UNIVERSITY OF NAPOLI FEDERICO II

Department of Veterinary Medicine and Animal Production

PhD in Model Organisms in Medical Research and Veterinary
XXV cycle

Analysis of a novel thyroid-specific DNA binding activity in the Pax8 promoter

Coordinator
Prof. Paolo De Girolamo

Candidate

Olga Spadaro

Tutor
Prof. Roberto Di Lauro

INDEX

INTRODUCTION

CHAPTER 1.

Transcriptional regulation of gene expression

The development of eukaryotic organisms requires the differential transcription of thousands of genes in precise spatial and temporal patterns. Cells accomplish this task by means of a limited repertoire of activators which link related signaling pathways and integrate diverse regulatory cues employing the principle of cooperativity and transcriptional synergy on regulatory elements along the DNA (Michael Carey, 1998).

1. Transcription factors

Any protein that is needed for the initiation of transcription, but which is not itself part of RNA polymerase is defined a transcription factor. They can be divided in three main classes: basal factors interact with RNA polymerase at the startpoint; activators recognize specific short response elements (REs) in the promoter or enhancers increasing the frequency of transcription; coactivators provides a connection between activators and the basal apparatus. Transcription factors can be furtherly grouped into classes that use related structural motif for recognition: helix-turn-helix (HTH) proteins, the homeodomains, zinc finger proteins, the steroid receptors, leucine zipper proteins and helix loop helix proteins.

For example the homeodomain is a DNA-binding domain of 60 amino acids that has three α-helices; it was firstly found in homeotic genes which are one of three groups of genes which control *D. melanogaster* development; in particular these genes determine the identity of body structures, i.e. impose the program that determines the unique differentiation of each segment. Sequence related to the homeodomain are found in several types of animal transcription factors, for example in the paired box transcription factors *(Pax family)* which control embryonic development within a variety

of cell lineages. The leucine zipper (ZIP), instead, is an amphipathic helix that dimerizes forming the bZIP motif in which two adiacent basic regions bind inverted repeats in DNA. This motif is found in the AP-1 protein, a transcription factor composed of Jun and Fos family members; diverse combinations of AP-1 components mediate various biological processes, for example the heterodimer AP-1$^{cJun-cFos}$ is required for the expression of activin-responsive organizer genes (Sung-Young Lee et al., 2011).

As last protein domain example, the helix loop helix motif is a stretch of 40-50 amino acids containing two amphipatic α-helices separated by a linker region (the loop) of varying length which role is probably important for allowing the freedom for the two helical regions to interact indipendently of one another. Proteins belonging to this class have important roles in differentiation and development as shown by MyoD which is involved in the formation of muscle cells together with several basic HLH proteins; MyoD interaction with the non basic HLH Id prevents myogenesis, while MyoD association with basic HLH E12 or E47 proteins is the trigger for muscle differentiation in that the complex activate transcription (Lewin Gene VIII, Gary Carlson Editor).

The overall structure of transcription complexes is that of a controlled mechanisms with a modular architecture both in proteins structure and in regulatory DNA elements which they bind to. The modular structure in proteins is found in distinct regions dedicated to different functions: a DNA-binding domain that directs the protein to a specific DNA site, a multimerization domain that allows assembly of either homo- or hetero-multimers, and an effector domain that can modulate the rate of transcription (Menie Merika et al., 2001). These domains act as a bridge between the basal transcription apparatus and activators; the former is infact not adequate to initiate more than low levels of transcription which significant amount need transcriptional activators; these are in turn targets of multiple signaling pathways providing notions on how cells adapt to external cues by controlling gene expression.

2. Cis-regulatory elements

The elements on which the transcriptional apparatus assembles thus responding by means of gene expression to external environment is made of cis-acting control elements that specify the location, timing and magnitude of the response (Menie Merika et al., 2001).

Each gene has in fact a particular combination of regulatory elements, whose nature, number and spatial arrangement determine the gene's unique pattern of expression: in that consist the above mentioned modularity in DNA elements. They can be simply divided in two main categories: promoters which comprise core promoter and proximal promoter elements, and distal regulatory elements made up of enhancer, silencers, insulators, or locus control regions.

2.1 Promoter and upstream promoter elements

The "gene promoter" is a collection of cis-regulatory elements required for initiation of transcription; three classes of promoters drive different classes of genes. Polymerase I promoters are used by genes that encode ribosomal RNAs, polymerase II promoters are used by genes that are transcribed to yield mRNAs and hence, proteins; polymerase III promoters are used by genes that encode small RNAs.

Pol II core promoter is an approximately 60 bp DNA sequence overlapping the transcription start site (+1) that serves as the recognition site for RNA polymerase II and general transcription factors. Core promoters elements consist of: BRE, TATA box, Inr-Initiator, MTE-motif ten element, DPE-Downstream promoter element and DCE-downstream core element (Fig. 1).
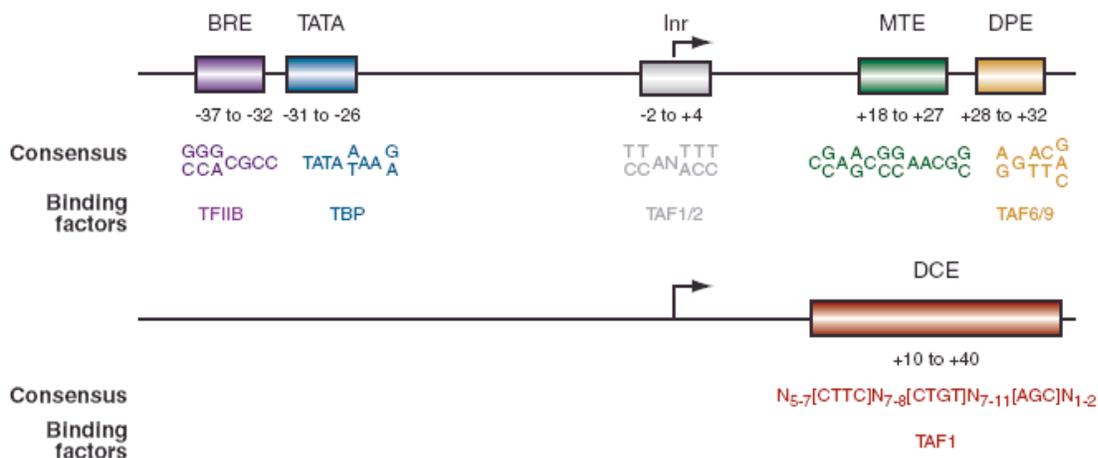

Fig. 1: Core promoter elements.
Metazoan core promoters are composed of a number of elements that may include a TATA box, an Initiator element (Inr), a Downstream Promoter Element (DPE), a Downstream Core Element (DCE), a TFIIB-Recognition Element (BRE), and a Motif Ten Element (MTE). The human consensus sequence of these elements, their relative positions, and the transcription factors that bind them are shown. The DCE is shown on a separate core promoter for illustration purposes only. Although the DCE can be present in promoters containing a TATA box and/or Inr, it presumably does not occur with a DPE or MTE (Glenn A. et al., 2006).

The TATA box is an AT reach region whose consensus sequence is TATAAA, located about 25-30 nt upstream of the transcription start site, but even upstream promoter sequences that resemble TATAAA may be functional TATA motif.

In humans it was found that about 32% of 1031 potential promoter regions contain a putative TATA box motif. Recognition of the TATA box is conferred by the TATA-binding protein (TBP), a small protein of 30 kDa, which constitutes the general transcription factor TFIID together with other subunits called TAFs (for TBP associated factors).

The BRE sequence is a TFIIB binding site that is located immediately upstream of some TATA boxes; in vitro transcription experiments with purified basal transcription factors revealed that the BRE facilitates the incorporation of TFIIB into productive transcription initiation complexes (Lagrange et al. 1998, Suzuki et al. 2001).

The Inr encompasses the transcription start site and was identified in a variety of eukaryotes (Breathnach and Chambon, 1981).

Inr elements are found both in TATA-containing as well as TATA-less core promoters. The consensus for the Inr in mammalian cells is Py-Py(C)-A-N-T/A-Py-Py. The A+1 position is designated at the +1 start site because transcription commonly initiates at this nucleotide. More generally, however, transcription initiates at a single site or in a cluster of multiple sites in the vicinity of the Inr (and not necessarily at the A+1 position).

Different sequence-specific DNA-binding factors have been found to interact with the Inr, there is considerable evidence that TFIID binds to the Inr in a sequenc specific manner (Roy et al., 1991).

Aside from them, it has also been observed that RNA Polymerase II is able to recognize the Inr and to medite transcription in an Inr-mediated manner in the absence of TAFs (Carcamo et al., 1991; Weis and Reinberg, 1997). These results suggest that recognition and interaction with the Inr occurs at different steps in the transcription process.

The DPE is located precisely at +28 to +32 relative to the A+1 position; it was identified as a downstream core promoter-binding site for purified *Drosophila* TFIID (Burke and Kadonaga, 1996). TFIID binds cooperatively to the Inr and DPE motifs, as mutation of either the Inr or the DPE results in loss of TFIID binding to the core promoter (Burke and Kadonaga, 1996). Naturally occurring in TATA-less core promoters, mutation of the DPE motif results in a 10- to 50-fold reduction in basal transcription activity, as observed in the analysis of about 18 *Drosophila* core promoters (Burke and Kadonaga 1997; Kutach and Kadonaga 2000). All of the known DPE-containing promoters possess identical spacing between the Inr and DPE motifs, and the alteration

of the spacing between the Inr and DPE by a single nucleotide causes a several-fold reduction in TFIID binding and basal transcription activity (Burke and Kadonaga 1997; Kutach and Kadonaga 2000). The consensus sequence for the DPE is estimated to be A/G+28-G-A/T-C/T-G/A/C (Kutach and Kadonaga 2000). Although the DPE consensus sequence is somewhat degenerate, it should be considered that both DPE and Inr motifs are required in DPE-dependent promoters and that the spacing between the DPE and Inr is invariant (which enables the cooperative binding of TFIID to the two motifs). Thus, the functional consensus for DPE-dependent core promoters consists of the Inr and DPE motifs with the DPE located at +28 to +32 relative to A+1.

The MTE is a new core promoter element conserved from Drosophila to humans, which promotes transcription by RNA polymerase II when it is located precisely at positions +18 +27 relative to A+1 in the initiator element. The MTE requires the Inr, but functions independently of the TATA-box and DPE. Notably, the loss of transcriptional activity upon mutation of a TATA-box or DPE can be compensated by the addition of an MTE. In addition, the MTE exhibits strong synergism with the TATA-box as well as the DPE (Lim et al. 2004).

The DCE element firstly found in the human β-globin promoter is present in a large number of promoters and with high incidence in promoter containing a TATA motif; it is constituted by three sub elements of which one is independent of the others; it is reported to bind solely the TAFs within the TFIID complex (Dong-Hoon Lee et al., 2005).

Promoter elements become non-functional when moved even a short distance from the start of transcription or if their orientation is altered.

Besides the aforementioned core promoter elements, the region immediately upstream (up to a few hundred base pairs) from the core promoter is defined proximal promoter and contains primary regulatory elements defined as specific transcription factors binding sites.

These sequences determine whether the promoter is expressed in all cell types or is specifically regulated. Promoters that are constitutively expressed (their genes are sometimes called housekeeping genes) have upstream sequence elements that are recognized by ubiquitous activators.

2.2 Distal regulatory elements

As for distal regulatory sites, enhancers are defined as *cis*-acting DNA sequences that stimulate transcription independently of the sequence orientation or transcriptional start site position (Banerji

et al., 1981), indeed they can be found far upstream, downstream or even within the gene.

However these elements function in ways that go beyond this molecular shorthand. As a matter of fact the nucleoprotein complexes on enhancers can constitute the "enhanceosome" in which there's a high degree of cooperativity between enhancer-bound proteins, such that alterations in individual binding sites, or the absence of one regulatory protein, can have drastic effects on enhancer output.

Nevertheless enhanceosome may represent only a subclass of genetic switches and an alternative mechanisms, the "billboard enhancer", could take account for complex patterns of expression during development; in this model the entire element need not to function as a cooperative unit, but rather as an ensemble of separate elements that independently affect gene expression.

Enhancer associated proteins can bind in sequence-specific or sequence-non-specific manners, indirectly through protein-protein contacts, promoting the decondensation of repressed chromatin and facilitating the assembly of the transcription machinery at gene promoters.

One class of proteins, represented by the SWI/SNF complexes, modifies the chromatin structure non-covalently in an ATP-dependent fashion (Kingston et al., 1999). These proteins, once recruited to enhancer elements, can reposition specific nucleosomes along the DNA. Consequently core promoters may be exposed to allow transcription to start.

One more regulatory sequence at service of genes to ensure their proper temporal and spatial transcription, in the same class as enhancers and promoters, is the insulator.

Insulators have been characterized by two experimentally defined properties involving altered gene expression; the first one is called "enhancer blocking or promoter decoy" and occurs when an insulator is positioned between an enhancer and a promoter acting as a barrier against a signal propagated from the enhancer, as it interact with enahncer-bound factors and prevents the proper interaction of enhancer with its target promoter (Geyer and Clark, 2002). The second property is referred as "barrier activity" and occurs when insulators flank a transgene and prevent its transcriptional repression from positioning effects of heterocromatin by recruiting histone modifying or gene-activating factors (Capelson et al., 2004).

In this manner insulators are able to exert two opposing effects on transcription.

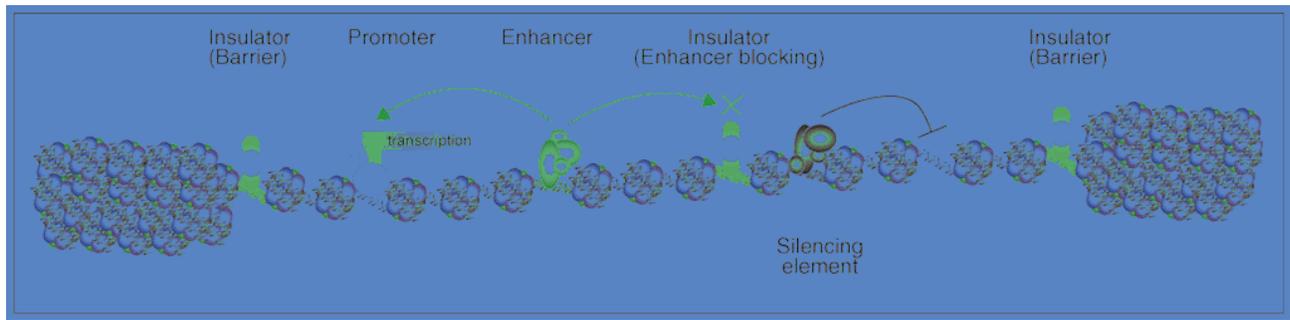A view of cis- regulatory elements analyzed so far is schematized (Fig. 2)

Fig.2: Different classes of *cis*-regulatory elements in a typical human genome. Transcription initiates at promoters (blue DNA), which are further activated by enhancers (green DNA) or repressed by silencers (red DNA). The activity of enhancers and silencers may be confined by insulators (yellow DNA), which also prevent the spreading of repressive condensed chromatin structures (shown at each end of this chromosomal region). This model depicts
nucleosomes as DNA (gray helix) wound around histone proteins (various colors), which are less dense at exposed DNA regulatory elements when bound by various transcription factors (blue ovals), activator and repressor proteins (green and red ovals, respectively), and CCCTC-binding factor (yellow oval) (Heintzman and Ren, 2009).

