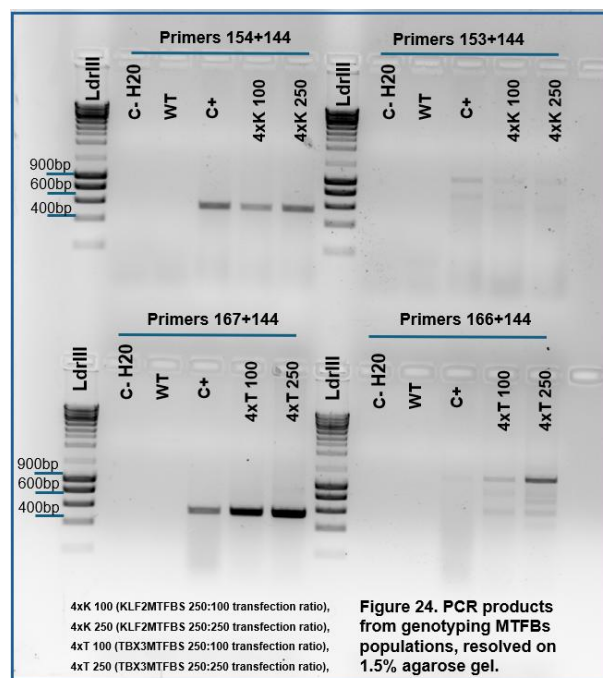
153+144: expected amplicon size of 1143 bp

- For constructs containing the TBX3 enhancer (TBX3MTFBS), the following primer pairs were used:

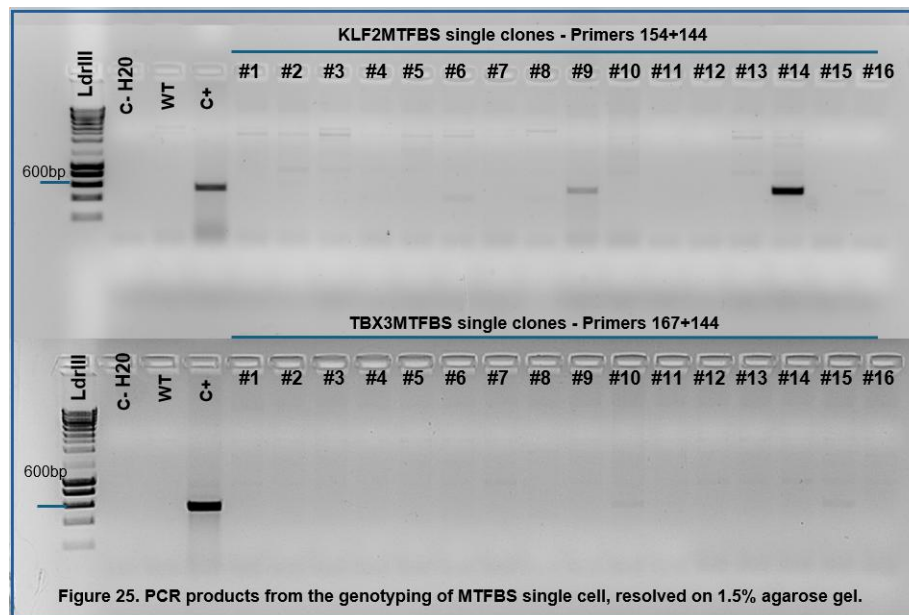
167+144: expected amplicon size of 628 bp

166+144: expected amplicon size of 1312 bp

In both cases, the first primer pair targets the right flank of the construct, spanning from the fourth spacer to the genomic region downstream of the RHA, fully encompassing the latter. The second primer pair include the same region as the first, with the addition of the fourth enhancer copy (Fig. 4). In this case, the positive control used corresponded to genomic DNA from transfected cell populations derived from a previous transfection with the same constructs. Although these bulk populations tested positive in the initial genotyping assay, they were ultimately discarded due to the failure to isolate individual positive clones.



Following confirmation of transfected cells within the populations, simultaneous seeding was performed using both 10-cell and single-cell approaches. Although genotyping initially focused on the 10-cell populations, several of which tested positive, the subsequent single-cell seeding derived from these positive 10-cell populations was not ready in time for inclusion in the current project. Therefore, the genotyping data from those derived single-cell clones is not presented in this manuscript. However, the positive 10-cell populations are included in the final summary of generated lines, presented as an annexed table (Fig.27). Consequently, attention was directed to the single-cell populations seeded in parallel on the same day as the 10-cell seeding. These clones underwent preliminary genotyping using primer combinations 154+144 for KLF2MTFBS and 167+144 for TBX3MTFBS, as these pairs had previously showed the most reliable results in the analysis of total transfected populations. This approach enabled timely identification of positive clones for both constructs (Fig. 25).



