University of Eastern Piedmont

School of Medicine

Department of Health Sciences

Master Degree in Medical Biotechnologies

Labelling and functional studies of the transcription factor FOXE1

Internal Supervisor: Prof. Paolo Marzullo

Candidate: Carlo Corneo

Matr. Number: 20023457

External co-supervisors:

Dr Elisa Stellaria Grassi

(Università degli studi di Milano - Milano)

Dr Francesca Coscia

(Human Technopole, Milano)

Academic Year: 2022-2023

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my internal supervisor, professor Paolo Marzullo, who always supported me promptly. His lectures and professionalism guided me in choosing the field of endocrinology for my Master's thesis.

I would like to express my deepest gratitude to professor Luca Persani for giving me the opportunity to work on my experimental thesis in his Experimental Endocrine-Metabolic Research Laboratory in the Auxologico Biomedical Research and Technology Centre. Additionally, I would like to thank to my external supervisor, Dr. Francesca Coscia, for welcoming me into her research group in the field of structural biology at Human Technopole. Their care and encouragement helped me to realize my project.

I would like to really thank to my other external supervisor, Elisa Stellaria Grassi, who followed me in every single step of each experiment making me feel her support. Her critical spirit and meticulousness will help me in my professional growth and as a person.

I would like to thank the whole Auxologico lab team. They were always kind and ready to help.

I would like to really thank Laura Tosatto for her teachings and the time she gave me. I was her "shadow" and I could grasp how to move in a laboratory. She spurred me to give my best.

I would like to thank also the whole Human Technopole team. They were always kind and helpful.

I would like to express my deepest gratitude to my parents for their support in each step of this course of study.

I would like to thank my aunt Silvana for her support.

I would like to thank my best friend Aurora for her for his patience in putting up with my paranoia. I love you and we will still have plenty of moments to share. Essential.

I would like to thank my best friend Mirko for putting up with my freaks and for giving me his support. I love you.

I would like to thank Federica, Francesco, Maurizio, Riccardo, Federico and Savio for their friendship and moments of leisure that they give me.

In the end, I would like to say that I know that you are proud of me.

TABLE OF CONTENTS

ABSTRACT

FOXE1 is a pioneer transcription factor involved in thyroid gland development and it is composed of 373 amino acids coded in only one exon on chromosome 9q22. It is formed by a DBD (DNA Binding Domain) followed by a poly alanine (poly-ala) tract of variable length and a disordered region at the C-terminal. From the manuscript published by Grassi et al, 2023 it was suggested that there are some functional roles of pathogenic mutation p.L107V combined with length of poly-ala tract. With the long-term hypothesis of performing dynamic studies in cells for this protein and to further investigate the role of the poly-ala tract, we added via genetic engineering a fluorescent tag to the protein and performed functional *in vitro* studies in thyroid cells. We wanted to understand whether labelled FOXE1 constructs behaved like the ones without fluorescent proteins to evaluate protein expression, localisation and activity after transient transfection in human thyroid NTHY-ORI cells. Firstly, we designed mammalian plasmids with the different variants in terms of number of alanines (14/16) and with or without L107V mutation of FOXE1 in which we introduced either the fluorescent proteins mCherry or mNeon Green. We tested the expression of proteins with Western blot and we had a production similar to the ones without fluorescent proteins addition. Secondly, we conducted localisation experiments with confocal microscopy to see the nuclear localization of the proteins. Our results confirm a reduced expression and modified nuclear localisation of the 14 Ala FOXE1 compared to the 16 Ala variant in NTHY-ORI cells and the role of p.L107V variant. Cells did not have the nuclear rim morphology as expected, but only nuclear diffuse signal or dense spots suggesting the presence of aggregates. As last, to test effects on transcription, we did preliminary experiments of Dual Luciferase Assay to study the activity of the tagged proteins. The study was performed on the promoter of thyroglobulin gene (a main thyroid gene under control of FOXE1) with or without the interaction with other co-factors important for transcription. The p.L107V variant indeed affected negatively the activity of both the versions with 14 or 16 Ala variants as without label. The preliminary Luciferase tests confirmed preservation of the activity when combined to NKX2-1. In conclusion the added label had similar features as the unlabelled constructs, apart the different nuclear localisation. This may be due to the presence of the bulky fluorescent tags that have approximately the same molecular weight of FOXE1, determining a different conformation and interaction with protein and DNA partners at the N-terminal side. Our study suggests that for live-imaging studies the label could be put instead at the C-terminus and/or or replaced by smaller fluorescent tags.

I. INTRODUCTION

A. The thyroid gland

The thyroid is an endocrine gland located in the inferior, anterior neck. It is responsible for the production of the thyroid hormones (TH) and of the maintenance of iodine balance in theorganism. These hormones are triiodothyronine (T3) and thyroxine (T4) with a prevalence of 90% of production of T4 that is then converted peripherally to the active form [1].

The precursor of TH is thyroglobulin (TG). It is synthesised by thyrocytes and stored in the follicle lumen as colloid. The fundamental step for the production of TH is the iodination of TG [2]. This is performed by two different enzymes, Dual Oxidase (DUOX) and Thyroid peroxidase (TPO). TPO is responsible of the oxidation of iodide (I^-) into iodine (I_2) using hydrogen peroxide (H_2O_2) generated by DUOX. Other important player in TH production is the Sodium/Iodide Symporter (NIS, also known as SLC5A5), fundamental for the concentration of iodide (I^-) in the follicles. Then, TPO mediates the formation of monoiodotyrosine (MIT) and diiodotyrosine (DIT) linking I_2 to tyrosine residues of TG. In the end, it combines MIT and DIT to form T3 and T4 [2].

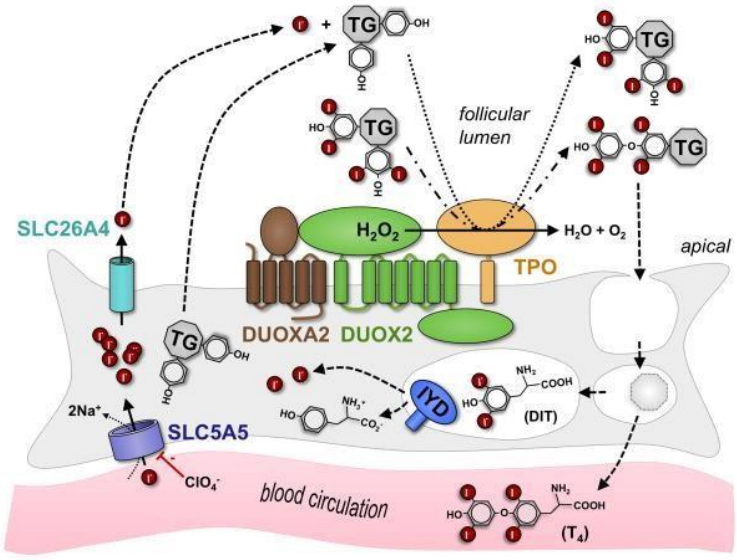

Figure 1 Schematic view of a thyrocite with illustrated the key players involved in TH production. *Image fromGenetic causes of congenital hypothyroidism due to dyshormonogenesis of Helmut Grasberger and Samuel Refetoff* [3].

In the above Figure 1 is shown the fundamental mechanisms regarding the TH production. The production of TH is strictly regulated by the hypothalamic-pituitary-thyroid axis. The following image (Figure 2) explains the mode of action of this regulation.

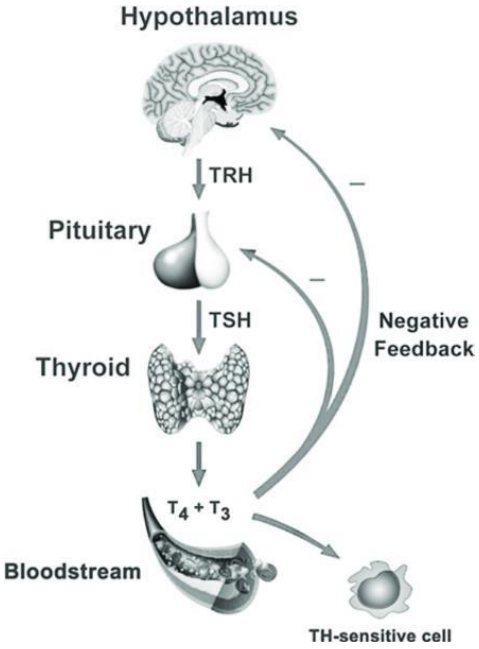

Figure 2 Hypothalamic-pituitary-thyroid axis. *Image from Jaime Freitas et al Development and Validation ofIn Vitro Bioassays for Thyroid Hormone Receptor Mediated Endocrine Disruption.*

TRH is secreted from the hypothalamus and its activity results in binding to its transmembrane receptor on the thyrotropes, endocrine cells on anterior pituitary gland, where it stimulates the TSH release [4]. Then TSH binds to TSH-receptor expressed on the thyrocytes where it stimulates the production and release of TH [5]. The amount of T3 and T4 provides a negative feedback as indicated in the figure above acting on the pituitary and hypothalamus [6].

TH are important in the regulation of different processes, among them:

- Bone homeostasis
- Glucose metabolism
- Lipid homeostasis
- Insulin secretion

Other processes are illustrated in Figure 3 and it is noteworthy that TH can act throughout the organism.

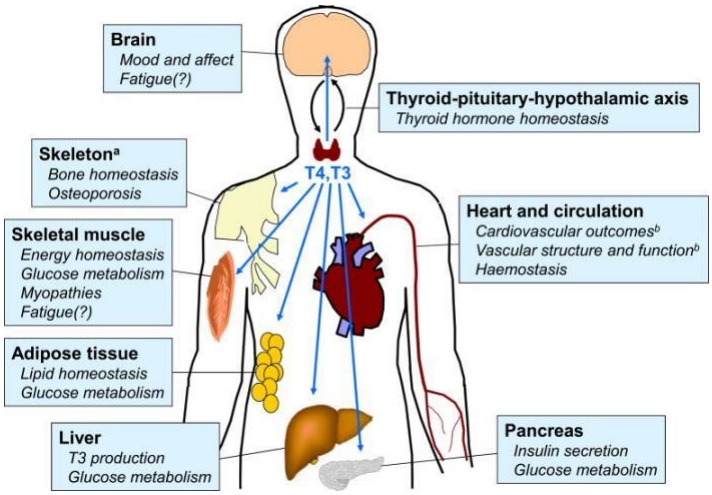

Figure 3 TH action in the organism. *Image of George J Kahali et al in Use of levothyroxine in the management of hypothyroidism: A historical perspective.*

B. Hypothyroidism and congenital hypothyroidism

Hypothyroidism is a common disorder in the general population and clinically it presents as a condition of thyroid hormone (TH) deficiency [7].

Commons symptoms of hypothyroidism are: fatigue, cold intolerance, dry skin, constipation, bradycardia, weight gain and muscle weakness.

On other hand, hyperthyroidism is the condition in which high levels of TH are present. The most known symptoms that are manifested are anxiety and irritability, difficulty in sleeping, heat intolerance, tachycardia, diarrhoea, fine tremor and weight loss [2].

A particular type of hypothyroidism is congenital hypothyroidism (CH). It is the most common congenital endocrine disease in the childhood and it is one of the most common treatable causes of intellectual disability [8]. The overall incidence of CH ranges from 1 in 3000 to 1 in 4000 newborn infants [9]. CH may be due to incorrect hormone synthesis and functionality, named Dyshormonogenesis with Gland *in Situ* (GIS) or to a defective thyroid gland development, named Thyroid Dysgenesis (TD). The most common form of TD (66%) is thyroid ectopy, an abnormal gland localisation generally characterized by small size. Another type of TD is athyreosis that affects 20-30% of patients and consists of complete absence of the gland [10]. Moreover it is possible to have hypoplasia or hemiagenesis (hypoplasia of only one thyroid lobe) [11] [12].

The more common clinical manifestations of CH comprehend myxedematous face, macroglossia, large fontanels, cold dry skin, hoarse cry, constipation, distended abdomen, umbilical hernia, decreased activity, hypotonia, feeding difficulty and prolonged jaundice. The symptoms of CH are added to ones related to hypothyroidism [13]. The most frequent symptoms of CH are present in the following image. (Figure 4)

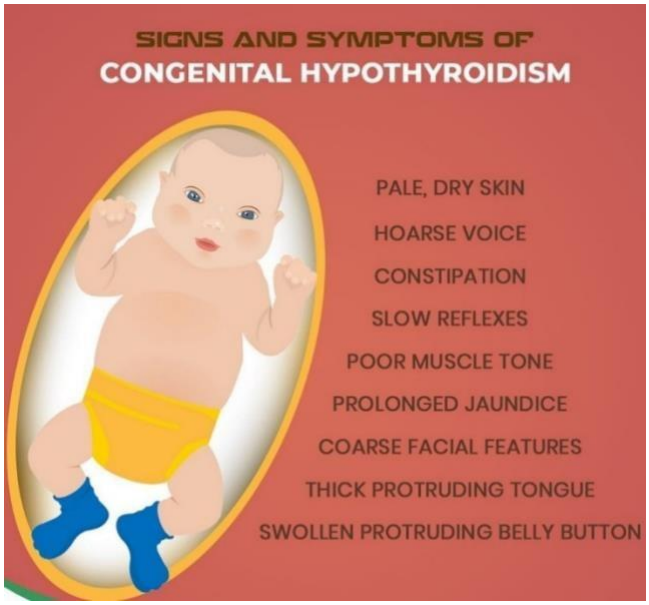

Figure 4 Signs and symptoms of CH. *Adapted from a2zmedicalnote hypothyroidism-clinical-features.*

In order to identify the suitable treatment, it is important to have a quick neonatal screening through the measurement of serum fT4 (free T4, not bound to plasma proteins) and TSH. It allows the diagnosis and the monitoring of the alterations of the thyroid functionality. It is generally low in the hypothyroidism condition. If the serum fT4 concentration is below and TSH level clearly above the age-specific reference interval, a hormone replacement therapy should be started immediately and not later than 2 weeks after birth [14].

Different outcomes must be considered during the treatment. It is important to evaluate psychomotor development and school progression periodically. Speech delay, attention and memory problems and behavioral problems are conditions to be evaluated. The early detection and screening of the patients is very important to guarantee them the best possible treatment [14].

C. Thyroid transcription factors

Transcription factors (TFs) are proteins able to bind DNA in specific sequence motifs located at regulatory elements, like promoters and enhancers [15]. The major TF families are: C2H2- zinc-finger (ZF), Homeodomain, basic helix-loop-helix (bHLH), basic leucine zipper (bZIP) and nuclear hormone receptor (NHR) [16]. TFs can respond to intercellular or environmental signals. The same TF can regulate different genes in different cell types, indicating that regulatory networks are dynamic even within the same organism [17]. Some genes are always actively transcribed in all cells, for example the ones that encode structural proteins and enzymes catalysing the reactions of basic metabolism. Other genes are only transcribed in one or a few cell types, usually only during a particular stage of development or under extracellular and/or intracellular signals regulation.

The principal functions of TFs are the following:

- basal transcriptional regulation: general transcription factors (GTFs) are necessary for transcription [18],
- differential enhancement of transcription: TFs can bind to enhancer regions of DNA to regulate the expression of genes in the right cell at the right time in the right amount [19],
- development: TFs can respond to stimuli (internal or external) activating or repressing transcription of certain genes. They can change cells morphologies or alter the cell fateand the cellular differentiation [20],
- response to environment: they are sensible to external stimuli, such as higher temperature (ex. heat shock factor HSF) [21], low-oxygen environment (hypoxia inducible factor HIF) [22],
- cell cycle control: some TFs are proto-oncogenes or tumor suppressors and can regulate the cell cycle. An example is Myc oncogene that can regulate cell growth and apoptosis [23].

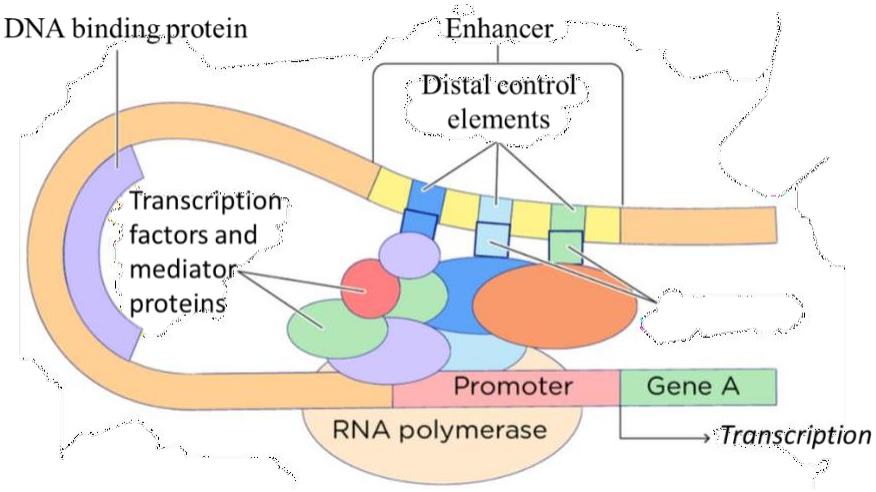

Figure 5 Mode of action of a transcription factor. Image from Jack Westin in the section DNA binding proteins, transcription factors.