2.3 Mutations in regulatory elements

Cis-regulatory polymorphisms

A recent survey on known functional cis-regulatory variants has revealed a number of distinctive characteristics of the model of action of this variants, with implications for their evolution and their role in phenotypic variation.

First, transcription factors typically regulate many downstream target genes whose identities differ among individuals on the basis of the variation in cis-regulatory sequences.

Consequently, changes in expression or structure of a transcription factor, will be highly pleiotropic, influencing many downstream loci, but also highly epistatic, dependent on genetic background.

Second, "epistasis in cis" where the effect on transcritpion of one cis regulatory variant depends on covariation at a linked cis regulatory site.

Third, many of the genes with regulatory polymorphisms interact with one another in regulatory, metabolic and physiological networks. As a consequence, moderate levels of variation at any one locus will be amplified by the numer of genes in the network to yield high levels of polymorphism at the network level. (Rockman and Wray, 2002).

Cis-regulatory elements mutations

Mutations in transcriptional regulatory elements have been found associated with numerous human disease (Table1). In some cases the defect is known, as in the Bernard-Soulier Syndrome where mutatons in GpIbβ gene proximal promoter result in reduced GATA-1 binding and GpIbβ gene expression. In other cases the underlying defect is less well defined. For instance, a 12-mer repeat expansion in the promoter of cystatin B gene has been proposed to cause progressive myoclonus epilepsy, presumably by altering the spacing of element in the promoter.

A variety of cancers result from chromosomal rearrangements involving either regulatory elements or transcription factors. For example, promoter and/or enhancer elements of one gene may become aberrantly linked to a proto-oncogene, thereby causing altered expression of an oncogenic protein.

CHAPTER 2

Transcriptional regulation of thyroid development in mouse

2.1 Thyroid gland development

The thyroid gland in mammals is located in the neck region. The gland produces thyroid hormones and calcitonin in two distinct cell types, the thyroid follicular cells (TCFs) and the parafollicular or C cells, respectively; all mammals share the same endocrine cell types and histoarchitecture.

The TFCs, the most numerous cell populations in the gland, form the thyroid follicles which serves as storage and release of the thyroid hormone thyroglobulin (Tg).

Moreover these cells are highly polarized, with thyroperoxidase (TPO) on the apical membrane, and TSH receptor and the sodium/iodide symporter (NIS) on the basolateral membrane.

The expression of all these genes represents the characteristic phenotype of a terminally differentiated thyroid cell.

The C cells, scattered in the interfollicular space, produce calcitonin, an hormone implicated in calcium homeostasis.

The thyroid gland originates from the ventral floor of the foregut first as a thickening, the thyroid anlage at E8.5; then this thickening proliferates, deepens and invades the surrounding mesenchyme forming a small pit, the thyroid bud (E9), which then becomes an outpouching of the endoderm, adjacent to the distal part of the outflow tract of the developing heart.

By E10 the thyroid primordium resembles a narrow neck that rapidly becomes a diverticulum.

The expansion of the thyroid primordium correspond to the lost of connections with the floor of pharnynx, the thyroglossal duct which connects the remnant of anlage with the thyroid primordium, in fact, disappears.

By E15-E16 the thyroid lobes expand considerably and the gland exhibits its definitive shape: two lobes connected by a narrow isthmus.

The onset of folliculogenesis is evident at E15.5 as well as the detection of calcitonin producing C cells derived from the fusion of ultimobranchial bodies with the primitive thyroid at E14.

The completion of organogenesis is in the functional differentiation of thyroid follicular cells that express a series of proteins, mentioned above, that are typical of TFCs and that are essential for thyroid hormone biosynthesis. Genes typical of this stage appear according to a given temporal

pattern: Tg, TPO and TSH receptor genes are expressed by E14.5; sodium iodide symporter (NIS) is detected by E16. T4 is first detected at E16.5. The differentiative program of TFCs is completed only when the gland reaches its final location; this point raises the question of whether a time or space-dependent signal is responsible for it. The last hypothesis is rejected by observations that patients with sublingual thyroid do produce in low amounts thyroid hormones.

The precise timing in the start of gene expression program necessary for thyroid hormone biosynthesis indicates that an ordered genetic mechanism must be responsible for it.

2.2 Thyroid transcription factors and their role in development

At the thyroid anlage stage (specification stage) thyroid cells express a specific combination of transcription factors: Titf1/Nkx2-1 (*thyroid transcription factor-1*), Hhex (*hematopoietically expressed homeobox*), Pax8 (p*aired box gene 8*) and TTF-2/Foxe1 (*thyroid transcription factor-2*).

The combination of these proteins is a unique feature of thyroid precursors cells and their descendants as fully differentiated thyroid follicular cells producing thyroid hormones.

All of these transcription factors are indispensable for normal thyroid, whose morphogenesis is severely impaired in their absence.

Titf1 is a homeodomain-containing transcription factor, firstly identified as a thyroid-specific DNA binding factor recognizing thyroglobulin and thyroperoxidase promoters; it belongs to the Nkx2 family of transcription factors, since it is named also Nkx2-1.

Since the expression in thyroid anlage in primitive pharynx, Titf1 remains expressed in thyroid follicular cells during all stages of development and in adulthood, being responsible of organogenesis and of the expression of TFC specific genes in adult life.

The thyroid primordium in *Titf1/Nkx2-1 -/-* embryos is in correct position but then undergoes degeneration at E10.5 and eventually disappears by E11.

These data, together with the presence of apoptotic cells, suggest that Titf1/Nkx2-1 is required to prevent the initiation of an apoptotic process, being essential for the survival of thyroid cell precursors but not for their initial formation.

This gene is also expressed in the trachea, lung epithelium and some areas of the forebrain including developing posterior pituitary; its expression in adulthood remains unvaried in lung epithelium and

posterior pituitary, while in the brain it becomes restricted to the paraventricular regions and some hypothalamic nuclei.

Titf1 regulates different genes in different cell types; for example the bone morphogenetic protein (Bmp)4, expressed in the growing tip of the branching lung epithelium in a normal embryo, is undetectable in the lung of *Titf1/Nkx2-1 -/-* embryos, while in the developing posterior pituitary which express Bmp4 and fibroblast growth factor (Fgf)8 in normal conditions, the lack of Titf1 influences only Fgf8 expression, leading to apoptosis in the anterior pituitary bud and later on in the posterior bud as well.

There's a sharing of Titf1 regulated mechanisms between the thyroid and pituitary cells that in both are required for the survival of cells.

Hhex is a homeodomain containing transcription factor first identified in hematopoietic cells (Crompton et al., 1992), characterized by a N-terminal prolin-rich region, involved in regulating the transcription of the target genes.

The homeobox gene *Hhex* is expressed in the anterior visceral endoderm (AVE) and definitive endoderm of the early mouse embryo, two tissues implicated in patterning the anterior neural plate.

In particular AVE is derived from a handful of proliferating cells in the distal visceral endoderm, which alone express Hhex.

Moreover graft experiments in the mouse demonstrates that the AVE, the node and its derivative, the axial mesendoderm, cooperate in the induction and patterning of the brain, while a complete neural axis is to be induced by a combination of the AVE, anterior epiblast and the early node.

Later in embryo development Hhex is localized in the ventral gut and then marks the primordium of several organs derived from the foregut such as thyroid, liver, thymus, pancreas and lungs.

In *Hhex* null embryos at E9 the thyroid anlage is present and comparable with the wild type and the expression of the other thyroid markers Titf1, Pax8 and Foxe1 is not affected.

Thus, specification of the thyroid cells does not require Hhex.

At E10 in mutant embryos the thyroid bud is impaired with alterations in the number of thyroid cell precursors and in its morphology, the expression of Pax8 and Foxe1 proteins is strongly reduced, while the expression of Titf1 protein is not significantly reduced at this stage.

A probably role of Hhex could be the maintenance of Titf1, Pax8 and Foxe1 expression in the developing thyroid.

Foxe1/TTF-2 is a member of winged helix/forkhead family of transcription factors first identified as a thyroid specific nuclear protein recognizing specific DNA sequence on both *thyroglobulin and thyroperoxidase* promoters under hormone stimulation.

Foxe1 mRNA is detected at E8.5 in all the endodermal cells of the floor of the foregut, including the thyroid anlage; hence at variance with Titf1/Nkx2-1 and Pax8, whose expression is limited to the thyroid anlage, Foxe1has a wider domain of expression.

In ectoderm derived structures beyond the posterior stomatodeum and buccopharyngeal membrane, TTF-2 is localized to Rathke's pouch, an ectodermal diverticulum in the roof of the primitive oral cavity that forms adenohypophysis, but not in the neuroectodermal derivatives that give rise to neurohypophysis; it is not expressed in later stages over E11.5 in this gland. TTF-2 is also expressed in two other ectoderm derivatives: nasal choanae (the openings of the olfactory pits into the oropharyngeal cavity), whiskers and hair follicles.

TTF-2 is expressed in the epithelial layer of the foregut lining the pharyngeal arches and in their derivative tissues thyroid, tongue, epiglottis, palate, and oesophagus, it is absent in the thymus,parathyroid, and ultimobranchial body, which are derivatives of the pharyngeal pouches free of TTF-2.

Expression of TTF-2 in thyroid cell precursors is maintained during development and persists in adult thyroid follicular cells.

Homozygous *foxe1 -/-* mice die within 48h, they display no thyroid in its normal location and the absence of thyroid hormones.

Despite the fact that the thyroid budding primordium is normal, at E9.5 thyroid precursor cells are still on the floor of pharynx, while in wild type embryos they are detached from the pharynx cavity and begin to descend. Mutant mice later in development could either exhibit no thyroid gland at all or a small thyroid remnant attached to the pharyngeal floor in which cells complete their differentiative process.

Pax8 is a member of a family of transcription factors characterized by a conserved DNA binding domain of 128-amino acids, the paired box, identified in the *Drosophila* segmentation genes *paired* (Kilchherr et al., 1986; Frigerio et al., 1986).

Like Titf1/Nkx2-1, Pax8 is detected in the developing thyroid from E8.5 at the time of specification and its expression is maintained in thyroid follicular cells during all stages of development.

In the nervous system *Pax8* mRNA is transiently expressed in the myelencephalon and through the entire length of the neural tube; at later stages of the development as well as in adulthood there are no more signals in the brain.

Pax8 mRNA is expressed in all stages of mammalian kidney development: from pronephric duct precursors that originates from nephrogenic mesenchyme to the metanephros, which originates from reciprocal inductive interactions between the ureteric bud and the metanephric mesenchyme; Pax8 persist in adult kidney. In particular Pax8/Pax2 coexpression in kidney shows redundant gene functions: the double mutant, unlike the single mutants, shows no early intermediate mesenchyme commitment to nephric duct lineage (Dahl et al., 1997).

Pax8 -/- homozygous mice born at the expected Mendelian frequency but show growth retardation and die within 2-3 weeks. These mice do not display any apparent defects in the spinal cord, midbrain/hindbrain boundary or kidneys; on the contrary the thyroid gland is severely affected: the thyroid diverticulum is able to evaginate from the endoderm but Pax8 is required for further development, as neither follicles nor TFCs can be detected and the rudimentary gland is composed almost completely of calcitonin-producing C cells. Hypothyroidism is the cause of death of the mutated animals: the administration of T4 allows the animals to survive.

Thus, like Titf1/Nkx2-1, Pax8 seems to be required for the survival of thyroid cell precursors and not for their specification.

2.3 Role of Pax8 in thyroid differentiation

Besides the just mentioned role in morphogenesis of the TFC component of the thyroid gland, there are at least two more evidence that point out Pax8 as a "master gene" for the regulation of the thyroid differentiated phenotype.

The first evidence has been obtained by using the sperimental model of the thyroid cell in colture.

The Polyoma antigen T transformed PCPy cell line, derivative of the differentiated thyroid cell line PC Cl3, loses the thyroid-differentiated phenotype both as tissue specific gene expression of Tg, TPO, NIS, and as TSH dependence for proliferation.

Among the thyroid-specific transcription factors discussed so far, only the expression of Pax8 is severely reduced.

The reintroduction of the gene in PCPy cells is sufficient to activate expression of the endogenous genes encoding Tg, TPO and NIS and suggest a fundamental role of this transcription factor in the maintenance of functional differentiation in thyroid cells (Pasca di Magliano et al. 2000).

The second line of evidence derives from patients suffering from congenital hypothyroidism, the most frequent endocrine disorder in newborn (Klett M. 1997), characterized by elevated levels of TSH in response to reduced thyroid hormone levels. In most cases congenital hypothyoridism is due to thyroid disgenesis, ie disturbance in the gland's organogenesis, which result in thyroid agenesis, ipoplasia or ectopia.

The involvement of Pax8 in thyroid dysgenesis is demonstrated by the fact that all affected individuals are heterozygous for loss-of-function mutations in its DNA binding domain and, in the familial cases, transmission is autosomal dominant.

Hence the importance of the role covered by Pax8 in the complex molecular mechanism of thyroid differentiation, though, in thyroid disgenesis, incomplete penetrance takes account of severe, mild, up to no phenotype, in patients carrying Pax8 inactivating mutations (Congdon T. 2001).

2.4 Identification of Pax8 thyroid specific far upstream element and proximal promoter

Comparative studies on human and mouse genomic sequences flanking the Pax8 gene revealed for conserved non coding sequences (CNS) instrumental in guiding the thyroid specific expression of Pax8: CNS 87 and CNS 14.

The CNS 87 (conserved non coding sequence 87) has all the features typical of an enhancer, being located 87 kb upstream the transcriptional start site and stimulating transcription of a reporter gene in a very strong and thyroid specific manner.

In particular Dnase-I footprinting analysis of CNS 87 revealed only six protected regions (denominated FT), which, incubated, with either thyroid (PC Cl3) or non-thyroid (HeLa) protein extracts gave different band retardation pattern in EMSA assay.

Among them, bioinformatic tools predicted FT-1 and FT-6 to have binding sites for the thyroid enriched transcription factor TTF-1, element subsequently confirmed by EMSA assay.

Taken together these data indicates that CNS87 is a distant regulatory element that controls Pax8 expression through TTF-1 binding on the sequence (Nitsch et al., 2010).