KLF2MTFBS#9, KLF2MTFBS#14, TBX3MTFBS#10, TBX3MTFBS#15, and were ultimately selected as positive clones for their respective constructs. However, the initial genotyping protocol used to identify these clones only covered a limited portion of the inserted sequence. To achieve a more comprehensive characterization of these lines, additional analyses were performed.

This strategy involved a second round of PCR-based genotyping using NZYTaq polymerase, with primer pairs specifically designed to span the largest possible region of the inserted construct (Fig.4).

The selected primer combinations were as follows:

- For KLF2 Multi Transcription Factor Binding Site (MTFBS):

137 + 183 (expected band size: 1265 bp), covering the amplicon from LHA to spacer2, including enhancers 1 and 2.

151 + 176 (expected band size: 1066 bp), covering the amplicon from spacer1 to spacer3, including TF 2 and TF 3.

152 + 138 (expected band size: 1397 bp), covering the amplicon from spacer2 to RHA, including TF 3 and TF 4.

- For TBX3 Multi Transcription Factor Binding Site (MTFBS):

143 + 185 (expected band size: 1059 bp), covering the amplicon from LHA to spacer2, including TF 1 and TF 2

164 + 177 (expected band size: 1395 bp), covering the amplicon from spacer1 to spacer3, including TF 2 and TF 3

166 + 144 (expected band size: 1312 bp), covering the amplicon from spacer3 to RHA, including TF 4

As a positive control, genomic DNA from bulk transfected cell populations was used, despite not having been previously tested with these specific primer combinations.

The results of these PCR reactions (Fig. 26) were not the expected in all cases. However, considering that variability may stem from incomplete optimization of the applied protocols, and given that the tested clones have produced positive results in some genotyping assays, they were ultimately included in the expression analysis.