In the above Figure 5 is shown the binding of a TF to a promoter that allows the transcriptionof the desired gene. Anyway, a TF can block the transcription acting as a repressor [24]. TFs maybe activated through their signal-sensing domain in different ways: through ligand binding, after phosphorylation or interacting with other TFs or coregulatory proteins [25]. Important actors in the transcription machinery are the chromatin remodelers. They can alter the locationof nucleosomes on DNA to let the binding of TFs to the promoters of specific genes. They have to act upstream because they are responsible for the opening of the chromatin. TFs usually binds DNA in euchromatin state. This conformation presents the DNA free from the nucleosomes and so ready to be transcribed [26]. A particular type of TFs called pioneer transcription factors are able to overcome

the nucleosomes barrier and bind to nucleosomal DNA [27]. They can access genes that are inaccessible to other transcription factors enabling other transcription factors, nucleosome remodeling complexes and histone modifiers to engage chromatin causing the activation or repression of gene expression [28].

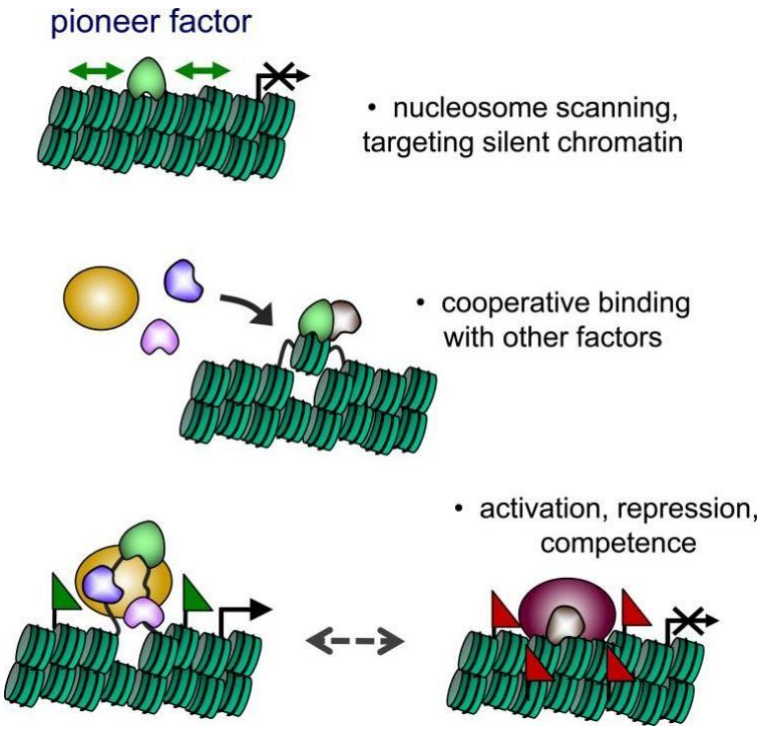

Figure 6 Activities of pioneer transcription factors. *Image from Zaret and Mango in Pioneer TranscriptionFactors, Chromatin Dynamics, and Cell Fate Control [29].*

In Figure 6 are illustrated the activities of pioneer transcription factors in order to activate or repress transcription.

Pioneer transcription factors can open the chromatin of differentiated cells to allow the binding of hormones-responsive TFs to promote the development process [30]. Due to this ability, they exhibit cell-specific DNA binding. They show a major stability when act in cooperation with other TFs [29].

Due to the important roles above mentioned in development, signaling and cell cycle regulation, some human diseases have been associated with mutations in TFs [31]. Among others, CH can also be due to mutations in the genes of certain transcription factors [32]. Among these TFs, HHEX, PAX8, NKX2-1 and FOXE1 are considered pivotal in the development of the thyroid gland starting from embryogenesis [33].

Our study is focused on FOXE1, a pioneer transcription factor composed of 373 amino acids coded in only one exon on chromosome 9q22 [34]. It is formed by a globular DNA Binding Domain (DBD)

followed by a poly alanine (poly-ala) tract of variable length and a disordered region at the C-terminal (Figure 7a). There are different studies regarding the variable lengthof the poly-ala tract in FOXE1. This condition has been documented with an involvement in thyroid-related pathologies [35] [36] [37]. The disordered unstructured region has unknown significance up to now [38].

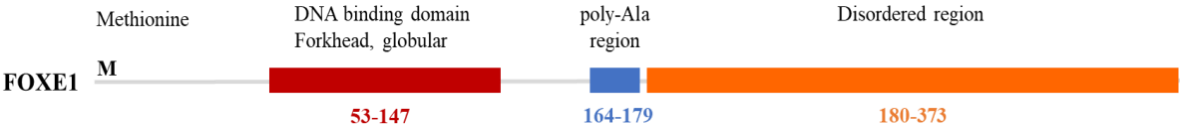

Figure 7a Schematic view of FOXE1. *In the figure are highlighted the upper-mentioned domain. With the letter M is indicated the amino acid methionine that is the first amino acid that will be translated.*

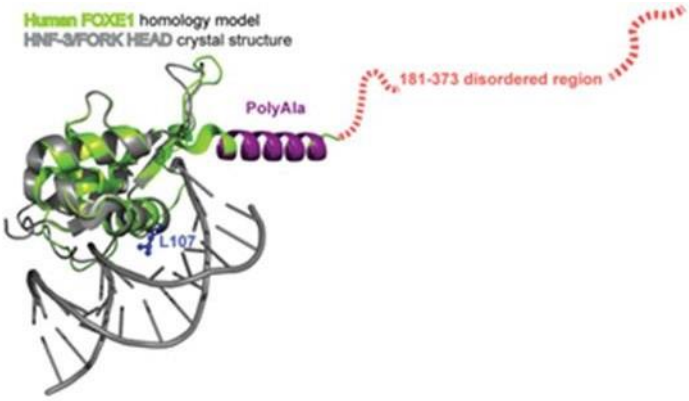

Figure 7b AlphaFold in silico model of human FOXE1. *Image readjusted from Elisa Stellaria Grassi et al Thelength of FOXE1 polyalanine tract in congenital hypothyroidism: Evidence for a pathogenic role from familial, molecular and cohort studies* [38].

It is known that FOXE1 is involved in thyrocyte precursors migration, differentiation and in thyroid hormone production. FOXE1 may act alone or together as co-regulator of other important thyroid transcription factors like PAX8 and NKX2-1 [39]. The expression of FOXE1 is important for the maintenance of the differentiated thyroid, in particular for the expression of *TG* and *TPO* genes [40]. The existing networks among the thyroid transcription factors are visible in the following figure (Figure 8).

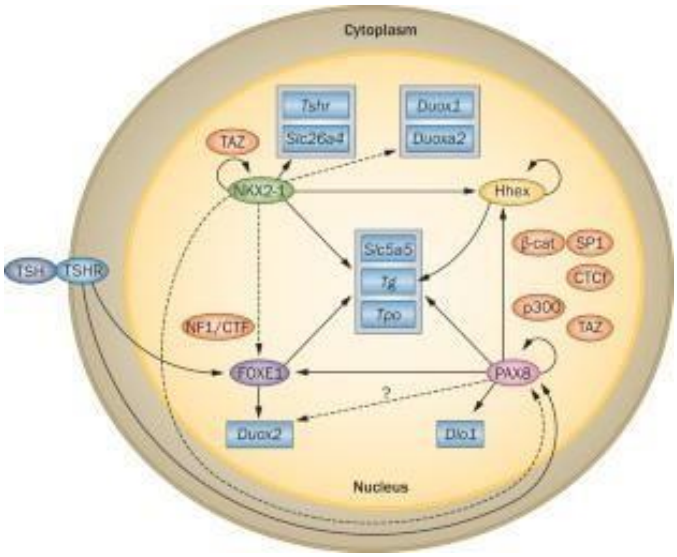

Figure 8 *Image adapted from Lara P Fernández et al in Thyroid transcription factors in development, differentiation and disease* [41].

FOXE1 is not expressed only in the thyroid gland during the development. It is expressed also in tissues derived from the pharyngeal arches and wall, such as the tongue, palate and oesophagus, as well as in anterior pituitary, choanae and hair follicles [42].

Almost all the mutations in FOXE1 that have already been documented in Bamforth-Lazarus syndrome, marked with thyroid dysgenesis, cleft palate, spiky hair and bifid epiglottis, are located in the DBD and they must be present in homozygosis [43]. An example is the mutation A65V that cause a significant reduction of DNA binding and transcription activation [44]. Recently a new *FOXE1* mutation detected in heterozygosis has been described as a possible concause of congenital hypothyroidism [38]. It is located in the gene at the level of the DBD at the position 319 (a guanine replaces a cytosine) of the sequence resulting in a Leucine substituted by a Valine (*FOXE1* p.L107V) [38]. This mutation cannot cause the disease when present alone in heterozygosis, but the combination with the poly-ala variant with 14 alanines is needed. This particular segregation was observed in the components affected in the family studied [38]. In addition to this family, other studies were conducted on cohorts of patients to evaluate the association between the number of alanines and CH both by Persani group and by other groups [45 37 35].

Massimo Tonacchera et al conducted a study on children with CH, TD, isolated cleft palate orcleft lip and thyroid hemiagenesis. Sequencing analysis was performed of the entire coding region of *FOXE1*. The most frequent poly-ala length resulting from the analysis was the 14 Ala (71% in the patients and 85% in the control). The version with 16 Ala was found in heterozygosis in the cases (26%) and in the normal control subjects (7%) [45].

The study of Aurore Carré et al was performed with *in vitro* studies with HEK293 cells for allthe

experiments conducted. The 14/14 genotype was the most frequent both in cases (63.5%) and in controls (45.0%) whereas the 14/16 and 16/16 were less frequent among patients (24.3% in patients vs. 39.5% in controls and 4.3% vs. 13.2%, respectively). Statistical analysissuggests that the presence of a 16 Ala tract could protect against the occurrence of TD in comparison to the 14/14 genotype. Patients with ectopy presented the 16 Ala tract less often than patients with athyreosis. Luciferase assays revealed that the plasmid containing FOXE116 Ala induced stronger transactivation than with 14 Ala when combined to constant amountsof PAX8 and NKX2-1 [37].

An Italian study of genetic analysis and DNA sequencing conducted in 2007 by Libero Santarpia et al showed that the most frequent Ala polymorphism in the patients' group was Ala14/14 (67%). Also in the control group this polymorphism was the most common (89%).The next most frequent polymorphism was Ala14/16. It had a percentage of 26% in the patients' group and of 11% in the control one. They found that patients with associated extra-thyroidal malformations carried either Ala 14/14 or Ala 14/16. Patients with or without malformations were equally distributed in both genotypes. Fewer CH patients than controls had the Ala 14/14 polymorphism and they could hypothesize that this variant was protective for CH but not for extra-thyroidal malformations [35].

D. Poly-alanine tract in proteins

A poly-alanine tract is constituted by the repetition of translated GCN trinucleotide repeats (Nstands for one of the four different bases, with GCC as the most abundant) [46].

There are different mechanisms that lead to poly-ala expansion. Examples are unequal allelic homologous recombination during meiosis processes, in-frame duplication and DNA polymerase slippage during translation [47 48].

Proteins containing poly-ala also present a nuclear localization motif that is typically contained in transcription factors, such as FOXE1 [46 49]. In fact, transcription factors accounts for 36% of human proteins with poly-ala tracts [50]. These repetitions are present in mammals with a role of regulator of transcription activation [46 51].

Poly-ala structures are not only spacers in proteins but they have a role in protein-protein and protein-DNA interactions [52].

An increase in the content of alanine in a protein can bring to cellular dysfunction causing pathological or non-pathological conditions [46 47].

In the tables below the most important proteins with an augmented poly-ala length are present. It is also present the related disease for each protein and the associated phenotypes. The molecular mechanism leading to each disease is as well mentioned. The longer poly-ala tracts tend to form stable beta sheets that are resistant to degradation [53].

GENE	GENE NAME	GENE LOCUS	UNIPROT	ALANINES	DISEASE RELATED
ARX	Aristaless related homeobox	Xp21.13	Q96QS3	12-36	Non-syndromic X-linked intellectual disability due to mis-localization of the proteins in the interneurons cytoplasm
FOXL2	Forkhead box L2	3q22.3	P58012	14-19/24	BPES (Blepharophimosis syndrome) caused by abnormal cytoplasmic sub-cellular localization of the proteins
HOXA13	Homeobox A13	7p15.2	P31271	18-32	Hand-foot-genital syndrome caused by no protein-protein interactions
HOXD13	Homeobox D13	2q31.1	P35453	15-22	Synpolydactyly caused by cytoplasmic retention of the proteins
PABPN1	Poly(A) binding protein nuclear 1	14q11.2	Q86U42	10-17	OPMD (oculopharyngeal muscular dystrophy) due to protein sequestration after aggregates formation
PHOX2B	Paired-like homeobox 2b	4p13	Q99453	20-33	Congenital central hypoventilazion syndrome (CCHS) due to cytoplasmic aggregation for the longest expansions while the shortest form multimers
RUNX2	Runt-related transcription factor 2	6p21.1	Q13950	17-27	Cleidocranial dysplasia caused by intracellular aggregation of the protein resulting in the exclusion from the nuclei
SOX3	SRY (sex determining region Y)- box 3	Xq27.1	P41225	15-22/26	X-linked hypopituitarism (XH) due to cytoplasmic and large perinuclear aggregates
ZIC2	Zinc family member 2	13q32.3	O95409	15-35	Holoprosencephaly (HPE) due to nuclear and cytoplasmic aggregation of the proteins

Table 1 *Readapted from Cheryl Shoubridge and Jozef Gecz in Polyalanine tract disorders and neurocognitive phenotypes* [54].

GENE	DISEASE	PHENOTYPE	REFERENCE
ARX	Non-syndromic X-linked intellectual disability	Early-onset cognitive impairment as a sole disability	María Isabel Tejada et al [55]
FOXL2	BPES (Blepharophimosis syndrome)	Narrowing of the eye opening, droopy eyelids, formation of an upward fold of the inner lower eyelid, increased distance between the eyes.	Adam J. Neuhouser et al [56]
HOXA13	Hand-foot-genital syndrome	Abnormally short thumbs and first (big) toes, small fifth fingers that curve inward (clinodactyly), short feet, and fusion or delayed hardening of bones in the wrists and ankles	Jeffrey W Innis et al [57]
HOXD13	Synpolydactyly	Fusion of third and fourth fingers, and/or second and third toes	S Malik et al [58]
PABPN1	OPMD (oculopharyngeal muscular dystrophy)	ptosis or droopy eyelids dysphagia (problems with swallowing) limb weakness in the muscles around the shoulders and hips problems with eye movements	Satoshi Yamashita [59]
PHOX2B	Ohtahara Syndrome	Infants most often have tonic seizures (stiffening of the muscles, upward eye gaze, dilated pupils, and altered breathing), but may also experience focal seizures (involving only one area or side of the brain), and rarely, myoclonic seizures (sudden jerks or twitches of the upper body, arms, and legs)	Piero Pavone et al [60]
RUNX2	Cleidocranial dysplasia	Delayed closing of the soft spots on the skull (fontanelles) and connecting joints of the skull (sutures). Underdeveloped or missing collarbones (clavicles). Permanent teeth emerging late, failure to lose primary teeth, extra teeth and dental crowding and abnormal alignment of the teeth (malocclusion).	Emilie Farrow et al [61]
SOX3	X-linked hypopituitarism (XH)	Hypothyroidism, hypogonadism, growth retardation and short stature, and secondary adrenal insufficiency	Marijke Bauters et al [62]
ZIC2	Holoprosencephaly (HPE)	Abnormal facial shape, abnormal nervous system morphology, bilateral cleft lip or median cleft lip and palate	Kristen S Barratt et al [63]

Table 2 *Readapted from Cheryl Shoubridge and Jozef Gecz in Polyalanine tract disorders and neurocognitive phenotypes [54].*

E. Fluorescent proteins

The first fluorescent protein was discovered in the 1960's. The researchers were studying the bioluminescent properties of a jellyfish *(Aequorea victoria*). It produced a blue-light-emitting bioluminescent protein called aequorin and a so named green-fluorescent protein (GFP) [64]. This jellyfish had the ability to convert Ca^{2+}-induced luminescent signals into the green luminescence [65]. GFP was cloned by the American Douglas Prasher and the first application was in the tracking of gene expression in bacteria [66]. The uniqueness of this protein is the internal localization of modified amino acid residues within the polypeptide [66]. Then some derived-GFP were engineered to produce fluorescent protein with other colours (blue, cyan, yellow, orange and red) [67] [68] [69]. The GFP-derived proteins have a molecular weight around 25kDa [70].

For our studies we decided to use two fluorescent proteins deriving from different substrates. We wanted to perform localisation assays firstly through confocal microscopy and then with live imaging. mCherry derived from DsRed of *Discosoma* sea anemones and mNeon Green derived from *Branchiostoma lanceolatum.* We chose these two fluorescent proteins rather than GFP or RFP because they are more photostable (important for live imaging) and less sensitive to laser induced bleaching. They are brighter 3 to 5 times than GFP and RFP. Additionally, this choice was dictated by the lower toxicity of these fluorescent proteins and by their monomeric structure instead of the tetrameric conformation of GFP and RFP. In this way we minimize any possible alteration of the characteristics of FOXE1.

The adding of a fluorescent protein *in frame* (the cloning created an open reading frame with both the fluorescent protein and the gene of interest) allows the tracking in real time of the protein expression, both *in vivo* and *in vitro.* In the Figure 9 are illustrated stainings with mCherry.