CNS 14 (conserved non coding sequence 14) spans from nucleotides -369 to + 680 (relative to transcriptional start site) and directs thyroid specific transcriptional activity, features proper of a promoter; in particular this activity is retained by the subregion 14.2 (-369 +100) thus identified as the Pax8 proximal promoter. Exhaustive analysis of the entire sequence, by EMSA assays on small 19 nucleotides sequences, has led to the observation of different DNA complexes, both thyroid and non-thyroid specific (Fig. 3).
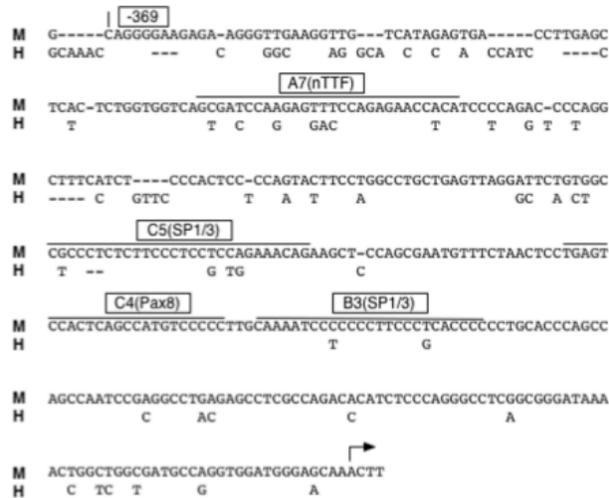

Fig.3: Architecture of mouse Pax8 promoter
The sequence of the mouse Pax8 promoter (M) is shown, aligned with the human sequence (H). Dashes indicate spacing introduced to maximize the alignement between the two sequences. In the human sequence, only the nucleotide changes with the mouse sequence are reported. Squares indicate binding factors above the corresponding recognition sequence (Di Gennaro et al., 2012)

The thyroid specific regulatory elements named as C4 and A7, from the initial nucleotide subdivision, have been further analyzed.

In the C4 sequence there's a Pax8 consensus strongly conserved in *thyroglobulin* and *tireoperoxidase* promoters; this datum has been confirmed by the purified, bacterially produced Pax8 paired domain (bPD-PAX8) binding complex onto the C4 oligonucleotide and by the Pax8-FLAG fusion protein, expressed in HeLa cells, giving the same band retardation in EMSA assays.

As for A7 sequence no binding was validated upon prediction of the bioinformatic tools (TRANSFAC-Genomatix) of the thyroid specific factors Nkx2-1, Foxe1, Pax8 and Hhex.

This sequence recognizes a novel thyroid specific DNA binding protein not characterized yet (Di Gennaro et al., 2012).

RESULTS

Previous studies on the *Pax8* promoter reveal a novel thyroid-specific DNA binding activity.

The *Pax8* minimal promoter has been recently described (Di Gennaro et al., 2012) as the CNS-14.2 (conserved sequence) region spanning the positions -369 to +1 relative to the gene transcriptional start site and expressed specifically in thyroid cells. The entire promoter was subdivided in several oligos and each was challenged with nuclear extracts derived from FRTL-5 and HeLa cells to detect thyroid-specific DNA binding activity. I studied the binding to one of these oligonucleotides that was proved to recognize a thyroid specific transcription factor; from now on I will refer to this sequence as A7 (Fig. 4 A,B).

A first attempt aimed at identifying the activity capable of binding to the A7 sequence, included the purification of the proteins involved and their identification by mass spectrometry (Di Gennaro, unpublished).

-370
CAGGGGAAGAGAAGGGGTTGAAGGTTGTCATAGAGTGACCTTGAGCTCACTCTGG

———————— A7 ————————
TGGTCAGCGATCCAAGAGTTTCCAGAGAACCACATCCCCAGACCCCAGGCTTTC

ATCTCCCACTCCCCAGTACTTCCTGGCCTGCTGAGTTAGGATTCTGTGGCCGCC

CTCTCTTCCCTCCTCCAGAAACAGAAGCTCCAGCGAATGTTTCTAACTCCTGAGT

CCACTCAGCCATGTCCCCCTTGCAAAATCCCCCCCCTTCCCTCACCCCCCTGCAC

CCAGCCAGCCAATCCGAGGCCTGAGAGCCTCGCCAGACACATCTCCCAGGGCC

TCGGCGGGATAAAACTGGCTGGCGATGCCAGGTGGATGGGAGCAAACTT +1

Fig.4A: Thyroid specific transcription factor bind to the A7 sequence
Radio-labeled A7 was incubated with nuclear extracts from different cell lines. Radio labeled Sp1 oligonucleotide was used as control to monitor the quality of the nuclear extracts used. Line 1 and 7, free probe; line 2 and 8, FRTL-5 nuclear extract; line 3 and 9, HeLa nuclear extract; line 4 and 10, MDCK nuclear extract; line 5 and 11, SK-N-BE nuclear extract; line 6 and 11, ES nuclear extract.
4B: Position of A7 sequence within the 14.2 promoter

Several proteins were identified, among these I tested whether Gtf2i could be the protein involved in binding to A7.

I first tested whether Gtf2i is involved in the regulation of the Pax8 promoter.

Expression of Pax8 mRNA in FRTL-5 is not inhibited by Gtf2i downregulation

To prove the involvement of Gtf2i on Pax8 transcriptional regulation, I performed the silencing of the protein in FRTL-5 cell line by RNA interference and look at the effect on Pax8 transcription.

An effective reduction in the levels of Gtf2i protein was confirmed by western blot analysis, using an anti-Gtf2i antibody (Fig. 5).
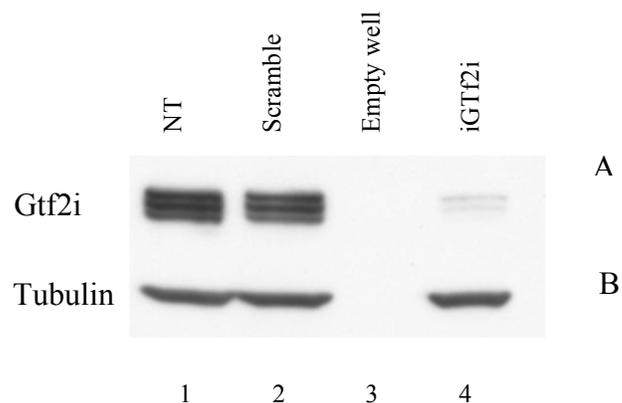
Fig.5: Downregulation of Gtf2i expression in FRTL-5
Western Blot analysis of Gtf2i protein expression 96 h after RNA interference transfection; panel A, Gtf2i expression; panel B, Tubulin expression.
Line 1, non transfected (NT); line 2, scramble 50 nM; line 3, empty well; line 4, 50 nM RNAi targeting Gtf2i rattus mRNA sequence.

Quantitative real time RT-PCR experiment was then performed to quantify Pax8 mRNA expression level in Gtf2i partially depleted samples.

Pax8 mRNA level in FRTL-5 cell line was found to be 1,25 fold less compared to the scramble (Fig. 6).

This mild reduction could be due to low Gtf2i protein levels still able to regulate the promoter or to the Pax8 autoregulation of its own promoter as previously demonstrated (Di Gennaro et al. Thyroid 2012).

Hence the need to develop a new experimental system able to confirm the data obtained by mass spectrometry.
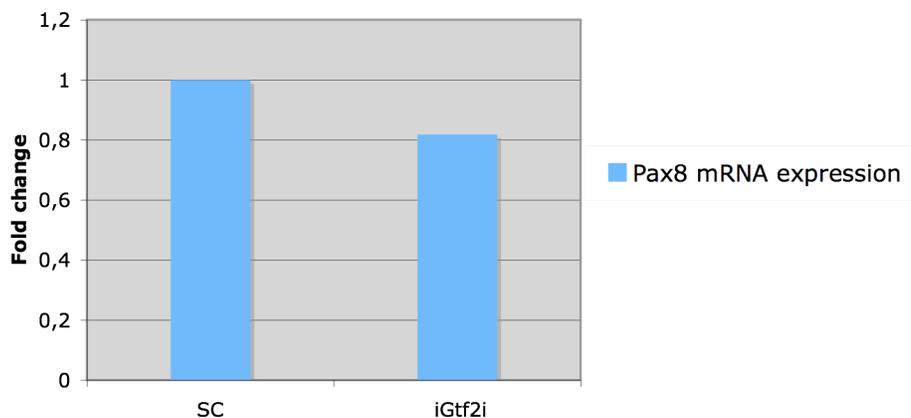
Fig.6: The down regulation of Gtf2i gene expression by RNAi does not decrease significantly Pax8 transcript expression.
Pax8 transcript abundance was measured by qRT-PCR as fold of induction on scramble.

Immunodepletion-EMSA does not confirm Gtf2i involvement in A7 transcriptional regulation.

To further test the involvement of Gtf2i involvement in the binding to the A7 sequence, I set up an immunodepletion assay for removing this protein from the extracts.

Briefly total protein extracts were submitted to immunoprecipitation with the specific antibody and the antigen-antibody complexes were isolated by protein A agarose beads.

The same procedure was simultaneously repeated with an unspecific antibody taken as a control.

Immunocomplexes and supernatants were both tested by western blotting.

Immunodepletion with the specific antibody removes efficiently Gtf2i from the extract, as all the protein is efficiently immunoprecipitated and no more visible in supernatant, while the control has just the opposite trend (Fig. 7).
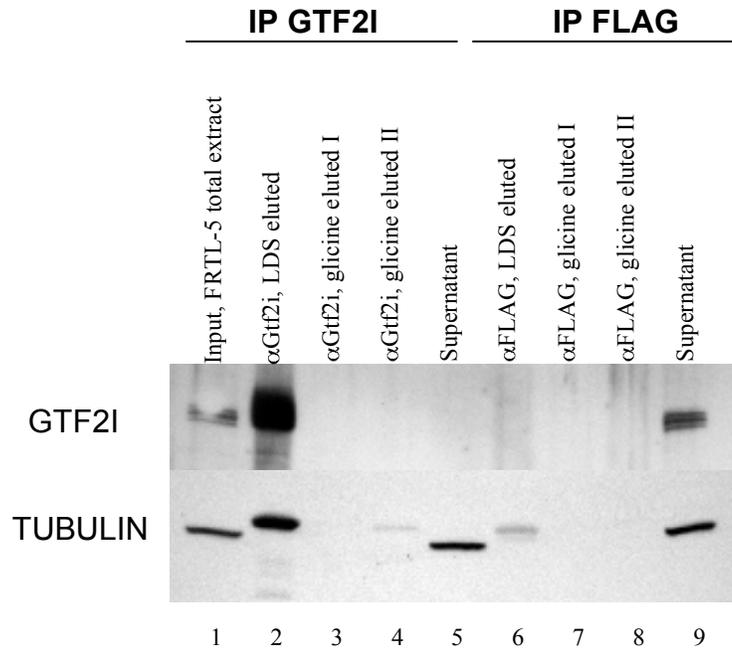
Fig.7: Immunodepletion of Gtf2i in FRTL-5 cells
Lane 1, input total FRTL-5 extract; lane 2-4, immunoprecipitation of Gtf2i: immunocomplexes eluted in denaturing condition with litium dodecyl sulfate (LDS) (2), or in non denaturing glicine elution (3-4); lane 6, supernatant immunodepleted of Gtf2i; lane 6-8, control immunoprecipitation with αFLAG antibody: immunocomplexes eluted in LDS denaturing conditions (7) and in non denaturing glicine elution (7-8); lane 9, supernatant of αFLAG immunoprecipitation

The residual activity on A7 after Gtf2i depletion was tested by EMSA assay (Fig. 8).

As shown the complex with the A7 oligonucleotide is not affected by Gtf2i depletion from the protein mixture, in fact comparing αGtf2i supernatant with αFLAG supernatant the complex have roughly the same trend.

Moreover, as a further confirmation, anti Gtf2i antibody was incubated in the binding mix with FRTL-5 extract and A7 probe, to detect a possible supershift as evidence of Gtf2i binding onto the sequence.

The absence of a supershift is consistent with the immunodepletion experiment and thus support the notion that Gtf2i is not the protein responsible for binding to A7 detected in thyroid cells extracts.

	A7					SP1			
FRTL-5	-	+	-	-	+	-	+	-	-
Supernatant IP GTF2I	-	-	+	-	-	-	-	+	-
Supernatant IP FLAG	-	-	-	+	-	-	-	-	+
Anti GTF2I	-	-	-	-	+	-	-	-	-

1 2 3 4 5 6 7 8 9
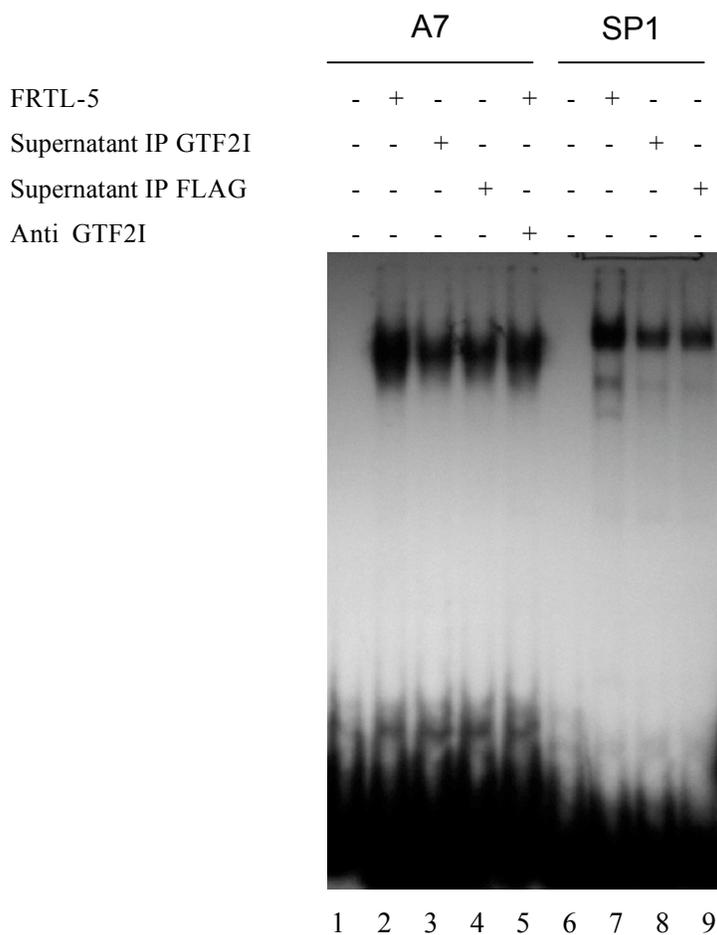
Fig.8: EMSA assay to test the binding activity of FRTL-5 extract immunodepleted of GTf2i.
Lines 1-5 binding assay on A7 probe; line 1, free probe; line 2, FRTL-5 input total extract; line 3, Gtf2i-immunodepleted FRTL-5 extract; line 4, supernatant of control αFLAG immunoprecipitation; FRTL-5 extract incubated with αGtf2i in the binding assay;
Lines 6-9 binding assay on Sp1 probe; line 6 free probe, line 7, FRTL-5 input total extract; line 8, Gtf2i-immunodepleted FRTL-5 extract; line 9, supernatant of control αFLAG immunoprecipitation.

A bioinformatic approach identifies one factor that recognizes a consensus on A7 sequence

In order to identify other candidates the sequence A7 was analyzed by MatInspector Professional program, which retrieved a binding consensus for the transcriptional factor NFAT5 (Fig. 9).
Interestingly this sequence is part of the core recognized by the NFAT family of transcription factors (Table 1) to which NFAT5 belongs.

<div align="center">gcgatccaagagtTTCCagagaac</div>

Fig.9: NFAT5 consensus binding motif found on A7 sequence.