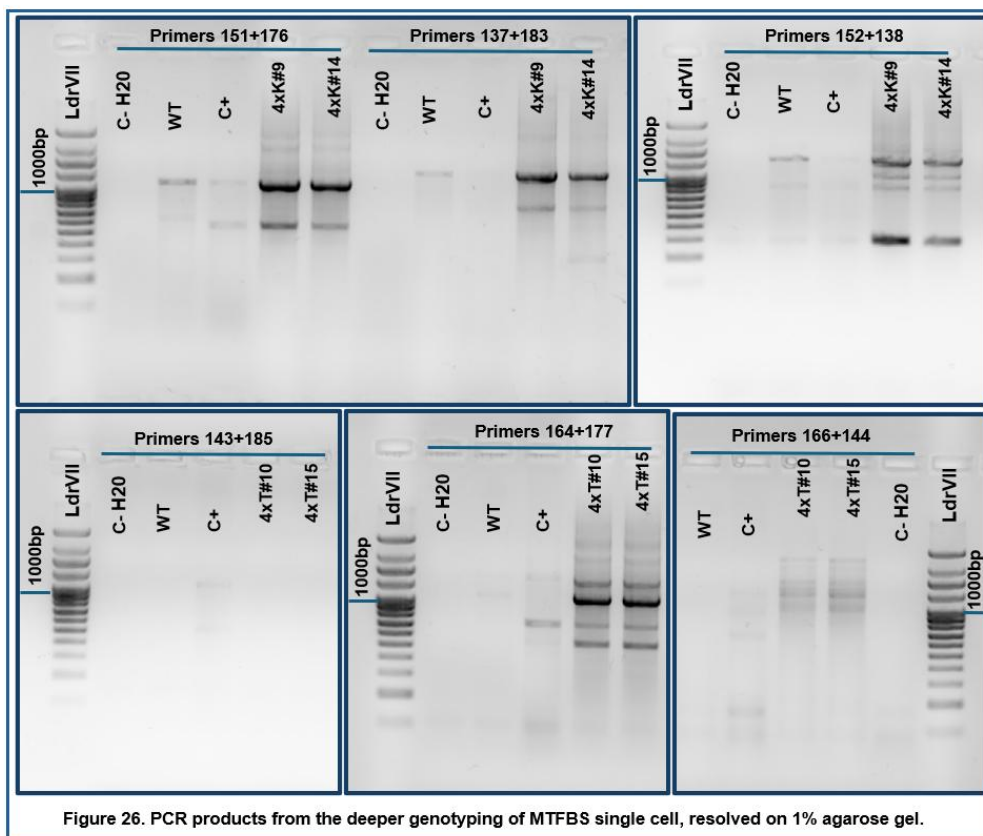


Figure 26. PCR products from the deeper genotyping of MTFBS single cell, resolved on 1% agarose gel.