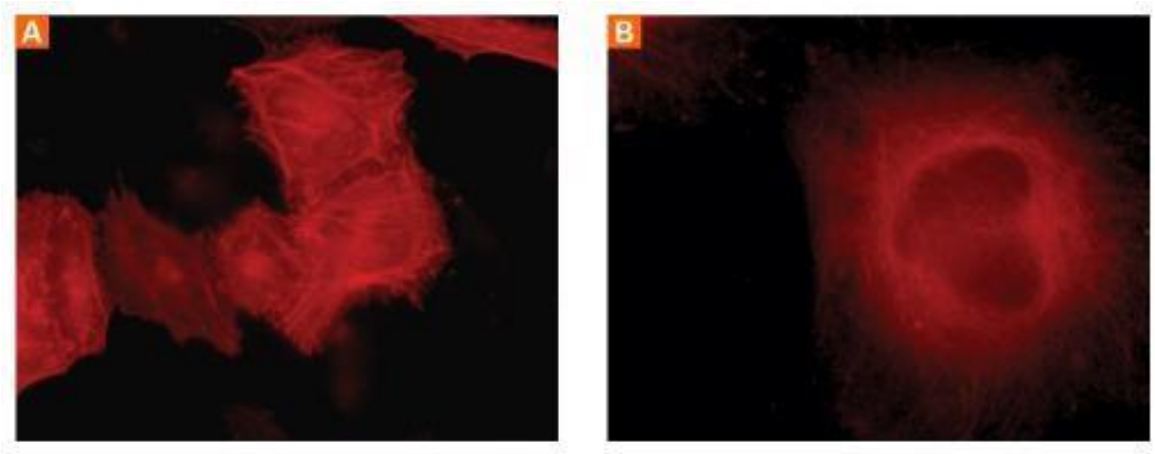

Figure 9 Example of stainings with mCherry fluorescent protein. *Image from Takara Bio.*

In the below Figure 10 are illustrated the main areas for the application of fluorescent proteins.

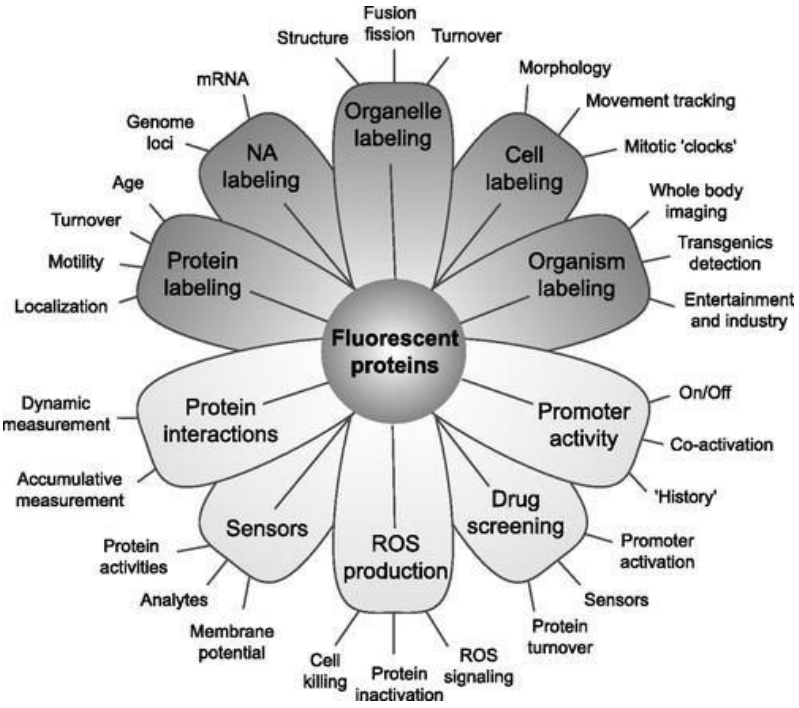

Figure 10 Main areas of application of fluorescent proteins. *Image from Dmitriy M Chudakov et al in Fluorescent proteins and their applications in imaging living cells and tissues* [71].

F. Aim of the thesis

The poly-ala role regarding the functionality of FOXE1 is not well known. The main hypothesis are that it can influence FOXE1 activity itself of its regulatory activity towards the other transcription factors.

The most frequent version of FOXE1 in humans is with 16 alanines while the presence of 14 alanines has been described as a predisposing factor for CH. Moreover, a study from 2023 described for the first time how the association of a heterozygous mutation in *FOXE1* DBD and the 14 alanines stretch may be responsible of CH onset [38]. The sequencing results of this study showed that the homozygous 14 Ala and the heterozygous p.L107V *FOXE1* variants were present in all the five siblings of the family studied with athyreosis. Instead the hypothyroid mother carried the 14 Ala/p.L107V combined with the 16 Ala WT allele. The euthyroid father carried the 14 Ala in homozygosis. Regarding the *in vitro* studies in Western blot experiments, a reduced expression of the 14 Ala FOXE1 versions were detected when compared to the 16 Ala versions in NTHY-ORI cells. With confocal microscopy experiments they realised that the poly-ala length and the presence of p.L107V variant could influence FOXE1 nuclear localisation. They detected three main different patterns: nuclear diffuse (more diffuse in the 16 Ala), nuclear rim and nuclear aggregates (more frequent in the 14 Ala and in presence of p.L107V). Functional assays indicated that the 14 Ala and

the 16 Ala alone had similar activity. The presence of p.L107V affected negatively both the versions with 14 or 16 Ala. Then they saw that only the 16 Ala significantly enhanced NKX2.1. In conclusion they had been able to ascertain that the presence of 14 Ala and p.L107V mutation in the DBD could negatively affect FOXE1 expression pattern and functionality.

Given these premises, we decided to further study the role of the alanine stretch in FOXE1 functionality. We thought to proceed with differentiated approaches:

- Study of how the different domains of FOXE1 influence the DNA binding and the functionality. To assess this, constructs with different poly-ala lengths and constructs with truncated forms with different combination of the functional domain are created. Through a previous purification of the constructs will be possible to perform Electrophoresis Mobility Shift Assay (EMSA) to evaluate the DNA binding of each construct.
- Study of how the different nuclear morphologies, explained in the study mentioned above of Persani group [38], correlates with the activity of FOXE1 through cell imaging with transfection of single constructs or with co-transfections.

I was mainly involved in the creation of FOXE1 tagged plasmids with the fluorescent proteins mCherry and mNeon Green for *in vitro* experiments. We then wanted to see whether there were any differences about protein production, cellular localisation and activity of these new FOXE1 plasmids rather than the results already obtained by Persani group with the FOXE1-FLAG without fluorescent proteins.

II. MATERIALS AND METHODS

A. Cloning

The vectors used in this work are produced by restriction-ligation cloning. We used restriction enzymes to cut open a plasmid (backbone) and insert a linear fragment of DNA (gene of interest) that has been cut by the same couple of restriction enzymes. Then an enzyme, DNA ligase, covalently binds the backbone to the gene of interest generating a circular plasmid with the desired building.

Vectors nature

- Mammalian: p3xFlag-CMV-7.1 (Sigma-Aldrich, catalogue number E7533). The presence of CMV promoter is dictated because it is a strong viral (Herpes virus 5) promoter. It has to the ability to drive a constitutive expression of genes under its control. It is an important presence that allows the final production of the proteins for the *in vitro* experiments with mammalian cells. The sequence of 3x FLAG was essential for the recognition of the proteins through the binding of anti-FLAG antibody in Western Blot experiments. It also contains the ampicillin resistance gene that was essential for the selection of colonies transformed with properly cloned plasmids.

General procedure

- PCR: to amplify the different forms of FOXE1 and the fluorescent proteins that will be cloned. It is performed to generate a good amount of DNA for the successive steps. To run a PCR is important to set a tailor-made programme for each fragment that we want to amplify. The duration of each step is recommended in the protocol related to the DNA polymerase (Q5 in our case) that we wanted to use. We used Q5 High-Fidelity 2X Master Mix from New England Biolabs (NEB, catalogue number: M0492L). It contains: Q5 polymerase, dNTPs, $MgCl_2$ and reaction buffers. Q5 polymerase had the task of synthesizing new strand of DNA complementary to the template. dNTPs were the components necessary to the building of the new strand of DNA. They had similar structures of the nucleotides that forms dsDNA. The presence of $MgCl_2$ helped the DNA polymerase too boost the amplification. Even the primers were necessary to mark the exact portion of DNA that we wanted to amplify. We designed them to have the insertion of the restriction sites presented in the backbone in their sequences for the further cloning experiments.

It presents 6 different steps that are the following:

1. First denaturation: consists in an incubation at high temperature (98°C in our case) to allows the separation into single strands (ssDNA) of the double- stranded DNA (dsDNA). In this way the primers sequences can bind to the correct portion of the obtained DNA single helix. Duration of 30 seconds recommended.
2. Denaturation: the temperature remains high as in the first denaturation (98° in our case) to disrupt the dsDNA as in the step above mentioned. Duration of 5-10 seconds recommended.
3. Annealing: the annealing temperature is determined from the primer sequences. It is related to the melting temperature (Tm) that is the temperature at which 50% of the dsDNA takes on a ssDNA conformation. To calculate the exact degree of Tm we used this formula:
   Tm = [4(G + C) + 2(A + T)] °C. The letters mean the different nucleotides (guanine, cytosine, adenine and thymine).
   Since we used the Q5 polymerase is recommended to set a temperature of annealing greater than one degree of the Tm. The duration recommended had a range of 10-30 seconds and it depends on the length of the fragment that we wanted to amplify.
4. Elongation: the temperature is raised to 72°C recommended by the activity of Q5 polymerase. It is suitable for the extension of the hybridized primers. It had a range of 20-30 seconds/kb.
5. Final elongation: the temperature was the same of the 4 step of elongation. It allows the polymerase to synthesize uncompleted amplicon to finish the process. The duration was of 2 minutes recommended by the protocol.
6. Incubation: the temperature recommended had a range of 4-10 °C for a correct conservation of DNA and could be prolonged indefinitely.

The step from 2 to 4 were repeated 35 times (35 cycles) to obtain a good amount of fragments amplified.

- Gel Electrophoresis: the whole amount of PCR product amplified is loaded on an agarose gel. We used a low melting agarose (LM agarose) to create a 1% agarose gel. The choice of the LM agarose helped us for the successive step of purification. It was easier to cut bands to purify from a LM agarose gel than from a normal agarose gel. This step is necessary to be sure about the gene loaded. If the band on the gel has the right height related to its base pair composition, it can be cut. To visualize the band is necessary the presence of Sybr™ Safe DNA gel stain (Invitrogen, catalogue number: S33102) that was able to interact with the DNA and allowed the detection of the band after UV stimulation.
- Gel band purification: using the NucleoSpin Gel and PCR Clean-up (Macherey-Nagel,

catalogue number 740609.50) the cut bands from the gel are purified and at the end ofthe protocol kit the DNA of the gene of interest is obtained. To quantify the amount of DNA we opted for a spectrophotometer quantification with Thermo Scientific™ Invitrogen™ Nanodrop™. After reading the absorbance at 260 nm (typical for nucleic acids), at 280 nm (typical for the proteins) and at 230 nm (typical for the chemical substances that could remain in the samples after purification protocol). A good sample has the ratio 260/280 absorbances between 1.8 and 2.1 and 260/230 between 2.0 and 2.2. The concentration of the samples was indicated in ng/μL.

- Digestion: the samples (backbones and gene of interest) are cut using restriction enzymes. A restriction enzyme is a protein that cleaves DNA sequences in specific sites giving DNA fragments with a known sequence at each end. The choice of these is driven by the knowledge of the plasmids maps. They have to cut both the gene of interest and the backbone in the desired site only. The reaction was run at 37°C for 1 hour.
- Gel Electrophoresis: the products of digestion of both the backbone and of the gene of interest are loaded on a gel and then visualized through fluorescence when irradiated with UV light as mentioned above to detect the different digested bands at a determined height.
- Gel band purification: as mentioned above in the second point but paying attention to the correct band. After digestion there are two bands from the backbone and the one that we wanted is the heaviest on the top of the gel because it is the one with the digested backbone portion ready for ligation with our desired insert. There was only one band related for the fragment, so no difficulty in selection.
- Ligation: assembly of the fusion plasmid through the association of the two different digested samples thanks to the same endings cut by the restriction enzymes. To be sure of the correct process we opted for a 1:3 proportion because they were different in lengths. The formula to obtain the quantity of insert to use was: required mass insert (g) = desired insert length /vector molar ratio x mass of vector (g) x ratio of insert to vector lengths. The aim is to create a plasmid with inserted the portion of interest inside the backbone. The enzyme that makes this union possible, in our case, was T4 DNA Ligase of NEB (New England BioLabs). The process requires 1 hour on the bench at room temperature.
- Transformation: ligation product is added to homemade bacterial (E.coli) competent cells. They are initially put on ice and then after heat shock (45'' at 42°C) they are grown in a shaking thermoblock at 37°C. The sudden increase of temperature during the heat shock allows to create pores in the cell membranes of bacteria and the plasmids DNA can enter easily in the bacterial cells. Before the one-hour growth was added S.O.C medium

(inVitrogen, catalogue number: 15544034) to obtain maximal transformation efficiency. After the growth, the cells are plated on a LB agar plate with ampicillin 1:1000 (stock of 100 mg/mL) over night (o/n) at 37°. The presence of ampicillin is fundamental for the selection of the bacteria transformed with plasmids with the inserted gene of interest. We used two different plates, one with the backbone alone that closes up in the absence of the digested fragment (autoligation) plate and the other with the desired cloning product. If the cloning worked well, the expectations is to find on the plate many colonies of cells transformed with the new cloned plasmids and no or few colonies in the autoligation plate. If this scenario was not revealed with also many colonies in the autoligation plate we had to perform colony PCR with the colonies in the plate with the cloned plasmids. Colony PCR is performed picking up single colonies from the plate to allow the PCR run with the DNA presented in each colony. After successive electrophoretic run, through UV visualization of the agarose gel, would be possible to determine whether the desired plasmid was present in each colony thanks to the known size of this.

- The day later, the plates are controlled to assert the growth of bacterial colonies. If there is a good ratio between colonies in the desired plate with the plasmids cloned and colonies in the autoligation plate) it is possible to move forward. A single colony is chosen and picked up in order to be expanded o/n in LB medium (prepared dissolving the powder of Luria Bertani (Lennox) of LLG in deionized water, catalogue number: BOD02549) with ampicillin selection to let the growth only for the correct plasmids. Then through the use of kit of NucleoSpin Plasmid from Macherey-Nagel was possible to isolate any plasmid from E. coli hosts. At the end of the protocol was possible to determine the concentration of the sample with the same method above mentioned with Nanodrop. After obtaining the isolated plasmids would be possible to analyse them through sequencing to assess the presence of the cloned plasmid.

Details of the techniques

1) Insertion of the fluorescent proteins in the empty mammalian vector

The first important step was to identify which restriction enzymes to use to cut the plasmids in order to have the same ends. In this case we used EcoRI-HF (NEB, catalogue number: R3101L) and HindIII-HF (NEB, catalogue number: R3104L) as restriction enzymes. We had to include the sequences of cut of the restriction enzymes in the primers sequences to allow the formation through PCR of the suitable sequences that will be cleaved during the digestion process. We had to add the restriction sites in the primers because they were not present in the original sequence of the

plasmids containing the gene that we wanted to insert in the mammalian vector. The choice of these two restriction sites was headed by the position of them inside the empty mammalian vector. They were right after the 3x FLAG sequence and it was a comfortable position for the insertion of the desired sequences. We started amplifying *mCherry* and *mNeon Green* with PCR reactions. These two were sequences coding for fluorescent proteins and the plasmids were provided to us by a group of collaborators.

In both the forward (FWD) and reverse (REV) primers sequences were added a tail of guanine (G) to clamp the primer due to the strong bonding of G through three hydrogen bonds helping the stability. The sequences immediately after the restriction sites were the complementary sequences to the desired fragments that we wanted to amplify. For the FWD primers we had ATG codon that coded for methionine that is the first amino acid present in a coding sequence. In the REV primers instead, we had the sequences complementary to the end of the desired insert. This helped the correct amplification of the sequences through PCR. We had to insert also in the REV primer sequences a stop codon (in yellow) to prevent the translational fusion of the vector sequences. Underlined we have the primer sequences that paired with the *mCherry* and *mNeon Green* sequences.

The primer sequences with the pairing on the sequences to be amplified were:

- mCherry FWD: GGGG **A|AGCTT** ATGGTGAGCAAGGGCGAG
- mCherry REV: GGGG **G|AATTC**CCGCTGCCTCACTTGTACAGCTCGTCCATGC
- mNeon Green FWD: GGGG **A|AGCTT** ATGGTCTCGAAGGGTGAGG
- mNeon Green REV: GGGG **G|AATTC**CC GCT GCC TCA CTTGTACAGCTCATCCATACCCATC

In bold character are highlighted the consensus sequences for the cleavage of the restriction enzymes (blue for EcoRI-HF and brown for HindIII-HF). The font "|" indicates the exact site of cleavage.

mCherry PCR programme

The PCR mix, with a total volume of 25 µL, was the following:

- 12,5 µL of Q5 High-Fidelity 2X Master Mix
- 1,25 µL of FWD primer
- 1,25 µL of REV primer
- 1 µL of template DNA concentrated 1 ng/µL
- 9 µL of H_2O

	Temperature	Time
First denaturation	98°C	30''
Denaturation	98°C	10''
Annealing	69°C	20''
Elongation	72°C	20''
Final elongation	72°C	2'
Incubation	4°C	∞

Table 3 mCherry PCR programme.