Since NFAT5 was shown to be expressed in Jurkat cells, we tested an antibody directed against NFAT5 on extracts of this cell line. As reported the antibody recognizes a band of about 170 kDa (Fig. 10). I then used the same antibody on FRTL-5 and HeLa extracts: as shown NFAT5 is present in both cell lines with the same pattern of protein expression observed in Jurkat.

NFAT family members	Alternative names	mouse chromosome location	human chromosome location
NFATc1	NFAT2 and NFATc	Chr18:80606205-80713071 bp	Chr18:77155772-77289323 bp
NFATc2	NFAT1 and NFATp	Chr2:168476410-168601657 bp	Chr20:50007765-50179168 bp
NFATc3	NFAT4 and NFATx	Chr8:106059603-106130537 bp	Chr16:68119375-68260837 bp
NFATc4	NFAT3	Chr14:55824795-55833943 bp	Chr14:24836145-24848810 bp
NFAT5	TonEBP and OREBP	Chr8:107293470-107379517 bp	Chr16:69598997-69738553 bp

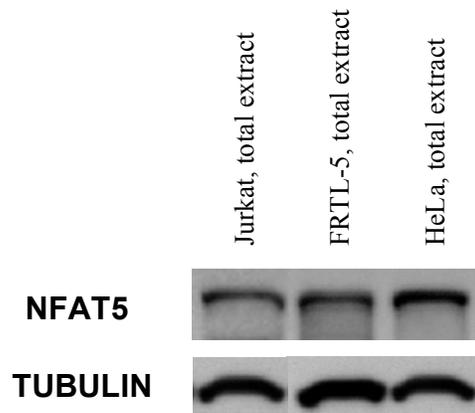**Table 1: NFAT family of transcription factors**

Fig.10: NFAT5 expression in Jurkat, FRTL-5 and HeLa cells
NFAT5 is expressed one band of about 170kDa in all cell lines
Line 1, Jurkat total extract; line 2, FRTL-5 total extract; line 3 HeLa total extract.
Tubulin was hybridized onto the same filter to test the quality of the extract and the relative distribution in the three cell lines.

To test if NFAT5 binds A7 I decided first to demonstrate whether extracts of Jurkat cells contain an activity that recognizes this oligo. As a control for the quality of Jurkat extracts we used an oligo derived from tumor necrosis factor α promoter that is known to be also regulated by NFAT5 (I'll refer to this sequence as NFAT5RecognitionSite).

As shown (Fig. 11) by the protein complex bound to the NFAT5RS, the quality of Jurkat extracts is good, in addition in this extracts there is also an activity that binds A7.

Conversely I tested whether proteins from FRTL-5 recognize the NFAT5RS or HeLa extracts taking as control the A7 sequence (Fig. 12).

Since the NFAT5RS is recognized by FRTL-5 and not by Hela, I conclude that the activity binding to this oligo is common to FRTL-5 and Jurkat and absent in HeLa cells. However in both experiments there is not supeshift probably due to the well-known fact that antibodies in supershift assays don't have assurance of working.

These binding experiments, together with western blot, can't definitely state NFAT5 uninvolvement in activity on A7, as it could be negatively regulated in HeLa.

In addition these data cannot either exclude the eventual binding of other members of the NFAT family that share a highly conserved DNA-binding domain conferring common DNA-binding specificities.

Fig.11: Gel Shift of Jurkat extract incubated with NFAT5RS and A7 sequences.
(A) Jurkat nuclear protein extract was incubated, alone or together, with an antibody anti-NFAT5, in EMSA assays with radiolabeled A7 and NFAT5RS probes. Line 1, A7 free probe; line 2, Jurkat nuclear extract; line 3, supershift analysis with NFAT5 antibody; line 4, supershift control analysis with HA antibody; line 5, NFAT5RS free probe; line 6, Jurkat nuclear extract; line 7, supershift analysis with NFAT5 antibody; line 7, supershift control analysis with HA antibody;

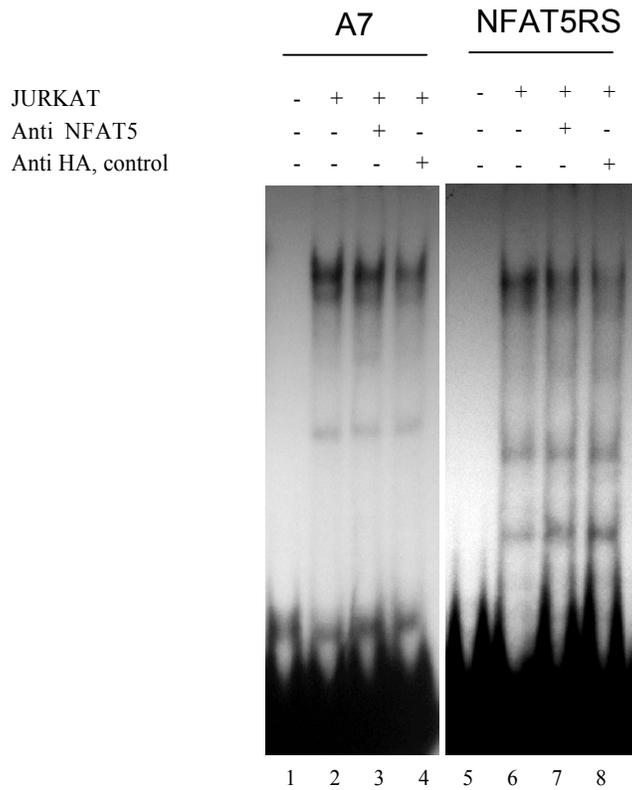|	NFAT5RS					A7				
FRTL-5	-	+	+	+	-	-	+	+	+	-
HeLa	-	-	-	-	+	-	-	-	-	+
Anti NFAT-5	-	-	+	-	-	-	-	+	-	-
Anti HA, control	-	-	-	+	-	-	-	-	+	-

1 2 3 4 5 6 7 8 9 10
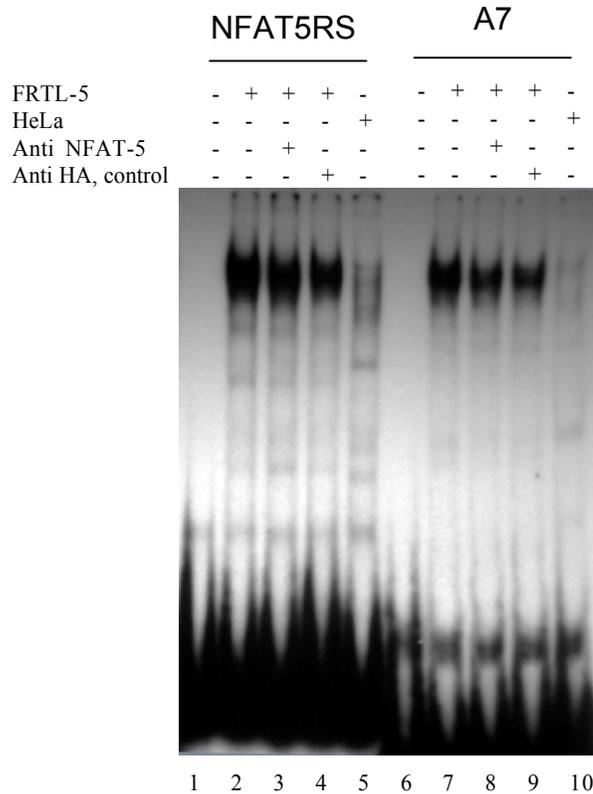
Fig.12: Gel Shift of FRTL-5 and HeLa extracts incubated with NFATRS and A7 sequences.
FRTL-5 nuclear protein extract was incubated, alone or together, with an antibody anti-NFAT5 in EMSA assays with radiolabeled NFAT5RS and A7 probes; HeLa nuclear cell lysate as binding negative control incubated with NFAT5RS and A7 radio-labeled probes.
Line 1, NFAT5RS free probe; line 2, FRTL-5 nuclear extract; line 3, supershift analysis with NFAT5 antibody; line 4, supershift control anlysis with anti HA antibody; line 5, HeLa nuclear extract; line 6, A7 free probe; line 7, FRTL-5 nuclear extract; line 8, supershift analysis with NFAT5 antibody; line 9, supershift control analysis with HA antibody; line 10, HeLa nuclear extract.

Immunodepletion proves that NFAT5 is not involved in binding to A7.

As for the transcription factor Gtf2i, even in this case, we sought confirmation for NFAT5 bound on A7 by immundepletion-EMSA assay in FRTL5 total extracts.

Briefly a first round of immunoprecipitation, which included the interaction with the only antibody, was followed by the separation of immunocomplexes from the lysate by the addition of protein A agarose beads.

As procedural control it was performed a simultaneous immunoprecipitation with an unrelated antibody.

The band at 150 kDa recognized by our antibody was efficiently depleted from total protein mixture (Fig. 13).
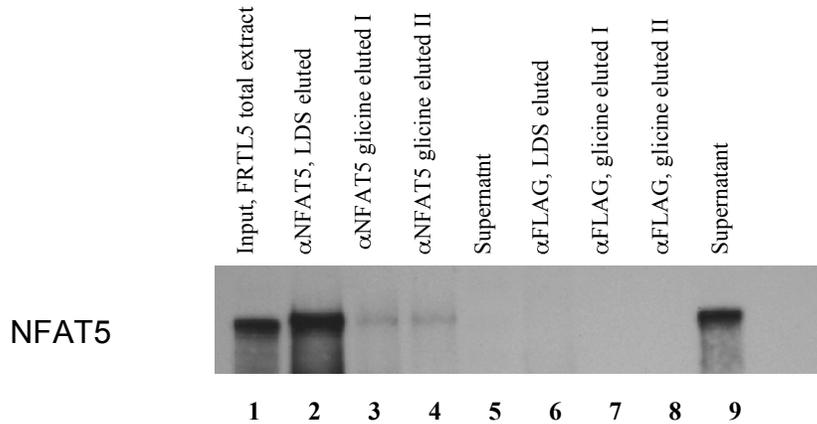
Fig.13: Immunodepletion of NFAT5 in FRTL5 cells
FRTL-5 total extracts (1) were challenged with anti NFAT5 antibody (2-5) and with anti FLAG control antibody (6-9).
The 150 kDa protein is visibile in denaturing elution of immuno-complexes (2) and mildly visibile in non denaturing glicine elution (3-4), the full depletion is visible as no band at 150 kDa in anti NFAT5 supernatant (5).
Denaturing elution of control anti-FLAG immunoprecipitation shows no detectable band on western (6) as well as in non denaturing glicine elution of anti FLAG immunocomplexes (7-8); anti FLAG supernatant show the band (9) as expected.

A7

	1	2	3	4
FRTL5	-	+	-	-
Supernatant IP NFAT5	-	-	+	-
Supernatant IP FLAG	-	-	-	+
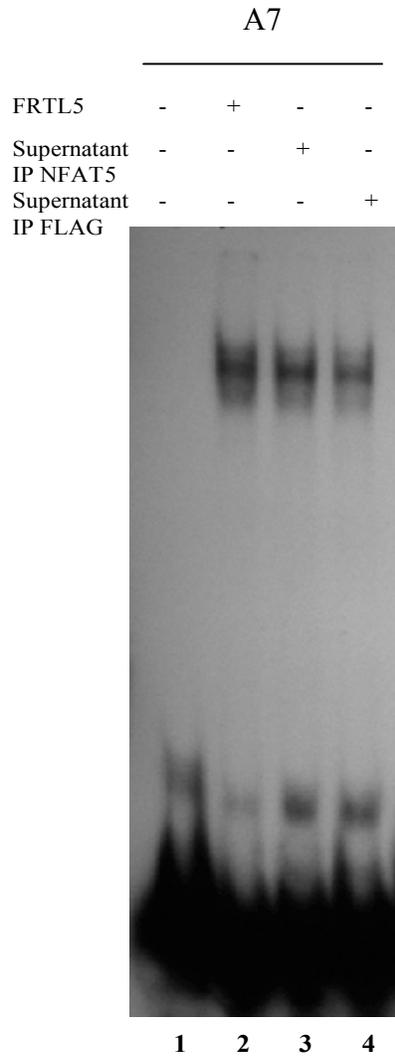
Fig.14: FRTL-5 extract immunodepleted of NFAT5 still bound to the A7 sequence. Line 1, free probe; line 2, FRTL-5 nuclear extract; line 3, supernatant of αNFAT5 immunoprecipitation; line 3, supernatant of control αFLAG immunoprecipitation.

The depleted extracts were then tested for binding activity on A7 sequence by EMSA assay (Fig. 14).

As shown the binding complex is not affected by the lack of the NFAT5 protein, both the αNFAT5 and αFLAG depleted lysates show activity on A7.

We can conclude that NFAT5 is not the candidate protein in binding A7.

Purification of transcription factor(s) binding to the A7 promoter region

As we did not get any result in the identification of the transcriptional activity on Pax8 promoter so far, we attempted its purification once again from FRTL-5 cells. The only difference between the previous procedure (Di Gennaro, unpublished data) was the use of the blue sepharose chromatography in place of the heparin in one of the purification steps.

As described in the Materials and method section 90 mg of nuclear extracts were loaded onto DEAE chromatography at a molar KCl concentration useful to prevent the binding from the resin itself, the flow through extract was then purified by Blue Sepharose chromatography.

Elution of discrete fractions was obtained by increasing salt concentration and the result was analyzed by EMSA (Fig. 15)

A7

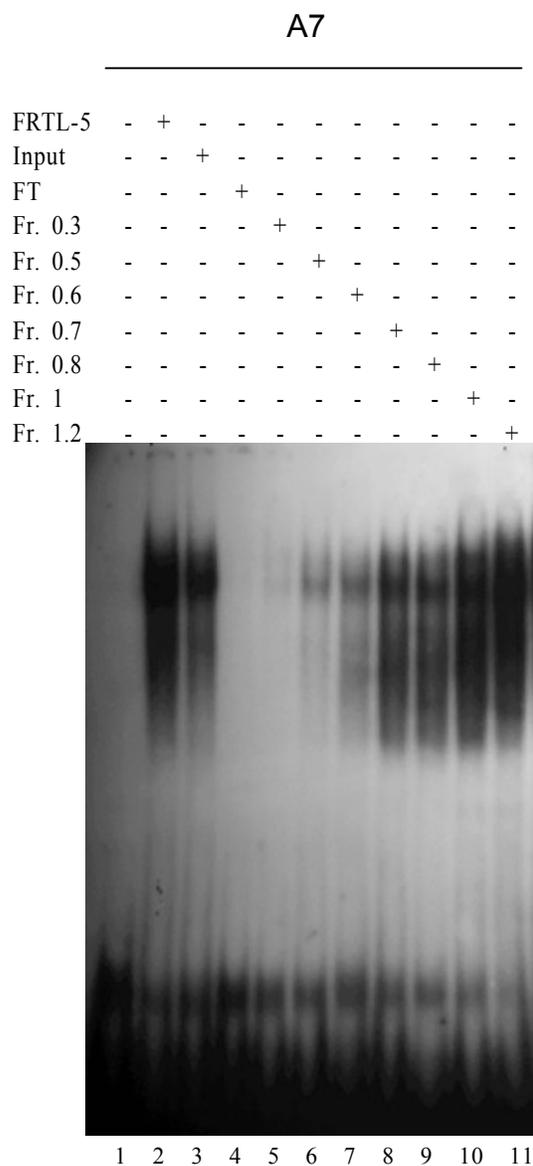|	1	2	3	4	5	6	7	8	9	10	11
FRTL-5	-	+	-	-	-	-	-	-	-	-	-
Input	-	-	+	-	-	-	-	-	-	-	-
FT	-	-	-	+	-	-	-	-	-	-	-
Fr. 0.3	-	-	-	-	+	-	-	-	-	-	-
Fr. 0.5	-	-	-	-	-	+	-	-	-	-	-
Fr. 0.6	-	-	-	-	-	-	+	-	-	-	-
Fr. 0.7	-	-	-	-	-	-	-	+	-	-	-
Fr. 0.8	-	-	-	-	-	-	-	-	+	-	-
Fr. 1	-	-	-	-	-	-	-	-	-	+	-
Fr. 1.2	-	-	-	-	-	-	-	-	-	-	+

Fig.15: EMSA of fractions eluted from Blue sepharose chromatography, binding on A7 sequence.
Lane 1, free probe; lane 2 FRTL-5 nuclear extract; lane 3, input Blue sepharose; lane 4, flow through (FT) Blue Sepharose; lane 5-11, fractions (Fr.) of Blue sepharose eluted with a linear KCl gradient from 0.3 to 1.2M

1 2 3 4 5 6 7 8 9 10 11

As shown the active complex spans on different elutions but focuses mainly on higher saline concentrations.