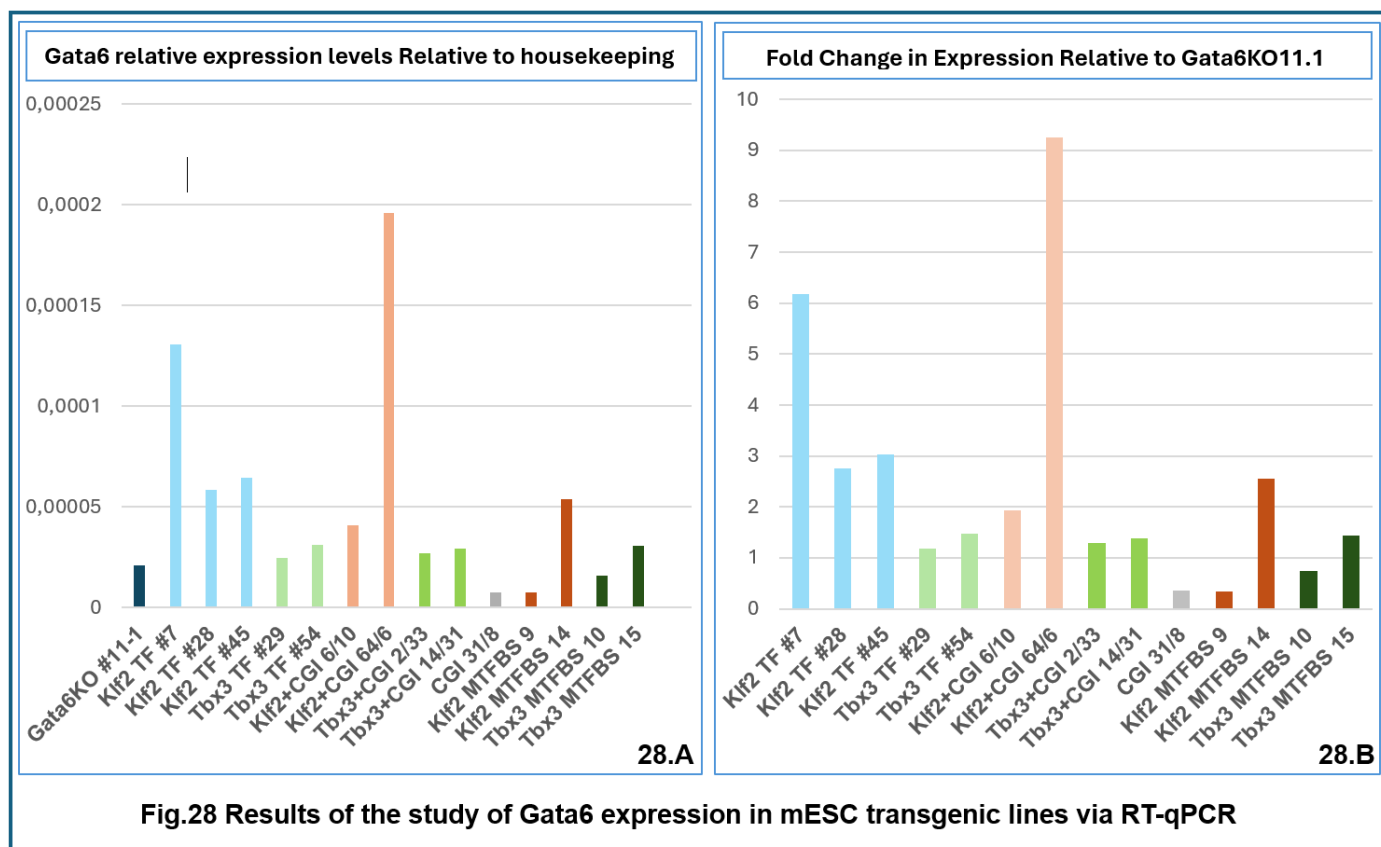
Genotype	Cell Input	Cell Line	Status	Genotype	Cell Input	Cell Line	Status
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #6	(+)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC #9	(+)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #10	(-)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 3 clone #11	(-)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #22	(-)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 8 clone #2	(-)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #29	(-)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 8 clone #6	(+)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #23	(-)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 8 clone #17	(+)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #25	(-)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 2 clone #3	(+)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #26	(-)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 2 clone #33	(+) Q
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #38	(-)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 14 clone #15	(+)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #42	(-)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 14 clone #25	(+)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #51	(+)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 14 clone #31	(+) Q
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #56	(+)	Tbx3 + CGI	Single clone	Tbx3 + CGI SC pop 14 clone #36	(+)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #64	(+)	OnlyCGI	10 cells	CGI 10 cells pop #31	(+)
Klf2 + CGI	10 cells	Klf2 + CGI 10 cells pop #96	(+)	OnlyCGI	Single cell	CGI SC pop 31 clone #5	(+)
Klf2 + CGI	Single clone	Klf2 + CGI SC #1	(-)	OnlyCGI	Single cell	CGI SC pop 31 clone #8	(+) Q
Klf2 + CGI	Single clone	Klf2 + CGI SC #5	(-)	Klf2 MTFBS	10 cells	Klf2 MTFBS 10 cells pop #3	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC #8	(-)	Klf2 MTFBS	10 cells	Klf2 MTFBS 10 cells pop #21	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC #12	(-)	Klf2 MTFBS	10 cells	Klf2 MTFBS 10 cells pop #28	(+)
Klf2 + CGI	Single clone	Klf2 + CGI SC #19	(-)	Klf2 MTFBS	10 cells	Klf2 MTFBS 10 cells pop #31	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC #21	(-)	Klf2 MTFBS	Single cell	Klf2 MTFBS SC clone #9	(+) Q
Klf2 + CGI	Single clone	Klf2 + CGI SC #22	(-)	Klf2 MTFBS	Single cell	Klf2 MTFBS SC clone #14	(+) Q
Klf2 + CGI	Single clone	Klf2 + CGI SC #32	(-)	Klf2 MTFBS	Single cell	Klf2 MTFBS pop21 clone #30	(+)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 6 clone #10	(+) S Q	Tbx3 MTFBS	10 cells	Tbx3 MTFBS 10 cells pop #10	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 64 clone #2	(+)	Tbx3 MTFBS	10 cells	Tbx3 MTFBS 10 cells pop #22	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 64 clone #6	(+) Q	Tbx3 MTFBS	10 cells	Tbx3 MTFBS 10 cells pop #24	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 64 clone #20	(+)	Tbx3 MTFBS	10 cells	Tbx3 MTFBS 10 cells pop #26	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 64 clone #24	(+)	Tbx3 MTFBS	10 cells	Tbx3 MTFBS 10 cells pop #28	(+)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 96 clone #2	(+)	Tbx3 MTFBS	10 cells	Tbx3 MTFBS 10 cells pop #36	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 96 clone #16	(+)	Tbx3 MTFBS	10 cells	Tbx3 MTFBS 10 cells pop #43	(+)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 96 clone #25	(+)	Tbx3 MTFBS	10 cells	Tbx3 MTFBS 10 cells pop #44	(-)
Klf2 + CGI	Single clone	Klf2 + CGI SC pop 96 clone #26	(+)	Tbx3 MTFBS	Single cell	Tbx3 MTFBS SC clone #10	(+) Q
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #3	(+)	Tbx3 MTFBS	Single cell	Tbx3 MTFBS SC clone #15	(+) Q
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #7	(-)	Tbx3 MTFBS	Single cell	Tbx3 MTFBS SC clone #23	(+)
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #8	(+)	Tbx3 MTFBS	Single cell	Tbx3 MTFBS pop28 clone #17	(+)
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #10	(-)	Tbx3 MTFBS	Single cell	Tbx3 MTFBS pop28 clone #23	(+)
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #20	(-)	Tbx3 MTFBS	Single cell	Tbx3 MTFBS pop28 clone #27	(+)
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #1	(-)	Figure 27. Summary of All Transgenic Lines Generated. Legend: (+) Line has shown positive results in all genotyping assays to which it was subjected, using validated protocols. (-) Line has shown negative results in at least one assay with proven reliability. S: Line has been sequenced. Q: Line was selected for <i>Gata6</i> expression analysis by RT-qPCR.			
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #2	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #8	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #12	(-)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #13	(-)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #14	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #15	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #19	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #24	(-)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #26	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #29	(-)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #35	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #38	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #40	(+)				
Tbx3 + CGI	10 cells	Tbx3 + CGI 10 cells pop #42	(-)				

3) Preliminary analysis of *Gata6* expression in asynchronous cells.