The coloured steps are repeated for 35 cycles.

mNeon Green PCR programme

The Melting temperature changed for *mNeon Green* due to different primers, so they had a different annealing temperature.

The PCR mix was the same of mCherry with the differences in the template added with mNeon Green that replaced mCherry and the two primers.

	Temperature	Time
First denaturation	98°C	30''
Denaturation	98°C	10''
Annealing	67°C	20''
Elongation	72°C	20''
Final elongation	72°C	2'
Incubation	4°C	∞

Table 4 mNeon Green PCR programme.

The coloured steps are repeated for 35 cycles.

Then we digested the PCR products and the backbone p3xFlag-CMV-7.1 empty with restriction enzymes EcoRI-HF and HindIII-HF.

At the end of ligation (mentioned in the general procedures) we had mammalian plasmids with *mCherry* and *mNeon Green*.

2) Formation of the fusion plasmids with fluorescent proteins and the different versions of FOXE1

We then used the obtained plasmids for the creation of FOXE1 tagged with fluorescent proteins. We wanted to put *mCherry* and *mNeon Green in frame* with the FOXE1 plasmids. For this type of PCR we planned to remove the stop codon at the end of *mCherry* and *mNeon Green* gene insertion. This because we wanted the expression of these kinds of plasmids in mammalian cells. Plasmids with an open reading frame with the fluorescent proteins and the different version of FOXE1. So, we designed two new REV primers. A linker sequence was added in the REV primers between the fluorescent proteins and the FOXE1 versions. It is important the choice of the right linker to ensure that proteins are *in frame*. It also provided many other functions, such as maintaining cooperative inter-domain interactions and preserving biological activity.

- mCherry FOR: GGGG **A|AGCTT** ATGGTGAGCAAGGGCGAG
- mCherry REV: GGGG **G|AATTC** **CCGCT** GCC CTTGTACAGCTCGTCCATGC
- mNeon Green FOR: GGGG **A|AGCTT** ATGGTCTCGAAGGGTGAGG
- mNeon Green REV: GGGG **G|AATTC** **CCGCT** GCC CTTGTACAGCTCATCCATACCCATC

Underlined we have the primer sequences that paired with the *mCherry* and *mNeon Green* sequences. In bold character are highlighted the consensus sequences for the cleavage of the restriction enzymes (blue for EcoRI-HF and brown for HindIII-HF). The font "|" indicates the exact site of cleavage. The tail of G is present again with the same intent already explained. The linker sequence is coloured in pink.

We digested the whole amount of PCR and the backbone p3xFlag-CMV-7.1 with FOXE1 14 Ala WT (provided by the group of prof. Persani from IRCCS Istituto Auxologico Italiano) with EcoRI-HF and HindIII-HF.

At the end of ligation, we had the first mammalian plasmids with the fusion of mCherry and mNeon Green with FOXE1 14 Ala WT.

In the end we completed the sequence of plasmids with mCherry and mNeon Green with the other version of FOXE1 (14 Ala L107V, 16 Ala WT and 16 Ala L107V). We decided to cut these vectors

provided by the group of prof. Persani from IRCCS Istituto Auxologico Italiano and then to insert the digested fragments of mCherry and mNeon Green sequences into the vectors carrying the different FOXE1 variants.
Digestion worked as above for the version with 14 Ala WT. HindIII-HF and EcoRI-HF was used as restriction enzymes.
After ligation, we had finally all the version of FOXE1 with mCherry and mNeon Green fluorescent tag.

B. Cell line and transfection

Cell line

- NTHY-ORI 3-1 is a cell line derived from immortalized human thyroid follicular epithelial cells (ECACC 90011609).

  NTHY-ORI 3-1 cells were grown in RPMI-1640 medium, supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. Cells were cultured at 37°C in humidified 5% CO2 environment and were routinely tested for Mycoplasma.

Transient transfection

Transient transfection consists in a process that allows the insertion of foreign nucleic acids into cells in order to obtain the desired product. This process is called transient because the inserted nucleic acids remain extra-chromosomal and are maintained for few days. It is performed usually to study functionality and regulation or protein production [72]. Different methods can be used to transfect cells. Liposomes are vesicles that can interact with the cell membrane helping the nucleic acids transfer to cells [73]. Lipofection is one of the most used methods for transient transfection.
400.000 cells were seeded in 6 well plates. To obtain this desired amount we had to spread a defined amount (volume or cell number) of cells. The number of wells is determined based on how many conditions we wanted to study. We started removing the medium from the cells and we added trypsin to detach them. Then we resuspended in fresh medium (like the one we had just removed) and centrifugated to obtain a cell pellet. Then we counted the cells to then plate 400.000 cells in each well. For the cell counting was necessary to mix 1:1 (20 uL each) the cells in the medium and trypan blue. Trypan blue is a stain able to bind selectively to dead cells. It was useful to discern between living and dead cells. 10 uL of the mixtures were placed in a Burker chamber placed under the microscope and then the living cells were counted by the operator. Knowing how many living cells are present we then were able to relate the quantity of living cells in a mL and then determine how many mL of cells and of medium weneeded to seed. The day after the medium was changed to OPTI-MEM. This choice is driven by the recommendation to use OPTI-MEM (due to the serum-

free nature of the medium) during transfection with Lipofectamine 2000 (Invitrogen, catalogue number: 11668019) and this was our scenario. Moreover, the "transfection mixture" (1 ug of plasmid DNA, 3 uL of Lipofectamine 2000) was diluted in 200 uL of OPTI-MEM. Cells were incubated with transfection mixture for 5 hours. Then medium was replaced by RPMI-1640 supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin and cells were cultured for further 24 hours before processing. Then the day after all the following experiments were performed with their related protocols.

C. Western blot

Western blot is a semi-quantitative technique used to assess the presence of specific proteins and it is possible to provide a relative comparison of proteins levels compared to a known standard [74]. The cells for these experiments were prepared as explained in the transient transfection section of Cell line and transfection (point B of Materials and methods). The process requires 6 steps:

- Sample preparation: firstly the medium was removed from the cells and a quick wash with PBS was performed. Then the transfected cells expressing the protein of interest were lysed with RIPA Lysis Buffer (0,5M Tris-HCl, pH 7,4, 1,5M NaCl, 2,5% deoxycholic acid, 10% NP-40, 10mM EDTA) (Millipore, catalogue number: 20-188) supplemented with Complete Mini protease and phosphatase inhibitor cocktails (Roche, catalogue numbers: 11836153001) to preserve the integrity of proteins. The lysate was subsequently sonicated to obtain the complete breakdown of cell membranes. Sample's protein amount was determined with Pierce bicinchonic acid (BCA) assay kit. A purple water-soluble complex is formed and can be quantified with a spectrophotometer by reading absorbance at 560 nm. The concentration of proteins was measured to be sure of loading the same amount for each sample.
- Protein Electrophoresis: this method exploits the use of a positive electrode to attract the negatively charged proteins that migrates differently according to their molecular weights. Proteins are all negatively charged due to the SDS reagent in the RIPA Lysis Buffer. SDS can interact with proteins with a constant ratio. Bigger proteins will migrate slowly in the gel respect to smaller ones (in terms of molecular weights). The samples are loaded into a polyacrylamide precast gradient gel NuPAGE™ 4 to 12%, Bis-Tris, 1.0–1.5 mm (ThermoFisher, catalogue number: NP0321BOX) inserted in the inner chamber of the apparatus and filled with NuPAGE™ MOPS SDS Running Buffer (ThermoFisher, catalogue number: NP0001) and 500 uL of NuPAGE™ Antioxidant (ThermoFisher, catalogue number: NP0005) to facilitate the electricity transmission. Also, the outer

chamber needs to be fully covered by the buffer. The scheme of the apparatus is illustrated in Figure 8. Bis-Tris chemistry provides a neutral (pH 7.0) environment during electrophoresis. This condition may provide better sample integrity and stability of the gel. This helped to reduce protein modifications and to produce sharp band resolution (ref. from https://www.thermofisher.com/it/en/home/life-science/protein-biology/protein-gel-electrophoresis/protein-gels/nupage-bis-tris-gels.html). The choice of a gradient gel (4 to 12%) is dictated by an accurate protein separation. The larger pore size toward the top of the gel (4%) permits resolution of larger molecules, while pore sizes that decrease toward the bottom of the gel (12%) restrict excessive separation of small molecules. We wanted to see clearly the difference between the bands that were at 45,48 and 54 kDa. These molecular weights are the ones related to the proteins of our interest.

- Membrane transfer: the different proteins are transferred out of the gel into a solid nitrocellulose membrane with the same rationale regarding electric conductivity of electrophoresis. The only difference in the methodology was that this step was conducted in a dry environment.
- Immunoblotting: composed of three stages: blocking, incubation with primary antibody and incubation with secondary antibody.

  Blocking helps to prevent non-specific antibodies-binding to the membrane. It also reduces the background signal. It is composed by 5% milk in TBST solution.

  The primary antibody, dissolved in the blocking buffer is added before an overnight incubation at 4°C with soft shaking.
- Detection: the day after a one-hour incubation step with HRP conjugated secondary antibody (Merck Millipore) in 5% milk is performed. The detection is achieved with Westar Supernova (Cyanagen, catalogue number: XLS3P) chemiluminescent substrates with Azure Biosystem C400 camera. It is based on chemiluminescence. HRP conjugated to the secondary antibody reacts with its substrate luminol that is contained in the Westar Supernova. This reaction brings to a luminol product in excited state that is able to emit light.
- Quantification: the analysis of densitometry was performed with FIJI [75]. This programme is able to quantify each gel band relatively to the levels of black. Bands more intense are related to more expressed proteins.

Before performing the control on the membrane, stripping is necessary. It consists of the elimination of residues of antibodies before another incubation. The product used is Restore™ Western Blot Stripping Buffer (ThermoFisher, catalogue number: 21059) and the ideal time of incubation is maximum 30 minutes. Loading control antibody is useful to assess the efficiency of

the process of western blot. It is possible to compare the amounts of protein loaded in each well across the gel. This control helps to determine whether the expression level differences are due to the real activity of the protein transfected or from loading variances.

The process was the same compared to the ones regarding the visualization of the proteins of interest. So, the step from immunoblotting above mentioned to the quantification were repeated in the same ways.

The primary antibodies used in this study were:

- Anti-FLAG (M2, Sigma RRID:AB_259529) diluted 1:2000.
- Anti-GAPDH (sc-25778, Santa Cruz Biotechnology RRID: AB 10167668) diluted 1:1000.

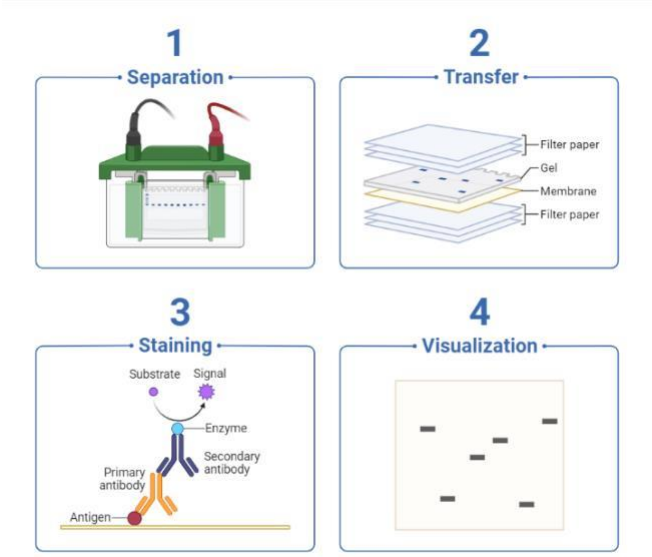

Figure 11 Western blot workflow. *Image from BioRender.*

D. Cell imaging and confocal microscopy

Cell imaging enables the study of cellular dynamics to better understand biological functions. Labelling proteins or intracellular molecules with fluorescent proteins allows the detection through microscope visualisation. When fluorescent proteins are excited by a beam of light or by a laser, they are able to emit the energy absorbed. This energy could be detected by fluorescent channels (in the order of nanoseconds). The fluorescent proteins that we used were mCherry and mNeon Green. They were cloned in frame with the different versions of FOXE1 to obtain the entire fusion proteins for experiments after transient transfection. The visualization of this fluorescence component can be performed with microscopy [76]. We decided to use a confocal and not an epifluorescence microscopy because of its characteristics. In a confocal microscope the light is emitted by a laser. The light with a specific wavelength, related to the fluorescent proteins excited, passes through a "light source pinhole" to become punctiform and is reflected by dichroic filters. These dichroic filters are able to convey emissions with specific wavelengths. Then the excited fluorescent protein emits its fluorescence that passes through the dichroic filters and is directed to the "detector pinhole" in front of the camera. The presence of a pinhole allows to only light from the plane of focus to reach the camera. This is fundamental to reduce the capture of out-of-focus light and to acquire images with improved quality. If the pinhole is more closed, the detected image is more accurate at the expense of intensity. Moreover, it provides an optical sectioning for 3D reconstructions of imaged samples due to the possibility of moving across different layers of the cells in the z-axis thanks to a motorised table.

We performed our experiments with a Nikon Eclipse Ti-E inverted microscope with implemented the confocal setting.

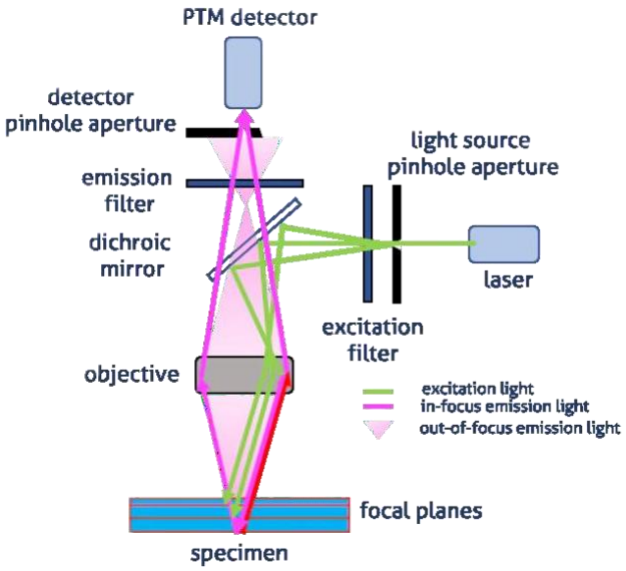

Figure 12 Pinhole effect in confocal microscopy. *Image from Proteintech, IF imaging: Widefield versus confocal microscopy.*

In the above Figure 12 is illustrated the light channelling due to the pinhole.

A confocal microscopy is able to detect images in different channels depending on the wavelengths of excitation and emission. This ability is due to the presence of dichroic filters. They are able to minimise the absorption of light radiation and allow the passage of light at a certain wavelength, i.e. of a very narrow range of frequencies. DAPI's stainings are possible to detect due to emission in the blue range (at a maximum of 461 nm) after absorption of light at a specific wavelength (at a maximum of 358 nm). DAPI is a dye that binds strongly to DNAregion enriched of A-T sequences, so it is a perfect nuclear staining. Then our proteins tagged with fluorescent protein could be detected due to the range of emission related to the fluorophores. In our case we did not need the incubation with dye-labelled antibodies because we set the cloning experiments with the insertion of fluorescent proteins directly in the plasmids with the different versions of FOXE1. So, the ones with mCherry could be detected in the red channel with an absorption of light between 540-590 nm and emission of light in the range of 550-650 nm wavelength. The ones with mNeon Green would be visible in the green channel after the process of excitation (with a maximum at 506 nm of wavelength) and emission (at a maximum of 517 nm of wavelength) of lights of the fluorophores.

The cells for these experiments were prepared as explained in the transient transfection section of Cell line and transfection (point B of Materials and methods).

After the transfection protocol already mentioned we performed this workflow:

- Washing: after the removal of medium from the transfected cells, PBS washes were performed.
- Fixation: it is an important passage because helps to avoid autolysis and maintain the cellular architecture. Samples were fixed in 4% paraformaldehyde (PFA) for 10 minutes.
- Washing: three washes with PBS of 5 minutes each.
- DAPI: it was added for 10 minutes and then removed.
- Washing: three washes with PBS of 5 minutes each as above.
- Mounting: the samples had to be mounted on slides to be viewed under a microscope. The mounting with Vectashield Hard Set with DAPI (DAKO) on a slide avoids dehydration and increments the refractive index [76]. The use of a mounting with additional DAPI helped to get better images. Less background noise was present.
- Acquiring: the images were acquired with Nikon EclipseTi-E inverted microscope with IMA10X Argon-ion laser System (Melles Griot).

E. Dual Luciferase Assay

Luciferase assay is used to understand the activity of a promoter of interest. It is an assay performed in transfected cells. In the transfection cocktail are present different reagents. The most important reagent is characterized by a fusion protein. It is always composed by the gene responsible for the luciferase reporter (firefly), that is placed under the control of the promoter region of the target gene. The target gene was thyroglobulin in our case, so the *TG* promoter was fused with the luciferase reporter. This choice was dictated by the knowledge of FOXE1 as transcription factor is involved in thyroglobulin transcription. In the same transfection cocktail, another plasmid for Renilla constitutive expression as internal control is also present [77]. In the end, the plasmids containing the gene of interest is present too.