This is why we then decide to proceed to the next purification step by pulling active fractions from 0.6 to 1.2M KCl and bringing them to lower salt concentrations to allow the binding to the DNA affinity column, made up of purified A7 oligonucleotide multimers coupled to a CNBr activated resin.

Retained proteins were eluted by increasing KCl concentration and fractions thus obtained were analyzed by EMSA assay.

The majority of the binding activity is retained by the 0.4M fraction, with a smaller proportion in the 0,3M fraction as well (Fig. 16).

A7

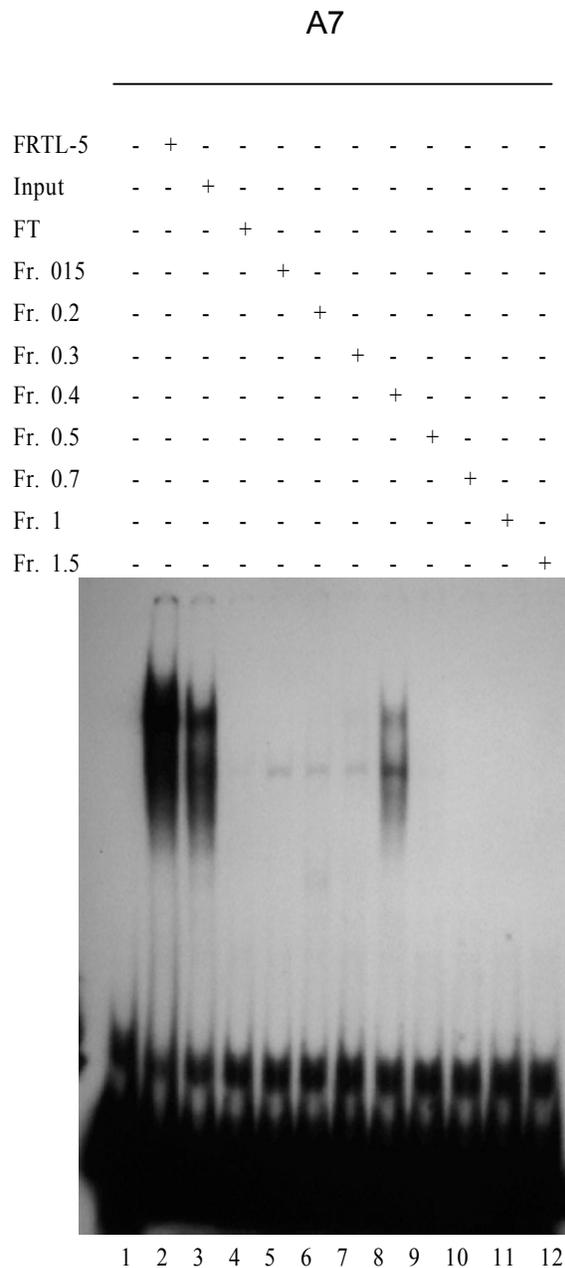|											
FRTL-5	-	+	-	-	-	-	-	-	-	-	-	-
Input	-	-	+	-	-	-	-	-	-	-	-	-
FT	-	-	-	+	-	-	-	-	-	-	-	-
Fr. 015	-	-	-	-	+	-	-	-	-	-	-	-
Fr. 0.2	-	-	-	-	-	+	-	-	-	-	-	-
Fr. 0.3	-	-	-	-	-	-	+	-	-	-	-	-
Fr. 0.4	-	-	-	-	-	-	-	+	-	-	-	-
Fr. 0.5	-	-	-	-	-	-	-	-	+	-	-	-
Fr. 0.7	-	-	-	-	-	-	-	-	-	+	-	-
Fr. 1	-	-	-	-	-	-	-	-	-	-	+	-
Fr. 1.5	-	-	-	-	-	-	-	-	-	-	-	+

Fig.16: EMSA of fractions eluted from A7/DNA affinity chromatography, binding on A7 sequence.
Lane 1, free probe; lane 2 FRTL-5 nuclear extract; lane 3, input DNA affinity; lane 4, flow through (FT) DNA affinity; lane 5-12, fractions of DNA affinity eluted with a lineat KCl gradient from 0.15 to 1.5M KCl

1 2 3 4 5 6 7 8 9 10 11 12

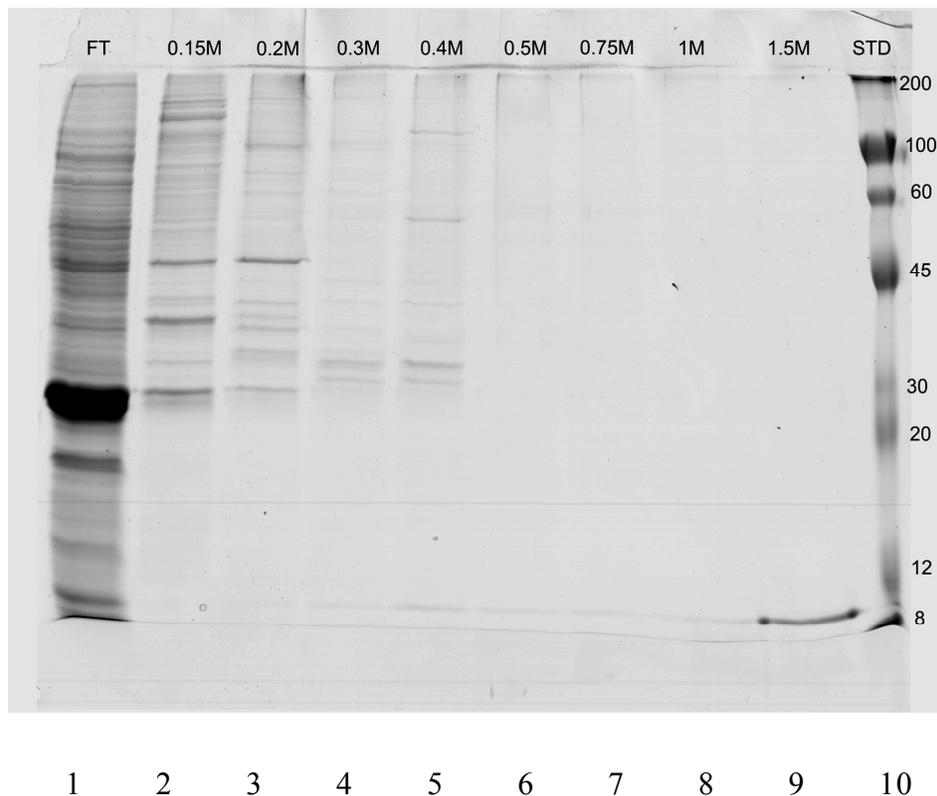All the fractions were TCA precipitated, resolved by SDS PAGE and colloidal Comassie stained (Fig. 17).

Fig.17: SDS PAGE of DNA affinity fractions, stained by Colloidal Comassie
Line 1, flow-throug; lanes 2 to 9, fractions from DNA-affinity
chromatography eluted with a linear KCl gradient from 0.15 to 1.5M.

There has been an enrichment in discrete bands of the proteins stained in the active fractions 0.3-0.4 compared to the non active fractions 0.2-0.15 and to the FT (Fig.17).

Lines corresponding to 0.2-0.3.0.4M fractions eluted, were excised, divided into bands (Fig. 18) and subjected to sequence analysis by Mass Spectrometry, at the ISPAAM CNR of Naples, Dott. Scaloni laboratory.
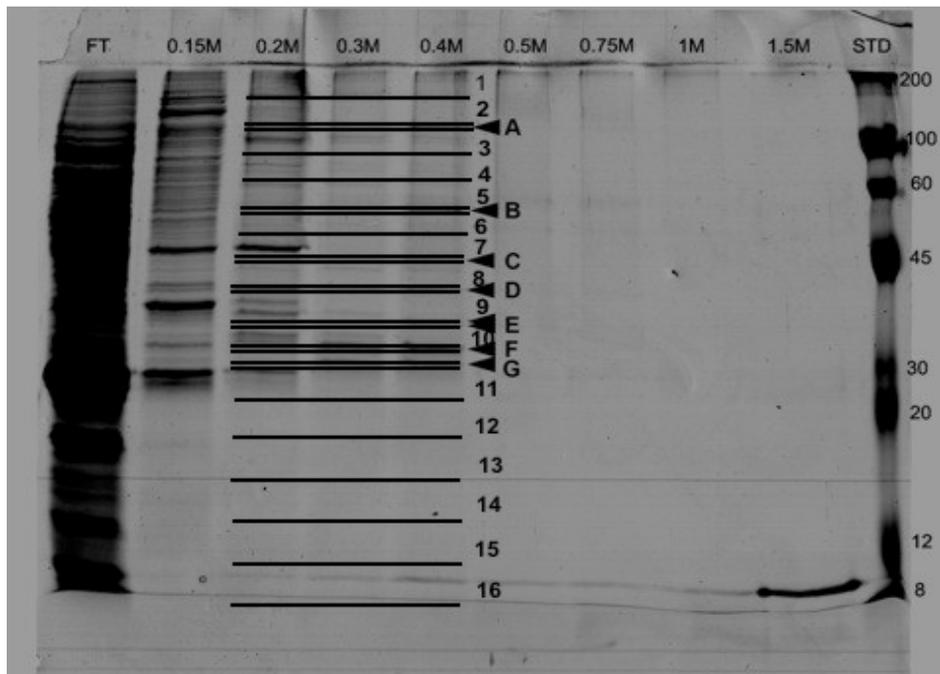
Fig.18: SDS PAGE of DNA affinity fractions, subdivision of the lanes analyzed.
The lines indicate the slices in which every lane has been subdivided; the arrows indicate a further subdivision according to the major bands visible onto the gel.

The analysis output returned a large number of proteins.

Two bioinformatic approaches were used to highlight the most significant candidates among those identified by mass spec.

The first one consisted in identify proteins from the active fractions absent in the HeLa proteome. The second approach consisted in rearrange the data in order to highlight only the DNA binding proteins and transcription factors in the active fractions and, among these, were taken into consideration those with the greatest number of peptides.

Different candidates match the above criteria and in particular we focused on three of these: NFATc1 meets the first point being found absent in the HeLa proteome (Table 2).

NFATc2 and NFATc3 meet the second one being the transcription factors most represented by the total number of peptides in active fractions (Table 3).

Uniprot AccNum	Protein symbol	Gene symbol	Description	Active fraction	Peptides	DNA binding domain	Transcription factor	GO Molecular Function
D4A8C8	GCFC1	GCFC1	GC-rich sequence DNA-binding factor 1	0,4M	2	YES	YES	transcription factor activity
Q3SWT1	NABP2	NABP2	nucleic acid binding protein 2	0,4-0,3M	9	YES	NO	DNA repair
D3ZE20	RGD1560225	NFATC1	nuclear factor of activated T-cells	0,4M	2	YES	YES	transcription factor activity
D3ZZQ2	NFATC2	NFATC2	Protein Nfatc2	0,4-0,3M	179	YES	YES	transcription factor activity
D3ZU59	NFATC3	NFATC3	Nuclear factor of activated T-cells, cytoplasmic, ca	0,4-0,3M	82	YES	YES	transcription factor activity
G3V740	NKX2-1	NKX2-1	NK2 homeobox 1	0,4-0,3M	15	YES	YES	transcription factor activity
P02768	ALB	ALB	albumin	0,4M	28	NO	NO	ion, drug, fatty acid binding
D4ACW8	APLF	APLF	aprataxin and PNKP like factor	0,4M	2	NO	NO	DNA repair
D3ZEJ9	RGD1310429	C17orf75	chromosome 17 open reading frame 75	0,4M	2	NO	NO	
D3ZR08	RCG_26646	LOC100362121	Protein LOC100362121	0,4-0,3M	16	NO	NO	ubiquitin protein-ligase activity
B2RYT0	MRPS21	MRPS21	mitochondrial ribosomal protein S21	0,4-0,3M	5	NO	NO	ribosomal protein
B2RZD5	RPL22L1	RPL22L1	ribosomal protein L22-like 1	0,4M	2	NO	NO	ribosomal protein
F1SW39	*SBP75*	*SBP75*	*PC4 and SFRS1 interacting protein 1*	*0,4M*	*2*	*NO*	*NO*	
G3V9I9	SREK1	SREK1	splicing regulatory glutamine/lysine-rich protein 1	0,4M	11	NO	NO	nucleotide binding
Q4KLK1	SRSF11	SRSF11	serine/arginine-rich splicing factor 11	0,4M	3	NO	NO	nucleotide binding
G3V798	SRSF4	SRSF4	serine/arginine-rich splicing factor 4	0,4M	12	NO	NO	nucleotide binding
Q68FR8	TUBA3a	TUBA3A	Tubulin alpha-3 chain	0,4M	2	NO	NO	structural molecule activity, GTPase activity
G3V7C6	TUBB4b	TUBB4B	Tubulin alpha-3 chain	0,4-0,3M	26	NO	NO	structural molecule activity, GTPase activity

Table 2: List of proteins from the active fractions 0.3M-0.4M not found in the HeLa proteome