Following the characterization of the generated transgenic lines, the following clones were selected for the analysis of *Gata6* expression via RT-qPCR: Klf2+CGI pop64 clone #6, Klf2+CGI pop64 clone #20, Tbx3+CGI pop2 clone #3, Tbx3+CGI pop14 clone #15, Tbx3+CGI pop14 clone #31, Klf2+CGI pop6 clone #10, OnlyCGIpop31clone#8, KLF2MTFBS#9, KLF2MTFBS#14, TBX3MTFBS#10, and TBX3MTFBS#15. The *Gata6*KO11.1Hemizygous line was included as a control, as it represents the parental line into which the constructs were introduced.

Additionally, the following clones were incorporated: Klf2 TF #7, Klf2 TF #28, Klf2 TF #45, Tbx3 TF #29, and Tbx3 TF #54. These clones had been previously generated in the laboratory using the same parental line (*Gata6*KO11.1Hemizygous), but the introduced constructs contained only the corresponding enhancer sequences. This design allows for a more robust assessment of transcriptional responsiveness.

Data processing involved normalization against housekeeping gene expression, followed by graphical representation. Two types of data visualization were performed: (Fig. 28.A) shows the relative expression levels across all lines, while (Fig. 28.B) displays the fold change in expression relative to *Gata6*KO11.1Hemizygous)



Discussion

In the initial phase of this study, the focus was placed on generating a library of plasmids carrying various combinations of regulatory elements and their subsequent cloning into *E. coli* DH5 α . The establishment of this vector library enables the long-term preservation of these constructs for future applications in regulatory genomics, extending beyond the scope of the current project. Furthermore, the plasmids designed here provide a foundation for generating novel combinations of regulatory sequences.

Sanger sequencing confirmed the correct assembly of the constructs, validating the efficiency of the cloning strategy employed. Given that most primers were custom designed specifically for this project, their optimization was carried out de novo and proved successful in all cases.

This was particularly critical for the assembly of vectors containing tandem arrays of four enhancer repeats (KLF2MTFBSVector and TBX3MTFBSVector), which are structurally more complex and larger than the other constructs generated (KLF2+CGIVector, TBX3+CGIVector, OnlyCGIVector). Although the latter constructs also posed challenges, primarily due to the presence of GC-rich sequences (CpG islands), which are known to impair DNA polymerase efficiency (Zhu et al., 2016), the complexity of the MTFBS vectors required a more elaborate approach.

Initially, the cloning strategy for the MTFBS vectors involved outsourcing the synthesis of the entire enhancer tandem arrays as gBlocks (Integrated DNA Technologies). However, this approach was abandoned due to size limitations inherent to the synthesis platform. Attempts to reduce the sequence length were insufficient, as spacer regions between enhancers were necessary to facilitate primer design for downstream genotyping in cell lines.

This constraint led to the development of the final cloning strategy (Fig. 9, Fig. 10), which, although effective in generating the desired constructs, introduced certain drawbacks. Chief among these was the need to clone and sequence each intermediate construct in *E. coli* DH5 α , increasing both the time and resources required, as well as the risk of introducing mutations during repeated amplification cycles. Although no mutations were ultimately detected in the final vectors, this risk should be considered when employing similar strategies, given the extensive replication steps involved.

Altogether, all cloning strategies developed for this project were successfully implemented. The resulting vector library serves as a robust platform for future genetic engineering applications in cell culture systems, enabling functional studies and the

construction of new regulatory architectures (e.g., MTFBS+CGI combinations or alternative enhancer sequences).

Thanks to the successful construction of all vector architectures, transfection of mESCs was carried out with each of them. However, not all resulting cell lines could be reliably sequenced, complicating the interpretation of downstream expression analyses and necessitating more advanced genotyping strategies. Although many of the generated lines have shown promising results at the current stage of characterization, full sequencing would be ideal to enable accurate interpretation of their phenotypic traits.