The luminometric signal that was detected was proportional to the amount of firefly present. After this, another reagent simultaneously quenches the firefly signal and let the Renilla's related luminescence to be measured.

The cells for these experiments were prepared as explained in the transient transfection section of Cell line and transfection (point B of Materials and methods).

FOXE1 activity was measured with the Dual-Luciferase Reporter Assay System from Promega (catalogue number: E1910). 250 ng of FOXE1 variants were co-transfected with 250ng of Firefly Luciferase reporter under the control of human thyroglobulin promoter, 80 ng of Renilla construct from Promega and if necessary (as in the second experiment) also 250 ng of NKX2-1 expression vector. The addition of NKX2-1 was related to the study of co-activity of FOXE1. These quantities were determined based on the amount plated cells. Empty vector was used as needed to maintain constant total amount of DNA in the transfection experiments. The day after transfection the medium was removed and the cells were firstly washed in PBS. Then they were lysed using the lysis buffer provided by the kit. Passive Lysis Buffer is concentrated 5x, so before using it needs to be diluted. Then 100 µL of diluted Passive Lysis Buffer is added to each well. After harvesting and collecting the lysate, the loading plate is prepared with 30 µL of lysate in each well of the 96-multiwell plate. A duplicate is loaded for each sample.

Then the prepared plate is placed in the Fluoroskan Ascent FL multiplate reader. Luciferase Assay Reagent II allows luciferase reaction while Stop & Glo is responsible for Renilla reaction. The instrument can detect two readings of luminescence. The first is related to luciferase and the second to Renilla. Firstly, the Luciferase Assay Reagent II was injected into the first well and the relative reading was given. The emitted photons had a wavelength of 560 nm. Then the reaction of the firefly luciferase was quenched by the injection of the second reagent Stop & Glo. It was also responsible for the stimulation of the Renilla, so the second reading could be obtained. The related emitted

photons of Renilla had a wavelength of 482 nm.

In the Figure 13 is illustrated the workflow of Dual Luciferase Assay in our case with the relatives substrates.

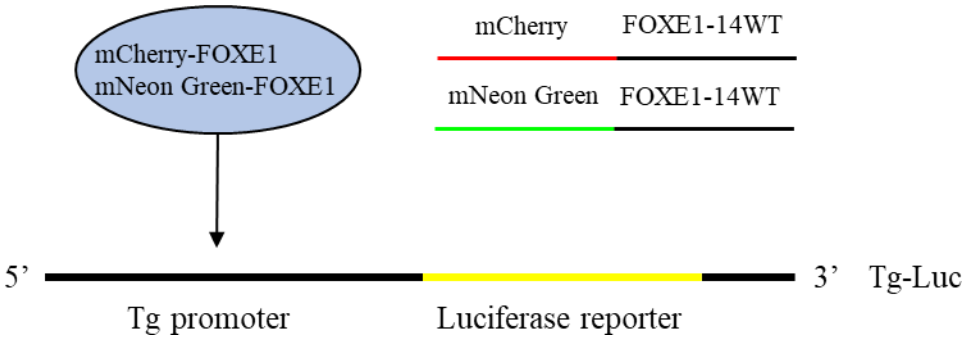

Figure 13 Luciferase assay mode of action.

F. Statistics

The first statistic test that we applied was Shapiro-Wilk normality test. This kind of test showed the data distribution (Gaussian or not). Since our data were not normally distributed, the applied statistics results were made with a one-way Anova test using the variance analysis to determine the existence of a significant difference among medians deriving from more groups. Kruskal-Wallis test is used to obtain the significancy (P value) of the data. A P value < 0.05 was set for the acceptance of significancy.

III. RESULTS

A. Cloning process to obtain the plasmids for the further *in vitro* experiments

The first important step for our studies was to create constructs for the expression in mammal cells of FOXE1 (14 Ala WT, 14 Ala p.L107V, 16 Ala WT and 16 Ala p.L107V) tagged with either mCherry or mNeon Green. Cloning techniques were used to pursue this goal and it was the starting point for the further experiments. The process has been already mentioned in the Material and methods section.

It consisted of two different processes:

1) insertion of the fluorescent tags mCherry and mNeon Green in the mammalian vector pCMV-7.1 in order to create the respective empty plasmids controls (Fig. 14A, B),

2) cloning of the fluorescent tags into the different FOXE1 variants plasmids (Fig. 15A, B).

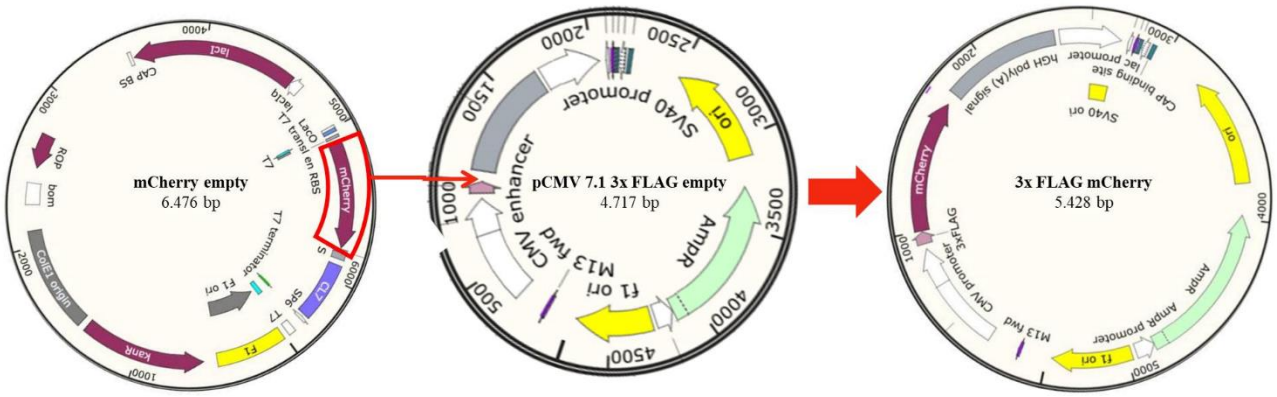

Figure 14A Creation of pCMV 7.1 mCherry empty backbone. *Image readapted from SnapGene.*

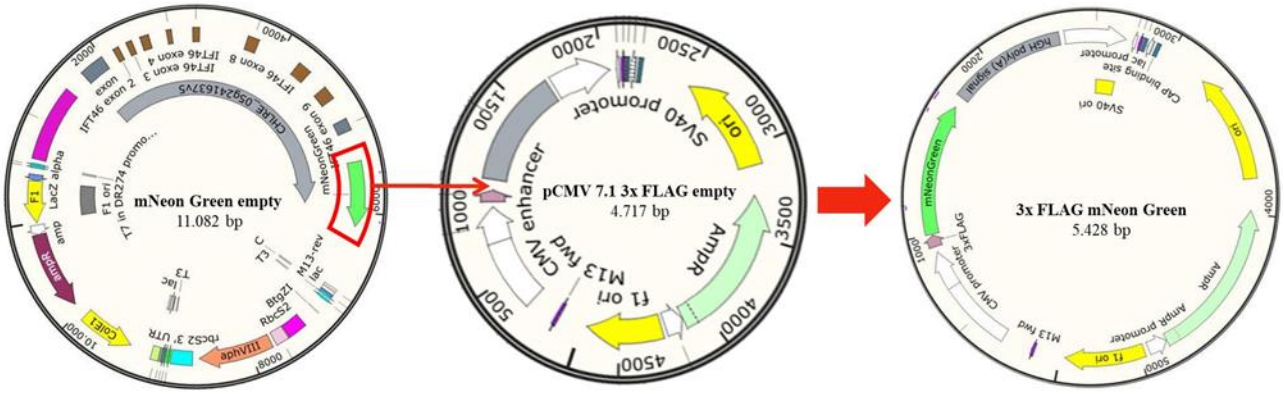

Figure 14B Creation of pCMV 7.1 mNeon Green empty backbone. *Image readapted from SnapGene.*

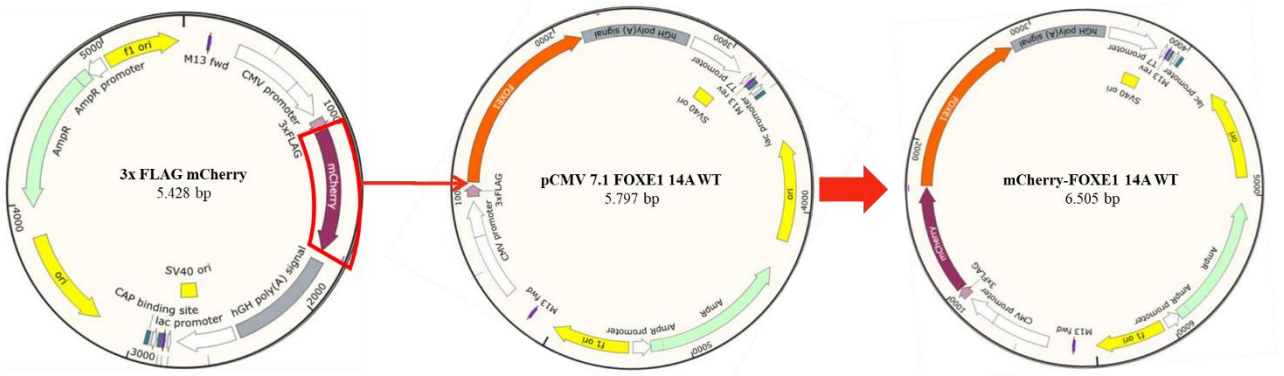

Figure 15A Cloning of mCherry in pCMV7.1-FOXE1 plasmids. *Image readapted from SnapGene.*

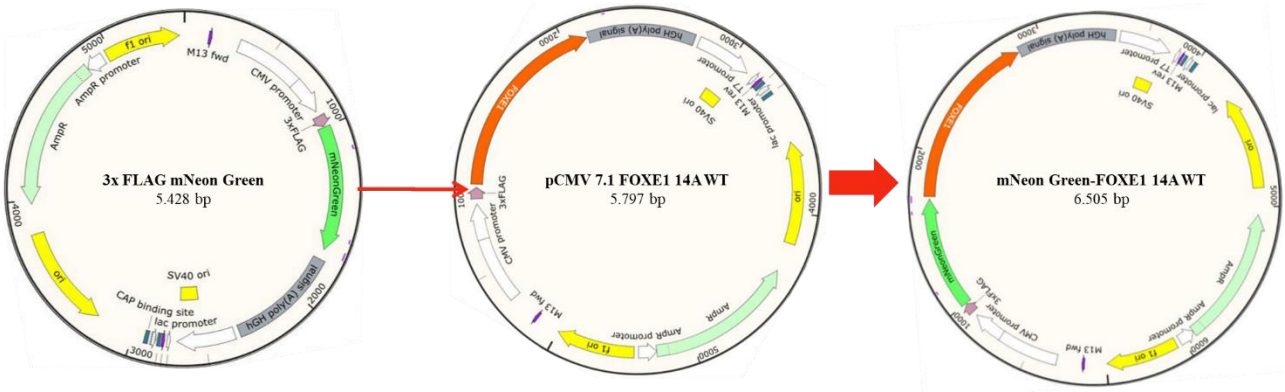

Figure 15B Cloning of mNeon Green in pCMV7.1-FOXE1 plasmids. *Image readapted from SnapGene.*

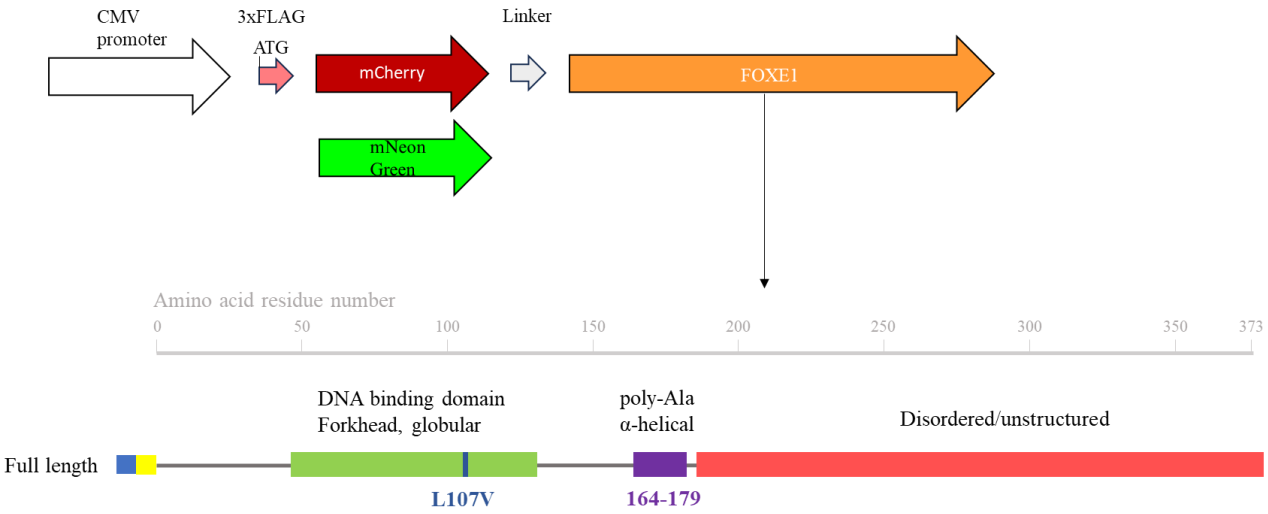

Figure 16 Scheme of the final plasmids cloned. *Illustration of the different portions of the final plasmids with highlighted the CMV promoter, the fluorescent proteins and FOXE1 with its domains.*

The scheme of the obtained plasmids is shown in the Figure 16 with highlighted the most important domains.

3xFLAG-mCherry pCMV7.1	**3xFLAG-mNeon Green pCMV7.1**
3xFLAG-mCherry-FOXE1 14 Ala WT pCMV7.1	**3xFLAG-mNeon Green-FOXE1 14 Ala WT pCMV7.1**
3xFLAG-mCherry-FOXE1 14 Ala L107V pCMV7.1	**3xFLAG-mNeon Green-FOXE1 14 Ala L107V pCMV7.1**
3xFLAG-mCherry-FOXE1 16 Ala WT pCMV7.1	**3xFLAG-mNeon Green-FOXE1 16 Ala WT pCMV7.1**
3xFLAG-mCherry-FOXE1 16 Ala L107V pCMV 7.1	**3xFLAG-mNeon Green-FOXE1 16 Ala L107V pCMV7.1**

Table 5 List of the obtained plasmids after cloning.

The list of the obtained plasmids is summarized in the above Table 5.

All the FOXE1 expression experiments were performed in cells with the tagged plasmids above mentioned. NTHY-ORI cells were used because they are an immortalized line derived from human thyroid follicular epithelia, so they were an ideal model for our experiments.

B. Western Blot analysis did not show any significant differences among the FOXE1 fluorescent tagged plasmids

After we obtained the desired plasmids, we performed *in vitro* experiments to determine if the presence of the fluorescent tags might have induced possible alterations of expression, localisation and/or functionality. As first step we wanted to verify whether the produced protein levels with fluorescent tags were comparable to the FOXE1-FLAG without fluorescent tags. So, we decided to perform different experiments that consisted of transient transfection followed by Western Blot analysis.

We planned different sets of Western Blot on the basis of what we wanted to compare and analyse.

First of all, we wanted to compare the levels of the protein FOXE1 14 Ala WT with FLAG, mCherry and mNeon Green tagged versions (Figure 17). The results showed that the different constructs had the same quantity of proteins produced, as significant differences were not detected among the 14 Ala WTs (P value of 0.0560).