		Slice	Accession	Description	Score	calc. pI	biological actvity	Total Peptides
	0.3M 4		D3ZZQ2	Protein Nfatc2 OS=R	142,84	6,61	transcr. Factor	179
	0.3M 5		D3ZZQ2	Protein Nfatc2 OS=R	188,09	6,61	transcr. Factor	179
	0.3M 6		D3ZZQ2	Protein Nfatc2 OS=R	72,85	6,61	transcr. Factor	179
	0.3M 7		D3ZZQ2	Protein Nfatc2 OS=R	118,28	6,61	transcr. Factor	179
	0.3M A		D3ZZQ2	Protein Nfatc2 OS=R	189,92	6,61	transcr. Factor	179
	0.3M B		D3ZZQ2	Protein Nfatc2 OS=R	130,90	6,61	transcr. Factor	179
0.4M 1			D3ZZQ2	Protein Nfatc2 OS=R	112,94	6,61	transcr. Factor	179
0.4M 2			D3ZZQ2	Protein Nfatc2 OS=R	283,06	6,61	transcr. Factor	179
0.4M 3			D3ZZQ2	Protein Nfatc2 OS=R	1052,26	6,61	transcr. Factor	179
0.4M 4			D3ZZQ2	Protein Nfatc2 OS=R	711,39	6,61	transcr. Factor	179
0.4M 5			D3ZZQ2	Protein Nfatc2 OS=R	311,37	6,61	transcr. Factor	179
0.4M 6			D3ZZQ2	Protein Nfatc2 OS=R	340,80	6,61	transcr. Factor	179
0.4M 7			D3ZZQ2	Protein Nfatc2 OS=R	437,26	6,61	transcr. Factor	179
0.4M 8			D3ZZQ2	Protein Nfatc2 OS=R	191,31	6,61	transcr. Factor	179
0.4M B			D3ZZQ2	Protein Nfatc2 OS=R	209,99	6,61	transcr. Factor	179
0.4M C			D3ZZQ2	Protein Nfatc2 OS=R	444,17	6,61	transcr. Factor	179
0.4M D			D3ZZQ2	Protein Nfatc2 OS=R	183,16	6,61	transcr. Factor	179
	0.3M 3		D3ZU59	Nuclear factor of acti	139,33	6,37	transcr. Factor	82
	0.3M 4		D3ZU59	Nuclear factor of acti	130,88	6,37	transcr. Factor	82
	0.3M 5		D3ZU59	Nuclear factor of acti	62,48	6,37	transcr. Factor	82
	0.3M 6		D3ZU59	Nuclear factor of acti	155,66	6,37	transcr. Factor	82
	0.3M 7		D3ZU59	Nuclear factor of acti	270,82	6,37	transcr. Factor	82
	0.3M B		D3ZU59	Nuclear factor of acti	190,40	6,37	transcr. Factor	82
	0.3M C		D3ZU59	Nuclear factor of acti	134,16	6,37	transcr. Factor	82
0.4M 2			D3ZU59	Nuclear factor of acti	89,56	6,37	transcr. Factor	82
0.3M 3			D3ZU59	Nuclear factor of acti	118,97	6,37	transcr. Factor	82
0.4M 4			D3ZU59	Nuclear factor of acti	256,08	6,37	transcr. Factor	82
0.4M 5			D3ZU59	Nuclear factor of acti	177,48	6,37	transcr. Factor	82
0.4M 6			D3ZU59	Nuclear factor of acti	142,17	6,37	transcr. Factor	82
0.4M 7			D3ZU59	Nuclear factor of acti	389,20	6,37	transcr. Factor	82
0.4M 8			D3ZU59	Nuclear factor of acti	249,43	6,37	transcr. Factor	82
0.4M B			D3ZU59	Nuclear factor of acti	75,26	6,37	transcr. Factor	82
0.4M C			D3ZU59	Nuclear factor of acti	183,56	6,37	transcr. Factor	82
0.4M D			D3ZU59	Nuclear factor of acti	118,41	6,37	transcr. Factor	82
0.4M 3			G3V817	Protein Xrcc5 OS=Ra	352,84	5,12	DNA helicase activity	44
0.4M 4			G3V817	Protein Xrcc5 OS=Ra	288,92	5,12	DNA helicase activity	44
	0.3M 3		G3V817	Protein Xrcc5 OS=Ra	177,21	5,12	DNA helicase activity	44
	0.3M 4		G3V817	Protein Xrcc5 OS=Ra	101,19	5,12	DNA helicase activity	44
	0.3M 5		G3V817	Protein Xrcc5 OS=Ra	86,74	5,12	DNA helicase activity	44
0.4M 5			G3V817	Protein Xrcc5 OS=Ra	35,58	5,12	DNA helicase activity	44
0.4M 6			Q9EPH8	Polyadenylate-bindin	159,63	9,50	transcr. factor,DNA b	22
	0.3M 4		Q9EPH8	Polyadenylate-bindin	307,76	9,50	transcr. factor,DNA b	22
0.4M 4			Q9EPH8	Polyadenylate-bindin	88,07	9,50	transcr. factor,DNA b	22
0.4M 2			G3V8T4	DNA damage-binding	357,47	5,26	damage DNA binding	22
	0.3M 2		G3V8T4	DNA damage-binding	284,88	5,26	damage DNA binding	22
		0.2M D	D3ZQV8	RCG47045, isoform	292,73	8,73	double-stranded D	21
		0.2M 12	D3ZQV8	RCG47045, isoform	200,86	8,73	double-stranded D	21
		0.2M 13	D3ZQV8	RCG47045, isoform	155,75	8,73	double-stranded D	21
0.4M 8			D3ZQV8	RCG47045, isoform	55,10	8,73	double-stranded D	21
		0.2M 11	F1M1H9	Uncharacterized prot	463,16	9,69	transc. Factor HMG	19
0.4M 11			F1M1H9	Uncharacterized prot	179,35	9,69	transc. Factor HMG	19
0.4M 8			D3ZR08	Protein LOC1003621	93,78	9,16	damage DNA binding	16
0.4M D			D3ZR08	Protein LOC1003621	61,07	9,16	damage DNA binding	16
	0.3M D		D3ZR08	Protein LOC1003621	55,22	9,16	damage DNA binding	16
0.4M C			D3ZR08	Protein LOC1003621	50,27	9,16	damage DNA binding	16
	0.3M 8		D3ZR08	Protein LOC1003621	46,35	9,16	damage DNA binding	16
	0.3M G		G3V740	Homeobox protein N	62,04	9,69	transcr. Factor	15
0.4M 11			G3V740	Homeobox protein N	165,78	9,69	transcr. Factor	15
	0.3M F		G3V740	Homeobox protein N	61,72	9,69	transcr. Factor	15
		0.2M B	O35986	Zinc finger Ran-bindi	89,14	9,89	transc. Factor zinc	13
		0.2M 7	O35986	Zinc finger Ran-bindi	128,31	9,89	transc. Factor zinc	13
0.4M 1			O35986	Zinc finger Ran-bindi	62,06	9,89	transc. Factor zinc	13
0.4M 14			F1LP73	Uncharacterized prot	193,26	10,46	histone,DNA binding	6
	0.3M 13		F1LP73	Uncharacterized prot	175,38	10,46	histone,DNA binding	6
	0.3M 16		F1LP73	Uncharacterized prot	141,15	10,46	histone,DNA binding	6
		0.2M 2	F8WGA1	Cullin-associated NE	98,61	5,83	transc. Factor	6
0.4M 2			F8WGA1	Cullin-associated NE	68,41	5,83	transc. Factor	6
0.4M 8			F1LPS8	Transcriptional activa	275,29	6,44	transcr. factor,DNA b	6
0.4M 4			Q99PK0	Pre-mRNA-splicing fa	76,58	6,23	transcr. Factor,dama	3
0.4M 11			B0BN99	Hmgb3 protein OS=	72,26	8,37	transcr. Factor HMG	2
0.4M 7			B5DFL5	Protein Sap30bp OS=	59,01	4,87	transcr. factor	2
0.4M 2			D4A8C8	Uncharacterized prot	65,22	5,67	transcr. factor	2
0.4M 4			F1SW39	PC4 and SFRS1 inter	120,65	9,13	transcr. Cofactor	2
0.4M 16			Q5RK03	CDKN2AIP N-termina	91,72	5,29	transcr. Actvity,DNA	2

Table 3: Transcription factors identified by mass spectrometry ordered according to the total number of peptides of each.
The lines are divided by color: orange indicates the proteins present in the active fractions 0.4-0.3M, pink stands for 0.4-0.2M, violet for the proteins in 0.4M

From here onwards the involvement of these factors to the transcriptional regulation of the *Pax8* promoter should be confirmed just as done previously.

Transcript abundance of *Nfatc* genes identified in FRTL-5 and HeLa cells

The different binding activity observed in the different extracts, may be due to the differential gene expression in the cell lines analyzed.

So the first point I assessed was the differential *Nfatc2/c3* gene expression in FRTL-5 versus HeLa cells, one of the cell lines that show no activity on the A7 sequence.

As shown *Nfatc2* mRNA is more expressed than its human counterpart expression (Fig.22), while *Nfatc3* mRNA has roughly the same low levels of expression between the two cell lines.
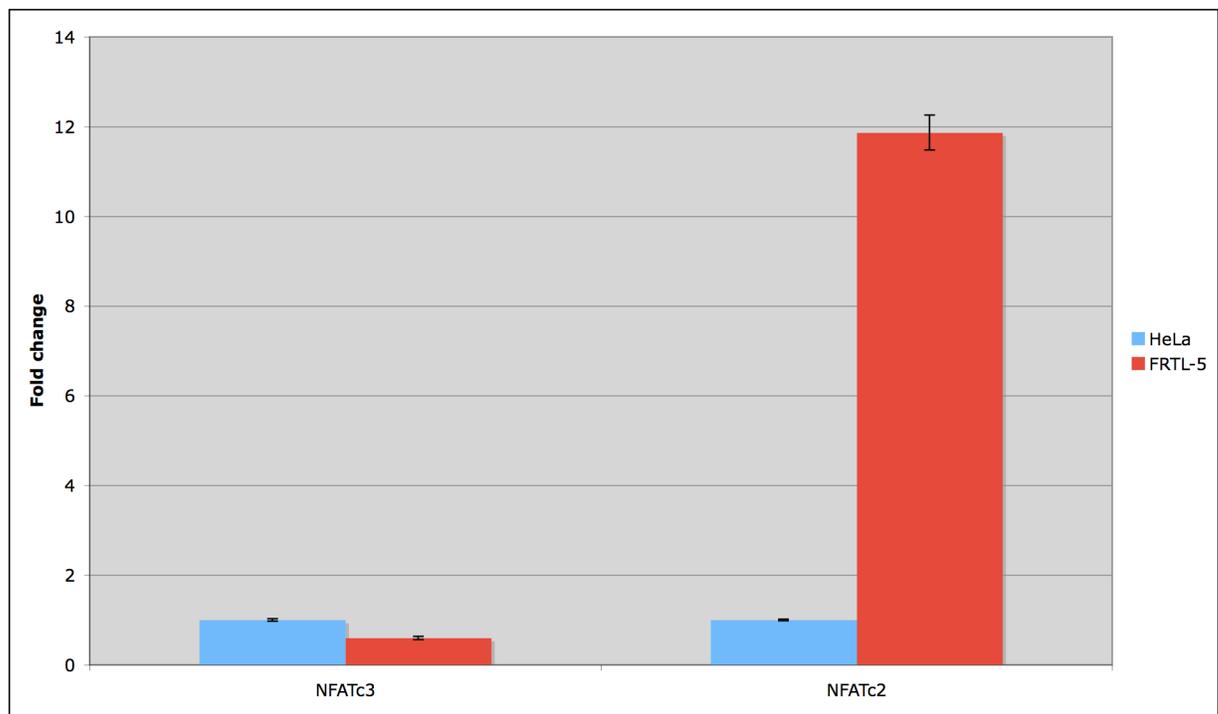
Fig.19: Comparison of Nfatc2/c3 mRNA gene expression by qRT between HeLa and FRTL-5.
Quantitative real time PCR was carried out using pair of primers common to all isoforms of each gene per cell line. The mRNA expression level was normalized to the tubulin mRNA expression level, in fold change to HeLa Nfatc2/c3 gene expression, mean value of technical triplicate.
Error bars represent the error standard.

For this reason I went on with my work by focusing on *Nfatc2* and *Nfatc1* gene expression including some of the cell lines analyzed by DNA binding activity on A7.

I included into the analysis the Jurkat cell line which showed activity on A7 and is reported to express NFAT family proteins, and the human neuroblastoma cell line SK-N-BE, one of the cell lines with no activity on A7.

Both *Nfatc1-Nfatc2* genes share the same trend of expression with the highest level in FRTL-5, in comparison with a very low expression value for HeLa and SK-N-BE, while the gene expression in Jurkat cells stands between the two. Even if these preliminary data need to be confirmed at the protein level it can be stated that *Nfatc1-c2* transcript expression is higher in those cell lines which show binding on A7.
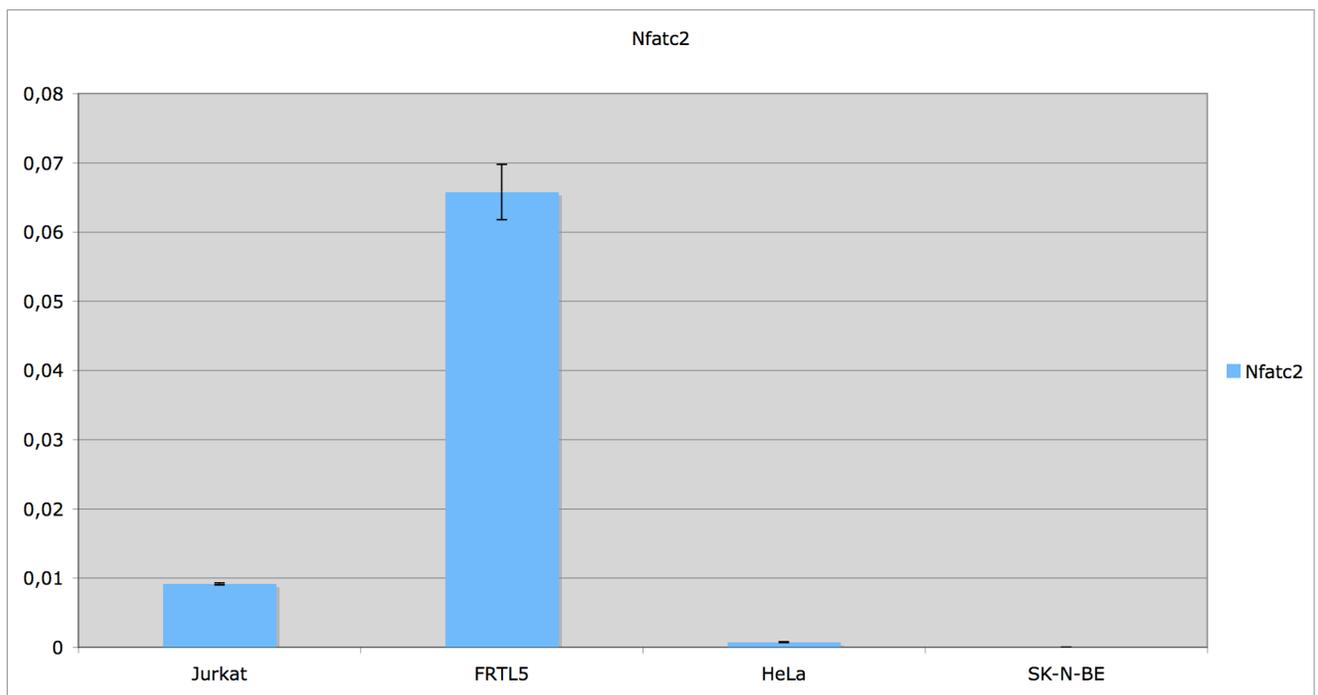