In our case, the clone Klf2+CGIpop6#10 was successfully confirmed by sequencing using KAPA polymerase, requiring only a preliminary superficial genotyping assay. However, the inability to sequence the remaining lines led to the development of alternative genotyping approaches aimed at deeper characterization. Establishing these protocols will allow for early exclusion of non-viable clones during the genotyping process, reducing the number of candidates to be sequenced in future experiments.

For transgenic lines carrying enhancer+CGI combinations, genotyping was performed from both the enhancer side and the CGI side (Fig. 20), and the primers designed for these assays proved effective. Adequate controls were included to ensure reliable interpretation of the results. However, in the case of the 10-cell populations subjected to duplication and duplication+inversion analysis (Fig. 19), the results were not entirely reliable, primarily due to the lack of appropriate controls. Specifically, no positive control was available for the duplication+inversion assay or for duplication in the *KLF2* context. This is particularly problematic given that the expected result is the absence of a band (indicating absence of duplication or inversion), which underscores the importance of including a positive control to confirm that the PCR reaction itself was successful.

Additionally, partial duplications may occur, leading to variable band sizes in the duplication assay (Fig. 18). In all cases, clones showing amplification were excluded. To further refine this protocol, it could be applied to monoclonal populations (we used it to select which 10-cell populations to plate as single cells), as some populations displayed ladder-like banding patterns. These could reflect either mixed genotypes within the population or nonspecific primer binding. Comprehensive genotyping will enable early exclusion of aberrant lines in future studies, but the protocols must be optimized with greater precision.

Regarding the characterization of OnlyCGI lines, the CGI-side genotyping protocol established for enhancer+CGI lines was applied, and its functionality was assumed.

While the protocol was effective, the gel bands produced were of low intensity (Fig. 22, Fig. 23). These PCR reactions were performed using DNA extracted via a crude Q&D protocol without purification or quantification, and the low band intensity was attributed to the phenotypic characteristics of the OnlyCGI lines, which typically exhibited slower proliferation, resulting in fewer cells per well and ultimately less DNA template for genotyping PCR.

It is critical in transgenic line genotyping that PCR assays include both part of the inserted construct and flanking genomic DNA, to confirm not only the presence of the transgene but also its integration at the intended genomic locus (Haraguchi & Nakagawara, 2009). Taking this into consideration, enhancer-side and CGI-side analyses were performed on enhancer+CGI lines. For MTFBS lines, previously validated enhancer-side primers were reused, producing positive results for several clones (Fig. 25). Clones TBX3MTFBS#10 and TBX3MTFBS#15 showed very low band intensity, which did not correlate with the cell density of the corresponding wells. Interestingly, these same clones produced strong bands in other assays (Fig. 26), suggesting that DNA concentration was not the limiting factor.

Since CGI-based genotyping was not applicable in this context, new primer pairs were designed and tested in various combinations (Fig. 26) to enable deeper characterization of these lines. The presence of repetitive sequences in the MTFBS constructs complicates amplification, promoting nonspecific primer binding and ladder-like banding patterns even in monoclonal populations (Fig. 25). At the sequencing level, these repeats hinder full-length amplification and interfere with polymerase performance due to secondary structure formation (Murat, Guilbaud & Sale, 2020).

To overcome this, the constructs could be amplified using multiple primer combinations to generate smaller fragments, which could then be sequenced and aligned to reconstruct the full sequence. However, this approach is essentially equivalent to performing multiple genotyping PCRs and given the low robustness of the current genotyping results (Fig. 26), sequencing of the resulting amplicons was deemed inadvisable.

Improving these genotyping protocols, through the inclusion of more reliable controls and further optimization of primer performance and thermocycling conditions, is essential for obtaining more consistent results. Despite these limitations, the lines that tested positive in the initial genotyping assays (Fig. 25) were included in the *Gata6* expression analysis.

Although fully characterized monoclonal populations could not be established for all constructs, a considerable number of transgenic lines were successfully generated and remain valuable for future studies (Fig. 27). Similarly, not all genotyping primers designed (Fig. 4) were tested prior to the development of reliable analytical protocols. Nonetheless, many of the cell lines produced in this project have shown positive outcomes at the current stage of analysis and are expected to be fully characterized in future work, thanks to the strategic framework developed here.