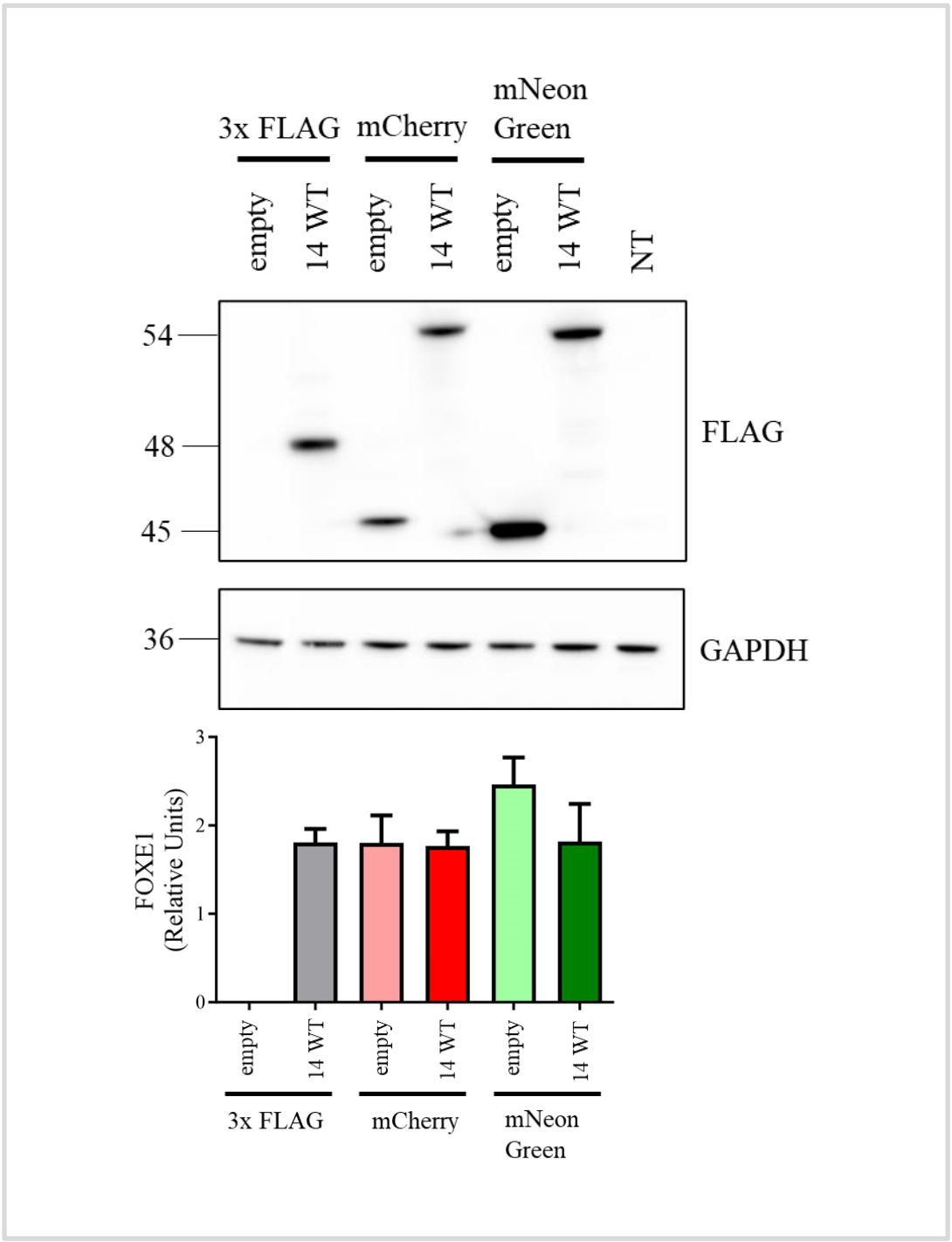

Figure 17 The different tags do not impact the expression of FOXE1 14 Ala. *Representative images and quantification of Western Blot experiments showing the different expression of FOXE1 14 Ala's after transient transfection (n=3).*

After asserting the similarity of the tagged FOXE1 14 Ala WT plasmids expression, we then compared the levels of FOXE1 16 Ala WT FLAG with the ones cloned with mCherry and mNeon Green (Fig. 18).

A P value of 0.07 was calculated and no significant difference was demonstrated. Anyway, a trend could be noticed with the FOXE1 16 Ala WT FLAG that had higher levels of production compared to the fluorescent tagged proteins. To overcome this doubt, it would be necessary to obtain more replicates.

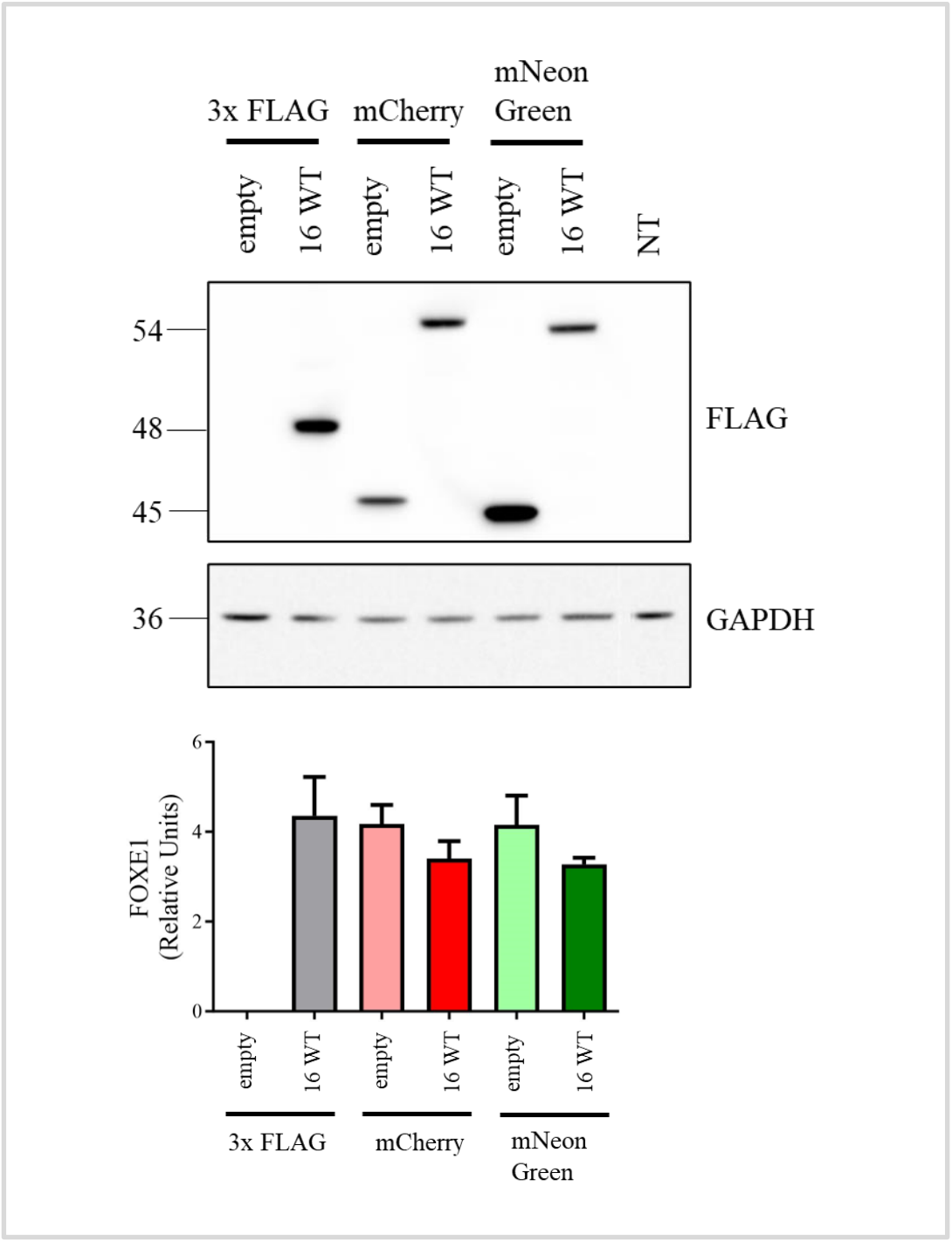

Figure 18 The different tags have a weak impact on the expression of FOXE1 16 Ala. *Representative images and quantification of Western Blot experiments showing the different expression of FOXE1 16 Ala's after transient transfection (n=3).*

As we did not detect any significant difference in the expression between the 14 and 16 Ala WT FOXE1s, we then moved forward and performed different experiments in order to verify if the previously detected differences in production, localisation and functionality among FOXE1 variants (Grassi et al [38]) were maintained also in the fluorescent-tagged proteins.

We first examined the mNeon Green constructs (Figure 19).

The mNeon Green tagged proteins shown a little, but not significant (P value of 0.4913) increase of the two mutants rather than the respective WT versions. It can also be noticed a major stability of the mNeon empty.

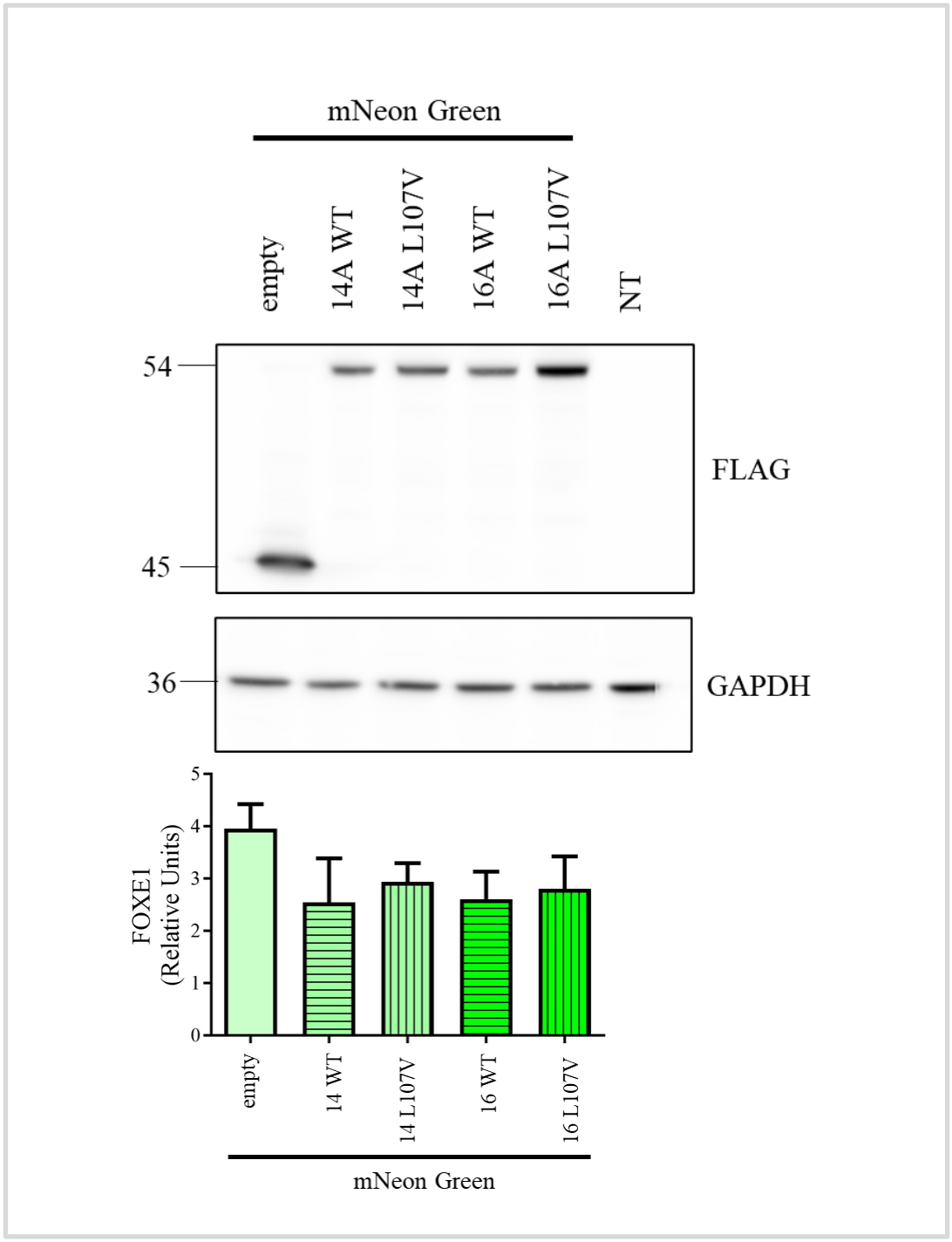

Figure 19 The mNeon Green tagged FOXE1 variants show only slight difference in protein levels. *Representative images and quantification of Western Blot experiments showing the different expression ofmNeon-FOXE1 variants after transient transfection (n=3).*

In the end we compared the levels of FOXE1 (14 Ala WT, 14 Ala L107V, 16 Ala WT and 16 Ala L107V) tagged with mCherry (Figure 20).

The mCherry tagged proteins did not show any variation among them (P value of 0.88789). As already mentioned for mNeon Green empty, there was a more stability also in the mCherry empty.

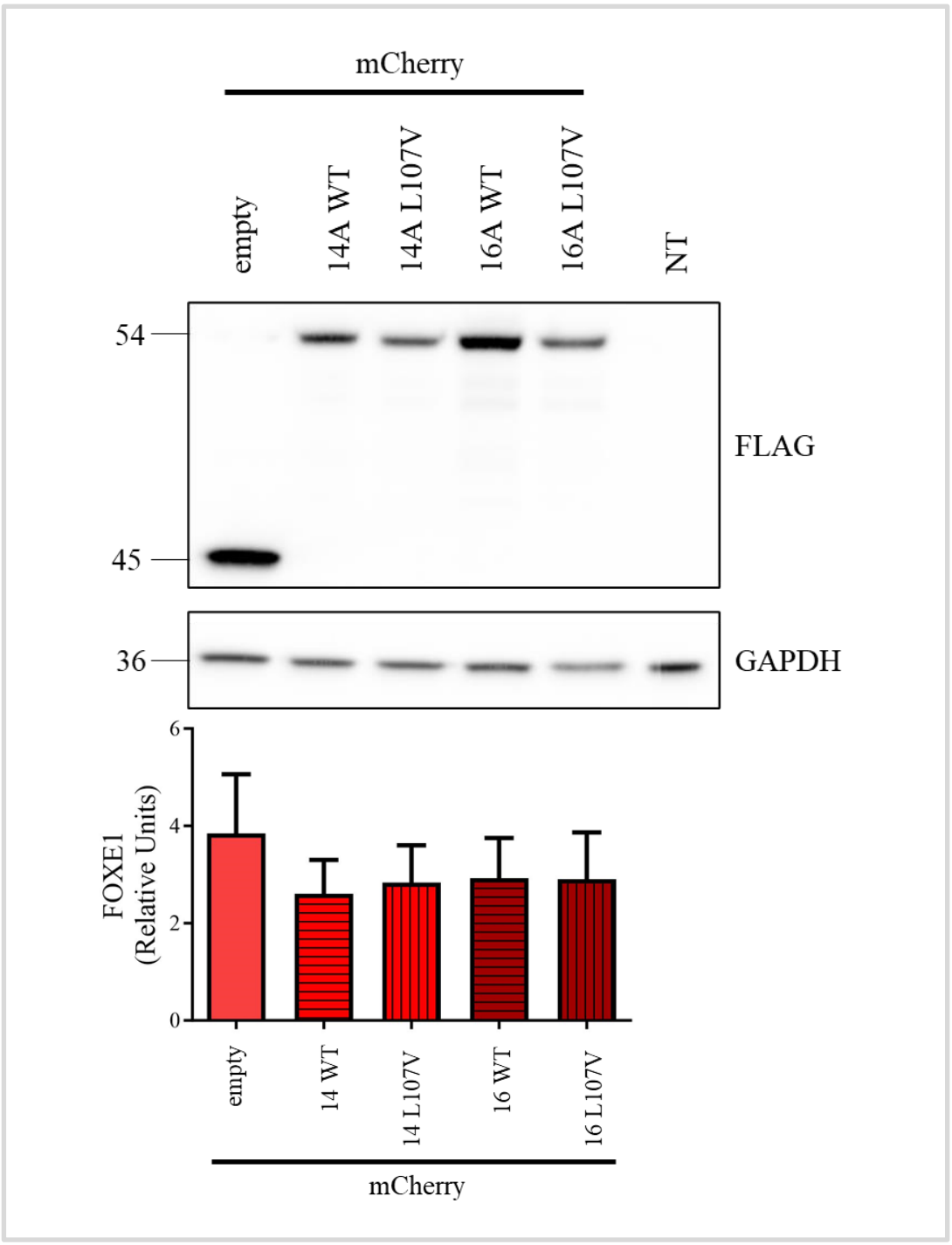

Figure 20 FOXE1 variants tagged with mCherry have a similar levels of protein expression. *Representative images and quantification of Western Blot experiments showing the different expression of mCherry-FOXE1 variants after transient transfection (n=3).*

C. Confocal microscopy demonstrated nuclear morphologies similar to those already seen

Since it was previously shown that variations in the FOXE1 alanine length may influence the protein subcellular localisation (Figure 21) (Grassi et al [38]) and given the slight variation among the fluorescent tagged FOXE1s detected at WB, we decided to investigate the intracellular localisation of the proteins after transient transfection in NTHY-ORI cells.

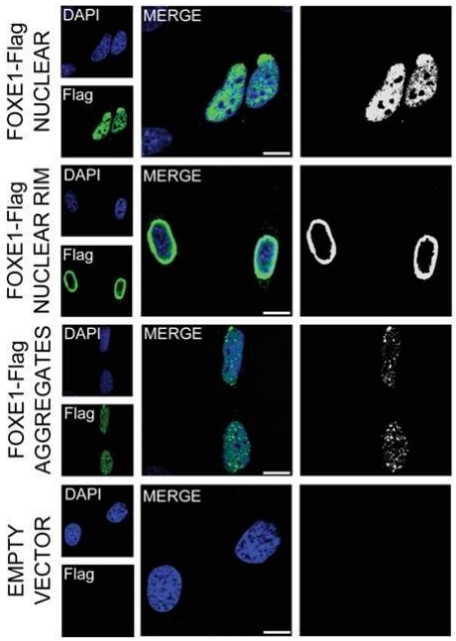

Figure 21 *Image showing the different FOXE1 nuclear morphologies. Readapted from Grassi et al in The length of FOXE1 polyalanine tract in congenital hypothyroidism: Evidence for a pathogenic role from familial, molecular and cohort studies* [38].

We wanted to discriminate the various nuclear morphologies for each version of FOXE1 tagged with mCherry or mNeon Green.