Fig. 20: Comparison of Nfatc2 transcript abundance in FRTL5, Jurkat, HeLa and SK-N-BE by qRT.
Quantitative real time PCR was carried out using pair of primers common to all isoforms of each gene per cell line. The mRNA expression level is the mean value of technical triplicate normalized to the tubulin mRNA expression level. Error bars represent the standard deviation.
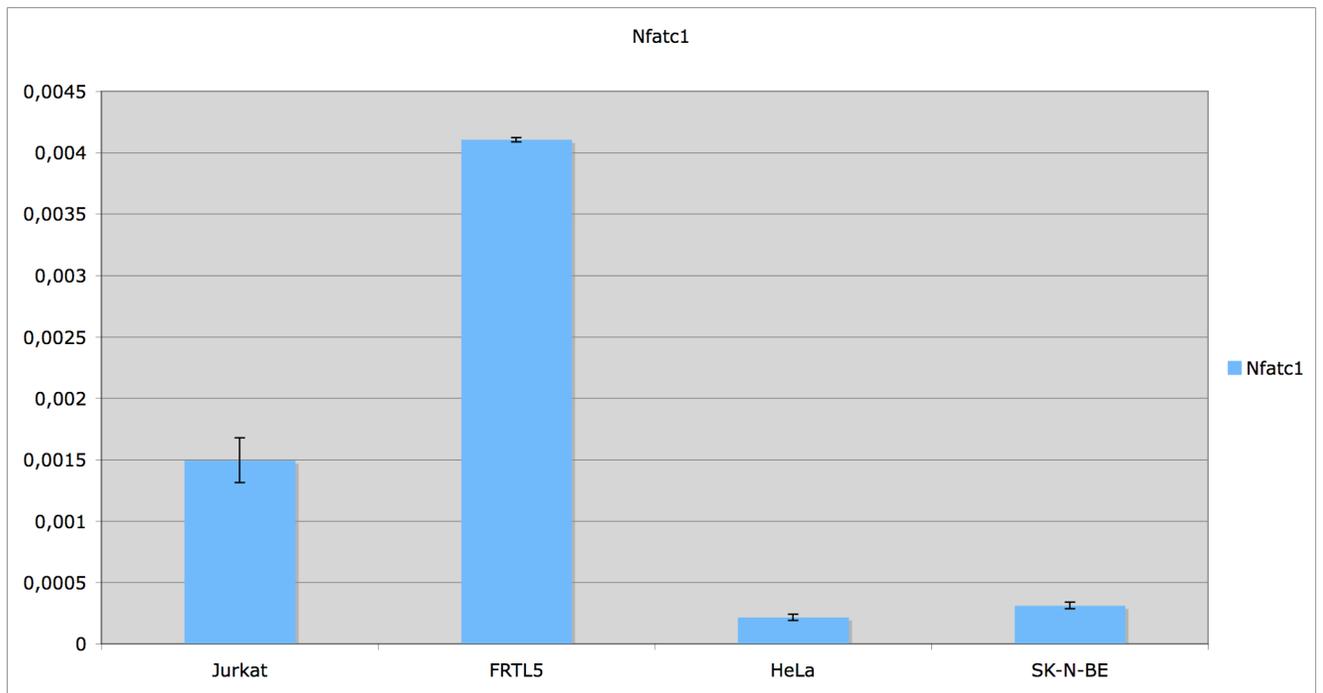
Fig. 21: Comparison of Nfatc1 transcript abundance in FRTL5, Jurkat, HeLa and SK-N-BE by qRT.
Quantitative real time PCR was carried out using pair of primers common to all isoforms of each gene per cell line. The mRNA expression level is the mean value of technical triplicate normalized to the tubulin mRNA expression level. Error bars represent the standard deviation.

DISCUSSION

My thesis work has been focused on the study of an unknown DNA binding activity that recognize the *Pax8* promoter and shows a preferential presence in differentiated thyroid cells.

The Pax8 transcription factor has an important role in thyroid organogenesis as it, together with Nkx2-1, Hhex and Foxe1 transcription factors, marks a restricted number of cells in the ventral wall of the primitive pharynx as precursors of thyroid follicular cells.

Each of these transcription factors is also expressed in other tissues, but such a combination is a unique hallmark of both differentiated thyroid follicular cells and their precursors. Studies in animal models have shown the relevance of these factors for thyroid development. In the absence of Nkx2–1, Hhex, Pax8, or Foxe1, the thyroid anlage is correctly formed but the subsequent thyroid morphogenesis is severely impaired. Thus, neither of this transcription factors is alone required for the specification of thyroid progenitors in the foregut endoderm and to date there are still no data available on the early events in thyroid organogenesis.

It is well known that cell determination is a process in which a cell's potency is progressively restricted during development involving both cell-autonomous mechanisms and inductive signals from a cell's surroundings. Combinations of these influences result in progressive alterations in the gene expression patterns of embryonic cells.

During morphogenesis, embryonic tissues sorrounding the developing thyroid become increasingly complex as suggested by the structure of the pharyngeal metamere; this reflects the complexity of the network signaling involved.

Recent studies have in fact demonstrated the importance of the expression of some genes in tissues adjacent to thyroid primordium and that can influence the morphogenesis of the thyroid gland although they are never expressed during the development of this tissue.

For example FGFs derived from the mesoderm evidently influence thyroid morphogenesis: *fgf*8 deficient zebrafish embryos show a severely hypoplastic thyroid (Wendl et al., 2007), in mice FGF8 was recently shown to stimulate the generation of endoderm progenitors committed to a thyroid fate (Lania et al., 2009) by means of the Tbx protein: the

overexpression of Fgf8 in the mesoderm partially rescues the thyroid defects observed in *tbx* null mice.

This provides a direct proof of a central role of mesoderm in thyroid development in higher vertebrates.

Among the cell-autonomous mechanisms in cell fate determination, the transcriptional regulation resembles some aspects of the embryonic regulatory programs, suggesting that studies of cells in culture may improve our understanding of embryonic development (Fagman 2011).

An example of this is given by the transcriptional coactivator Taz, recently shown to enhance in vitro the activity of Pax8 and Nkx2-1 on the thyroglobulin promoter (Di Palma et al 2009).

Taz is expressed in coincidence with the onset of Tg biosynthesis in mouse embryonic thyroid at E 14.5 and in its lack Pax8 and Nkx2-1 are unable to activate the *tg* promoter, suggesting a Taz-mediated coactivation of genes at earlier stages of development.

In general transcription factors exert their influence on the specification by cell-autonomous regulation of tissue specific genes; hence the study of cis-regulatory elements of thyroid specific selector genes would be the first step in the elucidation of the transactivating factors estabilishing the thyroid-specific molecular network.

In particular during these years our research group has been focusing on the regulation of the *pax8* gene that is the first of the four above mentioned selector genes to be expressed.

Pax8 is also considered a master gene in the maintenance of functional differentiation in thyroid cells, activating transcription of thyroid specific genes such as thyroglobulin and thyroperoxidase which are exclusively expressed in this organ; in the *tg* promoter Pax8 binding sites overlap with that of Titf-1, while in *tpo* promoter it was found a mutation that affected only Pax8 binding thus suggesting an important role for Pax8 in thyroid specific gene expression.

Moreover it was demonstrated that the introduction of Pax8 in the transformed PCPy cell line free of thyroid specific genes, was able to reactivate their transcription (Pasca di Magliano et al 2000).

Recently it has been reported that a transient overexpression of the transcription factors Nkx2-1 and Pax8 is sufficient to direct mouse embryonic stem-cell differentiation into thyroid follicular cells that organize into three-dimensional follicular structures when treated with thyrotropin (Antonica F. et al., 2012).

Hence the identification of the novel thyroid specific DNA binding activity in the pax8 promoter could elucidate the early events in thyroid commitment; we therefore carried out biochemical analysis on the binding complex and identified some probable candidates belonging to the family of transcription factors NFAT, integrating mass spectrometry data with bioinformatic data: NFATc2, NFATc3, NFATc1.

The NFAT (nuclear factor of activated T cell) family of transcription factors was initially identified as comprising inducible nuclear factors which bound the interleukin-2 promoter in activated T cell (Shaw, J. P. et al. 1988).

The family consist of five members: NFATc1- also called NFATc or NFAT2; NFATc2 - also called NFATp or NFAT1; NFATc3 - also called NFATx or NFAT4; NFATc4 - also called NFAT3 and NFAT5.

The primordial family member is NFAT5, the only NFAT-related protein represented in the *Drosophila* genome, while NFAT1-4 does not have homologs present in invertebrates and appear to have emerged simultaneously early in the course of vertebrate evolution.

These proteins are widely expressed but each of them has a different tissue specificity.

The hallmark of NFAT members (except for NFAT5) is their regulation by Ca^{++} and the Ca^{++}/calmodulin-dependent serine phosphatase calcineurin. NFAT proteins are phosphorylated and reside in the cytoplasm in resting cells; upon stimulation, they are dephosphorylated by calcineurin, translocate to the nucleus, and become transcriptionally active. The primary role for them in this pathway is the expression/repression of genes like *tnfα*, *IL-2*, *IL-4* involved in the differentiation program of Th1/Th2.

All members of the NFAT family share a conserved core region which consist of two tandem domains: a regulatory domain, the NFAT-homology region (NHR) which contains a transactivation domain, docking sites for calcineurin and NFAT kinases, and the REL-homology region (RHR) which binds DNA and is evolutionary related by structure similarity to the Rel/NFkB family.

However, until now no information is given on the regulation of these proteins in thyroid and therefore on the role that these domains, if present together, exercise in the thyroid.

Although the redundancy in some pathways due to common DNA-binding specificities, NFAT members can differ in their functions through unique partner interactions and post-translational modifications, result of further modulation by additional inputs from diverse signaling pathways, which affect NFAT kinases and nuclear partner proteins.

For example sumoylation has recently been identified as a new mechanism that regulates NFAT1-specific nuclear retention, in that sumoylation sites located in its C terminus are not conserved in other members.

In T cell, the specific expression of *IL-3* gene is partly controlled by cooperation between Oct and NFATp binding onto the enhancer.

Moreover the NFAT role in cell differentiation covers many other tissues beside lymphoid lineage; this is in accordance with the wide expression of this family.

Gene targeting has implicated NFATc1 in cardiac valve developmental process. Mice homozygous for two different single-exon deletions in the NFATc1 gene display defects in cardiac valve and septum formation; the cardiac failure is lethal by E13.5. (de la Pompa et al. 1998; Ranger et al. 1998a).

Nfatc3 -/- mutant mice exhibit skeletal muscle hypoplasia (Kegley et al. 2001) due to an impaired muscle development during embryogenesis.

Nfatc3/c4 null mice die at E10.5 with vasculature patterning defects not related to the inability of endothelial cells to differentiate but to the failure of responding and give signals essential for the assembly of vessels along specific pathways (Graef et al 2001).

To confirm the involvement of these proteins as effectors in the early events in thyroid commitment, I firstly carried out studies on their expression in FRTL-5 bringing Jurkat and HeLa cells respectively as positive and negative controls.

These studies have reported a good expression for the transcript of NFATc1 and NFATc2 in FRTL-5 when compared with that of some cell lines with no binding on A7, suggesting a positive regulation for these genes in FRTL-5. From here it will follow the setting up of further experiments to confirm that these proteins actually bind the Pax8 promoter regulating it.

Anyway these preliminary experiments, together with the role in development above mentioned, could lead us thinking of an involvement of the NFAT family to the thyroid commitment event as that of a tissue specific activity given by an integration of Ca^{++}/calcineurin and developmental signals on NFAT transcription complexes assembled on the Pax8 promoter.

MATERIALS AND METHODS

Cell cultures

The rat thyroid FRTL-5 cell line was grown in Coon's modified F-12 medium (Euroclone) supplemented with 5% v/v calf serum and a six-hormone mixture (Ambesi-Impiombato and Coon, 1979). The human epitheloid cervix carcinoma HeLa, the dog epithelial-like kidney MDCK cell line and the human neuroblastoma cell line SK-N-BE were grown in Dulbecco's modified Eagle's medium (Euroclone) supplemented with 10% fetal calf serum (Hyclone). The human T lymphoblastoid cell line Jurkat was grown in RPMI 1640 (Gibco) supplemented with 10% v/v fetal calf serum (Hyclone).

RNA interference

FRTL-5 cells were plated ($18x10^4$) in 6-well plate and were transfected with 50 nM Gtf2i siRNA (4390771- s165531 life technology) and 50nM siRNA negative control (AM4611 life technology); INTERFERin transfection reagent (Polyplus 409-10) was used following the manufacturer's protocol. Cells were harvested 96 h after transfection, proteins and total RNA were extracted.

RNA extraction, quantitative real time RT-PCR.

Total RNA was extracted using TRIzol reagent (Invitrogen) following the manufacturer's protocol. For qRT-PCR the cDNA was synthetized using Quantitect Reverse Transcription kit (205-313 Qiagen). Real time RT-PCR analysis was performed using SYBR Green PCR master mix (Applied Biosystem) in an iCycler-iQ real-time detection system. Reactions were carried out in triplicate. The specific primer sets used for real time analysis were: tubulin1a human-rattus (Fwd:

CAACACCTTCTTCAGTGAGAC Rev: TACATGATCTCCTTGCCAATGGT), Pax8 (Fwd: 5'CAGCTATGCCTCTTCCGCTATT3'; Rev: 5'TGTGGCTGTAGGCATTGCC3'), NFATc2 human (Fwd:TTCCCATCTGCAGCATCCCAG; Rev:GCTGCCTTCTGTCTCATAGTGG), NFATc2 rattus (Fwd: ATTCCGCTCCAGAGTCCATC Rev: CATAGGAGCCCGACTGGTTG), NFATc1 human (Fwd: GCCATCCTCTCCAACACCAA Rev TTCAGGATTCCGGCACAGTC), NFATc1 rattus (Fwd: AATAACCAGCCCCGTCCAAG Rev: GGTCAGAGCTGGCTCAAAGT), Nfatc3 human (Fwd:GCCCATTATGAAACTGAAGGTAGC Rev:CGATCATCTGCTGTCCCAAT), NFATc3 rattus (Fwd:GGTGGCCATCCTGTTGTGAAG Rev: TCCAGTAATGCGATGCACCT) .Expression values are means ± SE of technical triplicates, normalized by the expression of tubulin1a.

Nuclear extracts

Nuclear extracts were prepared by a modification of described procedures (Civitareale D. et al 1989). As for 100 mm dishes extraction cells were lysed starting from cells grown to near confluency, washed twice with ice-cold phosphate-buffered saline (PBS) and lysed in 400 ul of buffer I. After 10 min on ice, the cells were vortexed and sucrose restore solution was then added (1/10 of total volume). The homogenate was spun for 20 sec at 13.000 rpm and surnatant (containing cytoplasm) was removed. The nuclear pellet was resuspended in solution IS, spun as above, and resuspended in 50 ul of solution II. After 30 min on ice, the nuclear suspension was spun at 13.000 rpm for 5 minutes. The same procedure as above for cell grown in 150 mm dishes except for the volumes, 2.5x greater in all steps. The nuclear proteins present in the surnatant were frozen in liquid nitrogen and stored in aliquots at -80°C.