This collection of lines therefore represents a solid starting point for subsequent investigations into genetic regulation. It opens the possibility of exploring how the interaction between enhancers and CpG islands may represent a novel and potentially critical component in fundamental biological processes such as cell division.

At the beginning of this project, we hypothesized that the interaction between CpG islands and active enhancers could play a key role in the reactivation of specific genes upon mitotic exit, thereby maintaining the cellular identity of daughter cells. Since *GATA6* is mostly inactive in our study model (mESC), its increased expression could be striking in response to the different regulatory constructs to which we have exposed it. Therefore, by measuring its transcriptional profile in our selected positive lines, we begin to elucidate the regulatory action of the enhancer–CGI interaction.

It must be considered that these results represent only a preliminary study for several reasons. First, we performed RT-qPCR on asynchronous cells, meaning they were at different points in the cell cycle at the time of RNA extraction, preventing us from evaluating effects specifically derived from mitotic exit and requiring future studies in an exclusively mitotic context. Secondly, the genotypic characterization of our lines is imprecise, so the subsequent phenotypic changes cannot be clearly related to the action of each theoretically introduced regulatory structure. And finally, the number of replicates is insufficient to obtain statistically robust results. Taking all this into account, the transcriptional study via RT-qPCR was carried out as a preliminary analysis and to generate data that may be useful in future studies in this field.

The results of this analysis (Fig. 28) allow us to extract interesting information. The effect in the lines carrying only the enhancer sequence (provided by the laboratory as controls and fully characterized) is clearly greater for the KLF2 enhancer than for TBX3. This difference between both enhancers has been maintained in our enhancer+CGI constructs, with KLF2+CGI lines showing a greater increase in expression than TBX3+CGI.

It is also striking that the enhancer+CGI lines show increased expression while the onlyCGI clone has decreased expression. This could be a consequence of the perturbation derived from the transgenesis process, with the CGI alone lacking sufficient activating capacity to overcome this negative effect. Another possibility is that the three-dimensional reformation of the DNA resulting from the introduction of a GC-rich sequence prevents proper expression without an active enhancer alongside which to act.

The data derived from the MTFBS lines show contradictory effects between clones carrying the same tandem enhancer (e.g., KLF2MTFBS#9 showed decreased expression while KLF2MTFBS#14 showed increased expression). Such significant differences seem to indicate that they do not share the same genotype.

Finally, although KLF2+CGIpop6#10 is the best characterized and therefore the most reliable in terms of results, it does not show an increased response compared to its counterparts without CGI (KLF2 TF #7#28#45). Conversely, KLF2+CGIpop64#6 has shown a fold change in expression greater than any other line analysed. Therefore, if the correct phenotype of this clone is fully confirmed, it would represent a perfect example of how an active enhancer by itself promotes gene expression, but this transcriptional stimulation is even greater in the presence of a CpG island.

Conclusions

- The plasmid library containing different regulatory structures has been successfully established, demonstrating the effectiveness of the designed strategy.
- Transgenic mESC lines have been generated by introducing regulatory structures, although their characterization requires further in-depth analysis.
- The repressive influence of the CpG island on *GATA6* is reversed when paired with an active enhancer, revealing a synergistic effect that enhances gene expression.
- The enhancer associated with *Klf2* shows greater activating capacity than that of *Tbx3*, and this difference is maintained both in the absence and presence of the CpG island.