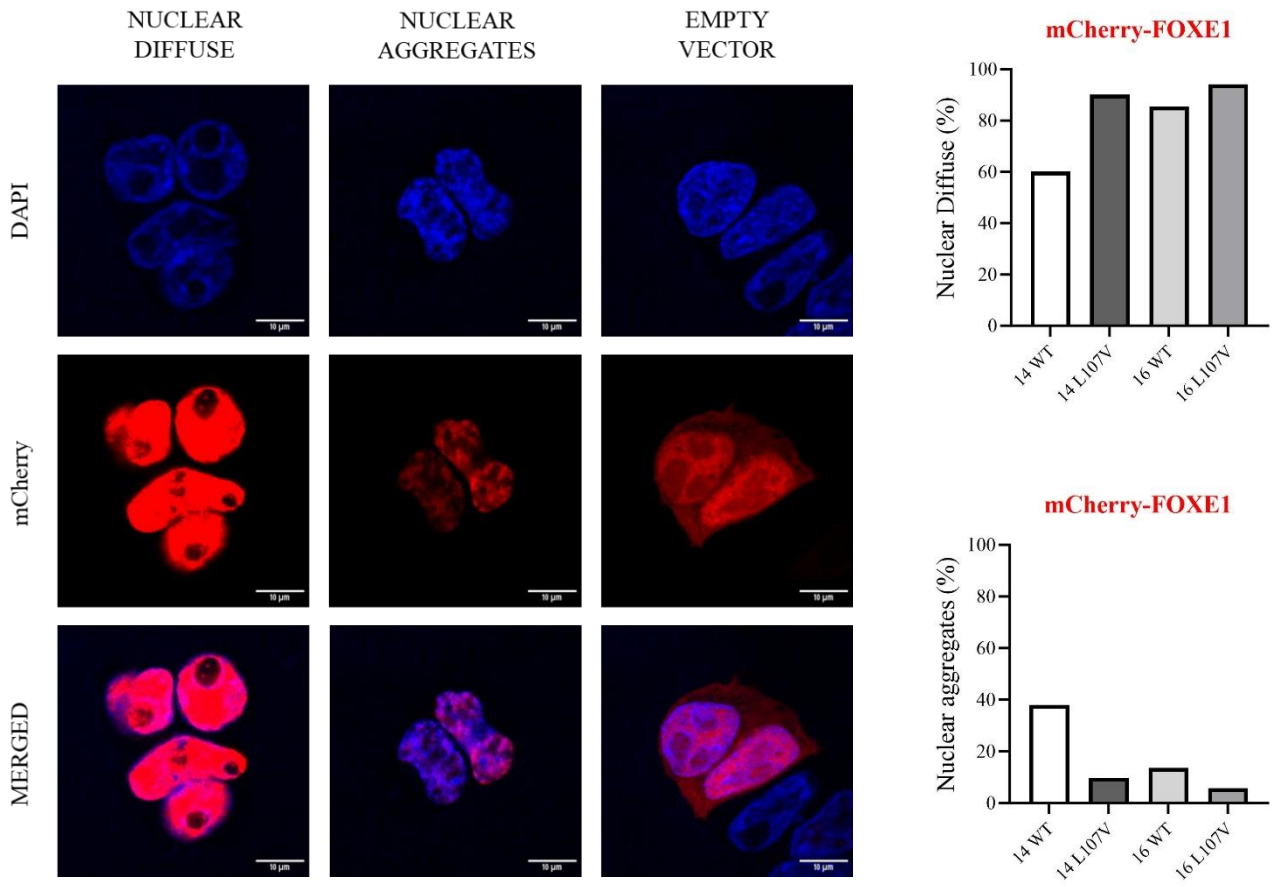

Figure 22 Cell imaging of mCherry-FOXE1 proteins. *Confocal microscopy images (with a scale bar of 10 µm) representing the different mCherry-FOXE1 nuclear pattern compared to the empty vector morphology and relative quantification (n=1; 843 cells analysed).*

Differently than previously reported in the FOXE1-FLAG, the obtained results showed that our proteins do not exhibit the nuclear rim morphology. Only the nuclear diffuse and the nuclear aggregates patterns were recognizable. The empty vector instead has the expected behaviour with a nuclear and cytoplasmic distribution (Figure 22).

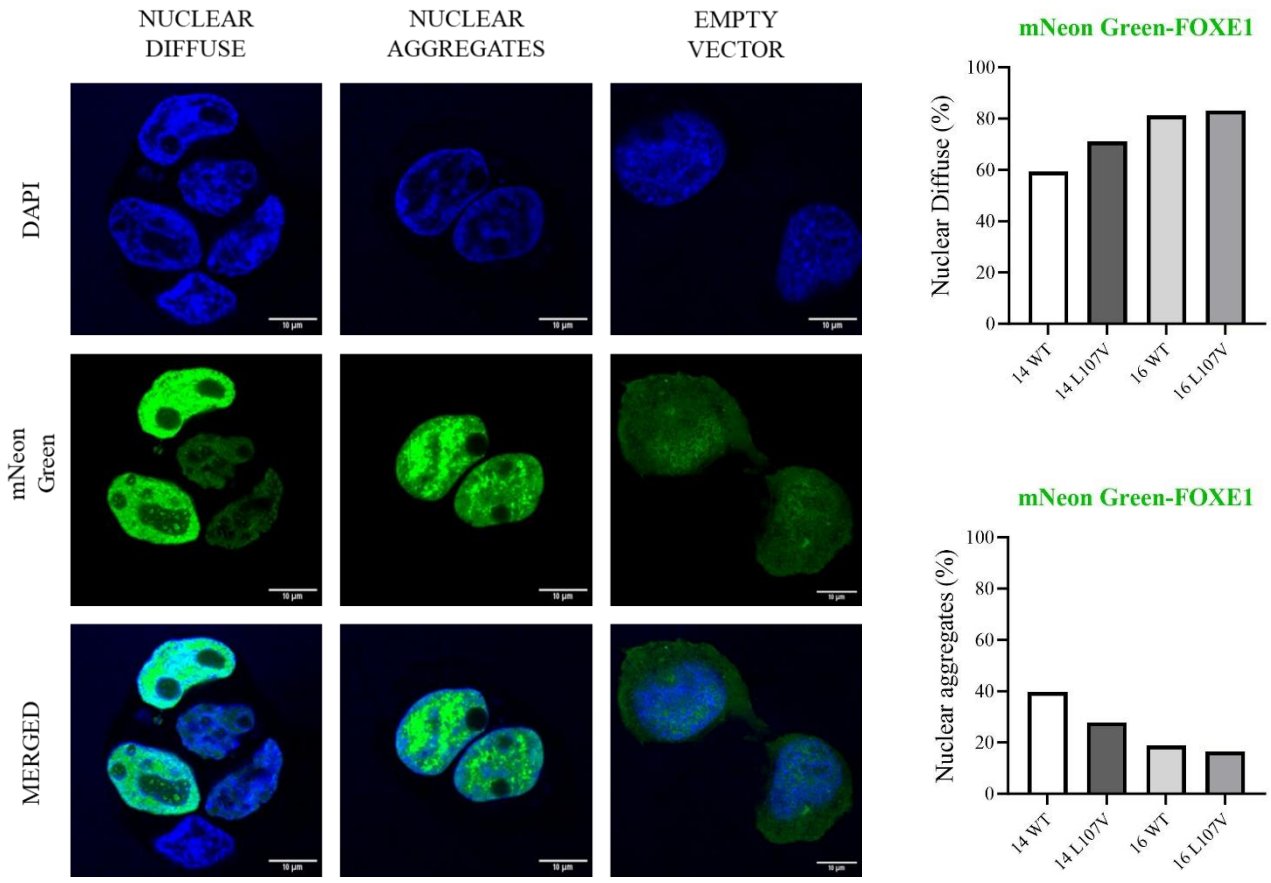

Figure 23 Immunofluorescence of mNeon Green-FOXE1 proteins. *Confocal microscopy images (with a scalebar of 10 µm) representing the different mNeon Green-FOXE1 nuclear pattern compared to the empty vector morphology and relative quantification (n=1; 865 cells analysed).*

The obtained results showed that our proteins do not exhibit the nuclear rim morphology but only the nuclear diffuse and the nuclear aggregates (Figure 23). This absence is similar to the mCherry tagged FOXE1 proteins (Figure 22). The empty vector instead has the expected behaviour with a nuclear and cytoplasmic distribution.

D. Dual Luciferase Assay confirmed the functionality of the constructs with fluorescent proteins

Given the previously detected differences among FOXE1 expression and nuclear morphologies (Grassi et al [38]), as last step, we analysed the transcriptional activity of our proteins by Dual-Luciferase experiments. The aim of this experiment was to test whether our clones with mCherry-FOXE1 and mNeon-FOXE1 had the same activity of the ones without fluorophores on the promoter of thyroglobulin (TG). Preliminary experiments were performed with the 14 alanines isoform WT (wild type). The first two graphs are related to experiments with only FOXE1 (mCherry-14 Ala WT or mNeon Green-14 Ala WT). The other two present the combination of our FOXE1 plasmids (tagged with mCherry or mNeon Green) co-transfected with NKX2-1 plasmid. We measured the ratio between luciferase and renilla activities (luciferase/renilla).

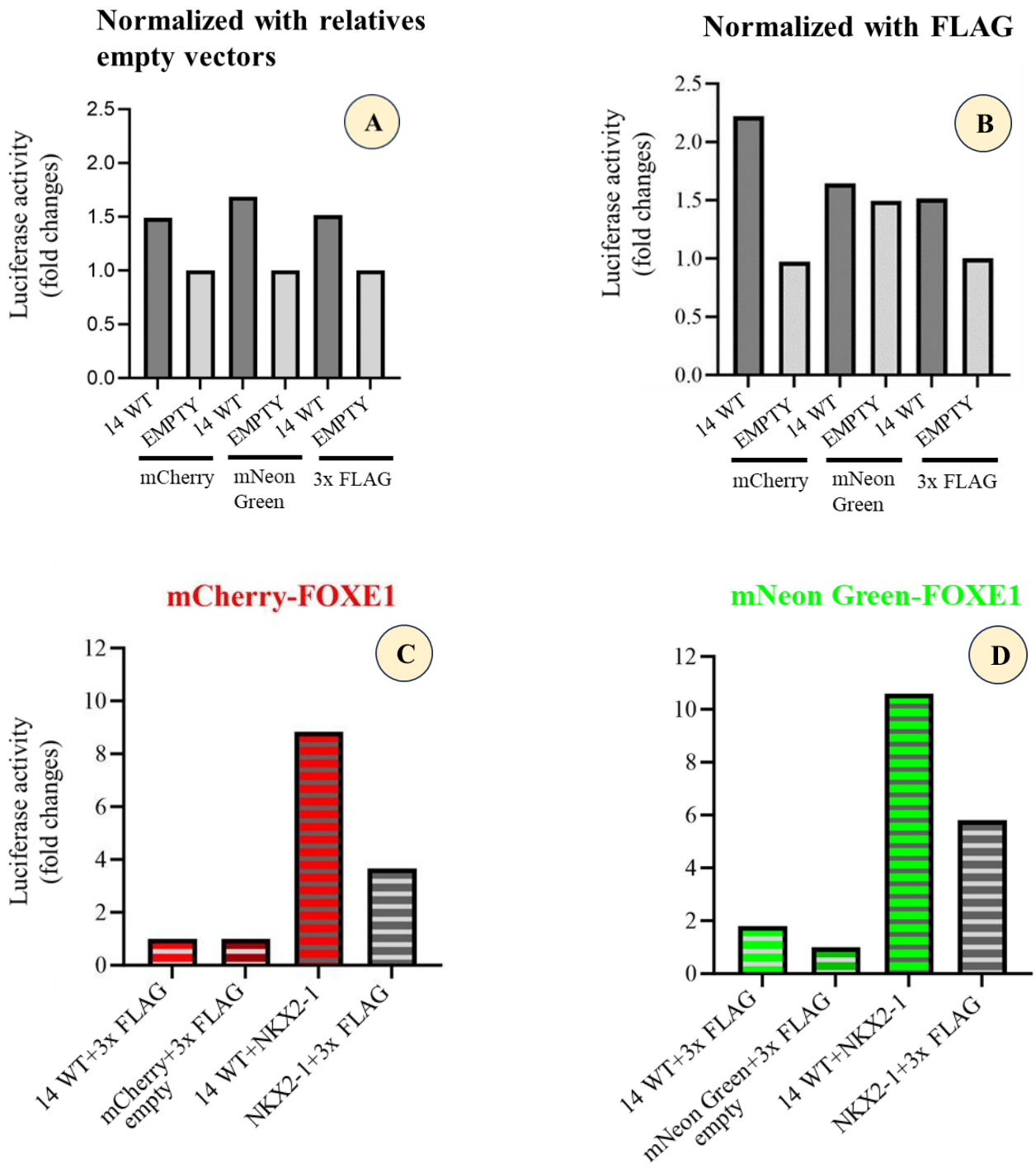

Figure 24 Dual Luciferase assay. *FOXE1 transcriptional activity measured by luc-TG reporter.*

Referring to the Figure 24A, the activity of FOXE1 14 Ala WT tagged with mCherry and mNeon Green is comparable to the 14 Ala WT-FLAG with major evidence for the mNeon tagged protein in relation with the relatives empties. When normalized with the 3x FLAG empty vector we see an augmented activity for the mCherry-tagged FOXE1 14 Ala WT, visible in the Figure 24B. Instead, the interaction with NKX2-1 shows a strong activity by more than 4 times of the mCherry tagged FOXE1 14 Ala WT compared to the FOXE1 14 Ala WT without tag (Figure 24C). The mNeon Green tagged FOXE1 14 Ala WT had about 5 times more activity when combined to NKX2-1

respect that alone (14 Ala WT + 3x FLAG). (Figure 24D)

Any statistics were performed because these results are only preliminary. Triplicates are needed to confirm these first evidence and to apply statistics tests. The first impression was that the activity was not being influenced by the presence of the fluorescent proteins.

IV. DISCUSSION

FOXE1 is an important transcription factor for the development of the thyroid gland, and it is involved in pathologies such as cancer and congenital hypothyroidism. As for other pioneer transcription factors, the DNA-binding-domain (DBD) of FOXE1 binds target genes to be transcribed within nucleosomes by conditioning chromatin structure and making them available for the binding of other transcription factors. This happens, for example, for the key thyroid hormone precursor gene, thyroglobulin [39]. The current thesis project was born from the discovery, by Persani group, of a family in which the variant p.L107V in the DNA-binding domain (DBD) of FOXE1 in heterozygosis seemed to cause CH when associated with a poly-ala domain of 14 alanines [38]. In this context we put our interest on the effects of poly-alanine variations in FOXE1 and with the hypothesis of performing more advanced live imaging studies in cells, we explored the effects on function of a fluorescent label (mNeon green and mCherry). Therefore, we engineer by genetic fusion at the N-terminus the fluorescent label on FOXE1 WT and variants vectors and checked for changes in expression, cellular localisation and transcriptional activity.

Our *in vitro* studies revealed that the addition of fluorescent tags to FOXE1 partially reproduces previously reported among the unlabelled FOXE1 variants [38]. We did not notice significant differences of expression in Western blot analysis and transcriptional activity, also in presence of the other co-transcription factor NKX2.1. However, we observed a different phenotype in nuclear localisation assays in the confocal microscopy images. In our FOXE1 version labelled with fluorophores (mCherry and mNeon Green) we did not find the nuclear ring distribution shown in WT FOXE1, but only a nuclear diffuse distribution and presence of local aggregates. This variation could be referred to the addition of the fluorescent proteins that have approximately the same molecular weight of FOXE1 at the N-terminus. The addition of the label could have partially impaired the conformation of the protein at the N-terminus and interaction with some nuclear factors, determining the nuclear rim localisation observed for the WT FOXE1.

We surmise that a possible future strategy to label FOXE1 could be the addition of the label at the C-terminus instead of the N-terminus (close to the DBD domain) or using smaller tags, for example alpha tags in combination with co-expressed fluorescent anti-alpha tag nanobodies.