Electrophoretic mobility-shift assay

Single stranded oligonucleotides (1 pmol) were end-labelled with 2ul of [γ-^{32}P]ATP (Perkin Elmer; specific radioactivity 3000 Ci/mmol) in the presence of 10 units of T4 polynucleotide kinase (New England BioLabs) at 37° for 1 hour and then annealed to antisense complementary sequences to make double-strand probes. The probes (80000 cpm) were mixed with nuclear cell extracts (4 µg) in

20 µl of reaction buffer containing 20 mM Tris (pH 7.5), 75 mM KCl, 1 mM dithiothreitol, 1mg/ml BSA, 10% glycerol, 3 ug poly (dI-dC) as a non-specific competitor, and incubated at room temperature for 30 min. DNA-protein complexes were resolved on a 5% (w/v) non denaturing polyacrylamide gel in 0.5x TBE buffer at 200 V; gels were dried and autoradiographed with intensifying screens at -70° from 12 to 24 hours. For supershift assays, 5µg of specific antibodies αGtf2i (H-58 Santa Cruz, CA), αNFAT5 (H-300 Santa Cruz, CA), αFLAG (clone M2 F3165 SIGMA) were incubated with protein extracts for 20 min before adding the probe.

Western Blot analysis

Cells were washed twice with ice-cold phosphate buffered saline (PBS) and lysed in a buffer containing Tris pH 8 50 mM, MgCl$_2$ 5 mM, NaCl 150 mM, deossicolic acid 0, 5%, SDS 0,1%, triton X-100 1%, and protease inhibitors cocktail (SIGMA), 1mM dithiothreitol (DTT), 1mM phenylmethylsulfonyl fluoride (PMSF). The protein concentration was determined by bicinchoninic acid protein assay (Pierce Biotechnology, Rockford, IL) and extracts were separated on NuPAGE® 4-12% Bis Tris gels (Invitrogen) under reducing conditions and transferred to polyvinylidene difluoride (PVDF) membrane Immobilon P (Millipore, Bredford, MA, USA) for 1,30 hour; membranes were blocked 1 hour or overnight in 5% nonfat dried milk in tris buffered saline plus Tween 20 0, 1%. Membranes were then hybridized with antibodies against Gtf2i (H-58 sc-28716x Santa Cruz, CA), NFAT5 (H-300 sc-13035x Santa Cruz, CA), tubulin (SIGMA). The filters were washed three times in Tris-buffered saline plus 0.1% Tween 20 before the addition of horseradish peroxidase-conjugated secondary antibodies for 45 min. Horseradish peroxidase was detected with ECL western Blot substrate (Pierce) and X-ray films (Fuji film).

Immunodepletion

For immunodepletion FRTL-5 cells were washed twice with ice-cold phosphate-buffered saline (PBS) and lysed in Tris pH 8 50 mM, MgCl$_2$ 5 mM, NaCl 150 mM, deossicolic acid 0, 5%, SDS 0,1%, triton X-100 1%, and protease inhibitors cocktail (SIGMA), 1mM dithiothreitol (DTT), 1mM phenylmethylsulfonyl fluoride (PMSF). 1 mg of total protein extract was adjusted to a final volume

of 300 ul and precleared with 50 ul slurry protein A agarose beads on a rotating wheel for 20 min at 4°C; discarded the beads the supernatant was incubated with 10ug of specific antibodies (monoclonal anti GTF2i WH0002969M1 SIGMA, anti NFAT5 H300 sc-13035x Santa Cruz) for 2h; immunocomplexes were captured by adding 50 ul slurry Protein A agarose beads and by gentle rocking on a rotating wheel for 1h at 4°C. Beads were collected, washed five times in lysis buffer and bound proteins were eluted in native state using 0.1 M glycine pH 2.7 for 5 min on ice then neutralized using 0,1 volume of 1M Tris-HCl pH 7.5; denaturing elution was performed using NuPAGE® LDS Sample Buffer and reducing agent boiling samples for five minutes. Equal volumes of supernatant and eluted proteins were separated by Western Blot according the above mentioned procedure; supernatant was resolved for DNA-protein complexes by EMSA in accordance with the described procedure.

Purification procedure

Nuclear extracts from FRTL-5 cell line (about 90 mg from 270 cell culture dishes of 150 mm) were loaded on a 2,5ml packed DEAE column (DEAE-Sepharose fast flow, GE Healthcare) at a flow rate of 2 ml/min, after 0.4 M to 0.3 M KCl salt dilution in buffer A (40 mM Hepes pH 7.9, 6mM $MgCl_2$, 0.2mM EDTA, 20% v/v glycerol, 1 mM DTT). This KCl molarity prevents proteins from being retained by the column thus separating them from the DNA which remains bound to the resin. The flow through is then loaded on 2 ml packed Blue Sepharose column (Capto Blue fast flow, GE Healthcare) at a flow rate of 0,5 ml/min. Retained proteins were eluted by raising the KCl concentration of buffer A (0.3-0.5-0.6-0.7-0.8-1-1.2M KCl). Fractions containing the activity span from 0.6 to 1.2 M KCl, as assessed by EMSA, were pooled and carried to 50 mM by dilution in buffer Z (25 mM Hepes pH 7.8, 12.5 mM $MgCl_2$, 0.1% NP40) upon concentration on centrifugal filter devices (Amicon Ultra-15). Extracts were then loaded on a 4ml slurry (2 ml packed) DNA affinity column (CNBr Sepharose fast flow, GE Healthcare) constructed with the multimeric A7 oligonucleotide (see below). Retained proteins were eluted by a 5 ml linear KCl gradient (0.15-0.2-0.3-0.4-0.5-0.75-1-1.5M KCl) and active fractions were assessed by EMSA. All fractions were trichloroacetic acid precipitated, resolved by SDS-PAGE 12% and Comassie colloidal stained. Lines corresponding to 0.2-0.3.0.4 M fractions eluted, were cut into 16 slices and 7-band, analyzed in duplicate by nLC-ESI-IT-MS/MS after digestion with trypsin. Research in database was made using

the software MASCOT (using Proteome Discoverer 1.3) of the UniProt database (Rattus) updated, considering significant identifications with Mascot score > 25 and at least two different peptides (ISPAAM CNR of Naples, Dott. Scaloni laboratory).

DNA affinity column

Chemically synthesized and 5' phosphorilated complementary oligonucleotides (250 ug each) with an extra BamHI recognition site at 5' ends, were annealed together with 5' $[\gamma\text{-}^{32}P]$ phosphorilated-BamHI recognition site-forward oligonucleotides (500000 cpm) in 0.4 M NaCl, 10 mM Tris pH7.5, at 95°C for 5 min, 65°C for 10 min, 37°C 10 min, and room temperature for 10 min. The DNA is ethanol-precipitated, resuspended in H_2O and subjected to ligation reaction over night at 16°C. DNA is then phenol-extracted, precipitated with ethanol, dried in vacuum and dissolved in water. Analysis of the resulting DNA by agarose gel electrophoresis typically shows oligomers of the basic oligodeoxynucleotide unit ranging from 3-mers to 75-mers. The DNA oligomers are covalently attached to CNBr Sepharose (0,5 g previously swelled in 25 ml HCl 1mM, then equilibrated in buffer K-phosphate 10 mM pH8 to give final 2ml packed) by mixing the DNA and resin on a rotating wheel over night at room temperature. After recovering the unbound DNA the unreacted CNBr Sepharose is inactivated by ethanolamine 1M pH 8 by mixing on a rotary shaker at room temperature for 2h. The resin is then washed with buffer K-phosphate 10 mM pH 8, K-phosphate 1M pH8 and KCl 1 M and kept in storage buffer (10 mM Tris pH7.6, 0.3 M NaCl, 1mM EDTA, 0.02% NaN_3).

BIBLIOGRAPHY

- Antonica F, Kasprzyk DF, Opitz R, Iacovino M, Liao XH, Dumitrescu AM, Refetoff S, Peremans K, Manto M, Kyba M, Costagliola S. Generation of functional thyroid from embryonic stem cells. 2012; Nature 491(7422):66-71.

- Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. 1981; Cell 27:299-308.

- Breathnach R, Chambon P. Organization and expression of eukaryotic split genes coding for proteins. 1981; Annu. Rev. Biochem. 50:349-83.

- Burke TW, Kadonaga JT. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev. 1997; 11:3020-31.

- Capelson Maya, Corces Victor G. Boundary elements and nuclear organization. 2004; Biology of the Cell 96(8): 617-29.

- Carcamo J, Buckbinder L, and Reinberg D. The initiator directs the assembly of a transcription factor IID-dependent transcription complex. 1991; Proc. Natl. Acad. Sci. 88: 8052–8056.

- Congdon T., Nguyen LQ, Nogueira CR, Habiby RL, Medeiros-Neto G, Kopp P. A novel mutation (Q40P) in PAX8 associated with congenital hypothyroidism and thyroid hypoplasia: evidence for phenotypic variability in mother and child. 2001; J. Clin. Endocrinol. Metab. 86(8):3962-7.

- Crabtree GR, Olson EN. NFAT signaling: choreographing the social lives of cells. 2002; Cell. 109Suppl:S67-79.

- Crompton MR, Barletti TJ, MacGregor AD, Manfioletti G, Buratti E, Giancotti V, Goodwin GH. Identification of a novel vertebrate homeobox gene expressed in hematopoietic cells. 1992 Nucleic Acid Res. 20:5661-5667

- De la Pompa JL, Timmerman LA, Takimoto H, Yoshida H, Elia AJ, Samper E, Potter J, Wakeham A, Marengere L, Langille BL et al. Role of the NF-ATc transcription factor in morphogenesis of cardiac valves and septum. 1998; Nature 392:182–186

- Di Gennaro A, Spadaro O, Baratta MG, Di Lauro R and De Felice M. Functional analysis of the Pax8 promoter reveals autoregulation and the presence of a novel thyroid-specific DNA binding actvity. 2012; Thyroid.

- Di Palma T, D'Andrea B, Liguori GL, Liguoro A, de Cristofaro T, Del Prete D, Pappalardo A, Mascia A, Zannini M. TAZ is a coactivator for Pax8 and TTF-1, two transcription factors involved in thyroid differentiation. 2009; Exp. Cell. Res. 315(2):162-75.

- Dong-Hoon Lee, Naum I. Gershenzon, Malavika Gupta, Ilya P. Ioshikhes, Danny Reinberg, Brian A. Lewis. Functional Characterization of Core Promoter Elements: the downstream core element is recognized by TAF1. 2005; Mol. Cell. Biol. 25:9674-9686.

- Edgar Dahl, Haruhiko Koseki, Rudi Ballino. Pax genes and organogenesis. 1997; BioEssays. 19(9):755-765.

- Esensten JH, Tsytsykova AV, Lopez-Rodriguez C, Ligeiro FA, Rao A, Goldfeld AE. NFAT5 binds to the TNF promoter distinctly from NFATp, c, 3 and 4, and activates TNF transcription during hypertonic stress alone. 2005; Nucleic Acids Res. 12;33(12):3845-54.

- Fagman H, Nilsson M. Morphogenetics of early thyroid development. 2011; J. Mol. Endocrinol. 46(1):R33-42.

- Frigerio G, Burri M., Bopp D, Baumgartner S and Noll M Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. 1986; Cell. 47:735-746.

- Geyer PK, and I Clark. Protecting against promiscuity: regulatory role of insulators. 2002; Cell. Mol. Life Sci. 59: 2112–2127.

- Graef IA, Chen F, Chen L, Kuo A, Crabtree GR. Signals transduced by Ca(2+)/calcineurin and NFATc3/c4 pattern the developing vasculature. 2001; Cell. 105: 863–875

- Jong-Il Kim, Jae-Bong Park, Jae-Yong Lee, SungChan Kim, Mae Ja Park, Zigang Dong, Jaebong Kim

- Kegley KM, Gephart J, Warren GL and Pavlath GK. Altered primary myogenesis in NFATC3(–/–) mice leads to decreased muscle size in the adult. 2001 Dev. Biol. 232:115-126.

- Kilcherr F, Baumgartner S, Bopp D, Frei E and Noll M. Isolation of the paired gene of Drosophila and its spatial expression during early embryogenesis. 1986; Nature 321:493-499.

▪ Kingston RE, and Narlikar GJ. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. 1999; Genes & Dev. 13:2339–2352.

▪ Kingston RE, Narlikar GJ. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. Genes Dev. 1999; 13:2339–2352.

▪ Klett M. Epidemiology of congenital hypothyroidism. 1997; Exp. Clin. Endocrinol. Diabetes 105:19–23.

▪ Kutach AK and Kadonaga JT. The downsteam promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. 2000; Mol. Cell. Biol. 18:233-239.

▪ Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev. 1998; 12:34-44.

▪ Lania G, Zhang Z, Huynh T, Caprio C, Moon AM, Vitelli F and Baldini A. Early thyroid development requires a Tbx1–Fgf8 pathway. 2009; Developmental Biology 328:109–117.

▪ Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. The MTE, a new core promoter element for transcription by RNA polymerase II. 2004; Genes dev. 18:1606-1617.

▪ Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev. 2004 Jul 1; 18(13):1606-17.

▪ Matthew V, Rockman and Gregory A Wray. Abundant Raw Material for Cis-Regulatory Evolution in Humans. 2002; Mol Biol Evol. 19(11):1991-2004.

▪ Menie Merika and Dimitris Thanos. Enhanceosomes. 2001; Current opinion in genetics & development; 11(2):205-8.

▪ Michael Carey. The Enhanceosome and transcriptional synergy. Cell. 1998 January 9; 92(1):5-8.

▪ Nitsch R., Di Dato V., Di Gennaro A., De Cristofaro T., Abbondante S., De Felice M., Zannini M.S., Di Lauro R. Comparative genomics reveals a functional thyroid-specific element in the far upstream region of the PAX8 gene. 2010; BMC Genomics. 11:306.

▪ Pasca di Magliano M, Di Lauro R, Zannini M. Pax8 has a key role in thyroid cell differentiation. 2000; Proc. Natl. Acad. Sci USA. 97(24):13144-9.

- Roy A.L., Meisterernst M., Pognonec P. and Roeder R.G. Cooperative interaction of an initiator-binding transcription initiation factor and the helix–loop–helix activator USF. 1991; Nature, 354:245–248.

- Shaw JP, Utz PJ, Durand DB, Toole JJ, Emmel EA, Crabtree GR. Identification of a putative regulator of early T cell activation genes. 2010; J Immunol. 185(9):4972-5.

- Sung-Young Lee, Jaeho Yoon, Hyun-Shik Lee, Yoo-Seok Hwang, Sang-Wook Cha, Chul-Ho Jeong,

- Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, Hata H, Ota T, Isogai T, Tanaka T, Nakamura Y, Suyama A, Sakaki Y, Morishita S, Okubo K, Sugano S. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. 2001; Genome Res. 11:677-84.

- The Function of Heterodimeric AP-1 Comprised of c-Jun and c-Fos in Activin Mediated Spemann Organizer Gene Expression. 2011; PLoS One 6(7): e21796.

- Weis L and Reinberg D. Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factor and initiator-binding proteins. 1997; Mol. Cell. Biol. 17:2973-2984.

- Wendl T, Lun K, Mione M, Favor J, Brand M, Wilson SW & Rohr KB. Pax2.1 is required for the development of thyroid follicles in zebrafish. 2002; Development 129:3751–3760.