Bibliography

- Angeloni, A., & Bogdanović, O. (2021). Sequence determinants, function, and evolution of CpG islands. *Biochemical Society Transactions*, 49(3), 1109–1119. <https://doi.org/10.1042/BST20200695>
- Bell, J. S. K., & Vertino, P. M. (2017). Orphan CpG islands define a novel class of highly active enhancers. *Epigenetics*, 12(6), 449–464. <https://doi.org/10.1080/15592294.2017.1297910>
- Calo, E., & Wysocka, J. (2013). Modification of enhancer chromatin: What, how, and why? *Molecular Cell*, 49(5), 825–837. <https://doi.org/10.1016/j.molcel.2013.01.038>
- Contreras, A., & Perea-Resa, C. (2024). Transcriptional repression across mitosis: Mechanisms and functions. *Biochemical Society Transactions*, 52(1), 455–464. <https://doi.org/10.1042/BST20231071>
- Coux, R.-X., Dubois, A., Chervova, A., Festuccia, N., Gonzalez, I., Vandormael-Pournin, S., Cohen-Tannoudji, M., & Navarro, P. (2024). Molecular and epistatic interactions between pioneer transcription factors shape nucleosome dynamics and cell differentiation. *bioRxiv*. <https://doi.org/10.1101/2024.05.27.596047>
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers. *Proceedings of the National Academy of Sciences*, 107(50), 21931–21936. <https://doi.org/10.1073/pnas.1016071107>
- Festuccia, N., Gonzalez, I., & Navarro, P. (2017). Mitotic bookmarking in development and stem cells. *Development*, 144(20), 3633–3645. <https://doi.org/10.1242/dev.146522>
- Haraguchi, S., & Nakagawara, A. (2009). A simple PCR method for rapid genotype analysis of the TH-MYCN transgenic mouse. *PLoS ONE*, 4(9), e6902. <https://doi.org/10.1371/journal.pone.0006902>
- Ito, K., & Zaret, K. S. (2022). Maintaining transcriptional specificity through mitosis. *Annual Review of Genomics and Human Genetics*, 23, 53–71. <https://doi.org/10.1146/annurev-genom-121321-094603>
- Kadauke, S., & Blobel, G. A. (2013). Mitotic bookmarking by transcription factors. *Epigenetics & Chromatin*, 6, 6. <https://doi.org/10.1186/1756-8935-6-6>
- Kulkarni, A., Anderson, A. G., Merullo, D. P., & Konopka, G. (2019). Beyond bulk: A review of single cell transcriptomics methodologies and applications. *Current Opinion in Biotechnology*, 58, 129–136. <https://doi.org/10.1016/j.copbio.2019.03.001>
- Murat, P., Guilbaud, G., & Sale, J. E. (2020). DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome Biology*, 21, 209. <https://doi.org/10.1186/s13059-020-02124-x>
- Pachano, T., Sánchez-Gaya, V., Ealo, T., Mariner-Faulí, M., Bleckwehl, T., Asenjo, H. G., ... Rada-Iglesias, Á. (2021). Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. *Nature Genetics*, 53(7), 1036–1049. <https://doi.org/10.1038/s41588-021-00888-x>
- Pelham-Webb, B., Polyzos, A., Wojenski, L., Kloetgen, A., Li, J., Di Giammartino, D. C., ... Apostolou, E. (2021). H3K27ac bookmarking promotes rapid post-mitotic activation of the pluripotent stem cell program without impacting 3D chromatin reorganization. *Molecular Cell*, 81(8), 1732–1748.e8. <https://doi.org/10.1016/j.molcel.2021.02.032>
- Proudhon, C., Snetkova, V., Raviram, R., Lobry, C., Badri, S., Jiang, T., Hao, B., Trimarchi, T., Kluger, Y., Aifantis, I., Bonneau, R., & Skok, J. A. (2016). Active and inactive enhancers cooperate to exert localized and long-range control of gene regulation. *Cell Reports*, 15(10), 2159–2169. <https://doi.org/10.1016/j.celrep.2016.04.087>
- Raccaud, M., & Suter, D. M. (2018). Transcription factor retention on mitotic chromosomes: Regulatory mechanisms and impact on cell fate decisions. *FEBS Letters*, 592(6), 878–887. <https://doi.org/10.1002/1873-3468.12943>
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), 279–283. <https://doi.org/10.1038/nature09645>
- Teves, S. S., An, L., Hansen, A. S., Xie, L., Darzacq, X., & Tjian, R. (2016). A dynamic mode of mitotic bookmarking by transcription factors. *eLife*, 5, e22280. <https://doi.org/10.7554/eLife.22280>
- Zhu, J., Adhikari, K., Kuo, C. C., Cecere, G., & Rada-Iglesias, A. (2023). Mitotic transmission of active enhancer states ensures faithful post-mitotic gene reactivation in human embryonic stem cells. *Nature Cell Biology*, 25, 100–112. <https://doi.org/10.1038/s41556-022-01054-0>
- Zhu, X.-J., Sun, S., Xie, B., Hu, X., Zhang, Z., Qiu, M., & Dai, Z.-M. (2016). Guanine-rich sequences inhibit proofreading DNA polymerases. *Scientific Reports*, 6, 28769. <https://doi.org/10.1038/srep28769>