V. BIBLIOGRAPHY

1. Armstrong, M., Asuka, E., and Fingeret, A. (2023). Physiology, Thyroid Function. In.
2. Shahid, M.A., Ashraf, M.A., and Sharma, S. (2023). Physiology, Thyroid Hormone. In.
3. Grasberger, H., and Refetoff, S. (2011). Genetic causes of congenital hypothyroidism due to dyshormonogenesis. Curr. Opin. Pediatr. *23*, 421–428. 10.1097/MOP.0b013e32834726a4.
4. Mullur, R., Liu, Y.-Y., and Brent, G.A. (2014). Thyroid hormone regulation of metabolism. Physiol. Rev. *94*, 355–382. 10.1152/physrev.00030.2013.
5. Di Cosmo, C., Liao, X.-H., Dumitrescu, A.M., Philp, N.J., Weiss, R.E., and Refetoff, S. (2010). Mice deficient in MCT8 reveal a mechanism regulating thyroid hormone secretion. J. Clin. Invest. *120*, 3377–3388. 10.1172/JCI42113.
6. Gereben, B., Zavacki, A.M., Ribich, S., Kim, B.W., Huang, S.A., Simonides, W.S., Zeöld, A., and Bianco, A.C. (2008). Cellular and molecular basis of deiodinase-regulated thyroid hormone signaling. Endocr. Rev. *29*, 898–938. 10.1210/er.2008-0019.
7. Chaker, L., Razvi, S., Bensenor, I.M., Azizi, F., Pearce, E.N., and Peeters, R.P. (2022). Hypothyroidism. Nat. Rev. Dis. Prim. *8*, 30. 10.1038/s41572-022-00357-7.
8. LaFranchi, S.H., and Austin, J. (2007). How should we be treating children with congenital hypothyroidism? J. Pediatr. Endocrinol. Metab. *20*, 559–578. 10.1515/jpem.2007.20.5.559.
9. Haddow, J.E., Palomaki, G.E., Allan, W.C., Williams, J.R., Knight, G.J., Gagnon, J., O'Heir, C.E., Mitchell, M.L., Hermos, R.J., Waisbren, S.E., et al. (1999). Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. N. Engl. J. Med. *341*, 549–555. 10.1056/NEJM199908193410801.
10. Segni, M. (2000). Congenital Hypothyroidism. In, K. R. Feingold, B. Anawalt, M. R. Blackman, A. Boyce, G. Chrousos, E. Corpas, W. W. de Herder, K. Dhatariya, K. Dungan, J. Hofland, et al., eds.
11. Peters, C., van Trotsenburg, A.S.P., and Schoenmakers, N. (2018). DIAGNOSIS OF ENDOCRINE DISEASE: Congenital hypothyroidism: update and perspectives. Eur. J. Endocrinol. *179*, R297–R317. 10.1530/EJE-18-0383.
12. Persani, L., Rurale, G., de Filippis, T., Galazzi, E., Muzza, M., and Fugazzola, L. (2018). Genetics and management of congenital hypothyroidism. Best Pract. Res. Clin. Endocrinol. Metab. *32*, 387–396. 10.1016/j.beem.2018.05.002.
13. Rastogi, M. V, and LaFranchi, S.H. (2010). Congenital hypothyroidism. Orphanet J. Rare Dis. *5*, 17. 10.1186/1750-1172-5-17.
14. van Trotsenburg, P., Stoupa, A., Léger, J., Rohrer, T., Peters, C., Fugazzola, L., Cassio, A., Heinrichs, C., Beauloye, V., Pohlenz, J., et al. (2021). Congenital Hypothyroidism: A 2020- 2021 Consensus Guidelines Update-An ENDO-European Reference Network Initiative Endorsed by the European Society for Pediatric Endocrinology and the European Society for Endocrinology. Thyroid *31*, 387–419. 10.1089/thy.2020.0333.

15. de Mendoza, A., and Sebé-Pedrós, A. (2019). Origin and evolution of eukaryotic transcriptionfactors. Curr. Opin. Genet. Dev. *58–59*, 25–32. 10.1016/j.gde.2019.07.010.
16. Johnson, P.F., and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. *58*, 799–839. 10.1146/annurev.bi.58.070189.004055.
17. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J.,Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650– 665. 10.1016/j.cell.2018.01.029.
18. Duttke, S.H.C. (2015). Evolution and diversification of the basal transcription machinery. Trends Biochem. Sci. *40*, 127–129. 10.1016/j.tibs.2015.01.005.
19. Kawasaki, K., and Fukaya, T. (2023). Functional coordination between transcription factor clustering and gene activity. Mol. Cell *83*, 1605-1622.e9. 10.1016/j.molcel.2023.04.018.
20. Wilkinson, A.C., Nakauchi, H., and Göttgens, B. (2017). Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. Cell Syst. *5*, 319–331. 10.1016/j.cels.2017.07.004.
21. Shamovsky, I., and Nudler, E. (2008). New insights into the mechanism of heat shock response activation. Cell. Mol. Life Sci. *65*, 855–861. 10.1007/s00018-008-7458-y.
22. Benizri, E., Ginouvès, A., and Berra, E. (2008). The magic of the hypoxia-signaling cascade. Cell. Mol. Life Sci. *65*, 1133–1149. 10.1007/s00018-008-7472-0.
23. Evan, G., Harrington, E., Fanidi, A., Land, H., Amati, B., and Bennett, M. (1994). Integrated control of cell proliferation and cell death by the c-myc oncogene. Philos. Trans. R. Soc. London. Ser. B, Biol. Sci. *345*, 269–275. 10.1098/rstb.1994.0105.
24. Latchman, D.S. (1996). Inhibitory transcription factors. Int. J. Biochem. Cell Biol. *28*, 965– 974. 10.1016/1357-2725(96)00039-8.
25. Pan, Y., Tsai, C.-J., Ma, B., and Nussinov, R. (2010). Mechanisms of transcription factor selectivity. Trends Genet. *26*, 75–83. 10.1016/j.tig.2009.12.003.
26. Li, M., Hada, A., Sen, P., Olufemi, L., Hall, M.A., Smith, B.Y., Forth, S., McKnight, J.N., Patel, A., Bowman, G.D., et al. (2015). Dynamic regulation of transcription factors bynucleosome remodeling. Elife *4*. 10.7554/eLife.06249.
27. Yu, X., and Buck, M.J. (2020). Pioneer factors and their in vitro identification methods. Mol. Genet. Genomics *295*, 825–835. 10.1007/s00438-020-01675-9.
28. Zaret, K.S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. Annu. Rev. Genet. *54*, 367–385. 10.1146/annurev-genet-030220-015007.
29. Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr. Opin. Genet. Dev. *37*, 76–81. 10.1016/j.gde.2015.12.003.
30. Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241. 10.1101/gad.176826.111.
31. Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell

152, 1237–1251. 10.1016/j.cell.2013.02.014.

32. Kostopoulou, E., Miliordos, K., and Spiliotis, B. (2021). Genetics of primary congenital hypothyroidism-a review. Hormones (Athens). *20*, 225–236. 10.1007/s42000-020-00267-x.
33. Nilsson, M., and Fagman, H. (2017). Development of the thyroid gland. Development *144*, 2123–2140. 10.1242/dev.145615.
34. Clifton-Bligh, R.J., Wentworth, J.M., Heinz, P., Crisp, M.S., John, R., Lazarus, J.H., Ludgate, M., and Chatterjee, V.K. (1998). Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. Nat. Genet. *19*, 399–401. 10.1038/1294.
35. Santarpia, L., Valenzise, M., Di Pasquale, G., Arrigo, T., San Martino, G., Cicciò, M.P., Trimarchi, F., De Luca, F., and Benvenga, S. (2007). TTF-2/FOXE1 gene polymorphisms in Sicilian patients with permanent primary congenital hypothyroidism. J. Endocrinol. Invest. *30*, 13–19. 10.1007/BF03347390.
36. Szczepanek, E., Ruchala, M., Szaflarski, W., Budny, B., Kilinska, L., Jaroniec, M., Niedziela, M., Zabel, M., and Sowinski, J. (2011). FOXE1 polyalanine tract length polymorphism in patients with thyroid hemiagenesis and subjects with normal thyroid. Horm. Res. Paediatr. *75*, 329–334. 10.1159/000322874.
37. Carré, A., Castanet, M., Sura-Trueba, S., Szinnai, G., Van Vliet, G., Trochet, D., Amiel, J., Léger, J., Czernichow, P., Scotet, V., et al. (2007). Polymorphic length of FOXE1 alanine stretch: evidence for genetic susceptibility to thyroid dysgenesis. Hum. Genet. *122*, 467–476.10.1007/s00439-007-0420-5.
38. Grassi, E.S., Rurale, G., de Filippis, T., Gentilini, D., Carbone, E., Coscia, F., Uraghi, S., Bullock, M., Clifton-Bligh, R.J., Gupta, A.K., et al. (2023). The length of FOXE1 polyalanine tract in congenital hypothyroidism: Evidence for a pathogenic role from familial, molecular and cohort studies. Front. Endocrinol. (Lausanne). *14*, 1127312. 10.3389/fendo.2023.1127312.
39. Fernández, L.P., López-Márquez, A., Martínez, A.M., Gómez-López, G., and Santisteban, P. (2013). New insights into FoxE1 functions: identification of direct FoxE1 targets in thyroid cells. PLoS One *8*, e62849. 10.1371/journal.pone.0062849.
40. Ortiz, L., Zannini, M., Di Lauro, R., and Santisteban, P. (1997). Transcriptional control of the forkhead thyroid transcription factor TTF-2 by thyrotropin, insulin, and insulin-like growth factor I. J. Biol. Chem. *272*, 23334–23339. 10.1074/jbc.272.37.23334.
41. Fernández, L.P., López-Márquez, A., and Santisteban, P. (2015). Thyroid transcription factors in development, differentiation and disease. Nat. Rev. Endocrinol. *11*, 29–42. 10.1038/nrendo.2014.186.
42. Dathan, N., Parlato, R., Rosica, A., De Felice, M., and Di Lauro, R. (2002). Distribution of the titf2/foxe1 gene product is consistent with an important role in the development of foregut endoderm, palate, and hair. Dev. Dyn. an Off. Publ. Am. Assoc. Anat. *224*, 450–456. 10.1002/dvdy.10118.
43. Sarma, A.S., Banda, L., Rao Vupputuri, M., Desai, A., and Dalal, A. (2022). A new FOXE1 homozygous frameshift variant expands the genotypic and phenotypic spectrum of Bamforth-Lazarus syndrome. Eur. J. Med. Genet. *65*, 104591. 10.1016/j.ejmg.2022.104591.

44. Moreno, L.M., Mansilla, M.A., Bullard, S.A., Cooper, M.E., Busch, T.D., Machida, J., Johnson, M.K., Brauer, D., Krahn, K., Daack-Hirsch, S., et al. (2009). FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. Hum. Mol. Genet. *18*, 4879–4896. 10.1093/hmg/ddp444.

45. Tonacchera, M., Banco, M., Lapi, P., Di Cosmo, C., Perri, A., Montanelli, L., Moschini, L., Gatti, G., Gandini, D., Massei, A., et al. (2004). Genetic analysis of TTF-2 gene in children with congenital hypothyroidism and cleft palate, congenital hypothyroidism, or isolated cleft palate. Thyroid *14*, 584–588. 10.1089/1050725041692864.

46. Lavoie, H., Debeane, F., Trinh, Q.-D., Turcotte, J.-F., Corbeil-Girard, L.-P., Dicaire, M.-J., Saint-Denis, A., Pagé, M., Rouleau, G.A., and Brais, B. (2003). Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. Hum. Mol. Genet. *12*, 2967–2979. 10.1093/hmg/ddg329.

47. Amiel, J., Trochet, D., Clément-Ziza, M., Munnich, A., and Lyonnet, S. (2004). Polyalanineexpansions in human. Hum. Mol. Genet. *13 Spec No*, R235-43. 10.1093/hmg/ddh251.

48. Cohen, M.M.J. (2006). Holoprosencephaly: clinical, anatomic, and molecular dimensions. Birth Defects Res. A. Clin. Mol. Teratol. *76*, 658–673. 10.1002/bdra.20295.

49. Albrecht, A., and Mundlos, S. (2005). The other trinucleotide repeat: polyalanine expansion disorders. Curr. Opin. Genet. Dev. *15*, 285–293. 10.1016/j.gde.2005.04.003.

50. Yamasaki, M., and Kanemura, Y. (2015). Molecular Biology of Pediatric Hydrocephalus and Hydrocephalus-related Diseases. Neurol. Med. Chir. (Tokyo). *55*, 640–646. 10.2176/nmc.ra.2015-0075.

51. Radó-Trilla, N., Arató, K., Pegueroles, C., Raya, A., de la Luna, S., and Albà, M.M. (2015). Key Role of Amino Acid Repeat Expansions in the Functional Diversification of Duplicated Transcription Factors. Mol. Biol. Evol. *32*, 2263–2272. 10.1093/molbev/msv103.

52. Brown, L.Y., and Brown, S.A. (2004). Alanine tracts: the expanding story of human illness and trinucleotide repeats. Trends Genet. *20*, 51–58. 10.1016/j.tig.2003.11.002.

53. Shibata, A., Machida, J., Yamaguchi, S., Kimura, M., Tatematsu, T., Miyachi, H., Matsushita, M., Kitoh, H., Ishiguro, N., Nakayama, A., et al. (2016). Characterisation of novel RUNX2 mutation with alanine tract expansion from Japanese cleidocranial dysplasia patient. Mutagenesis *31*, 61–67. 10.1093/mutage/gev057.

54. Shoubridge, C., and Gecz, J. (2012). Polyalanine tract disorders and neurocognitive phenotypes. Adv. Exp. Med. Biol. *769*, 185–203. 10.1007/978-1-4614-5434-2_12.

55. Tejada, M.I., and Ibarluzea, N. (2020). Non-syndromic X linked intellectual disability: Current knowledge in light of the recent advances in molecular and functional studies. Clin. Genet. *97*, 677–687. 10.1111/cge.13698.

56. Neuhouser, A.J., and Harrison, A.R. (2023). Blepharophimosis Syndrome. In.

57. Innis, J.W. (1993). Hand-Foot-Genital Syndrome. In, M. P. Adam, G. M. Mirzaa, R. A. Pagon, S. E.

Wallace, L. J. H. Bean, K. W. Gripp, and A. Amemiya, eds.

58. Malik, S., and Grzeschik, K.-H. (2008). Synpolydactyly: clinical and molecular advances. Clin. Genet. *73*, 113–120. 10.1111/j.1399-0004.2007.00935.x.

59. Yamashita, S. (2021). Recent Progress in Oculopharyngeal Muscular Dystrophy. J. Clin. Med. *10*. 10.3390/jcm10071375.

60. Pavone, P., Spalice, A., Polizzi, A., Parisi, P., and Ruggieri, M. (2012). Ohtahara syndrome with emphasis on recent genetic discovery. Brain Dev. *34*, 459–468. 10.1016/j.braindev.2011.09.004.

61. Farrow, E., Nicot, R., Wiss, A., Laborde, A., and Ferri, J. (2018). Cleidocranial Dysplasia: A Review of Clinical, Radiological, Genetic Implications and a Guidelines Proposal. J. Craniofac. Surg. *29*, 382–389. 10.1097/SCS.0000000000004200.

62. Bauters, M., Frints, S.G., Van Esch, H., Spruijt, L., Baldewijns, M.M., de Die-Smulders, C.E.M., Fryns, J.-P., Marynen, P., and Froyen, G. (2014). Evidence for increased SOX3 dosageas a risk factor for X-linked hypopituitarism and neural tube defects. Am. J. Med. Genet. A *164A*, 1947–1952. 10.1002/ajmg.a.36580.

63. Barratt, K.S., and Arkell, R.M. (2018). ZIC2 in Holoprosencephaly. Adv. Exp. Med. Biol. *1046*, 269–299. 10.1007/978-981-10-7311-3_14.

64. SHIMOMURA, O., JOHNSON, F.H., and SAIGA, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J. Cell. Comp. Physiol. *59*, 223–239. 10.1002/jcp.1030590302.

65. Kremers, G.-J., Gilbert, S.G., Cranfill, P.J., Davidson, M.W., and Piston, D.W. (2011). Fluorescent proteins at a glance. J. Cell Sci. *124*, 157–160. 10.1242/jcs.072744.

66. Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. Gene *111*, 229–233. 10.1016/0378-1119(92)90691-h.

67. Matz, M. V, Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. Nat. Biotechnol. *17*, 969–973. 10.1038/13657.

68. Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotechnol. *22*, 1567–1572. 10.1038/nbt1037.

69. Shaner, N.C., Lin, M.Z., McKeown, M.R., Steinbach, P.A., Hazelwood, K.L., Davidson, M.W., and Tsien, R.Y. (2008). Improving the photostability of bright monomeric orange and red fluorescent proteins. Nat. Methods *5*, 545–551. 10.1038/nmeth.1209.

70. Patterson, G.H., Knobel, S.M., Sharif, W.D., Kain, S.R., and Piston, D.W. (1997). Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. Biophys. J. *73*, 2782–2790. 10.1016/S0006-3495(97)78307-3.

71. Chudakov, D.M., Matz, M. V, Lukyanov, S., and Lukyanov, K.A. (2010). Fluorescent proteins and

their applications in imaging living cells and tissues. Physiol. Rev. *90*, 1103–1163. 10.1152/physrev.00038.2009.

72. Kim, T.K., and Eberwine, J.H. (2010). Mammalian cell transfection: the present and the future.Anal. Bioanal. Chem. *397*, 3173–3178. 10.1007/s00216-010-3821-6.
73. Ruozi, B., Forni, F., Battini, R., and Vandelli, M.A. (2003). Cationic liposomes for gene transfection. J. Drug Target. *11*, 407–414. 10.1080/10611860310001655600.
74. Mahmood, T., and Yang, P.-C. (2012). Western blot: technique, theory, and trouble shooting. N. Am. J. Med. Sci. *4*, 429–434. 10.4103/1947-2714.100998.
75. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. 10.1038/nmeth.2019.
76. Im, K., Mareninov, S., Diaz, M.F.P., and Yong, W.H. (2019). An Introduction to Performing Immunofluorescence Staining. Methods Mol. Biol. *1897*, 299–311. 10.1007/978-1-4939-8935-5_26.
77. Shifera, A.S., and Hardin, J.A. (2010). Factors modulating expression of Renilla luciferase from control plasmids used in luciferase reporter gene assays. Anal. Biochem. *396*, 167–172. 10.1016/j.ab.2009.09.043.
78. Cuesta, I., Zaret, K.S., and Santisteban, P. (2007). The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. Mol. Cell. Biol. *27*, 7302–7314. 10.1128/MCB.00758-07.
79. Cole, R. (2014). Live-cell imaging. Cell Adh. Migr. *8*, 452–459. 10.4161/cam.28348.